Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Scienze Ambientali
Ciclo XXVII
Anno di discussione 2015

Chemical characterization of plant materials and development of analytical methodologies for metabolite determination

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: CHIM/01

Tesi di Dottorato di Elisa Scalabrin, matricola 825868

Coordinatore del Dottorato	**Tutore del Dottorando**
Prof. Gabriele Capodaglio	**Prof. Gabriele Capodaglio**

A Luca e ai nostri sogni

Contents

Summary

Plants synthesize a huge number of metabolites, involved in essential life functions (primary metabolites) or in particular defense, signaling and development roles (secondary metabolites). Plant metabolome remains for a large part uncharacterized, due to the high variability among species. The investigations of this rich variety of plant metabolites have been improved in the 20th century with the development of new analytical techniques, which permitted the identification of low-abundant molecules and their structural characterization. In this study we present different approaches for the characterization of the chemical composition of plant materials. Taking advantage from the use of HRMS technology, we developed both quantitative and qualitative methods, in order to obtain a comprehensive profiling of metabolites. In the first research chapter we describe a study of wild and transgenic *Nicotiana Langsdorfii* plants, exposed to different abiotic stresses. These plants have been traditionally used for genetic and physiologic studies. The plants were modified by the insertion of the Rol C gene, from *Agrobacterium rhizogenes*, and of the rat glucocorticoid receptor (GR). The aim of this study was the investigation of the metabolic changes associated with the genetic modifications and stress exposition, in order to highlight eventual advantages deriving from the inserted genes and to better understand the effects of abiotic stresses. The integration of different analytical approach permits to identify a high number of metabolites and to highlight the potential increased resistance induced by GR gene. The main molecules involved in plant stress response were identified: lipids (SQDG/DGDG and fatty acids), acylsucroses and glykoalkaloids were the main compounds involved in heat stress response; this finding suggests a role of the trichomes, important *Solanaceae* structures, in this kind of stress. The Cr(VI) and water stress application resulted mainly in enhanced antioxidants (HCAs, polyamine) levels and in the damage of lipids and fatty acids, as a consequence of ROS production. In the second study we considered two different species of the *Glycyrrhiza* genus, *G. Glabra* and *G. Uralensis*, commonly used for the production of licorice. The characterization of the two *G. Glabra* varieties, *glandulifera* and *typica*, was also carried out. The application of the metabolomic method aimed to the characterization of the chemical composition of each species, which is directly involved in the quality of the derived products. The main part of identified metabolites belongs to the flavonoid and triterpenic saponin classes, which are known to be the principal constituents of licorice roots. Key metabolites, specific for each licorice typology, were identified: glabridin and glycycoumarin were assessed as biomarkers for *G.Glabra* and *G.Uralensis* respectively. The most relevant compounds for *G. Glabra* resulted Licorice saponin H2/K2, glabridin, Licorice glycoside A/C1/C2 and Kanzonol Y. The *Uralensis* species showed a broad range of specific compounds, mainly flavonoids, and particularly high liquiritin/isoliquiritin intensities. We were able to identify some atypical metabolites in *G. Glabra*, which were not previously detected in licorice. Their presence could be relevant for further pharmacological applications. The third study report the analysis of the fruits of *Coffea Canephora* and *Coffea Arabica*, the most cultivated species for the production of coffee beans. The different parts of the fruits (perisperm, endosperm, pulp) were collected separately in two different harvest seasons. The metabolomic analysis was performed by integration of two analytical approaches: the use of HPLC-HRMS for the detection of secondary metabolites and the use of GC-TOF for the identification of primary metabolites. The results highlighted higher chlorogenic acids and ferulic derivatives content in *C. Canephora*; these compounds are known to possess important antioxidant properties and to contribute to the characteristic bitterness and acidity of coffee. *C. Arabica* resulted characterized by a wide variety of diterpenic glycosides, some of which already detected in this species, especially presents in the pulp and the perisperm tissues. *Coffea* fruits also showed the presence of flavan-3-ols and procyanidins, in higher number and intensities in *C. Canephora*

than in *C. Arabica*. Interestingly, *C. Arabica* resulted characterized by higher epicatechin levels while *C. Canephora* presents increased catechin amounts. The ripening process showed to be faster for *C. Arabica* than for *C. Canephora;* in *C. Arabica* the maturation resulted more gradual than in *C. Canephora*, in which metabolite composition changes abruptly during the last stage. The sugar content resulted higher for *C.Arabica* than for *C.Canephora*, with direct implication for coffee flavor and aroma.

This work showed the potential of the integration of different approaches in analytical chemistry, contributing to the comprehension of plant stress response and suggesting some possible application of the genetic modifications tested. Moreover we provide useful information about licorice and coffee, demonstrating the potential of metabolomic methodology as a tool in the food characterization and quality assurance. The metabolomic analyses also permit to suggest a few biosynthetic pathway regulations involved in stress response and ripening process. The study represents a good starting point for future works in the field of foodomics and of system biology, highlighting original findings, not reported before, which should be better investigated.

1 General introduction

Plants, as main primary producers, play the most important part in the natural cycle: without vegetation there could be no life on Earth. Plants, trough photosynthesis, are able to obtain energy from the sun and convert water and carbon dioxide into carbon sources, like sugars, proteins, fats, starch, which supply energy to all the living organisms. Plants products are vital to humans and include food, wood, fibers, drugs, oils, fragrances, pigments and many other raw materials [1]. Moreover, plants contribute to the regulation of Earth climate, strongly influencing atmospheric chemistry through carbon dioxide consumption and nitrogen fixation; plants also have an important role in the biogeochemical cycles, in the control of soil chemistry and water content [2]. Throughout history, plants have been collected, traded, adapted for new environments and bred for new combinations. From fundamental discoveries about plant life, technologies and capabilities in a wide range of practical applications arose; research with plants has strongly influenced the development of biology and has contributed to many important scientific advances: the discovery of the rules of genetic inheritance (Gregor Mendel's peas), the role of light in regulating the physiologic responses of higher organisms (phytochromes), the transposition of genetic elements, the protein nature of enzymes [3]. Nowadays, many studies have been developed on vegetal life, about biology, physiology, ecology and biochemistry but much work still remains in order to understand such complex organisms and their interactions with external environment. Particularly, many efforts have been made in plant genetic manipulation, in a perspective of climate changing conditions; genetic experiments aim to understand the factors involved in plant development, plant adaptation capabilities, the possibilities of improve vegetal products for human use; another essential and well studied field is the food characterization, which aim to improve the quantity and quality of vegetal products, to discover new applications and identify new commercial valuable molecules, considering the increased food request. Plants' world offers a wide range of new economical possibilities, especially for the high adaptation ability and the complexity of its systems. Plant genome contains around the same number of genes as that of the humans, but the metabolome comparison reveal a huge metabolic diversity of plants. This complex metabolism is the result of various developmental adaptations; plants, due to their lack of motility, evolved specific and advanced mechanisms which permitted their survival in a wide range of different environments and temporary extreme conditions [1, 4]. Evolution has driven organisms to increasingly sophisticated levels of chemical interaction in which each molecule have a role in the functioning of all the biologic system. Many metabolites are produced as essential compounds for plant growth, as structural materials and photosynthetic pigments; many others are biosynthesized for specific necessities, functioning as energy storage compounds, defensive substances, signaling molecules, and protective compounds [5]. The investigations of this rich variety of plant metabolites have been improved in the 20th century with the development of new analytical techniques, which permitted the identification of low-abundant molecules and their structural characterization [6]. Plant metabolite measurements have been carried out for decades owing to the fundamental regulatory importance of these molecules as components of metabolic pathways; however, only recently, the use of metabolites as diagnostic markers for a wide range of biological conditions has been recognized [7]. Historical approaches in metabolite analysis include metabolite profiling, metabolite fingerprint and targeted analysis; these strategies differ in the number of considered compounds, in the level of structural information obtained, and in their sensitivity [8]. The most common approach is the **metabolite profiling**, defined as the identification and quantification of a selected number of predefined compounds; in this case, sample processing aim to isolate the analytes from possible matrix effects and interferences, maximizing at the same time the extraction recovery. At the other extreme, the **metabolite fingerprint**,

which is widely used as a screening tool, quickly detects many compounds but with no clear identification of their structures; sample preparation is simple and need to avoid compound loss and sample contamination. **Metabolite targeted analysis** is the qualitative and quantitative determination of one or more metabolites related to a specific metabolic pathway or reaction; in this case, an extensive sample preparation protocol is required [9]. Today the development of the **metabolomic** methods typically allow measuring hundreds of compounds, with a small number being definitively identified, a larger number being identified as belonging to particular compound classes, and many remaining unidentified. The advances in analytical technologies, particularly the improvement of high resolution mass spectrometry (HRMS), and the increased availability of new informatics tools, make this new discipline rapidly grow and find application for a broad range of purposes [10]. Recently, plant biologists have disclosed metabolomic potential by using this high-throughput approach in the understanding of plant cell biochemistry, of the regulatory mechanisms which govern the system functioning and of the influences of mutations and environmental conditions on plant metabolism [11–15]; while primary metabolism in plant has been well understood, the secondary metabolite pathways, due to their diversity and number, are less characterized; the plant secondary metabolism is highly connected with abiotic or biotic stress resistance and organism interactions. Many secondary metabolites, moreover, play a central role in determining vegetal product quality, both in the alimentary and in the pharmaceutical field [8, 9, 16–19]. For these reasons, plant metabolomic, giving an overview of the molecular status and organization of the whole system, offers a new way of studying complex biological problems and results particularly applicable for food quality researches.

1.1 PRIMARY AND SECONDARY METABOLISM: BIOSYNTHESIS, METABOLITE TYPOLOGIES AND FUNCTIONS

Plant metabolism encompass many synthetic processes: growing plants but also the adult organisms has to continuously rebuild many organic structures. Anabolic processes are always endoergic, in contrast to catabolic reaction pathways, which consume large amounts of available free energy [20]. The metabolic pathways can be generally viewed as series of pool, metabolic intermediates linked by reversible enzymatic reactions, which direction depends from the metabolic requests [21]. Plants synthesize a high number of compounds, having many different chemical structures and functions. It is estimated that the whole plant kingdom can produce between 200000 and 400000 different metabolic compounds, whereby a single species may be producing about 500-10000 compounds in a certain time in a particular environment [1]. The whole plant metabolome could be classified into primary and secondary metabolome. Primary metabolites are generally distributed in all living organisms, involved in the basic life functions, like growth, development, respiration and biosynthesis of essential compounds [22]. Primary metabolites include ubiquitous compounds like carbohydrates, amino acids, fatty acids and organic acids [18]. Secondary metabolites are not directly needed for normal plant rising but their production can enhance the organisms' survival under particular stress conditions [23]. These compounds may serve as attractants for pollinators or symbionts, as deterrent against potential herbivores or pathogens or competing plant species. They are produced in specific parts of plants at defined developmental stages. The amounts of secondary metabolites are often low (less than 1% of dry weight) and highly variable [12, 24].

1.1.1 Plant primary metabolism overview

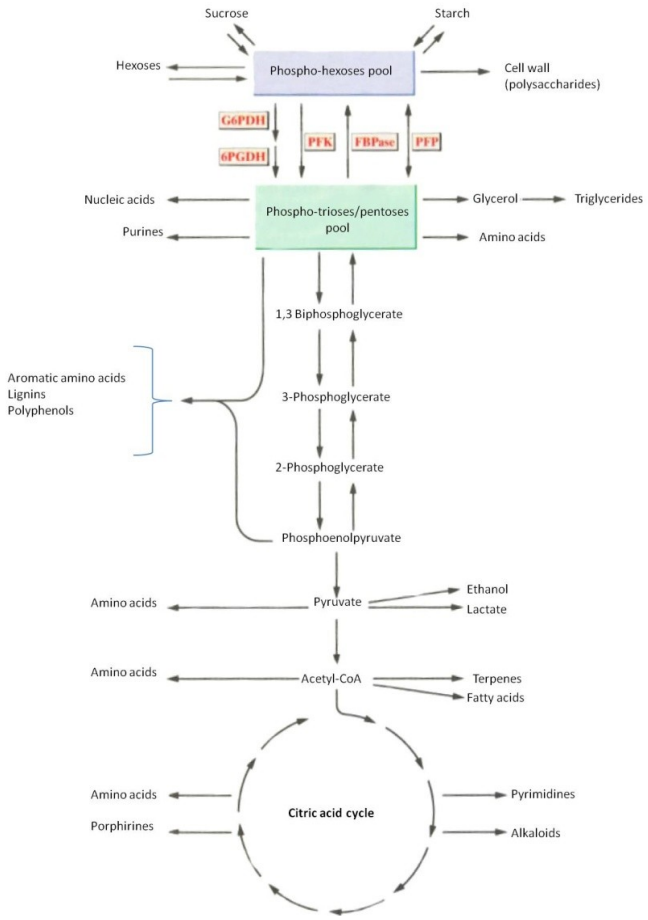

Figure 1 The central role of carbohydrate metabolism as a source of precursor for the synthesis of primary and secondary metabolites [21]

The highly diverse molecules of life are synthesized from a much smaller number of precursors, produced by primary metabolism. Sucrose is often the principal stable product of photosynthetic carbon assimilation and the principal form in which the energy is temporary stored and transported. Sugars, by themselves, are essential components of plant nutrition. The monosaccharides, the sugar alcohols, also called polyols, the sucrosyl oligosaccharides are other relevant sugars, sometimes used as transport compounds.

Carbohydrates metabolism in plants depends mainly from two pathways: the phospho-hexose pool, which lies at the junction between many different metabolic pathways (Figure 1), and the phospho-trioses/pentoses pool. From the phospho-hexose pool, sucrose, starch and cellulose are synthesized. From the triose/pentose pool, nucleic acids and glycerol are produced [21, 25].

Amino acids, which play a central role in both primary and secondary metabolism, are the main form of assimilation of nitrogen. Amino acids are synthesized from the phospho- triose/pentose,

glycolysis or citric acid cycle intermediates while the aromatic amino acids are synthesized via the shikimate pathway (see paragraph 1.1.1). Amino acids are involved not only in plant growth but also in plant defense [20, 26, 27].

The Acetyl-CoA is the substrate of fatty acid biosynthesis, key molecules in the production of various secondary metabolites, especially the lipids. More than 200 different fatty acids have been found in vascular plants; the most abundant, especially in the plant membranes, contains 16 or 18 carbon atoms and have different saturation degrees [28]. The term "lipids" is a general category used to indicate various structurally different compounds which share low water solubility. Lipids carry out many functions in plants: they are the main components of biological membranes, originating a hydrophobic barrier essential for life, and isolating cells from environment. Moreover, they represent an important free chemical energy reserve and important signaling molecules [21]. Many classes of lipids have been identified, each one with specific physiological functions. The **glycerolipids** are the most common lipids; they consists of fatty acids esterified with the glycerol or a derivative; they include the **triacylglycerols**, which are the principal energy storage form of plant cells, the **phospholipids**, which are the basis of cellular membrane bilayers, and the **galactolipids**, other important constituents of chloroplast membranes [21, 29].

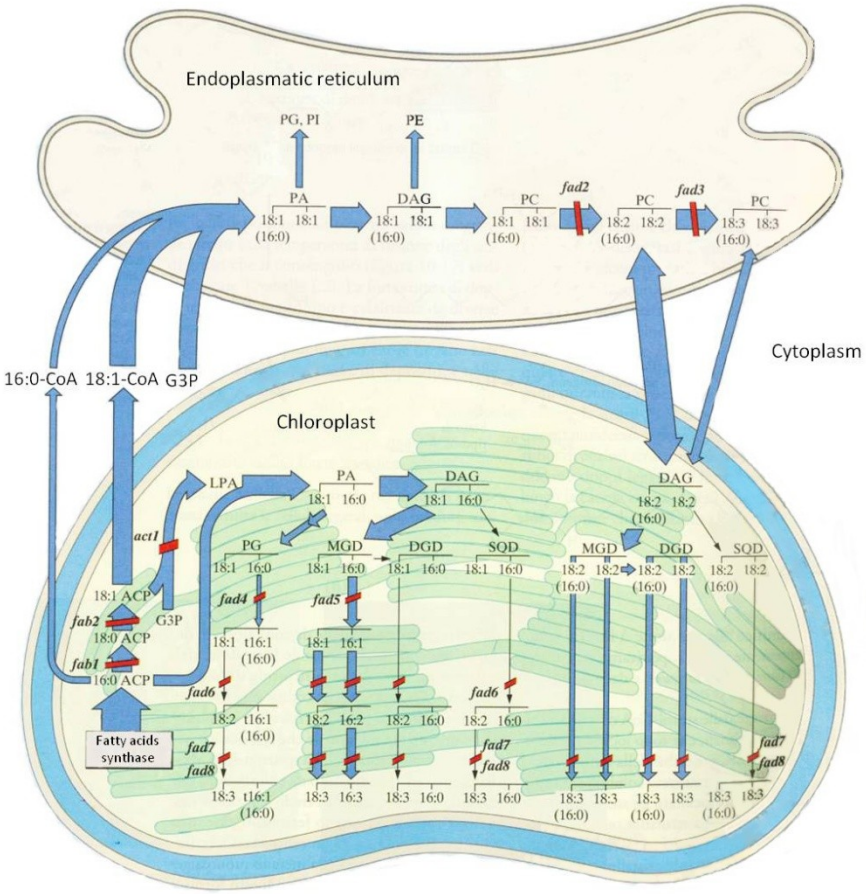

Figure 2 Simplified scheme of lipid synthesis in Arabidopsis, modified from [21].

Besides the glycerolipids, the plants synthesize other important compounds (Figure 2) as the **sphingolipids;** sphingolipids are generally found in the plasmatic membrane and are presents in low concentrations being potentially toxic compounds.

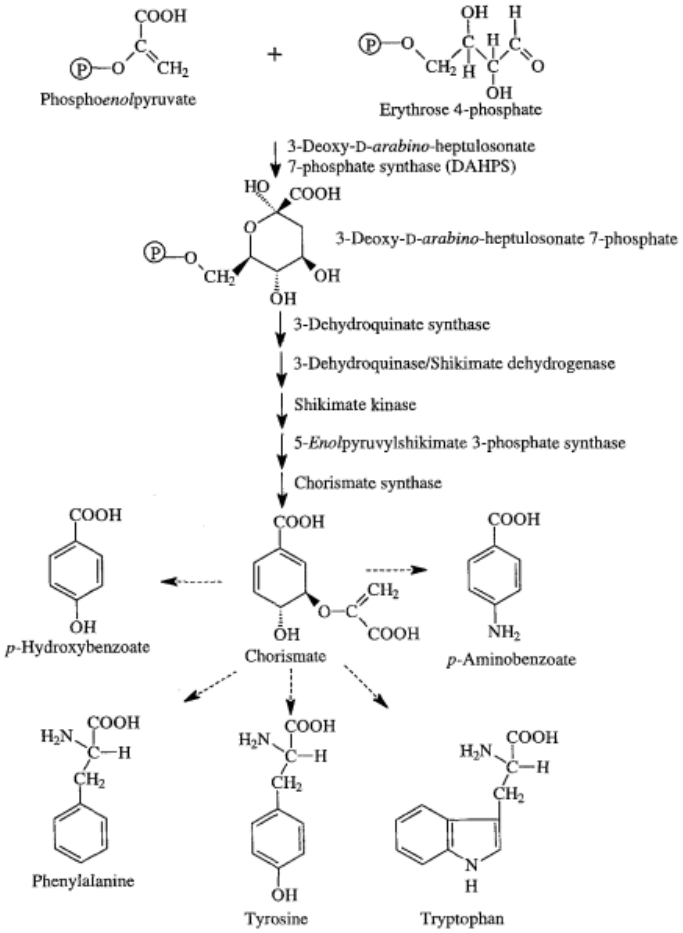

Figure 3 A schematic representation of the shikimate pathway in higher plants [27]

1.1.1 The shikimate pathway

The shikimate pathway represents the main connection point of primary and secondary metabolism. In plants, the shikimate pathway provides the precursors of phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) and of a very wide range of other aromatics. The chorismate, which is the end product of the pathway, is converted to quinones, indoles and the aromatic amino acids, all of which, in turn, are precursors of a multitude of secondary metabolites. Phe, Tyr and Trp are precursors in the biosynthesis of alkaloids. Furthermore, phenolic compounds derive from phenylalanine and, to a limited extent, from tyrosine, or directly from shikimate pathway intermediates [30]. The shikimate pathway can be divided into three parts: the first part, until chorismate production, is common for the synthesis of all the three aromatic amino acids, and hence is often called the prechorismate pathway (Figure 3). Beyond chorismate, the shikimate pathway branches into two pathways, one leading to phenylalanine and tyrosine and another producing tryptophan [27]. The enzymes of the shikimate pathway are highly affected by environmental factors therefore the synthesis of aromatic amino acids and secondary products is strongly dependent from plant growth conditions.

1.1.2 Secondary metabolites: main classes of interest and functions

As already outlined, plants are able to produce a high number of secondary metabolites which carry out different functions in the environmental context. Secondary metabolites can be classified on the basis of their structures, chemical properties, functions or biosynthetic pathway through which are produced. From the biosynthetic point of view, secondary metabolites can be divided into three major groups: **phenolics**, deriving mainly from the shikimic acid pathway, **terpenoids**, produced through the mevalonate pathway and **alkaloids,** which are amino acids derivatives [6, 17, 24, 31].

1.1.2.1 Phenolics

Phenolics represents one of the major class of secondary metabolites and their main common structure consists of an hydroxyl group bonded directly to an aromatic hydrocarbon group; phenolics include a high number of different compounds, ubiquitously found in plants and known to possess bioactive properties. From the biosynthetic point of view, phenols mainly derive from the shikimate pathway, through the phenylpropanoid way or from the polyketide acetate/malonate pathway. Hydroxycinnamic acids (HCAs) represent the first step of phenolic biosynthesis; they are synthesized by deamination from the amino acids phenylalanine and tyrosine, yielding the C6-C3 unit that serves as the core structure of many compounds, as phenyl-propanoids and lignins. HCAs can occur as free carboxylic acids, esters formed by condensation with hydroxylic acids, or amides. The most abundant HCA derivative in plants are chlorogenic acids, formed by the condensation of quinic acid with caffeic, p-coumaric and ferulic acids, which occur in many fruits and vegetables and especially in coffee and tobacco. HCAs are believed to posses antioxidant, antiallergic and antimicrobial activities and they are known to play an important role in the plant response to biotic and abiotic stress [32–34]. Linked to the synthesis of HCAs is the biosynthesis of another major group of phenolics, the flavonoids (Figure 4). Flavonoids can be included in the class of polyphenols, containing more than one phenol unit. Flavonoids are ubiquitously distributed in all plants and are presents in almost all green photosynthetic cells; the physiological roles of flavonoids in plants are diverse: they are important flower pigments, producing yellow or red/blue colors, they are involved in gene regulation and growth metabolism; flavonoids enhance tolerance to a variety of abiotic stressors and protect plants from UV radiation, acting as Reacting oxygen species (ROS) scavengers [35–37]. From the pharmacological point of view, flavonoids have been studied a lot and their antimicrobial, anti-inflammatory, estrogenic and UV-absorption activities have been assessed [38, 39]. Flavonoids include many subclasses of compounds which can differ for the degree and pattern of hydroxylation, methoxylation, prenylation, or glycosylation. Flavones, flavonols, flavanols, isoflavones and isoflavanones are probably the most abundant and widespread distributed in higher plants, including compounds as luteolin, apigenin, quercetin, kaempferol, catechins, and epicatechins. Flavonoids can occur in plants as aglycones but they are often found as glycosylated derivatives [40]. Moreover, some flavanols, namely catechin and epicatechin, are often found as dimeric/trimeric forms originating the proanthocyanidins (PAs), very abundant in fruits, bark, seeds and beans. PAs possess metal chelating and antioxidant activities; in plants they influence the color, bitterness and astringency of fruits and could participate in the defense mechanisms of the organisms [41]. Lignans and lignins directly derive from HCAs and probably take part in plant defense against pathogens and in regulation of plant growth [42, 43]. Lignins are the essential structural polymers of wood and, after

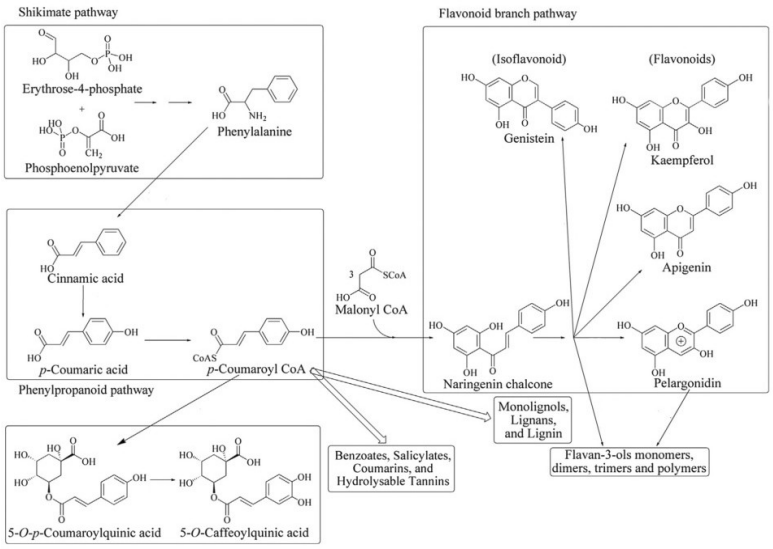

Figure 4 Scheme of the major branch pathways of phenolics biosynthesis modified from [350].

cellulose, the most abundant organic compound in plants. From the p-Coumaryl CoA other important secondary metabolites derive: the **coumarins**, which possess a high variety of pharmacological activities [44] **and salicylates,** which main role is plant cell signaling and immune response [45, 46]. Salicylic acid (SA) is first synthesized from the shikimic pathway and readily converted in its derivatives like methyl-SA, SA-glucosyl ester and glucosyl-SA; SA conjugates are mostly inactive compounds but they are believed to act as storage molecules with the potential to be converted back to SA when needed.

1.1.2.2 Terpenoids

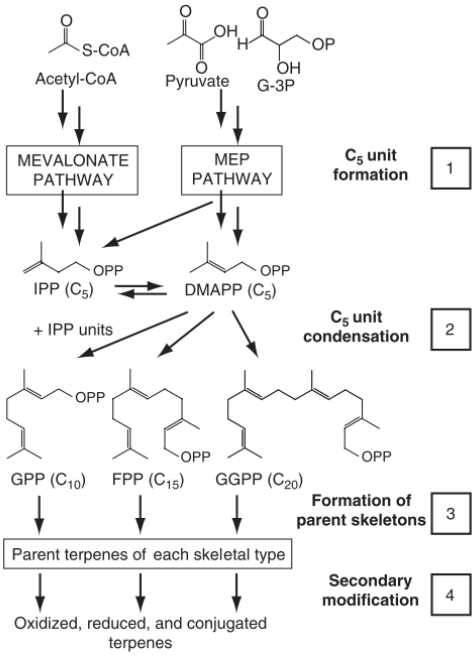

Figure 5 The four main phases of terpene biosynthesis [47]

Terpenes, also known as isoprenoids, are the largest group of plant secondary metabolites, comprising about 30 000 substances; in plants they take part both in primary and in secondary metabolisms: they function as membrane constituents, photosynthetic pigments, growth promoters, glucosyl carriers, plant hormones and protection substances. The building blocks of terpenes are the five-carbon isoprene units (C_5H_8) which are assembled and modified in thousands of ways[17, 47]. The biosynthesis of terpenes and terpenoids starts from primary metabolic intermediates as acetyl CoA, pyruvic acid and G3P, which are converted into the precursor five-carbon units, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) through both the mevalonate pathway and the methylerythritol phosphate pathway (Figure 5)[48]. Monoterpenes are produced from the geranylpyrophosphate (GPP) while the farnesylpyrophosphate (FPP) is the precursor of sesquiterpenes, the largest group of isoprenoids, which include many aromatic compounds and many precursors of important molecules. The metabolism of geranylgeranylpyrophosphate (GGPP) leads to the formation of diterpenes, including the plant hormones gibberellins, phytoalexins and pigments as carotenoids. Diterpenes are acyclic and cyclic compounds, widely presents in plants, especially in the glycosylated form; they often possess antimicrobial and anti-inflammatory activities. Triterpenes, which derive from FPP, are widely distributed in all higher plants classes. Squalene, a triterpene, is the precursor of **cholesterol**, sitosterols and **brassinosteroids**, an important class of phytohormones. Significant triterpenic derivatives are the **triterpenoid saponins**, glycosylated triterpenes found in dicotyledonous plant species; these compounds are cell membrane constituents, defense compounds and help in the maintenance of the fluidity of membranes. They are toxic substances in plants with hemolytic activity against herbivores and fungi. A saponin may contain more than one saccharide chains, and also an acyl group attached to the sugar moiety [49–51]. Due to the combination of the lipophilic properties of the aglycone and the hydrophilic sugars, many saponins are able to act as surfactants. In addition to triterpenic saponins **steroidal saponins**, where the aglycone is represented by a steroid derived from cholesterol, are very common in plants of the family *Solanaceae*. Triterpenic saponins are generally more widely distributed than steroidal saponins and are abundantly found in *Leguminosae*. More than 20 groups of triterpenic saponins can be distinguished, on the basis of the aglycon type, the most abundant of which belong to oleanane type. Many saponins are used in the pharmaceutical industries for their anti-inflammatory, anti-bacterial, anti-parasitic, anti-viral activities [52–55]. Tetraterpenes, biosynthesized from GGPP, are

important precursor of carotenoids, a group of pigments presents in flowers and fruits which play an important role in photosynthesis.

1.1.2.3 Alkaloids

Alkaloids are a wide group of low molecular weight organic molecules containing a heterocyclic nitrogen ring; they are structurally diverse and biogenically unrelated molecules produced by many different organisms, including animals and microbes. Plants produce the highest variety of substances which function principally in defense against herbivores and pathogens [56, 57]; alkaloids have been studied a lot for their pharmacological properties (analgesic, anesthetic, anti-cancer, narcotic activities). Many alkaloids are nevertheless toxic for other organisms like other plants and insects. Alkaloids are very species-specific compounds, indeed their biosynthetic genes are not conserved between different species [48]. The structural classes of alkaloids are typically defined by the substrate starting material: alkaloids mostly derive from the amino acids Phe, Tyr, Trp, Lys and Orn. **Isoquinoline** alkaloids represent the largest group of compounds which includes codeine, morphine and the antibiotic berberine. The **indole alkaloids** comprise over 2000 members, like strychnine and quinine, possessing a range of chemical structures and many biological activities [58, 59]. The **tropane alkaloids** are mainly biosynthesized in plants of the family *Solanaceae*, especially in roots, and contain a nitrogenous bicyclic organic compound. Tropane alkaloids are synthesized from Orn which is decarboxylated to putrescine, an important polyamine compound with a key role in plant response against abiotic stresses [11, 60]. Some tropane alkaloids, like cocaine and atropine, have pharmacological activities and are used as drugs, stimulants and anti-cholinergics. The class of **purine alkaloids** is biosynthesized from purine precursors, and is often referred as pseudo-alkaloids. The most important purine alkaloid is caffeine, found in seeds, leaves and fruits of many plants, especially in plants of genus *Coffea* and *Camellia*. In the coffee plant, caffeine is present in all parts over the ground. The ecological role of caffeine is probably the chemical defense against herbivores, insects, fungi or bacteria but its role is currently under discussion [58, 61]. Another interesting group of alkaloids, which classification is border-line with terpenoids, are the **glycoalkaloids** like solanine, tomatine, and chaconine. They are also called saponins due to their surfactant properties; several compounds are potentially toxic, remarkably the poisons found in the plant species *Solanum dulcamara* [18].

1.2 CHEMICAL CHARACTERIZATION OF PLANTS: OBJECTIVES, ANALYTICAL METHODOLOGIES

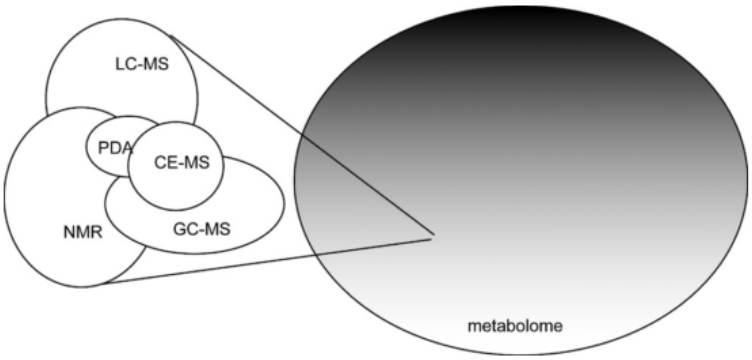

Figure 6 Schematic representation of the metabolome, indicating that only a small fraction of metabolites have been identified yet, the majority of naturally-occurring metabolites still being unknown [85].

The analysis of secondary metabolites in plants is a challenging task because of their chemical diversity, usually low abundance and high variability even within the same species. It is estimated that 100.000–200.000 metabolites occur in the plant kingdom [62]. Current metabolite profiling methods cover only a small part of the complete metabolite content of plants; therefore, the need of enlarge the range of the simultaneously detected compounds is a striking issue (Figure 6). Many approaches have been historically used in the metabolite profiling. Originally, the measurements of metabolites were achieved by spectrophotometric analysis, able to detect only single metabolite, or by simple chromatographic separations of low-complex mixtures [7]. A lot of work has been done in the field of **targeted quantitative analysis,** in which the analytical method is developed to determine specific compounds with high selectivity and reliability; this approach generally involves several steps in the sample processing as extraction, clean-up, extract fractionation, sample volume reduction, dilution; the process aims to exclude interfering substances and to maximize the recovery of the analytes. Moreover, quantitative analyses require method validation, with the estimation of the accuracy, precision and recovery percentage of metabolites, by means of reference materials with substance certified content [63–68]. Often, the evaluation of the matrix effect is also needed to guarantee a true quantification of the real analyte content and to avoid instrumental uncorrected estimations [69]. This approach is expensive and time consuming, requiring the use of analytical standards, which are not always easily found on the market; moreover targeted analysis could give information only about a limited number of substances [13]. Quantitative information about plant metabolite content are however useful, in order to compare results among different studies and methodologies, or to verify the fitting with legislation parameters, particularly in the field of food quality and security; moreover the quantitative analyses assure the reliability of results [70, 71]. Quantitative analyses are not always necessaries and qualitative approach could result more time-convenient and cheaper, depending by study aims. **Untargeted analysis** have been extensively used for the characterization of phytochemicals and includes the confirmation of known compounds and the elucidation of unknown compound structures; many studies focused on the isolation and purification of compounds from natural product extracts, which consists in a time-consuming and laborious process, often resulting in uninteresting or already characterized molecules [72]. Besides this selective approach, other studies report a more comprehensive characterization of metabolites through the **metabolic fingerprint,** which consist in the rapid and global analysis of biological matrices in order to provide sample classification and comparison. This kind of approach historically employed capillary electrophoresis (CE) [73], High performance liquid chromatography (HPLC) separations coupled with UV/DAD detection [74, 75] or the combination of more techniques [76], often revealing many compounds without providing their identification [8]. In the last decades metabolite fingerprinting has moved toward the use of more technological and effective instrumentations, like Nuclear magnetic resonance spectroscopy (NMR) [77], HPLC or Gas Cromatography (GC) coupled to mass spectrometry detection (MS) [33, 78–82], Fourier

transformed ion cyclotron resonance mass (FT-ICR-MS)[83] thus opening the road for the development of metabolomic discipline. In the last decades many steps have been done especially in the improvement of analytical instrumentations [4] and now scientists planning metabolomic analyses can choose among many different analytical techniques, able to provide adequate sensitivity, accuracy and resolution for the characterization of metabolites [1].

1.3 METABOLOMIC APPROACH: GOALS AND POTENTIAL

Plant metabolomics is a new discipline, which appeared in the early 21st century and provided an overview of metabolites in cells, tissues, organs, and whole organisms, resulting in succeeding innovative advances in genomics [84]. Probably the most outstanding feature of metabolomics is its interdisciplinary and integrative capacities, shifting from the pure analytical chemistry to a biochemical approach which take also advantage from bioinformatics [85]. Considering metabolomics as a combination of knowledge in biochemistry, analytical chemistry, biology, signal processing and data handling, the challenge lies in performing these resources in a organized and coordinated way, able to solve biological questions [8].The metabolomic aim at the comprehensive analysis of all metabolites presents in a biological system [86] and offers a new way of studying complex molecular problems, particularly applicable for natural products research [9]. As metabolites are the end products of biological processes, their comprehensive characterization offers a great opportunity in the understanding of cellular and system functioning. From this point of view, metabolomics is a powerful tool by which obtaining a wide perspective on metabolic network regulation. More than proteomics and genetics, metabolomic is capable to insight plant phenotypic changes, studying the interaction between different organization levels and providing conclusive and functional information [86, 87]. The integration of metabolomic data with transcriptomic and genetic information resulted particularly useful in the discovery of metabolic processes underlying the behavior of a biologic system. Metabolomics has found application in many fields, including natural product quality assessment and traceability [19, 88–96], food and nutrition [80, 97–99], in the comprehension of plant metabolism and pathway elucidation [100, 101] and in the study of plant response to environmental stresses [11, 13, 48, 102–104].

The metabolomic approach, ideally aiming to the complete characterization of plant metabolome, must avoid the exclusion of any metabolite by using simple, effective and well conceived sample preparation procedures; the assurance of sample product reproducibility, the high recovery and stability of most compounds are crucial issues in metabolomic method development. The choice of the instrumental technique must guarantee enough resolving power, maintain sensitivity, selectivity, matrix independence, and universal applicability [13, 105].

From the analytical point of view, two main techniques are used in the metabolomic field: NMR and HRMS. Hyphenation of chromatographic methods and MS and the simultaneous use of DAD and MS revelators enable the detection of several hundred of molecules with appropriate sensitivity and accuracy [84, 85]. NMR is essential for the unequivocal identification of unknown compounds, providing good chemical specificity and reproducibility for compounds containing elements with non-zero magnetic moments such as ^{1}H, ^{13}C, ^{15}N, and ^{32}P, that are commonly found in many biological matrices [15]. However, NMR lacks in sensitivity compared with MS and its capacity of annotating compounds is limited in complex mixtures [86]. MS approach is the most common employed in metabolomics: sometimes mass spectrometry has been used with direct flow injection, both at low and high resolution, especially for the detection of metabolites with low molecular masses [106]. The hyphenated approaches appear more advantageous, combining

separation and detection, thus permitting the analysis of complex biological mixtures. GC-MS allows the analysis of volatile compounds or of less volatile substances which need derivatization; up to 400 compounds can be separated and detected, including amino and organic acids, sugars, amines, fatty acids, and sterols [23]. GC coupled with electron impact (EI) and time of flight spectrometry (TOF) was the first approach used in large-scale plant metabolomics, covering mainly the metabolites involved in primary metabolism (in the polar fraction of extracts), as well as lipophilic compounds (in the apolar fraction) [105]. GC-TOF technology offers high mass resolution, high mass accuracy and fast scan speed, which are very important characteristics in the analysis of complex matrices [107]. The favored method for analyzing semi-polar metabolites, which represents the main part of plant secondary metabolites, is HPLC-MS which do not require derivatization prior to analysis and consents the detection even of thermolabile compounds. LC-MS technique has been successfully employed for the analysis of alkaloids [101, 108], saponins [55, 109–111], phenolic acids [34, 112], phenylpropanoids, flavonoids [63, 113], lipids [114]. LC-MS provides high selectivity and good sensitivity and can be simultaneously coupled with UV/PDA detection, offering multiple levels of information for chemical structure elucidation [15]. Modern MS often involves the use of two or more mass analyzers in tandem, in which a number of triggered mass selections are carried out, followed by subsequent fragmentation of the selected ions. The use of soft ionization sources, such as electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI), results in single-protonated (in positive mode) or deprotonated (in negative mode) molecular masses, which can be detected and simultaneously fragmented thus providing a complete fragmentation pattern of compounds [105]. Potentially, two (MS/MS) or more rounds (MSn) of this process are conducted, creating the so called spectral trees, an important feature for structural elucidation of unknown molecules [23, 115]. The use of the combination of low and high resolution mass spectrometry in tandem has become very common in metabolomic analyses; the Orbitrap mass analyzer, introduced in the early 2000s, has now reached the status of a mainstream mass spectrometry technique in combination with an external accumulation device such as a linear ion trap; the Orbitrap can support a wide range of applications including structure identification, analysis of trace-levels compounds, proteomics, foodomics [97, 101, 106, 116].

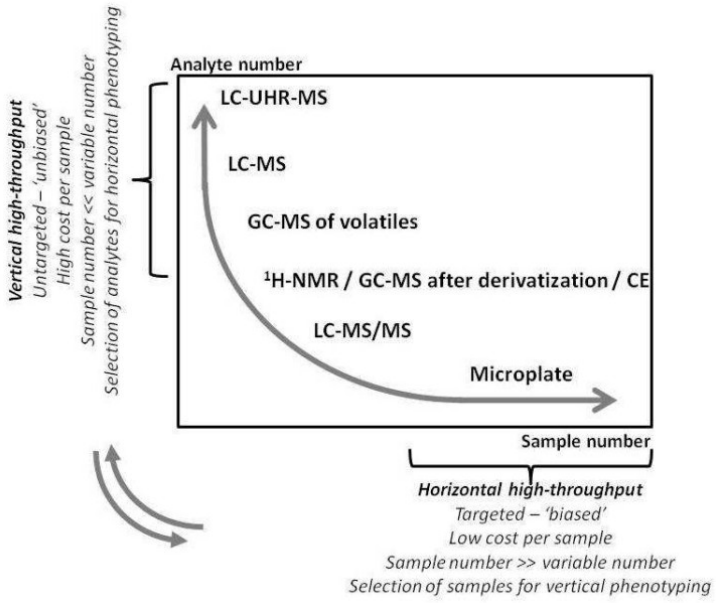

Figure 7 Characteristics of different metabolomic approach in term of costs and number of metabolites which can be detected [8]

Metabolomic analyses always produce huge data sets amount, consisting of several mass spectra which need to be processed with adequate tools; several software have been developed to fulfill metabolomics necessities, in order to compare mass spectra, remove instrumental noise, align spectra and consent the statistical analysis, obtaining a simplified description of the systemic response [105, 117–121]. Furthermore, true metabolomic approaches must also include strategies and criteria for metabolite identification, requiring appropriate tools like the calculation of the most probable molecular formula,

specific software for fragmentation simulations and adequate molecular and spectral libraries [10, 85, 122–124].

1.3.1 Main analytical technique: HPLC-LTQ-Orbitrap

The needs of bioanalytical chemistry in the last decades have forced the development of new technologies in mass spectrometry to permit the analysis of complex matrices, simultaneously detect a high number of molecules and unambiguously identify unknown compounds [125]. HPLC-MS is able to provide good sensibility, repeatability, accuracy in the detection of many plant metabolites and showed excellent results in untargeted, targeted and quantitative analysis of food and plants [63, 85, 87, 101, 126–128]. For this study the HPLC-LTQ Orbitrap technology was applied for both quantitative and metabolomic studies, demonstrating versatility and high quality results. RP-HPLC separation was employed, due to its flexibility in the analysis of compounds with different physicochemical properties by varying the stationary phases and the mobile phase composition [107]. RP-HPLC has become the preferred method in the analysis of complex matrices. The most common RP stationary phase is the octadecyl carbon chain (C18)-bonded silica which, when combined with other functional groups, permits the separation of both hydrophobic and polar compounds. In order to maximize the ionization of polar and high molecular weight molecules, the heated-ESI ionization technique was employed, coupled to an LTQ-Orbitrap XL mass spectrometer. This instrument joins a linear quadrupole ion trap to a high resolution mass analyzer permitting high accuracy mass measurements and MSn experiments. The HRMS measurements and the fragmentation patterns are necessaries in metabolomic analysis, for properly metabolite identification, but could result useful also in quantitative studies; the highly accurate measurements allow more reliable quantification, being able to distinguish the analytes from interfering molecules and matrix constituents with similar molecular masses. Moreover, the use of HRMS permits a simplification of the analytical procedures, thus not requiring sample clean-up for interfering substances removal. The linear ion trap integrated in the instrument has a greater ion storage capacity than conventional 3D ion trap devices. Ions can be stored and ejected, isolated, excited and then fragmented as necessary for MS/MS and MSn experiments. After ion trap mass filtering, ions can be injected in the Orbitrap or fragmented into the HCD collision cell. The ions reaching the Orbitrap mass analyzer are detected through ion image current, which measures the frequency of their harmonic oscillation; through the fast Fourier transform, the ion m/z value is calculated. The Orbitrap used in this work can reach the maximum resolving power of 100000 at m/z 400, with a mass range of m/z 50–2000 and the maximum mass accuracy of <2 ppm with internal calibration method [116, 125, 129].

1.3.2 Tools for metabolomic data analysis and metabolite identification

1.3.2.1 Data processing

The amount of data which results from metabolomic analyses need to be processed and cleaned, in order to extract the relevant information and transform raw data in a way in which they can be more easily handled. The quality of data processing is therefore an essential step for our ability to properly analyze and interpret metabolomic information [121]. Many tools, as elaboration software and mass spectral libraries, have been developed in the last years to support and simplify the metabolomic data processing, the molecule identification and data interpretation. MS manufacturers have produced available commercial tools, such as, for instance, Markerlynx (Waters) [130] and Sieve (ThermoFischer Scientific) [131] but many others can be used for free to preprocess and align data, as XCMS [132, 133], MZmine [134], MetaQuant [135], MET-IDEA [136], Tagfinder [137], IDEOM [118]. In this work we decided to use two free software developed at the Wageningen University (The Netherland): MetAlign (http://www.wageningenur.nl/nl/Expertises-Dienstverlening/Onderzoeksinstituten/rikilt/show/

MetAlign.htm, version 041012), developed by Arjen Lommen of the RIKILT-Institute of Food Safety and MSClust (http://www.wageningenur.nl/nl/Expertises-Dienstverlening/Onderzoeksinstituten/rikilt/show/MetAlign.htm), developed and implemented by Yuri Tikunov of the Plant Breeding department. These two complementary software, in comparison to many others, combine a user-friendly interface but, at the same time, work in extremely user-controlled conditions; the user can indeed set many parameters in order to find out the best processing method for its own data, with the assurance of efficient and statistical-based data processing. Moreover, MetAlign and MSClust can be used for both LC-MS and GC-MS data, with accurate mass and nominal mass measurements [117, 119, 120]. MetAlign is highly tested and implemented software, developed on the basis of metabolomic experience, especially in the field of plant and food analysis. MetAlign is constituted by a series of algorithms which permits the removal of noise and the alignment of spectra, on the basis of retention time or scan number. Further MetAlign information have been previously reported [119, 120, 138]. MSClust is an algorithm-based software tool for analysis of preprocessed GC-MS and LC-MS datasets. The main role of MSClust is the unsupervised clustering and extraction of putative metabolite mass spectra from ion-wise chromatographic alignment data. MSClust is able to elaborate tables (txt format) of ion mass with the relative retention time, scan number and intensities in samples, in order to remove metabolite signal redundancy and to recover mass spectral information of metabolites using mass peak clustering [117]. The free version includes the basic functions of the clustering algorithm while the complete version, which is in use only at the Wageningen University, provide many other tools especially for GC-MS data analysis [139].

1.3.3 Molecule identification

The identification of metabolites is considered the critical step of metabolomic analyses, due to sample complexity and of the lack of specific spectral and metabolite libraries [10, 85, 124]. The generally accepted levels of non-novel metabolite identification have been reported by Sumner et al., (2007) [140]. In this work the identification of metabolites, in the major part of cases, was performed at level 2, namely without comparison with chemical reference compounds. The identification was based upon physicochemical properties and spectral similarity with public or commercial spectral libraries. In some cases, when we lack information about the complete fragmentation pattern of compounds or when no structurally confirmations were found in databases and literatures, the identification level was established as 3. The first step of the identification process is the association of a molecular formula (MF) to the m/z ions. A single m/z ion could generally lead to many molecular formulas (MF), especially for high molecular masses. The presence of adducts, double-charges and dimers must even be considered. Using an instrument that can provide high mass accuracies, the range of possibilities of MFs is limited, thus increasing the probability to trace out the correct MF [85]. For molecular formula identification, we used the Xcalibur software 2.1 (Thermo Fisher Scientific, Germany) which is the manager software of the LTQ Orbitrap XL and that also provides some useful tools for qualitative analysis. We evaluated the accurate mass error, the “nitrogen rule”, when needed, and the isotopic pattern; the intensity of the second isotopic signal (^{13}C signal) provides an indication of the number of carbons atoms, on the basis of the natural abundance of ^{13}C (1.11%). We also considered that the main elements generally presents in plant metabolites are limited to C, H, N, P, S. After MF identification, the process continued with the identification of the class of compounds, and eventually the specific structure, through the use of metabolite libraries and ion fragmentation patterns deriving from data dependent analysis. The fragmentation pattern permits the deduction of the molecular structures, knowing that the breakages will occur at the weakest points of the ion. The ion fragmentation, especially when MSn experiments are performed with high resolution, can be highly informative for tracking functional groups and connectivity of fragments [85]. The mass fragments

can be compared with spectral libraries or previous analyses, unambiguously identify the metabolites. Moreover, in many cases, the ion fragments produced are the same for a whole class of compounds (diagnostic ions) and permit the identification of many structurally similar metabolites. Sometimes the fragments also help in the identification of isomers, as is in the case of chlorgenic acids, which fragment masses and relative abundances allow the discrimination among 3/4/5 caffeoylquinic acids, which have the same accurate mass [141].

In this work, beside the published literature, we used online libraries as a support for molecule identification:

- ✓ Metlin (http://metlin.scripps.edu/metabo_search_alt2.php): which is a is a metabolite database for metabolomics containing over 64.000 structures, providing public access to already identified metabolite library and MS/MS metabolite data with different searching entry options (ion mass, charge, adducts, mass error, molecular formula, compound name)
- ✓ HMDB (http://www.hmdb.ca/) a freely available database containing information about small metabolites found in the human body; it contains chemical, clinical and biochemistry data and resulted very useful mainly for the identification of bioactive compounds
- ✓ Dictionary of Natural Products (http://dnp.chemnetbase.com/dictionary-search.do;jsessionid=554FD79AF324919078FB74EF887BDB1D?method=view&id=10848469&si=) which is a partially open database (part of the CRCnetbase) of 170,000 natural products, organized into a key information product, providing names, formulae, chemical structures, extensive source data, uses and applications, searchable by text or by substructure
- ✓ The LIPID MAPS Structure Database (LMSD) (http://www.lipidmaps.org/data/structure/) which is a relational database encompassing structures and annotations of biologically relevant lipids, containing over 37.500 unique lipid structures
- ✓ SciFinder (https://scifinder.cas.org/scifinder/login?TYPE=33554433&REALMOID=06-b7b15cf0-642b-1005-963a-830c809fff21&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=-SM-R2FAJCY0X8RPk21U0oIg3O9nfTRvTz%2bpLDjik%2f7wsYs5xWzHaGMXe85LgdevI0IK&TARGET=-SM-http%3a%2f%2fscifinder%2ecas%2eorg%3a443%2fscifinder%2f), used only for the coffee data analysis, is a research discovery application that provides integrated access to the world's most comprehensive and authoritative source of references, substances and reactions

We also the employed the software ChemDraw (ChemOffice 2004, PerkinElmer, Waltham, Massachusetts, U.S.A.), a molecule editor which is very useful to visualize and test the hypothetic compound fragmentation

1.3.4 Statistical analysis

After metabolomic data processing, we also performed the statistical analysis, in order to extract from the complex matrices of data the most relevant information. The statistic methods applied were basically the t-test and the principal component analysis (PCA). The PCA permits to evaluate the correlation among the variables, reducing data dimensionality and synthesizing data description. By means of a mathematical model, a series of correlated variables (metabolites), which describe the system, are transformed in new orthogonal uncorrelated variables, called principal components. The principal components are linear combinations of the original variables and contain the most relevant information of the system, permitting the removal of the noise or of the non-systematic information. The principal components are extracted in the direction along which the data have the higher dispersion, thus collecting the highest variance. When

the correlation among the original variables is higher, a much more variance is explained in the first extracted components. The linear standardized coefficient of a original variable in a principal component is called loading value and it determines the relevance of a variable in a principal component; the results of the linear combination of the original variables, where the combination coefficients are the loading values in a certain component, are called score values and represent the new coordinates of a samples in the dimensional space of principal components. In this way, it is possible to visualize the data in the new space of principal components where each sample assume the score value, with respect to the component extracted (score plot), and the variables assume the loading values (between 0 and 1) in the respectively components [142, 143].

1.4 THESIS AIMS AND RESEARCH OVERVIEW

This work aims to the development and application of multiple metabolite analysis strategies, in order to characterize plant matrices. We would like to highlight the potential of both targeted and untargeted metabolomic analysis in different field of applications: firstly, to unravel the biochemical processes and their regulation in wild and OGM plants grown in different environmental conditions; secondly, to understand the biochemical differences between plant species and/or plant varieties, in relation to their commercial use. We optimized and integrated different metabolite analysis methodologies, in order to provide the most comprehensive characterization of the considered plants, fulfilling the specific aims of each study. In the second chapter we present a study about wild and genetically modified *Nicotiana Langsdorfii* plants, exposed to different abiotic stresses. These plants are traditionally used as plant models for genetic and physiological studies [83, 144–146]; these plants were selected due to the many available information about their growth and metabolism, in order to make comparison with earlier studies and link the results to previous findings. The metabolite characterization of *N. Langsdorfii* aimed to understand the effect of the insertion of genetic modifications on plant stress response; moreover, this experiment was also aimed to better understand the effect of environmental stresses on plant metabolism. The genetic modifications used - the insertion of the RolC gene from *Agrobacterium rhizogenes* and the rat glucocorticoid receptor (GR) gene - have been previously tested, demonstrating a potential role in the enhancement of plant resistance to abiotic stresses [147–154]. The experiment includes the exposition to chemical [Cr(VI)], water and heat stresses, which are the main concerning issue in relation to changing climate conditions [155, 156]. The analysis of target relevant phytohormones and metabolites (shikimic, jasmonic, salicylic and fatty acids), already known to be involved in plant stress response [146, 151, 157–168], was integrated with a more comprehensive untargeted metabolomic analysis. The combination of these approaches resulted in a wide characterization of stressed-non stressed metabolomes and permitted the comparison of the plant stress responses among wild and genetic modified plants. In the third chapter, a study of the metabolomic composition of different *Glycyrrhiza* species is presented; the roots of *G. glabra* (var. *typica* and var. *glandulifera*) and *G. uralensis* have been analyzed in order to highlight the different biochemical properties in relation to their commercial use in licorice production. The use of licorice roots is widely employed for alimentary and pharmacological purposes but a full chemical characterization of its constituent, in relation to species and variety, have been scarcely reported [94, 169, 170]. The metabolomic analysis showed the presence of known and not already-known components of these plants and permitted the identification of the species and variety peculiarities. Finally, in the fourth chapter, the untargeted metabolomic characterization of the fruits of two *Coffea* species is reported. The different part of the fruits were analyzed separately (perisperm, endosperm, pulp) at different grades of maturation. Two analytical strategies were employed: the use of HPLC-LTQ-Orbitrap for the profiling of secondary metabolites and the GC-TOFMS approach for the characterization of the primary metabolites. The integration of these two

techniques permitted a wide study of *C. Canephora* and *C. Arabica* metabolome composition during all the fruit ripening process. Variations in the profile of both primary and secondary metabolites were highlighted and differential regulations of the main metabolic pathways involved were suggested. These differences could directly lead to the characteristics and quality of coffee produced. The study of the different part of the fruits allows the characterization of the ripening mechanisms and the identification of specific metabolites in relation to plant tissue function. The results indicate that metabolomic analysis constitute an advantageous approach for the understanding of plant biochemical processes, taking advantages also from the integration of different analytical approaches; metabolomic demonstrated to be a valuable tool in the assessment of the specificity and quality of food.

2 Integration of targeted and untargeted metabolite analyses for stress response comprehension: *Nicotiana Langsdorfii* as a plant model

2.1 INTRODUCTION

Plants have a considerable ability to deal with highly variable environmental stresses through the activation of a complex system of responses [151, 171].When in adverse or limiting growth conditions, plants respond by activating stress-specific tolerance mechanisms at different levels of organization (transcriptomic, cellular, physiological) in order to minimize damages while conserving valuable resources for growth and reproduction [172] At a molecular level, abiotic stresses lead to the induction and repression of many genes, involving a precise regulation of stress gene network [60].Most of the products of these genes may function in stress response and tolerance [173].

Drought, high temperatures and pollution are important concerns regarding the current problem of global climate change [31]. Heat stress conditions are known to determine retardation of growth, damages to cell membranes, denaturation of proteins, increased transpiration and augmented stomata opening [172, 174]. During heat stress, one of the most important observed effect is the change in lipid composition of the cellular membranes: the levels of saturated fatty acids generally raise, both in the thylakoids and in the plasma, in order to fix the increased membrane fluidity [165, 174]. Water deficiency induces the inhibition of photosynthesis, the enhancement of respiration, the lack of mineral nutrients; water stress also affects membrane lipid composition, decreasing the levels of MGDG in favor of DGDG and phospholipids [175]. Environmental contaminants, especially heavy metals, due to their widespread distribution and their persistency, represent a relevant concern for agriculture and land use [153]. Heavy metals such as cadmium (Cd) and chromium (Cr) induce enzyme inhibition, cellular oxidation and the alteration of metabolism [11]. Cr(VI) is the most toxic oxidation state of chromium. The uptake of Cr(VI) showed to influence the concentration of many metabolites as chlorophyll, carotenoids, proteins, amino acids [176].

When subjected to stresses, plants activate a unique and integrated stress response, which is regulated by the hormonal network and, in combination with many metabolites and secondary messengers, play a pleiotropic and regulatory key role [151, 172]. Among the complex hormone signaling system of plants, some molecules have been recognized as central components of the defense responses, acting as up/down regulatory factors or being precursors of many secondary metabolites. Salicylic acid (SA)and Jasmonic acid (JA) are hormones involved in plant growth and development; recent studies demonstrated that there is a complex interplay between the two, which are mainly implicated in the signaling pathway in response to biotic stress [162, 172]. SA has a key role in the systemic acquired resistance (SAR) produced by pathogen's attacks while JA leads mainly to the induced systemic resistance (ISR) which is produced when roots are colonized by nonpathogenic rhizobacteria. JA concentration has been shown to increase when plants are subjected to wounding, UV light, water deficit, pathogens and ozone [162, 177]. SA concentrations increased in plants of the *Nicotiana* genus exposed to insect attacks [146], UV light and ozone[178]. While many SA and JA responses show mutual antagonism, some genes are induced by both compounds, revealing complexities in the network of defense pathways [179].Shikimic acid (SHA) is an important intermediate metabolite in plants and a key molecule in the biosynthesis of many secondary metabolites. The Shikimic acid pathway links the carbohydrate and aromatic biosynthetic pathways producing essential aromatic amino acids, lignins, and plant defense molecules (phytoalexins and alkaloids) [180]. The shikimic

acid concentration in plant organs is not constant and depends from the synthesis rate of aromatic compounds: studies reported a higher SHA concentrations in those tissue where the metabolic processes are slow or stopped (storage tissue or seeds and fruits) or where the metabolic process rate is high (young vegetative parts, leaves) [181]. Since the phytohormones are recognized as functioning in complex signaling networks, often with interactive effects, the simultaneous analysis of several compounds is fundamental to understand the effects of stress conditions on plant growth and metabolism [101, 182]. Moreover, since metabolites are the end products of cellular functions, their presence and relative concentrations may be regarded as the best indicator of an organism's physiological state [86]. Additionally, many secondary metabolites have a prominent role in environmental stress response, as signaling molecules, protective compound and toxins. Antioxidants as flavonoids, phenolics, anthocyanins are commonly induced during both heat and chemical stresses [31, 34, 35, 183] while compatible osmolytes as amino acids, fumarate, malate, γ amino-butyrate, sugars, putrescine [155, 184] are produced during water, heat and chemical stresses, balancing the plant osmosis.

The use of genetic engineering to produce transformed stress-resistant organisms is nowadays gaining an increased interest, especially in the context of expected climate changes and growing food request [31, 167, 185]. Among the studied genetic modifications, the integration of the rat glucocorticoid receptor (GR receptor) showed to induce pleiotropic morphological and physiological changes in *Nicotiana* plants, modifying the hormonal pattern [150]; moreover, the overexpression of GR in *Nicotiana tabacum* leads to an increased production of different secondary metabolites and higher resistance to nematode infections [186]. The *rol* genes from *Agrobacterium rhizogenes* have been identified as responsible of the hairy root disease in plants and show multiple physiological and biochemical alterations in transformed plants [145, 187–190]. The integration of the rolC genes activates secondary metabolic processes [188] and enhance plant response to abiotic and biotic stresses [153, 187, 191].

The *Nicotiana* genus (family of *Solanaceae*) includes small, well characterized plants, traditionally used as biological models for genetic and physiological studies [83, 144–146, 192]; the genetic rolC and GR modifications of *N. Langsdorfii* and *N. glauca* plants have been previously investigated, demonstrating interesting results for the production of resistant plants toward different stresses [150, 151, 153].

In this chapter we present a multi-approach study of *N. Langsdorfii* plants genetically modified for the rolC and GR genes, exposed to various abiotic stresses (chromium pollution, water deficiency, high temperature); the study aims to investigate the effects of genetic modifications on plant stress response and to better understand the metabolic influence of abiotic stresses on plants, through the combination of targeted and untargeted metabolite analyses. We developed a fast and sensitive quantitative method for the simultaneous determination of three key secondary metabolites (SA, SHA, JA) involved in plant growth and metabolic regulation; to our knowledge, no other quantitative methods which determine SHA, JA and SA have been previously reported. Moreover we monitored, through a semi-quantitative method, the changes of plant fatty acids, known to be involved in heat stress response. Furthermore, to get a more comprehensive characterization of the major metabolic changes associated with the genetic modifications and the stress exposition, we performed an untargeted metabolomic analysis. Metabolomics could offer an important contribution to understand the metabolic response of biological systems, providing an overview of the plant biological status during stress exposition. The integration of different approaches permitted the monitoring of key but low-abundant molecules and of the main secondary metabolites, allowing a more global evaluation of plant responses. The plant exposed to heat stress showed a unique set of induced secondary metabolites, both known and not earlier reported for this kind of stresses, and evidenced high changes in fatty acid concentrations; the results highlighted a clear influence of GR modification to the

plant stress response, especially toward water deficiency, which could be further investigated for its application.

2.2 MATERIALS AND METHODS

2.2.1 Sample preparation

2.2.1.1 Plant growth

Plants of *Nicotiana Langsdorfii Weinmann* were cultivated in vitro in the Laboratory of Plant Genetics, Department of Evolutionary Biology of the University of Florence, under the responsibility of the Dr. Patrizia Bogani. Plant genotypes were multiplied by withdrawing portions of stems containing the internodes. Plants were grown on a specific grown medium LS (MSM0-Sigma Aldrich®, Buchs, Switzerland) containing saccharose and Plant Agar (LAB Associated, BV, The Netherlands) in proportion of 3% and 0,6 % respectively; the pH conditions were comprised between 5.6 and 5.8. Plants were grown for 30 days in a growth chamber at the temperature of 24 ± 1 °C with a photoperiod of 16 h and 80% of relative humidity. Plants were then fast cleaned with distilled water to remove the grown medium residues and freezed in liquid nitrogen; samples were then freeze-dried in an Edward freeze-drying machine and, after complete water evaporation, maintained at environmental temperature. For phytohormone, fatty acid and metabolomic analyses, samples were grinded and homogenized by using a ball mill (MM 400, Retsch, Verder Scientific, Haan, Germany), equipped with two PTFE vessel and grinder balls; samples were grinded for 5 minutes with a vibration frequency of 20Hz to achieve a final fineness of ≈ 5µm.

2.2.1.2 Genetic modifications

Wild type plants were genetically modified by insertion of two kinds of genes: the gene codifying for the Glucocorticoid receptor (GR) deriving from rat and the rol C gene from *Agrobacterium rhizogenes*. These two genetic modifications had already been used for similar studies and were available at the laboratory of the University of Florence. The procedure for obtaining these genetic modifications is well described in the study of Del Bubba et al, 2013 [153] and Fuoco et al., 2013 [151]. Briefly, the primary transgenic plants (T0) were obtained by transformation of the wild genotype of *Nicotiana Langsdorfii* using the leaf disc infection technique, as described in the Horsch protocol. For rol C transformation, the leaf discs of *N. Langsdorfii* were incubated for 25 min in a culture of *Agrobacterium tumefaciens* strain LBA4404 containing the *A. rhizogenes* rol C gene. For GR insertion the procedure was the same and the vector pT118 containing the gene coding for the rat glucocorticoid receptor was used. After 48 h the discs were transferred onto a Petri dish containing the LS medium together with some phytoregulators and carbenicillin, which is necessary to eliminate *Agrobacterium*. The stable transformants were then transferred to Agar plates and grown as previously described. T0 transgenic plants were finally moved to a greenhouse and allowed to self-pollinate at a temperature ranging from 18 to 24◦C, thus obtaining T1 seeds. These seeds were used to grow the first generation of transgenic plants, the one used for this study. The transgenic plants have been screened for the presence and expression of the transgenes through PCR and RT-PCR amplification techniques; specific primers have been employed for the evaluation of the number and the expression level of GR and Rol C genes, on DNA and RNA deriving from the leaves.

2.2.1.3 Stress inductions

The plants analyzed in this study were exposed to three different abiotic stresses:

Heat stress: the heat stress was induced on wild type, Rol C and GR plants by exposition at a temperature of 50°C for two hours into a thermostated machine (SANYO mod.MIR-153, Panasonic, Moriguchi, Osaka,

Japan). After the treatment the plants were collected and freeze-dried. Some preliminary studies were conducted on *Nicotiana Langsdorfii* wild type plants in order to identify the best temperature for the heat treatment. Temperatures of 35, 44, 46, 48 and 50°C were tested for different periods of time and the final parameters were selected on the basis of the ionic release and the recovery ability, which is the percentage of survival of treated plants after their incubation at 24°C ± 1°C for 20 days.

Water stress: the hydric stress has been simulated by addition of polyethylene glicol 6000 (PEG 6000) into the grown medium, thus reducing the osmotic potential of the LS medium. Also in this case, different proportions of PEG have been tested and the addition of 20% was selected on the basis of the growth ability of wild type plants.

Heavy metal stress: The plants of *Nicotiana Langsdorfii* were incubated for fifteen days, under the normal temperature and humidity conditions, using a LS grown medium containing a solution of $K_2Cr_2O_7$ (Merck Tritisol) for a final Cr(VI) concentration of 50 ppm. This concentration was determined on the basis of previous studies, which include the use of other chemical agents as Cadmium (Cd(II)) and different heavy metal concentrations (0, 25, 50, 75 ppm); the evaluation of the better Cd(VI) concentration was based on the lethal-dosage 50 (LD50) test, on the survival and callus formation capacity of leaf tissues of wild type *N. Langsdorfii* plants [151].

In Table 1 we report a summary of the kind of samples analyzed in this work, with the abbreviations used to define them in the text.

Table 1 Summary of the *Nicotiana Langsdorfii* samples produced and analyzed in this work with the abbreviations used in the text

	Non-stressed	Water stressed	Heat stressed	Cr(VI) stressed
Wild	WT	NL-WS	NL-HS	NL-CR
Rol C modification	ROLC	ROLC-WS	ROLC-HS	ROLC-CR
GR modification	GR	GR-WS	GR-HS	GR-CR

2.2.2 Analysis of shikimic acids, jasmonic acid and salicylic acid

An analytical method has been developed for the quantification of Shikimic acid (SHA), Salicylic acid (SA), and Jasmonic acid (JA). We analyzed globally 82 samples of *N. Langsdorfii* (GR=10, WT=10, Rol C=10, GR-WS=6, WT-WS=5, Rol C-WS=4, WT-CR=10, GR-CR=11, Rol C-CR=7, WT-HS=3, Rol C-HS=2, GR-HS=2); the number of samples for each genotype and stress depended from the plant material availability (supplied by the University of Florence) and the weight of plants. In some cases was necessary to unify more than one organism to obtain enough sample amount to perform the analyses. In the case of WT-HS samples, we analyzed only one integer plant while the other two samples were made up only by the aerial part of the organism.

In this paragraph we report the final procedure for the analysis of SHA, SA and JA while in the paragraph 2.3.1 we will better describe the different phases of method development and validation. The sample treatment procedure and the instrumental method were optimized by testing different solvents, chromatography columns and eluent compositions. Then we evaluated the instrumental signal linearity, we estimated the matrix effect, we validated the method by the estimation of accuracy, precision, recovery and the limit of detection and quantification.

2.2.2.1 Theoretical definitions of quality control parameters

The **limit of detection** (LOD) is the lower analyte concentration which can be experimentally detected; the LOD is defined as the analyte concentration which produce a signal y, equal to the blank signal y_B, added three times the standard deviation determined on the medium signal of the blank S_B.

$$\mathbf{LOD = S_B + 3\sigma_B} \qquad \mathbf{2.1}$$

where σ_B = standard deviation of the medium signal of the blank

The **limit of quantification** (LOQ) is the lower analyte concentration which can be experimentally quantified; the LOQ is defined as the analyte concentration which produce a signal y, equal to the blank signal y_B, added ten times the standard deviation determined on the medium signal of the blank S_B.

$$\mathbf{LOD = S_B + 10\sigma_B} \qquad \mathbf{2.2}$$

It is necessary to distinguish between the LOD of the procedure, which is referred to the entire analytical process and correspond to the previous description, and the instrumental limit of detection, which is referred only to the instrumental analysis. The instrumental LOD is the lower analyte concentration which the instrument can detect and is generally calculated on the basis of the signal produced by analyte standard solutions; the instrumental LOD is evaluated by considering the ratio between the analyte signal and the noise. In this case we calculated the instrumental LOD as three times the signal to noise ratio in the chromatogram multiplied for the analyte concentration:

$$\mathbf{LOD_s = A_r C_n / A_n \ x3} \qquad \mathbf{2.3}$$

where A_r = noise area

A_n = chromatographic analyte peak area

C_n = analyte concentration

[193].

The recovery indicates the analyte quantity which the method could recover in comparison to the effective quantity presents. It is calculated as the percentage between the observed quantity recovered and the "true value":

$$\mathbf{R\% = x_i / x_t \ x \ 100} \qquad \mathbf{2.4}$$

where x_i = observed analytical value

x_t = true value

The accuracy indicates the proximity between the analyte value measured to the true value and can be expressed by the relative error (E%). It is calculated as:

$$E\%=(x_i-x_t)/x_t \times 100 \qquad 2.5$$

where x_i = observed analytical value

x_t = true value

The precision describes the repeatability of measurements and it is established evaluating the deviation of the data from the medium value. It can be expressed in terms of standard deviation, variance or as the coefficient of variation (CV%):

$$CV\%=\sigma/x \times 100 \qquad 2.6$$

where σ = standard deviation of the medium value of the analyte

x =medium observed value of the analyte

[194].

2.2.2.2 Internal standard and quantification

To minimize errors during the analytical procedure, the internal standard quantification method has been employed. To all samples and blanks, a known quantity of a reference compound has been added. The response signal of the analytes has been evaluated on the basis of the signal of the internal standard, in order to balance eventual casual errors or variations of the analytical parameters. Theoretically, the use of the internal standard allows the quantification even when the analyte recovery is very low; the internal standard, which must be a stable molecule with physicochemical properties similar to the target compounds, is subjected to all the analytical steps to which are exposed the analytes and thus its behavior is comparable to the one of the analytes. The use of the internal standard also permits the reduction of the matrix effect.

In this study, we used the labeled salicylic acid, with six ^{13}C carbons in the phenyl ring, as internal standard (Figure 8).

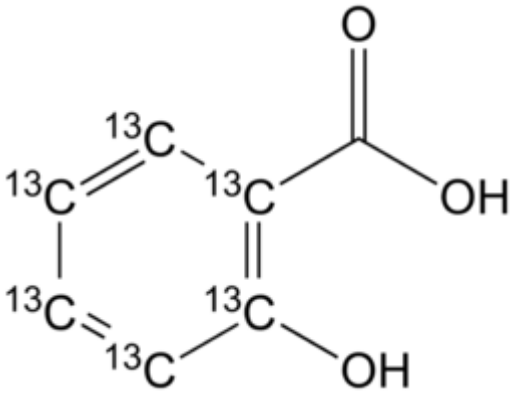

Figure 8 Structure of the salicylic acid phenyl-$^{13}C_6$

The quantification has been performed by means of an instrumental response factor (FR), a sort of synthetic sample, containing both the analytes and the internal standard; the FR allow to reduce the error deriving from instrumental signal variations.

The analyte concentration (C_i) is calculated as:

$$C_i = A_i C_s / A_s FR \quad (2.7)$$

where A_i = area of the chromatographic peak of the analyte

A_s = area of the peak of the internal standard

C_s = concentration of the internal standard

The FR considers the instrumental signal variations in relation to the variations of the analyte signal:

$$FR = A_i C_s / C_i A_s \quad (2.8)$$

2.2.2.3 Sample treatment procedure

The sample treatment procedure was developed in order to maximize the extraction efficiency and the recovery of analytes. It was however considered the big number of samples to be analyzed and therefore the procedure was kept as fast as possible. The use of a high resolution instrument permits to avoid the clean-up step, by means of solid-phase extraction (SPE), which was first tested and finally eliminated as unnecessary. Briefly, we weighted 0.1 of plant material, previously milled and homogenized, in an eppendorf tube with an analytical balance. We added to the matrix 200 µL of a water solution of internal standard (0,97 ng/ µL). We then added 1,5 mL of MeOH acidified with HCl (0,1%), we homogenized with the Vortex Agitator zx^3 (CDL, VELP Scientific Inc., New York, USA) and centrifuged the samples for 20 minutes at 14000 rpm with the centrifuge 5424 (Eppendorf AG, Hamburg, Germania). We recover the surnatant and we repeated this operation (addition of the solvent, centrifugation and recovery) for other two times. Then we evaporated the extracts unified in a thermostated bath Turbovap® II (Caliper Lifescience, Symark, Patterson, California, USA) at 30°C, under a gentle stream of nitrogen. When reached the volume of 0,5 mL, we filtrated the extracts with a syringe PTFE filter (Ø 25 mm, 0.45 µm), previously activated with 2 mL of Methanol and 1 mL of ultrapure water. Then we prepared two fractions for LC-MS analysis by diluting the extracts 1 to 20 times for SHA and SA determination and 1 to 5 times for JA. SHA and SA resulted too much concentrated in the undiluted matrix, producing an effect of signal saturation during MS analysis. JA concentration, on the contrary, was very low in the samples, but a dilution resulted necessary in order to reduce the high matrix effect and to avoid an excessive content of HCl in the sample extracts. A high HCl concentration in samples showed to influence the stability of the negative operating ion source, thus compromising the reproducibility of the instrumental signal.

Together with each set of samples, one blank sample (for a total amount of eight blanks) has been analyzed; the blanks were subjected to all processing steps, in order to evaluate the contamination deriving from sample handling and to calculate the LOD and LOQ values of the analytical procedure. For each sample batch, we also analyzed a calibration curve of SHA, in order to operate a correct quantification and to minimize instrumental signal variations. All the samples were analyzed the same day of their production, to prevent compound degradation.

2.2.2.4 Instrumental method

The mass spectrometry method was developed and optimized. The analytes were firstly detected separately, by direct infusion in the MS instrument, in order to find out the best parameters for the single substances; finally the best compromise in terms of signal intensity was selected. The final selected parameters are reported in Table A 1. The analysis was conducted in operating the source in negative

polarity, at an ionization temperature of 365 °C; the sheat gas flow was set at 41 µL/min and the auxiliary gas at 14 µL/min. The ion spray voltage was set at 4.0 kV and the capillary had a temperature of 275 °C and a voltage of -20 V. The tube lens was set at -57.97 V. The analyses were conducted in Full scan mode at the resolving power of 30000. The measurements of the accurate mass were performed with a maximum error of 5 ppm, by means of the external calibration method.

The cromatographical separation was performed on a C_{18} phase 4 µm Synergy Hydro-RP 80 Å, 50 x 2 mm (Phenomenex®, Torrance, California, USA) eluted with acetic acid 0.1% (Eluent A) and Methanol (Eluent B) at 200 µL/min. The stationary phase is a polar endcapped C_{18} phase, containing both apolar and polar groups, thus permitting the separation of both hydrophobic and hydrophilic compounds. The chromatographic run selected was composed of four steps: a first 2 minutes isocratic step (100% eluent A) for the elution of SHA, a 5 minutes gradient phase (20%/min) until reaching the 100% of eluent B, another 3 minutes of isocratic step (100% eluent B) for the elution of SA, the internal standard, and JA and a final 7 minutes step (100% eluent A) for column conditioning. The column was maintained under the controlled temperature of 30 °C over all the analysis time and the injection volume was 15 µL. The samples were stored in the refrigerated autosampler compartment at the temperature of 10°C, to prevent compound degradation.

2.2.3 Analysis of total fatty acids

2.2.3.1 Sample treatment procedure

The fatty acid analysis focused on the heat stress effects on *Nicotiana Langsdorfii* plants, in comparison to non-stressed plants. A semi-quantitative method was developed for the analysis of palmitoleic acid (C16:1), palmitic acid (C16:0), linolenic acid (C18:3), linoleic acid (C18:2), oleic acid (C18:1), stearic acid (C18:0) and arachidonic acid (C 20:4). To develop the method, we used fatty acid analytical standards, all purchased from Sigma Aldrich®(Buchs, Switzerland).

The method employed is based on the previous work of Weston et al., (2008) [195] and was developed by Martina Ranaldo of the Department of Environmental Science, Informatics and Statistics of the University of Venice, Ca'Foscari. The fatty acids presents in the samples, were trans-esterified by using methanol with 14% of BF_3 (Sigma Aldrich). The method permitted the trans-esterification of both free and lipid-contained fatty acids. Briefly, 100 mg of sample (already freeze-dried and milled as previously described) were weighted in an 7 mL amber vial and 1 mL of MeOH and 1 mL of 14% BF_3 MeOH were added. The vial was tapped and posed in a thermostated bath at 80°C for 1 hour. During all the reaction phases the vials have been frequently shaken. After cooling, 1 mL of ultrapure H_2O has been added and the solution has been shaken for 15 seconds with the Vortex agitator. We then added 1 mL of n-hexane, we homogenized the solution and finally centrifuged for three minutes at 3500 rps with the centrifuge 5424 (Eppendorf AG, Hamburg, Germania). At the end, the surnatant has been collected and analyzed by GC-MS instrument.

As internal standard the odd saturated fatty acid nonadecanoic acid (C19:0, Sigma, Sigma Aldrich®, Buchs, Switzerland), which is not naturally present in plants, has been used. It was added on the samples before extraction in concentration of 825 ng/g. The fatty acids were semi-quantified by means of external calibration curves and of an instrumental response factor (concentration of 100 pg/µL), as previously described in paragraph 2.2.2.2.

2.2.3.2 Instrumental method and quantification

The gas-chromatographic analysis have been conducted on a 5975 Agilent GC-MS equipped with a HP-5MS (60 m x 250 µm x 0.25 µm) column (Agilent technologies); the carrier gas used was helium. The chromatographic run included an initial equilibration time at 120 °C for three minutes, followed by a gradient of 10°C/min until 300 °C. The post run time was a 7 minutes isocratic at 305 °C. The injection volume was 2 µL at an injector temperature of 280 °C. The ions were identified by the Selection Ion monitoring (SIM) mode.

2.2.4 Metabolomic analysis

2.2.4.1 Sample treatment procedure

The sample treatment procedure was developed on the basis of the previous experience acquired with this matrix for quantitative analysis and with the reference protocol of De Vos et al., (2007) [105]. All the analytical phases were developed in order to assure the extraction and recovery of the highest number possible of compounds, avoiding the exclusion of potentially interesting substances. For this reason the method employed is very fast and easy, avoiding purification and concentration phases which can lead to the loss of some compounds. The principal quality criterion of metabolomic analysis is repeatability: the analytes are not quantified so, beside a simplification of methodology development, emerges the necessity to assure the comparability of samples, which are evaluated only on the basis of their relative metabolite content and intensity. For that reason, it is crucial to employ properly systems for repeatability control during all the analysis phases, from the extraction to the data processing.

We analyzed 9 different kind of samples (seeTable 1) produced as mixed plant material deriving from the single organism of the same typology; in this way, we reduced the possible biological variability associated with the single plants. We analyzed each sample in three replicates, to guarantee the repeatability of the procedure and of the instrumental analysis. Moreover, together with the samples, we analyzed three Mix samples, consisting in a pool mixed material of all the typologies of samples analyzed; these mix samples were used as repeatability controls during instrumental analysis and data processing. Finally, to verify the eventual contamination during the sample treatment, we analyzed three blank samples, subjected to all the phases of the sample processing (internal standard, extraction, filtration…). Briefly we weighted 50±0.5 mg of lyophilized plant materials (previously milled as described before) with an analytical balance in an eppendorf PTFE tube (2 ml). We added the internal standard phenyl-$^{13}C_6$ Salicylic acid in concentration of 19.4 µg/g in order to verify the reproducibility of the sample processing; we then added 1.5 mL of the extraction solution (MeOH:H_2O/75:25 acidified with Formic acid 0.1 %) and we homogenized the samples by mixing with the Vortex Agitator zx^3 (CDL, VELP Scientific Inc., New York, USA) for 10 seconds. The samples were extracted for 30 minutes in an ultrasonic bath and then centrifuged for 20 minutes at 14000 rpm with the centrifuge 5424 (Eppendorf AG, Hamburg, Germania). The surnatant was collected and filtered with PTFE syringe filters (Ø 25 mm, 0.2 µm), previously activated with 2 mL of MeOH and 1 mL of ultrapure water. Before filtering, the filters were dried under a N_2 flow. The solution obtained was analyzed by HPLC-HRMS methodology.

2.2.4.2 Instrumental method

Plant metabolome is composed by a diversity of small and less small molecules that differ greatly in their physical and chemical properties, as polarity/hydrophobicity and charge [196]. It is a huge challenge to develop a single analytical method which can provide enough chromatographic separation for all metabolites, good ionization, and enough sensibility. Whatever we choose, no single method could be ideal

for all classes of metabolites. We decided to follow the protocol of De Vos et al., (2007) [105] employing a chromatographical separation and spectrometric parameters which permitted the analysis of widest large range of molecules. The analyses were performed on a UltiMate 3000 (Dionex, Thermo Fisher Scientific, Waltham, USA) coupled to an ESI-LTQ Orbitrap XL (Thermo Scientific).The chromatographic separation was conducted on a SB-Aq Narrow Bore RR 2.1 x 150 mm, 3.5 µm (Agilent Technologies, Wilmington, USA) column which, being a C_{18} polar endcapped phase, consents the good separation of apolar and medium-polar compounds; the length of the column also favors the separation of a higher number of metabolites. The column was eluted with a mobile phase constituted by H_2O acidified with the 0.01% of formic acid (eluent A) and ACN acidified with 0.01 % of formic acid (eluent B). ACN is a favored eluent in metabolomic chromatographic separations because, with respect to methanol, it increases the retention of low polar metabolites and the peak resolution, which is an advantage in case of complex matrices analysis. The chromatographic run used in this study included a first 5 minutes isocratic phase at 100% of eluent A, followed by a 40 minutes gradient until 100% of eluent B composition; the elution continued then with another 15 minutes isocratic in organic phase and a final 15 minutes step of column conditioning at the initial eluent proportions. The column has been eluted with an eluent flow of 200 µL/min and during analysis it was maintained at a controlled temperature of 30 °C. The sample injection volume was as small as possible: 5 µL, in order to improve the peak resolution and to avoid peak enlargement due to the methanolic extraction solvent.

Upon starting a series of analyses, the chromatographic system is relatively unstable; to avoid an excessive retention time variability, before starting the sample sequence we include a few trial runs of samples, until the column resulted equilibrated with the sample matrix. Moreover, to check the technical instrument reproducibility, we analyzed the mix samples throughout the sample series, verifying the intensity of the signal.

With respect to the MS analysis, we employed a heated ESI ion source, which is known to provide better sensitivity [196] and we optimized the source and optic values; The parameters were chosen on the basis of a medium mass reference compound signal (caffeine in positive polarity m/z 524 and Taurocholic acid m/z 514 in negative polarity) and tested on real samples before analysis. The ESI source was operated in both negative and positive polarities; the capillary temperature was set at 275 °C, the vaporization temperature at 300 °C; the sheath, auxiliary and sweep gas were flow at 35, 5 and 0 µL/min respectively. The source voltage (kV) and current (µA), the capillary voltage (V) and the tube lens (V) for negative analyses were set at 3.5, 100, -40 and -100 respectively while for positive mode they were 4.5, 100, 24 and 100 respectively. The analyses were conducted in Full scan modality at the resolving power of 60000, with a mass range between 90 and 1500 m/z, in order to detect all the possible interesting metabolites. We also perform data dependent acquisitions on the mix samples, which are likely to contain all the metabolites presents in the samples; data dependent analyses are essential to obtain a complete fragmentation pattern for the molecules, which is an indispensable requirement for the metabolite identification. The LTQ-Orbitrap have the capability to perform MSn experiments by performing a MS_1 survey scan in full modality, selects among the m/z detected one or more ions for subsequent MS_2 or MS_n events, which include molecule fragmentation. We performed four different scan events, in addition to the first full scan acquisition, selecting the most abundant ion from the full scan (2^{nd} scan event) and the most, the second and the third abundant from the second scan event (3^{rd},4^{th} and 5^{th} scan events). MSn experiments were carried out at a resolution of 15000 and with normalized collision energy of 35 and an isolation width of 3. The minimum signal required for the selection and fragmentation was set on 50000, with an activation time of 30 and an activation Q of 0.25. To increase the mass measurements accuracy, until 2 ppm, we used the internal

calibration method: one or more known reference ions, presents as interferences in the Orbitrap analyzer, were employed as reference masses (Lock masses) to adjust the mass measure during the analysis thus reducing the m/z measure error.

The mass used for internal calibration are generally ions produced by solvents mixtures, interfering compounds, adducts or ubiquitous contaminants of which the exact mass is known.
The masses used for the internal calibration in this study are:

- Positive polarity: m/z 225.147000
- Negative polarity: m/z 112.985630; 226.978453; 265.147900; 281.248600; 283.264200

2.2.4.3 Data processing

The data processing includes the use of the previously described software (paragraph 1.3.2) and the statistical analysis. The parameters selected for MetAlign and MSClust processing are reported in Table A 4 and A 5.The best processing parameters were selected on the basis of the spectra characteristics and by several experimental tests in which we checked the presence of peak splitting, the incorrect removal of relevant peaks, the level of noise and the eventual misalignment. The list of masses produced as output by MetAlign was partially manually processed, in order to properly prepare it for the MSClust processing. We removed the metabolites which were not present in at least three samples (the three replicates), we randomized the noise values and we removed some detail columns which can't be further processed. The MSClust parameters were selected manually on the basis of the final number of metabolites produced, of the split peaks and of the significance of metabolites. After MSClust processing, we removed from the final mass list the compounds with the centroid factor (the correlation level to the centrotype of the cluster) under the level of 0.85. This arbitrary value can vary depending from the number of metabolites under the threshold and from their significance.

We finally performed the multivariate analysis, particularly the principal component analysis, by means of Statistica 8.0 (StatSoft, Inc., 2007); the metabolite intensities before PCA analysis were normalized using log2 transformation and standardized using range scaling. We selected, through the score plot and the loading plot, the most relevant variables to identify; we also decided to identify the most intense compounds in each group of three replicate samples, in order to give a complete metabolic characterization of each different group. We selected through the tools of Microsoft Office Excel (Version 2007) the fifty most abundant metabolites in each three replicates.

2.3 RESULTS

2.3.1 Development of a quantitative methodology for the analysis of Shikimic, Salicylic and Jasmonic acids in *Nicotiana Langsdorfii*

2.3.1.1 Instrumental methodology

A new method was developed for the quantitative measurement of SHA, SA and JA in plant samples. The analytical instrument chosen for the study was a HPLC (UHPLC Dionex Ultimate3000 lc system) coupled to a mass spectrometer ESI-LTQ Orbitrap (Thermo Fisher Scientific, Waltham, USA). This technique combines a good and fast separation with high resolution and high mass accuracy, which are essential requirements when analyzing complex matrices. The LC separation followed by ESI ionization was the most appropriate technique in this case, because of the high-medium polarity of the analytes. One of the most important issue in LC-MS quantitative analysis is the presence of interfering or co-eluting compounds, which could prevent the identification and quantification of analytes. These compounds are generally

removed with a specific sample clean-up step, by means for example of solid-phase extraction, leading to a longer and less reproducible sample processing phase. The use of high resolution instrumentation permits the identification of target molecules with high reliability, thus offering the advantage to shorten the sample treatment procedure without including a sample purification step. For the method development and validation we used high purity standard substances: Shikimic acid analytical standard purity >99% (Fluka, Sigma Aldrich®, Buchs, Switzerland), Jasmonic acid analytical standard grade (Sigma, Sigma Aldrich®, Buchs, Switzerland), Salicylic acid phenyl$^{13}C_6$ 99% (Aldrich, Sigma Aldrich®, Buchs, Switzerland), Salicylic acid purity >99% (Sigma Aldrich®, Buchs, Switzerland). First the spectrometric method was optimized by direct MS infusion of analytical standard solutions at level of 1 ng/µL. All the source parameters were optimized for all the compounds and finally the best compromise was chosen (see parameters details in Table A1). Then we developed the chromatographical separation; we tested different columns and different mobile phases. The best column for the three analytes, in terms of peak separation, peak shape and elution speed resulted the reversed C_{18} phase 4 µm Synergy Hydro-RP 80 Å, 50 x 2 mm (Phenomenex®, Torrance, California, USA). This column, with a C_{18} stationary phase containing also hydrophilic group, provides a high hydrophobic and slight polar selectivity, necessary characteristics when compounds with different polarities should be separated. Moreover this stationary phase is stable even in 100% aqueous mobile phase, which is an essential feature for the elution of highly polar compounds as SHA.

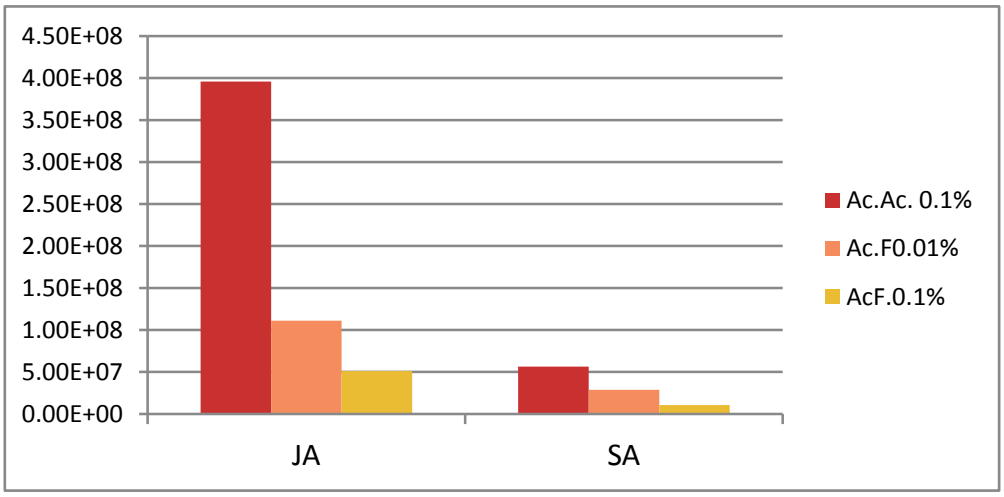

Figure 9 Relative signal intensity of JA and SA eluted with three different aqueous mobile phases: Acetic acid 0.1%, Formic acid 0.01 % and Formic acid 0.1%

The column, after testing different solvents, was eluted with an aqueous solution of acetic acid 0.1% (Eluent A) and MeOH (Eluent B). The use of an acidic modifier in the aqueous mobile phase was necessary to increase the retention of SHA, nevertheless suppressing its ionization and resulting in lower signal intensity than in neutral pH conditions. The relative intensity of analytes with the three acidic modifier tested are reported in Figure 9 and Figure 10.

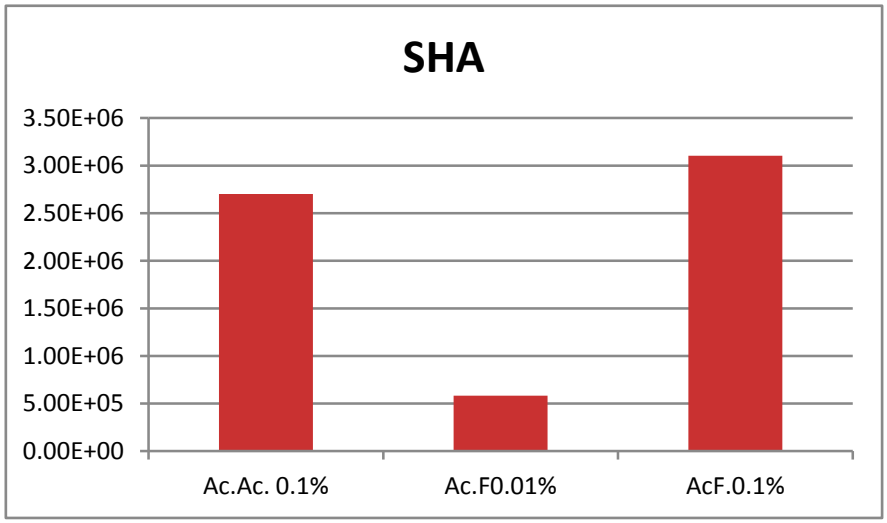

Figure 10 Relative signal intensity of SHA eluted with three different aqueous mobile phases: Acetic acid 0.1%, Formic acid 0.01 % and Formic acid 0.1%

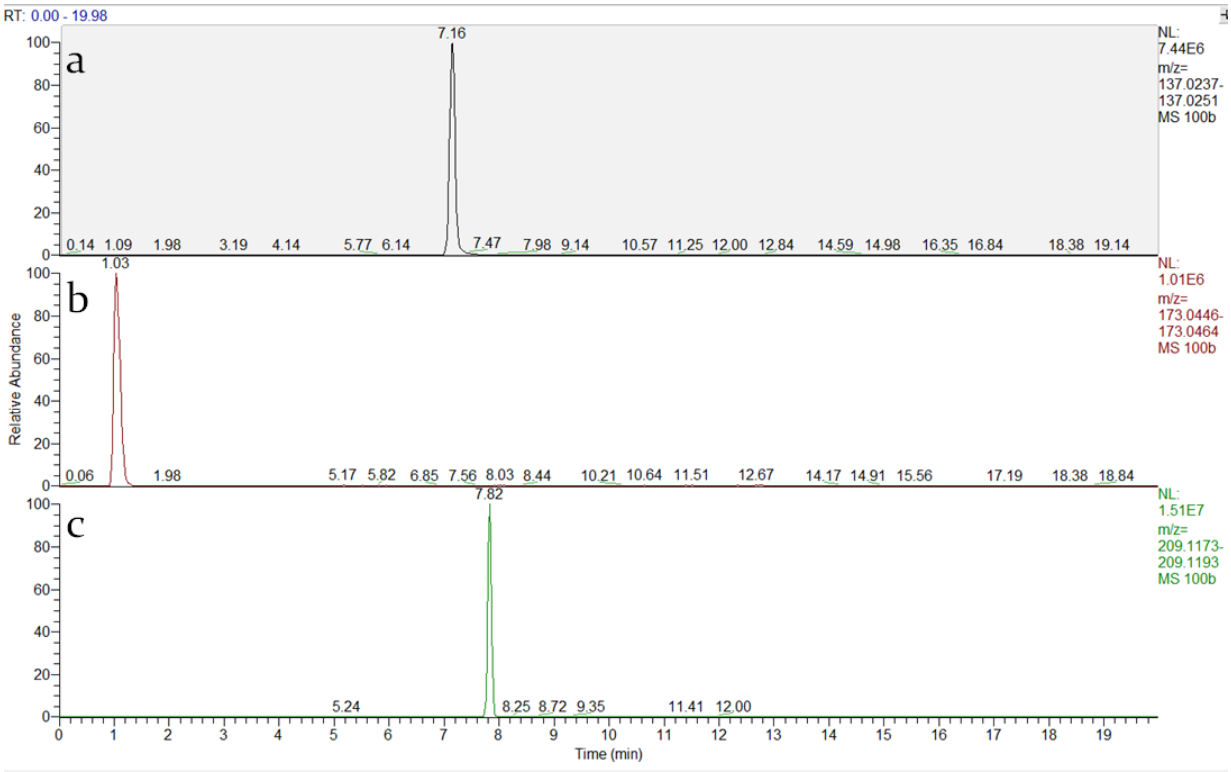

Figure 11 Chromatographic separation of SA (a), SHA (b) and JA (c).

The acetic acid addition, in comparison with the use of formic acid, resulted in a higher signal intensity for both SA and JA, however slightly decreasing SHA intensity. Therefore acetic acid was selected as the best modifier. The chromatographic separation was optimized and resulted in a two minutes isocratic at 100% of mobile phase A, in order to improve SHA peak shape, a 5 minutes gradient to 100% of mobile phase B, 3 minutes of isocratic at 100% of MeOH and finally 7 minutes of column conditioning at 100 % of aqueous mobile phase. The separation of the three compounds was achieved in 8 minutes (Figure 11).

2.3.1.2 Method validation and evaluation of matrix effect

Once developed, the method was validated by verifying instrumental linearity, method repeatability, precision and recovery. We also assess the matrix effect, defined as the effect of the components of the matrix on analyte detection; the mechanism of the matrix effect originates from the competition between an analyte and the other present constituents which react with primary ions formed in the HPLC-MS interface. This competition can influence analyte signals by suppressing or enhancing their ionization [69]. Analytical protocols generally require the validation of the method with a reference standard material which, however, is often not available, especially for biological matrices. For this reason we used the same samples analyzed for this study as a matrix for method evaluation; we make a homogenized pool of all the kind of samples and we spiked this matrix with the analytical standards. This expedient has a disadvantage: the matrix already contains a relative high amount of the analytes; it was therefore necessary to produce "blank" not spiked samples in order to evaluate the medium concentration of the three compounds in the pooled matrix, which was then subtracted to calculate the effective recoveries.

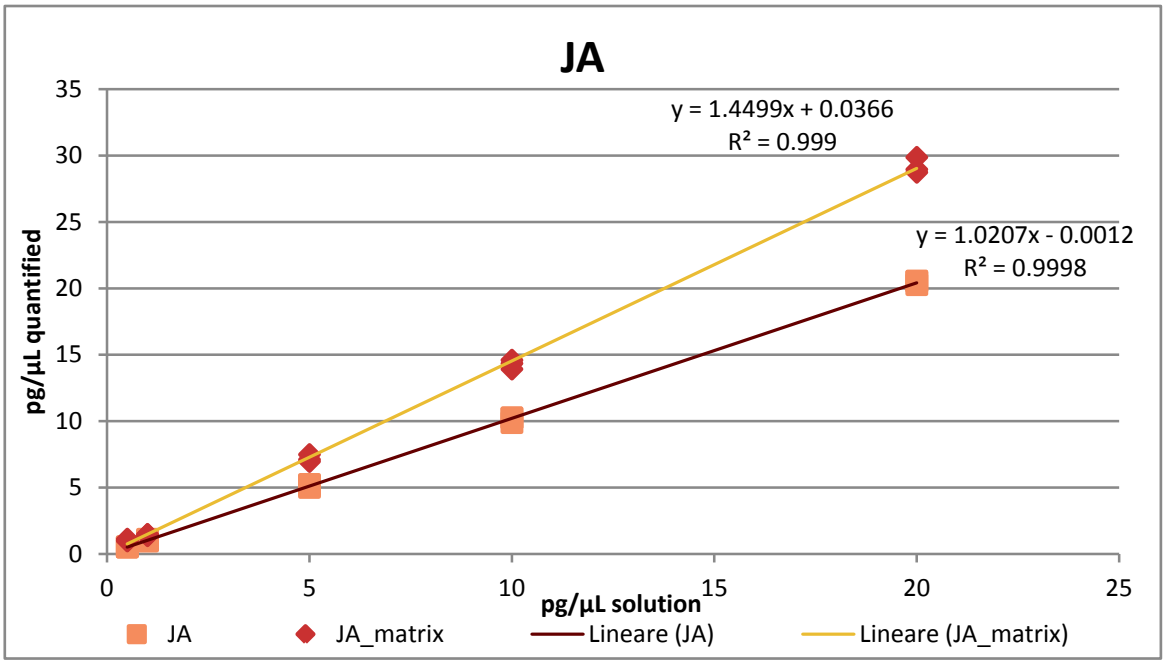

Figure 12 Calibration curves of JA prepared in water and in a matrix solution

The instrumental linearity was verified by means of calibration lines at 6 different level of concentration; the calibration curves were prepared in both pure aqueous solvent and in a diluted solution of the matrix (the same dilution used for the analysis of the samples) in order to verify the matrix effect. The results are reported in Figure 12, Figure 13 and Figure 14. Both the calibration curve in water and in the matrix solution fulfill the requirements of linearity; the linearity range cover all the analyte concentrations found in samples.

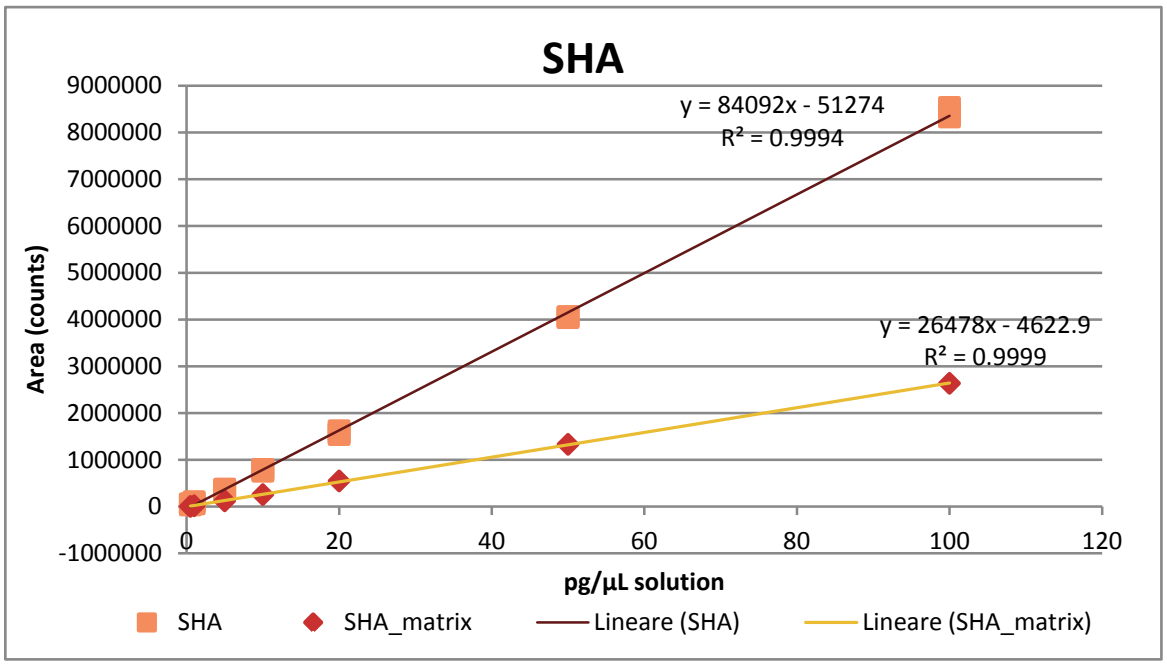

Figure 14 Calibration curves of SHA prepared in water and in a matrix solution

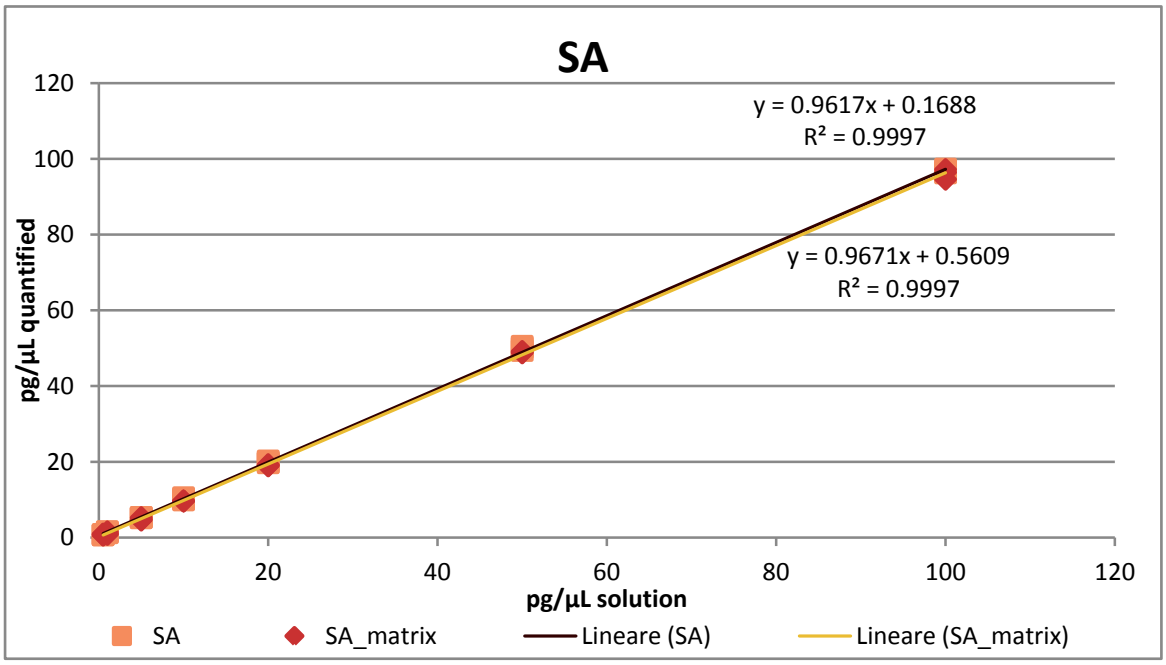

Figure 13 Calibration curves of SA prepared in water and in a matrix solution

As evident in the calibration curves, while SA instrumental response is not influenced by the presence of the matrix, SHA and JA signal intensities resulted very different in the two curves at the same concentration levels. The JA signal resulted enhanced by the matrix while the SHA signal is suppressed when analyzed in the presence of the matrix. For more accuracy we calculated the absolute matrix effect (ME) as described by [69]:

– **2.1**

where B is the calibration curve prepared in the matrix and A is the calibration curve prepared in water. The matrix effect values, calculated for each point of the calibration curves, are reported in Table 2.

Table 2 Matrix effect for SHA, SA and JA

Concentration level (pg/µL)	SHA	JA	SA
0.5	4	198	107
1	19	139	78
5	33	140	93
10	33	141	95
20	35	143	96
50	33	/	98
100	31	/	99

On the basis of these results we decide to quantify the three phytohormones by means of a factor response prepared in matrix, in order to perform a correct evaluation and avoid misleading results. The internal standard was used to quantify SA and JA while SHA was determined by means of an external calibration curve. The internal standard used for the quantification, phenyl-$^{13}C_6$ salicylic acid, resulted not appropriate for SHA quantification, because of the different physicochemical properties of the two substances which lead to different behaviors during sample treatment steps. Before selecting labeled salicylic acid as internal standard, we tested another labeled compound, Jasmonic-d_5 Acid (2,4,4-d_3; acetyl-2,2-d_2), thought to be more similar to SHA. Deuterium labeled substances are known to be less stable than carbon isotope labeled standards, because when dissolved in aprotic solvents (like methanol or water) could exchange deuterium atoms with hydrogen's. In our case Jasmonic-d_5 acid demonstrated to be unstable even for short period of time, when conserved in a concentrated solution (≈10 µg/µL in Acetonitrile for two weeks). Moreover the extraction step and the chromatographic separation include the use of Methanol, which is highly protic, and could compromise the stability of the standard compound. For these reasons we choose phenyl-$^{13}C_6$ salicylic acid as internal standard and we decide to quantify SHA by an external calibration curve.

Finally we carry out the method validation. We prepared five replicates of non spiked homogenized matrix, with the addition of internal standard before the extraction (≈ 1.94 µg/g) to assess the concentration of the analytes in the matrix used for the validation. Then, in order to evaluate the method repeatability, we prepared five replicates of homogenized matrix spiked before the extraction with both the internal standard (1.94 µg/g) and the standard solutions of the analytes (SA 50 µg/g and JA 500ng/g). To verify the recovery of the method we added in five independent replicate of the matrix the standard solution of analytes, in the same concentration as before; at the end of the procedure, directly in the vial for HPLC analysis, we added the internal standard. To validate the method for the SHA, which couldn't be quantify by means of the internal standard, we prepared three replicates of the matrix spiked with a SHA standard solution before extraction at four different concentrations levels (1.25 µg/g-2.5 µg/g, 5µg/g-7.5 µg/g); we

then calculated the recovery and accuracy of the method by means of an external calibration curves prepared in a matrix solution. The repeatability and recovery were calculated as previously explained (see paragraph 2.2.2.1). In Table 3 are reported the quality parameters obtained from the method validation.

Table 3 Validation results of SHA, JA and SA

	SHA	SA	JA
E% (repeatibility)	10	17	-16
Recovery%	99	71	58
CV% (Precision)	11	11	-18
LOD (ng ass)	54	19,4	0,2
LOQ (ng ass)	179	64,7	0,8
LODs (pg/µL)	41	7,5	0,5

Considering that biological matrices are generally characterized by a high variation between the samples and that for SA and JA we use an internal standard for quantification, we reputed the method validated for all the three compounds. The SHA, which showed high recovery and good repeatability, was not enough repeatable at the lower level of concentration of the calibration curve (1.25 µg/g); this is probably due to the variation of the concentration of the analyte in the matrix which, at low added standard concentrations, increase the error in the quantification. However sample never shown so low SHA concentrations. We also calculated the LOD, LOQ and LODs for the method.

2.3.2 Analysis of Shikimic, Jasmonic and Salicylic acids in *Nicotiana Langsdorfii*

For wild plants and for GR and Rol C plant sets, ten control samples were analyzed. SA showed the highest concentration in all the three groups ($1{,}1x10^3$-$8\ x10^3$ ng/g d.w.), followed by SHA ($3{,}9\ x10^3$ -$3{,}4\ x10^3$ ng/g d.w.) and JA ($7\ x10^1$-$2\ x10^1$ ng/g d.w.); JA showed the highest internal set variability, especially for GR plants, with RSD% 34-66. RSD% for SHA was 15-23 and for SA 32-35. The complete table of results for the single plants is reported in Appendix (Table A 1). As expected JA concentrations were extremely lower than SA and SHA concentrations, as previously reported for *Nicotiana tabacum* plants [146, 182]. SA concentrations for GR modified plants were consistent with what already reported by Fuoco et al. (2013) [151] but, in the wild control samples, we observed higher concentrations; on the contrary SHA levels were slightly lower in this study but globally in agreement with our findings. The SA and SHA profile resulted similar in the wild and transgenic controls, without showing statistically significant differences; only JA demonstrated significant low concentrations in Rol C samples, with an average fold of 0.45 in comparison to both WT and GR samples (Figure 15).

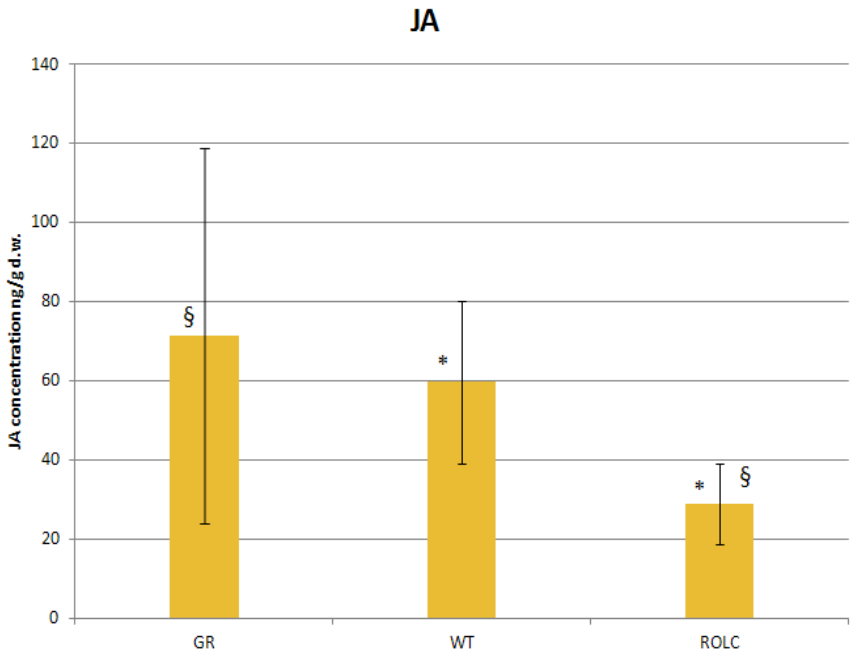

Figure 15 JA mean concentrations and standard deviations (error bar) in WT, GR and ROL C controls. The same symbol indicates a statistical significant difference between the two distributions, at a p value<0.05, according to the t-test

A few changes in plant responses were observed when plants were exposed to abiotic stresses (Figure 16 a,b,c). The SA levels resulted almost unchanged between the controls and both WS and HS samples; only WT plants, exposed to chromium stress, showed a statistically significant growth of 1.6 fold in average. With respect to SHA and JA concentrations, we observed significant increases in both WT and Rol C plants exposed to chemical stress. The highest increases were measured for Rol C samples, with a mean ratio between CR and the control samples of 1.9 for both SHA and JA. Interestingly, changes in phytohormone levels in GR plants never resulted significant, being in all cases statistically comparable with controls. The changes in SHA concentrations in CR stressed plants are consistent with what previously reported by Fuoco et al. (2013) [151] for both WT and GR plants; in this study some changes in SA levels were also observed: as in our work, an increase of SA in chromium stressed WT plants is reported, while a reduction of SA in GR-CR plants, not visible in our samples, is also noticed.

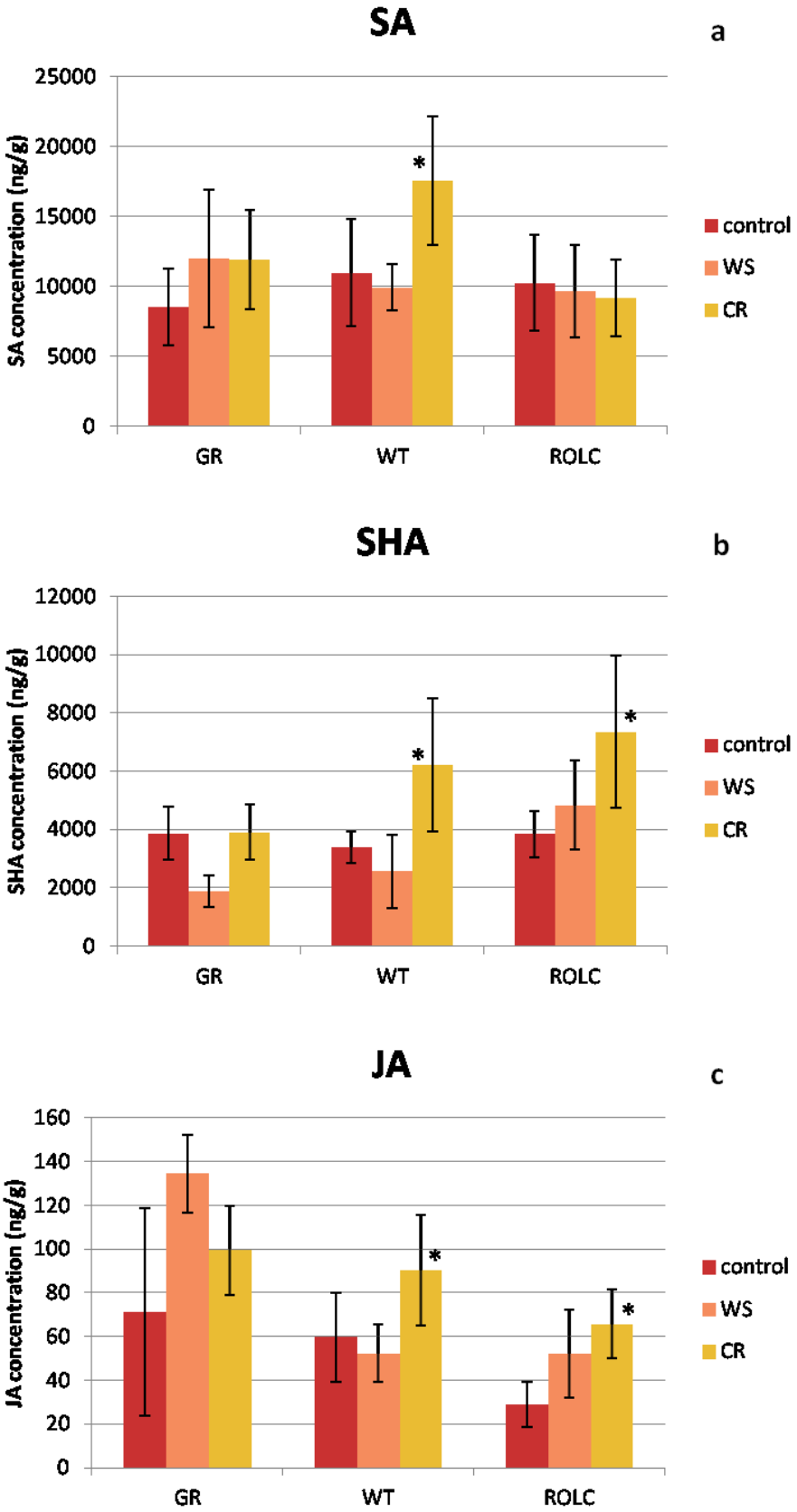

Figure 16 Mean phytohormone concentrations in controls, WS and CR samples in wild, GR, and ROLC modified genotypes. Error bars shows the standard deviation for each set of samples. Concentrations marked with an asterisk resulted statistically different at $p<0.05$ according to the t-test

With respect to the heat stressed samples (Figure 17 a,b,c), small but significant phytohormone changes were observed for SA and SHA, with respect to controls; the concentrations of these two compounds resulted increased of 1.3 fold in all HS samples, in comparison to non-stressed plants. An exception is represented by WTHS plants, in which the concentration of SHA decreased of almost 0.4 fold. Interestingly, only RolCHS samples showed a SHA concentration comparable to the controls. For what concern JA concentrations, we observed a not significant increase in GRHS samples, mainly due to the high internal variability of GR control subset. On the contrary JA concentrations in WTHS and RolCHS showed the highest increases, with ratio values of 2.1 and 6.8 fold respectively. No significant difference was observed in the phytohormone content between the aerial part and the total plant (WTHS samples), that were summed together.

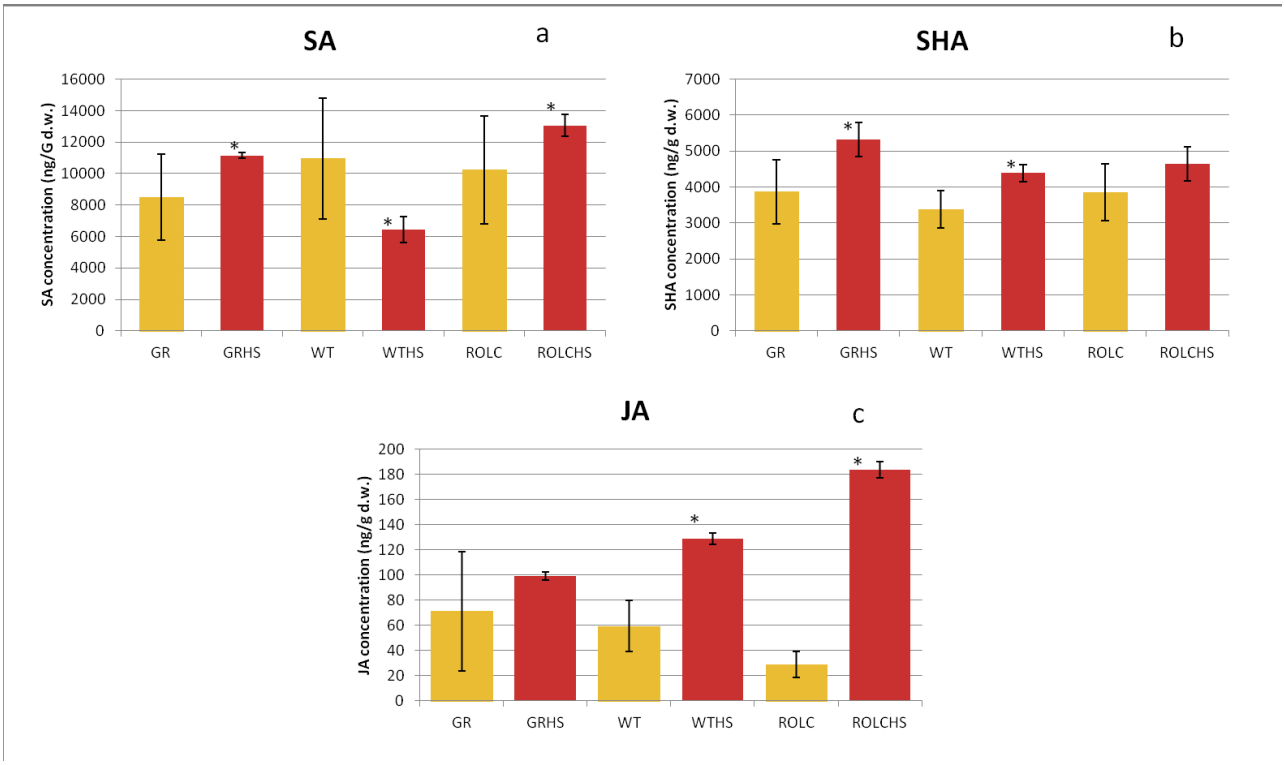

Figure 17 SHA, JA and SA concentrations in controls and HS samples for GR, WT and Rol C modified plants. Error bars represent the standard deviations. HS samples for which the increase, with respect to controls, is statistically significant (t-test at p value<0.05), are marked with an asterisk.

2.3.3 Analysis of fatty acids in *Nicotiana Langsdorfii*

The analysis of fatty acids was conducted on three independent replicates for each set of heat stressed and control samples. The complete concentrations of fatty acids are reported in Appendix in Table A 3. The measured concentrations ranged between 0.31 and 27701 ng/g. The most abundant compound in all kind of samples resulted the linolenic acid (18:3), which in average accounts for the 48% of the total content of fatty acids, with concentrations between 1015 and 27701 ng/g. Linolenic acid is followed by linoleic acid (18:2) and palmitic acid (16:0) with concentrations of 647-11656 ng/g and 612-8074 ng/g respectively. The monounsaturated palmitoleic and oleic acids and the arachidonic acid (20:4) were present in low concentrations, accounting in average for less than the 1% of the total content. Quantitative analyses of fatty acids in *Nicotiana* genus have been scarcely reported and the observations are generally limited to the relative abundances of compounds. In *Nicotiana* species the linolenic acid (18:3) generally represents more than the 50% of the total content of fatty acids in leaves; in agreement with our findings, the other two highly relevant compounds are the linoleic acid and the palmitic acid, accounting for more than the 15 % [197, 198].

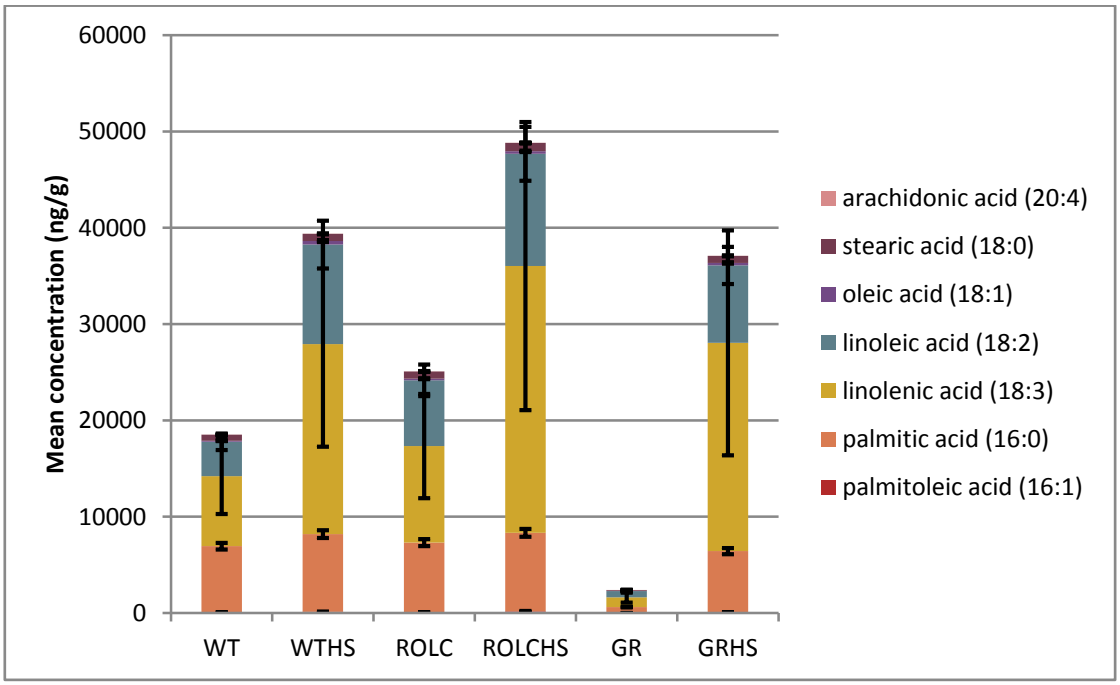

Figure 18 Concentrations of fatty acids in Nicotiana Langsdorfii samples. Error bars represents the mean RSD %.

The application of GR and Rol C genetic modifications demonstrates to influence more the fatty acid content than the saturation rate. As shown in Figure 18, the GR controls showed fatty acid concentrations extremely lower than the other samples, with a mean total content being the 13% of WT plants. On the contrary, the Rol C samples showed a slight enhancement (1.35 fold the WT controls) in fatty acid concentrations, especially of linolenic and linoleic acids; these changes strongly indicate an effect of GR and Rol C modifications on the fatty acid biosynthesis pathway and particularly on the acetyl-CoA metabolism. Despite the differences in fatty acid content in control samples, the concentrations in heat stressed samples resulted comparables; the concentrations of all compounds increased after stress exposition. The ratio values between HS samples and controls for WT and Rol C plants are similar, ranging from 0.8 to 6.8, with the highest increases for linolenic acid in both cases. The GR ratios resulted dramatically higher, with values included between 1.3 and 21. This result is in agreement with what reported in previous studies [174, 199]. Only arachidonic acid, which is highly unsaturated, slightly decreases in WT and Rol C heat

stressed samples, with ratios of 0.78 and 0.92 respectively. Generally, the fatty acid saturation degree is known to increase during heat stress conditions, as a defense mechanism to face protein denaturation and increased membrane fluidity, which could lead to plant death [60, 155, 165, 174, 200, 201]. In our case, however, we observed an increase in all fatty acids in stressed samples, especially in the level of linolenic acid. The huge increase of fatty acids in GR stressed samples clearly indicates an activation of fatty acid synthesis pathway in response to heat stress, in comparison to the pathway suppression in the controls. To explore the eventual influence of genetic modifications and heat stress on the inter-conversion between saturated and unsaturated compounds, we calculated the ratios between the unsaturated fatty acids and the respective saturated acids.(Figure 19).The levels of all unsaturated compounds in average increase in HS samples, with respect to controls; the highest increases were recorded for linolenic acid, which shifted from 10 to 29 fold in average for all kind of HS samples. On the contrary, the relation between the monounsaturated (palmitoleic and oleic acids), the linoleic acid (18:2) and the relative saturated remains similar in both Rol C and GR samples; higher changes were measured for oleic and linoleic acids between WT and WTHS samples. The content of monounsaturated acids resulted significantly lower than the one of saturated in all kind of samples, with mean levels between 0.008 and 0.49 fold. The increased levels of linolenic acid in HS samples may be related to an increase of JA synthesis, which derive from fatty acid metabolism.

With respect to the controls, the mean ratios of linolenic acids appear similar for WT, Rol C and GR samples, while linoleic and oleic acid levels resulted slightly reduced in WT, in comparison to GR and Rol C.

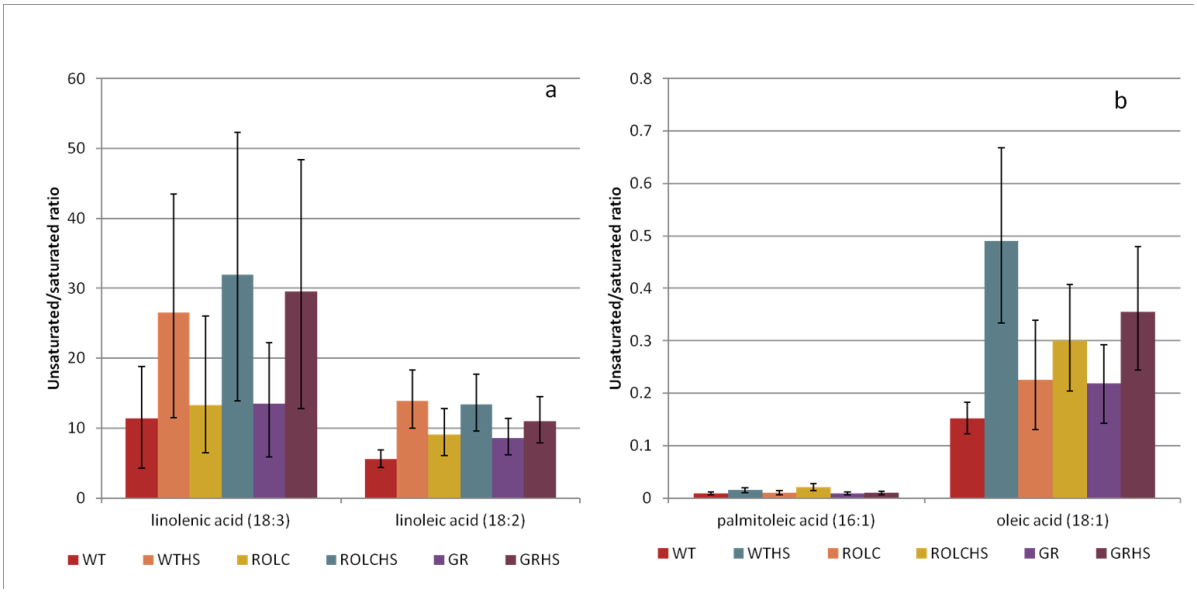

Figure 19 Ratios between unsaturated and the correspondent saturated fatty acids in *Nicotiana Langsdorfii* samples. Error bars represent the range of possible ratio values achievable according to standard deviations.

2.3.4 Metabolomic profile of *Nicotiana Langsdorfii* plants

The metabolomic analysis of *Nicotiana* samples was performed in order to obtain a complete profiling of the metabolites affected by stress exposition, in addition to the already quantified compounds. The chromatograms obtained from the HPLC-HRMS analysis were processed by means of MetAlign, for unbiased mass peak picking and baseline correction; the resulting peaks (more than 12000) were further handled with MSClust, grouping the mass signals in more than 700 putative metabolites. Finally, the most relevant metabolites for each sample typology were selected by means of statistical analysis, and 227 compounds were putatively annotated. The complete list of identified molecules is reported in Table A 6 and Table A 7. The principal component analysis was carried out, considering all the detected metabolites, and the resulting score plot between the first and second factors is presented in Figure 20. In Table 4 is reported the variance explained by the first three factors.

In the center of the score plot are visible the Mix samples, prepared as a pool of all the sample typologies, assuring the repeatability and correct data processing. The PCA highlighted a sample clustering mainly on the basis of the stress applied, as an indication that, despite the genetic modifications applied, the metabolic profile of plants was influenced in similar manner by the same stress. The heat stress samples resulted well separated from all the other samples, suggesting a significant metabolic change associated with this kind of stress. The plant not exposed to any kind of stress also clustered well, being separated from HS samples by the first component. The chemically stressed samples showed a higher internal variability and lied almost in the center of the plot. Interestingly, GR-WS and Rol C-WS and WT-WS samples presented high distance in both first and second factors (Figure 20), indicating clear differences in their metabolomic profile. GRWS samples seemed to be more similar to the non stressed or Cr-stressed samples, suggesting an influence of the GR genetic modification on the plant response to water stress.

With respect to the plants exposed to heat stress, the loading plot permitted the identification of the more relevant metabolites associated with this kind of stress. We considered all the variables of the corresponding loading plot with high positive coefficients (>0,5) for both the first and the second factors; this statistical selection permitted the identification of a few classes of metabolites, relevant for the heat stressed samples. Moreover, the metabolites with statistical significant differences were selected by the t-test with a p value<0.001.

Table 4 Variance explained in the first three factors of PCA including all the metabolites detected in *Nicotiana*

Factors	Variance explained (%)	Cumulative variance (%)
1	34,47	34,47
2	18,10	52,57
3	10,83	63,40

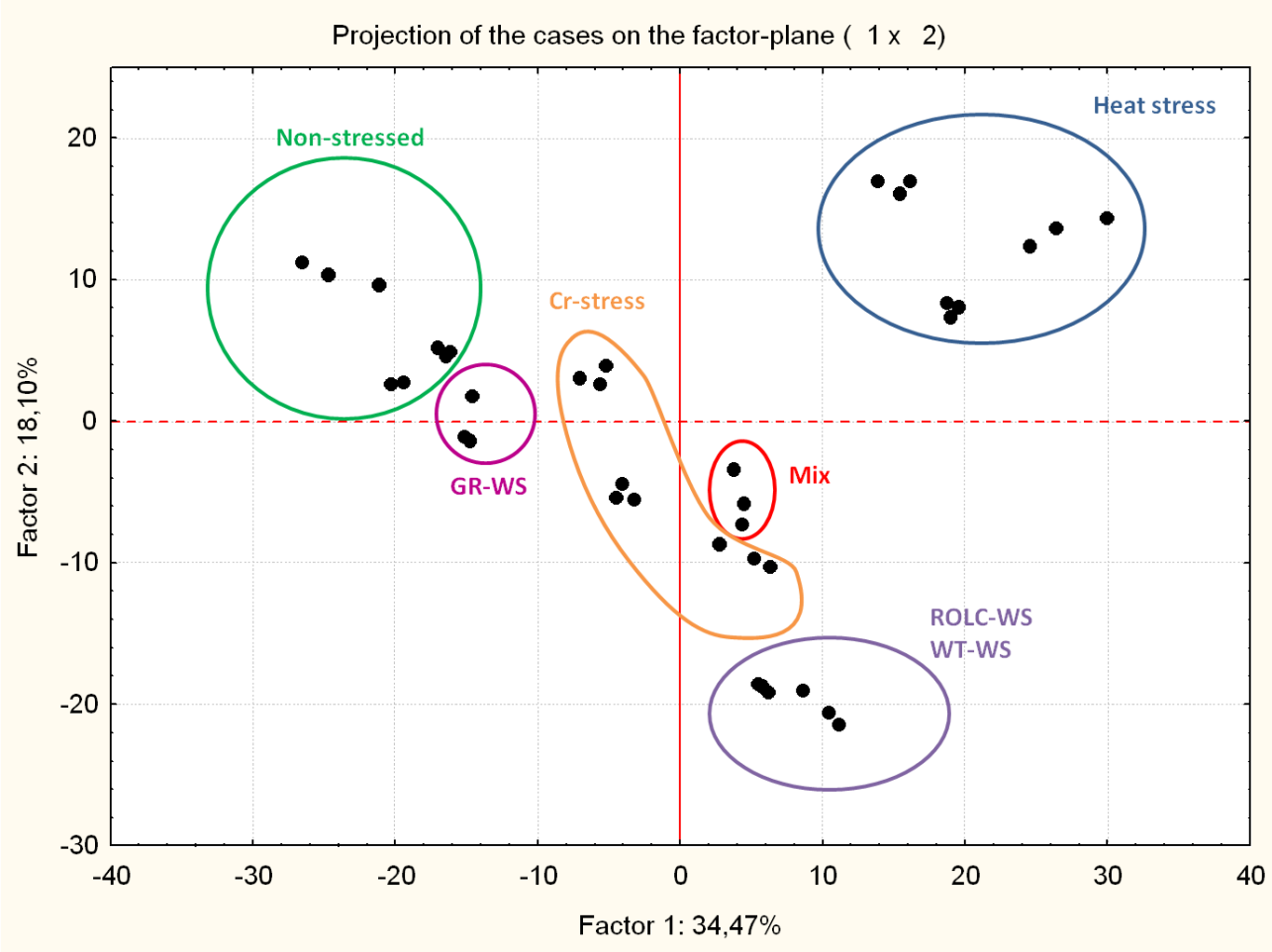

Figure 20 Score plot of principal component analysis of all the metabolites separated in *Nicotiana Langsdorfii* samples

2.3.5 Heat stress

The most important class of metabolites, which is already known to be influenced by heat stress, is the lipids. After long term exposure to heat stress (e.g. 1 month, 35 °C), the lipid saturation grade is known to increase, especially in the leaves; this response raises the temperature at which lipids separate into monolayer structures, avoiding the disruption of membrane organization and function [165, 174]. Our heat experiment consisted in a short-term exposure (2 hours) at very high temperature (50°C). Despite the relatively short time during which plants sustained the stress conditions, some changes took place in the metabolism. These changes can be seen as the first plant system response against temperature stress. Among the lipid classes, the typologies which showed the major changes were the sulpholipids (SQMG/SQDG), the MGDG/DGDG and the phosphatidylglycerol (PG), which are also the main constituents of the thylakoid membrane [21]. Sulpholipids were detected only in negative polarity while PG and PE only in positive polarity; MGDG and DGDG were identified in both source operating modality, as reported in Figure 42 and Figure 43.The levels of SQMG/SQDG and MGDG/DGDG showed the highest increases in WTHS and RolC-HS plants, with respect to the controls. The photosynthetic system stabilization function of these compounds during heat stress have been suggested by many studies [202–205]. Sulpholipids are known to accumulate during high/low temperature stress and water deficit conditions [206]. Our results are in agreement with what reported by Chen et al. (2006) [205] in *Arabidopsis* plants, excepting for the content of MGDG. Therefore the ratio between MGDG and DGDG is thought to decrease during heat stress, due to the smaller head group of MGDG and their less stable structure [207]. However a previous study [164] also demonstrated an increase of both MGDG and DGDG in *Atriplex lentiformis* exposed to high

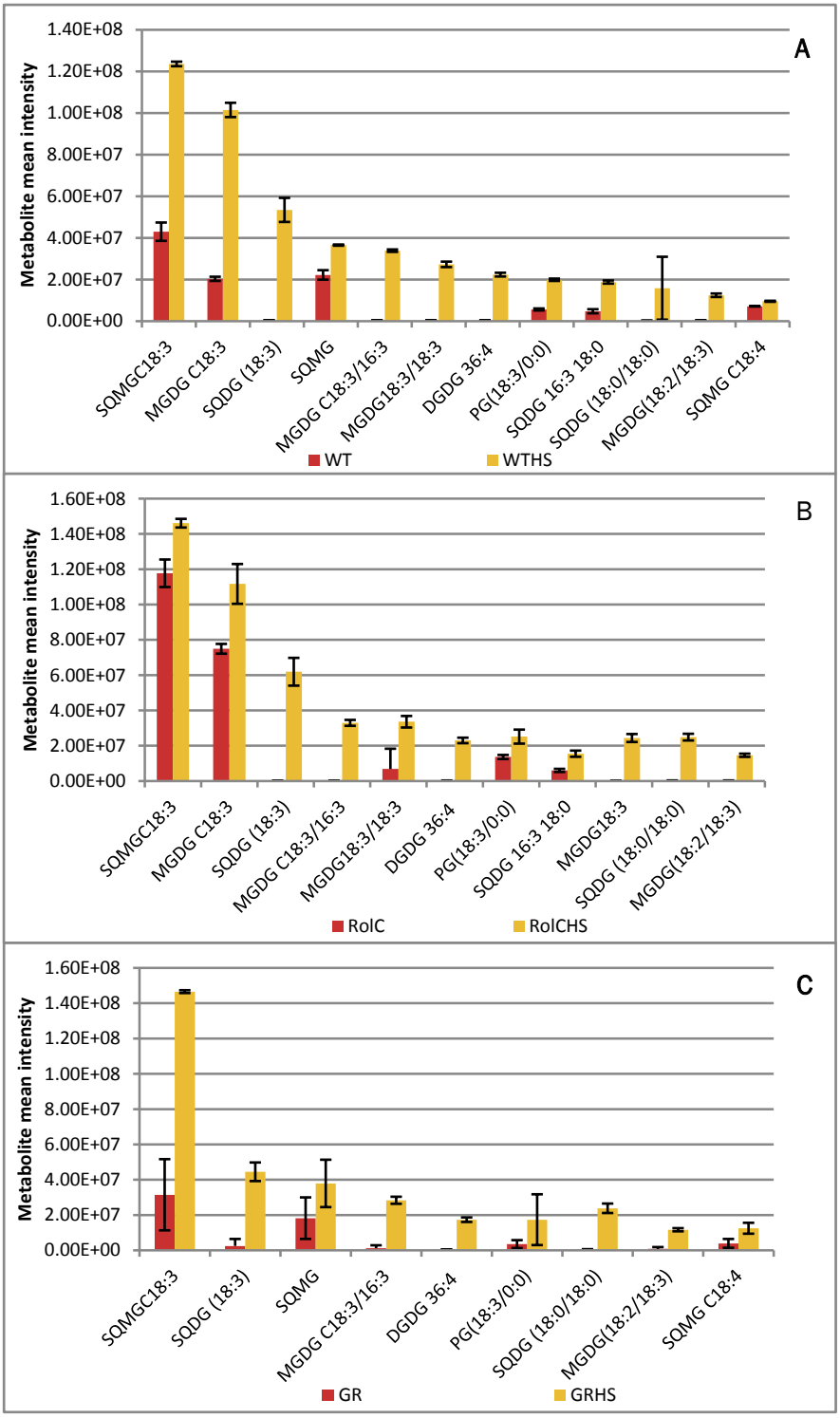

Figure 21 Levels of the principal SQMG, SQDG, MGDG, DGDG, identified in negative polarity, showing the highest increases in WTHS (A), RolC-HS (B), GRHS (C) samples with respect to the controls.

temperature stress. Another lipid which was found to increase in heat stressed samples was PE, which is one of the main constituents of the mitochondrial membrane, where the respiration is carried out. While the photosynthesis rate during high temperature stress is known to immediately declines, the respiration showed to have a different behavior: in the first phases respiration rate rapidly increases with temperature while finally drastically decrease, when an extreme tolerable temperature is overtaken [155, 201]. The observed increase in PE (Figure 21) with respect to non-stressed plants could therefore be associated to an enhanced respiration activity, as previously suggested by Otsuru et al., (2013)[208]. The ratio of lipids in heat stressed plants, with respect to the controls, vary from a minimum of 1 to a maximum of 54 in GR plants while in Rol C they range from 1 to 245 and in WT from 1 to 103. As already outlined, the GR genetic modification seem to influence the plant metabolism, conferring a higher resistance against various kind of stresses. Also in this case, the GR plants seem to

have a different behavior from both the WT and RoI C; the increases of lipids in GRHS samples resulted statistically different from WTHS plants for 15 compounds, in comparison to 7 statistically significant changes between RoIC-HS and WTHS. The lower increases in lipids, observed in GR plants, could signify a lower damage to the photosynthetic system and chloroplast structure; however more data are necessary to assess this effective difference. Another interesting finding is that the main part of the detected lipids has a high unsaturation grade, which is in contrast with what well known about heat stress effects. In *Nicotiana* species the linolenic acid (18:3) generally represent more than the 50% of the total content of fatty acids in leaves [197, 198], so it's not surprising that it is present in almost all lipids that we detected. We however expected to find a higher proportion of saturated or monounsaturated lipids which, also considering the fatty acid results reported in the paragraph 2.3.3, probably are produced during longer heat stress exposition, as reported in the major part of studies [174]. In the positive polarity mode (Figure 22) were also identified three acylglycerols which are minor cell components widely distributed in all plant tissues; these compounds resulted increased of 20-60 folds in HS samples but, differently from the other lipids, they presented the lower increases in the RoI C plants. These components generally are not very abundant in cells because of their detergent properties which can determine the cell membrane disruption. Mono-

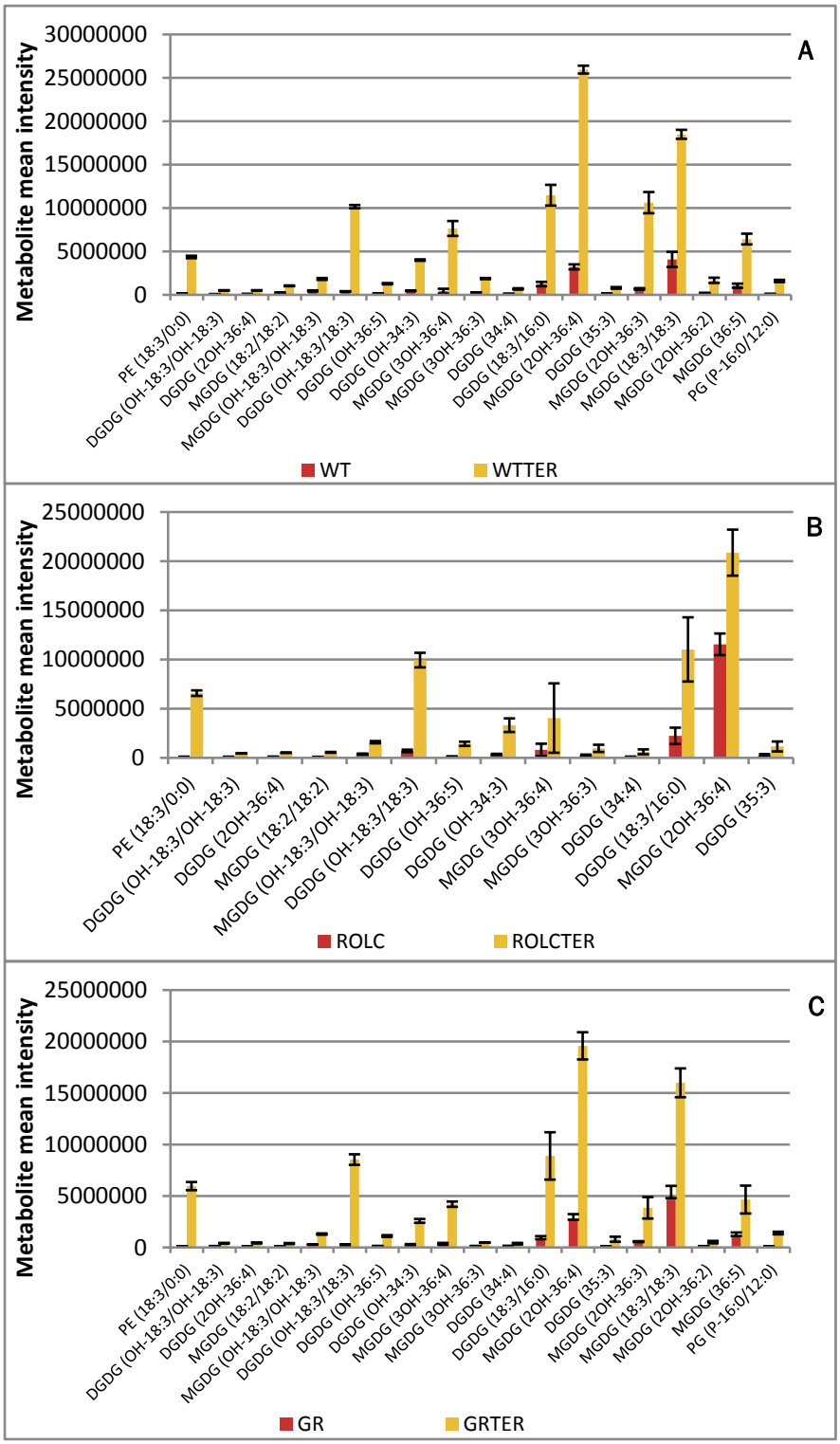

Figure 22 Levels of the principal MGDG, DGDG, PG and PE, identified in positive polarity, showing the highest increases in WTHS (A), RoIC-HS (B), GRHS (C) samples with respect to the controls.

and diacylglycerols are the precursors of MGDG/DGDG and SQDG in chloroplasts and of the PE production in the endoplasmatic reticulum [21]. These compounds could also be intermediates in the degradation of triacylglycerols. Their presence and function has not been related yet to any plant response against stress.

Another class of compounds which showed relevant changes in heat stressed samples was the acylsugar group. The identification of acylsugars was performed, as for all other metabolites, on the basis of the fragmentation pattern; in negative polarity all the compounds ionized as adducts with formic acid and the most part of compounds showed a diagnostic fragment at m/z 565.28375, in agreement with previous studies [209, 210]; in positive mode acylsugars ionized as adducts with Na^+ and their mass spectra showed a typical neutral loss of 162 Da, corresponding to a fructose, or a loss of 204 Da representing a fructose esterified with an acetate group (204 Da) [211].The acylsugars are non-volatile metabolites that constitute a significant proportion of leaf biomass in some Solanaceous species, as *Nicotiana*, known to be involved in pathogens and herbivory defense [209]. Acylsugars may be directly toxic to herbivores, but also operate as emulsifiers and surfactants leading to immobilization or suffocation of arthropods [212]. Acylsugars are mainly produced and accumulated in plants trichomes, thin outgrowing structures from leaves and stems, in which many specialized secondary metabolites and proteins are produced and stored, especially with defense functions [210, 213]. The increase in trichome density has already been positive correlated with temperature increases and water availability decrease, probably because trichomes favor sunlight radiation reflection, heating dissipation and water transpiration [214]; however, to our known, no studies investigated the possible relation between acylsugars and heat stress response yet. Our results suggest that these compounds, beside their role against biotic stress, could also take part in the response to abiotic stresses, especially heating or drought. We detected 20 acylsucroses, 17 of which demonstrated to significantly increase in the HS samples, with respect to the controls. The main part of these compounds was tetra-acylsucroses (detected in both positive and negative polarities), 7 tri-acylsucroses and two penta-acylsucroses. The higher increases during heat stress were detected for tetra-acylsucrose S4:20 (were 4 is the number of acyl groups and 20 the number of carbons in them) and penta-acylsucrose S5:25. The acylsugar ratios, with respect to the controls, resulted 20, 35 and 16 for penta-acylsucrose (S5:25) and 19, 6, 12 for tetra-acylsucrose (S4:20) in GRHS, RolCHS and WTHS respectively. The other compounds showed similar increases in the heat stressed samples of GR, Rol C and WT, excepting for tetra-acylsucrose (S4:17), tetra-acylsucrose (S4:18) and tetra-acylsucrose (S4:19) which demonstrated highest increases in GRHS samples.

Another class of relevant compounds identified in the heat stressed samples were the steroidal glycoalkaloid group of solasodine and tomatinedol; these compounds and their glycosides are known constituents of *Solanaceae* plants, studied for their presence and toxicity in potatoes [215, 216] and for their insecticidal potential [217, 218]. We identified in positive polarity 11 solamarines, solamargines and solasodines or their derivatives; the intensities of these compounds in WT, GR and RolC controls resulted comparable. The difference in the content of glykoalkaloids in control and HS samples resulted significant at a p value<0.05 for WT, Rol C and GR samples. δ-solamarine and δ-solamarine deoxyhexose showed the highest increases in WTHS and RolC-HS samples, with a ratio of 8 and 6 respectively, while solamargine presented the highest increase in GRHS samples, being 9 times higher than in controls. A study of Coria et al., (1998) [219] showed that the increase of the total glycoalkaloid content and of α-solanidine in potato tuber could be related to high temperature exposition (immersion in hot water at 35°C for 4 hours). Changes in glykoalkaloid levels in potatoes have also been related to wounding treatment and light exposition [220–222]. Glykoalkaloids are also known to be active against pathogen infestations [215, 217].

Here we suggest that the level of glykoalkaloids in *Nicotiana Langsdorfii* could be affected by stress conditions, especially heat stress..

2.3.6 Chemical and water stress

With respect to chemical stress, also in this case lipids and acylsucroses resulted included among the metabolites showing the highest increases, with respect to the controls. The tetra-acylsucroses (S4:19,S4:24, S4:26), detected in positive polarity, presented significant increases in all WTCR, GRCR and RolC-CR samples, with levels ranging from 9 to 136 fold the content of controls.

The WTHS samples proved the highest number of metabolites presenting significant changes (Figure 24); among these, a high number of structurally different lipids as sphingolipids, fatty acids, phospholipids and SQMG and lysoDGDG are observed. The presence of hydroxylated fatty acids, namely trihydroxyoctadecenoic acid and hydroxylinoleic acid was noticed. The increases of these compounds could be related to the production of ROS molecules, which is a well known consequence of heavy metal stress; the ROS, which are signaling compounds involved in stress response, generally determine the peroxidation of lipids generating the oxylipins as, for instance, the phytohormone jasmonic acid [223]. With respect to the other lipids, differently from heat stress experiments, an increase in some lysolipids is observed. Enhanced levels of lysophospholipids were previously observed in consequence to herbivore-stress and wounding [102, 224, 225]. This effect was explained as induction of the release of polyunsaturated fatty acids, which are known precursors of oxylipins. This suggestion is consistent also with the observed fatty acid hydroxylation in WTCR samples [226]. Interestingly, the RolC plants showed an opposite trend of fatty acid and hydroxyl fatty acid levels, with decreased intensities in the Chromium stressed samples with respect to the controls. The GR-CR plants don't showed significant changes neither in the lysolipid levels nor in the fatty acid content. These findings strongly indicate an effect of the RolC and GR genetic modifications on the chemical stress plant response. Besides lipids and acylsugars, two other compound classes appear to be relevant in plant response to chemical stress: amino acid derivatives, hydroxycinnamic acids and their polyamine conjugates. These classes of compounds are known to be involved in heavy metal stress response [11]. The 4-caffeoylquinic acid showed high induction in all CR stressed samples; interestingly, the other isomers do not present a similar behavior. Del Bubba et al., (2013) [153] reported similar changes in the shoots of *Nicotiana Langsdorfii* exposed to Cr(VI), while revealing increased levels also for 3-CQA and 5-CQA. In addition to 4-CQA several HCA-amino acid and HCA-amine conjugates resulted increases in CR samples, as feruloyl-aspartic acid, dicaffeoyl spermidine, glucopyranosyl-caffeoyl putrescine, bis-dihydrocaffeoyl spermine. Amino acids, especially proline, alanine, serine are known to accumulate in plants exposed to salinity stress as compatible solutes, in order to maintain the osmotic equilibrium of cells. A similar stabilizing role is also induced by the polyamines (spermidine, putrescine and spermine); the HCAs conjugates, moreover, have a high antioxidant potential and are known to be part of the plant defense system against ROS molecules and abiotic stresses [31, 184, 227]. Polyamines are known to be induced during water stress [34] and insect attacks [66].Interestingly, the GR modified Chromium stressed samples demonstrated the lowest number of affected metabolites, showing increased levels of principally conjugated amines The observed induction of grossamide, a lignin amide, in all the three kind of CR stressed plants, could moreover indicate an enhancement of the lignin biosynthetic pathway in consequence to chemical stress. A recent study of Van de Mortel et al., (2008) [228] suggested that the lignin biosynthesis have the function to prevent excess efflux of metals from the vascular system by formation of an extra endodermal layer.

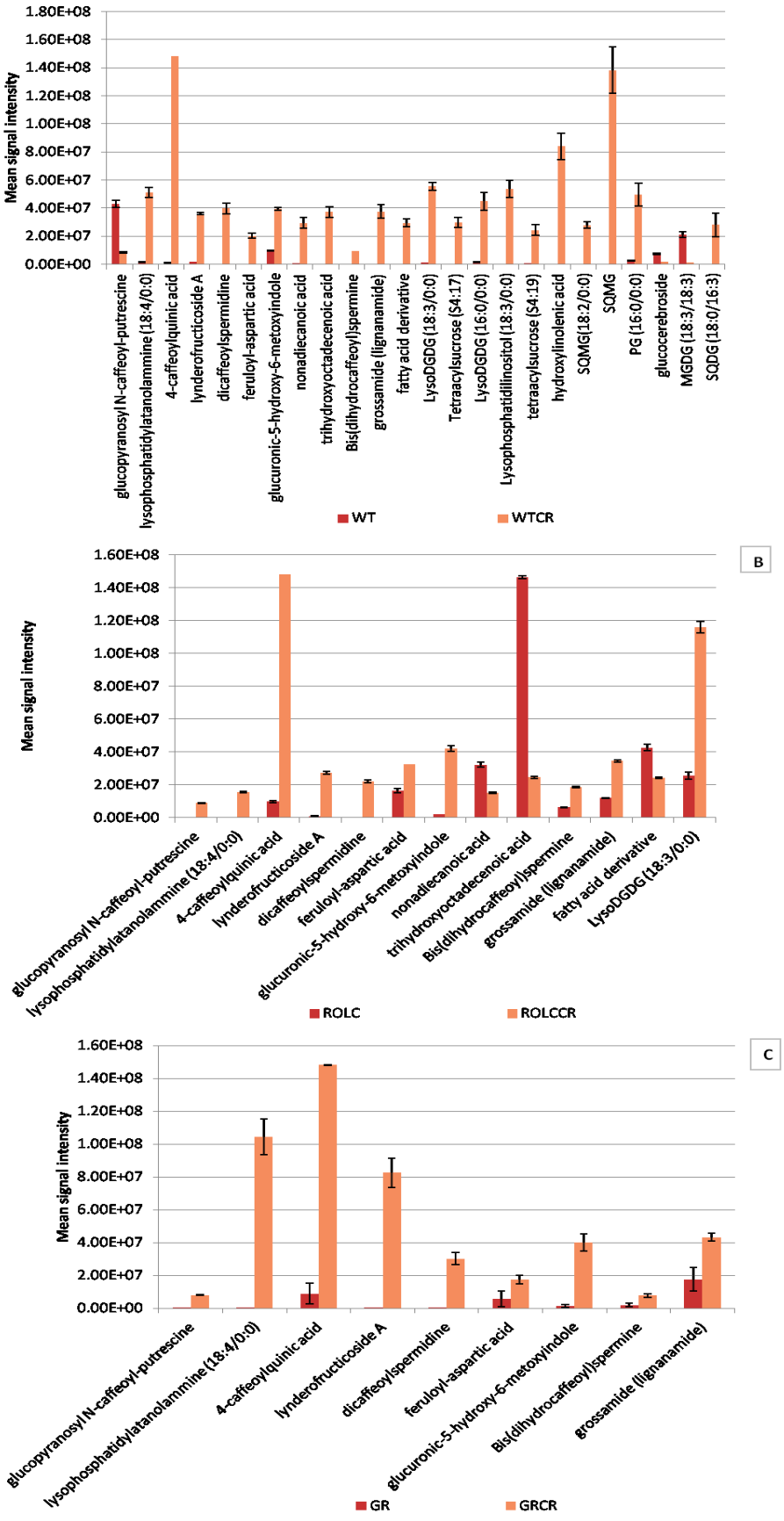

Figure 23 Compounds showing the statistically significant changes in WTCR (A), RolC-CR (B) and GRCR plants, identified in negative polarity. Error bars represent the standard deviation chemical stressed and not stressed samples detected in negative polarity

For what concern the water-stressed samples, their complete metabolite profiling in positive polarity was not possible because of the interferences caused by the PEG 6000, which prevented the detection of metabolites during the central 10 minutes of analysis (20-30 minutes of chromatographic run). However, the analysis in the negative polarity showed some interesting findings (Figure 25). The WT samples demonstrated the highest ratios of metabolites between stressed and non stressed samples; the highest changes were observed for different classes of compounds: lipids, hydroxyl and peroxyl fatty acids, lignins and lignanamides (syringaresinol and grossamide), terpenes (capsianoside II and prieuriarin), malonic acid derivatives, indolizidine alkaloids (ipalbidine), carbohydrate conjugates (primeveroside) and polyamines. This result could suggest that the response to water stress is less related to specific metabolic pathway than the one to chemical and heat stresses, whereas involving many processes. The effect of hydroxylation and peroxidation of lipids resulted similar to the one previously observed for Cr(VI) samples and it's probably related to the presence of ROS; the radical action is partly compensated for the presence of antioxidant compounds as terpenes and polyamines; the accumulation of sugars, amino acids and other small molecules as the malonic acid derivatives could have an osmotic regulating role, already observed during water stress [34]. The highest increases in WTWS samples were registered for the hydroxyperoxyoctadecatrienoic acid, the trihydroxyoctadecenoic acid, the glucopyranosyl-penta malonic acid and the bis-hydrocaffeoyl spermine. Rol CWS samples seemed to be mainly characterized by increased antioxidant levels, as the lignins, HCAs derivatives and terpenes. Also in the case of water stress, the GR response seems to be limited in comparison to WT and ROLC plants, showing statistically significant increases only in a few compounds with ratios ranging from 2 to 56. Interestingly, the water stress induced the production of 3-caffeoylquinic acid in all GRWS, WTWS and RolC-WS samples, while not resulting in increased amounts of 4-CQA.

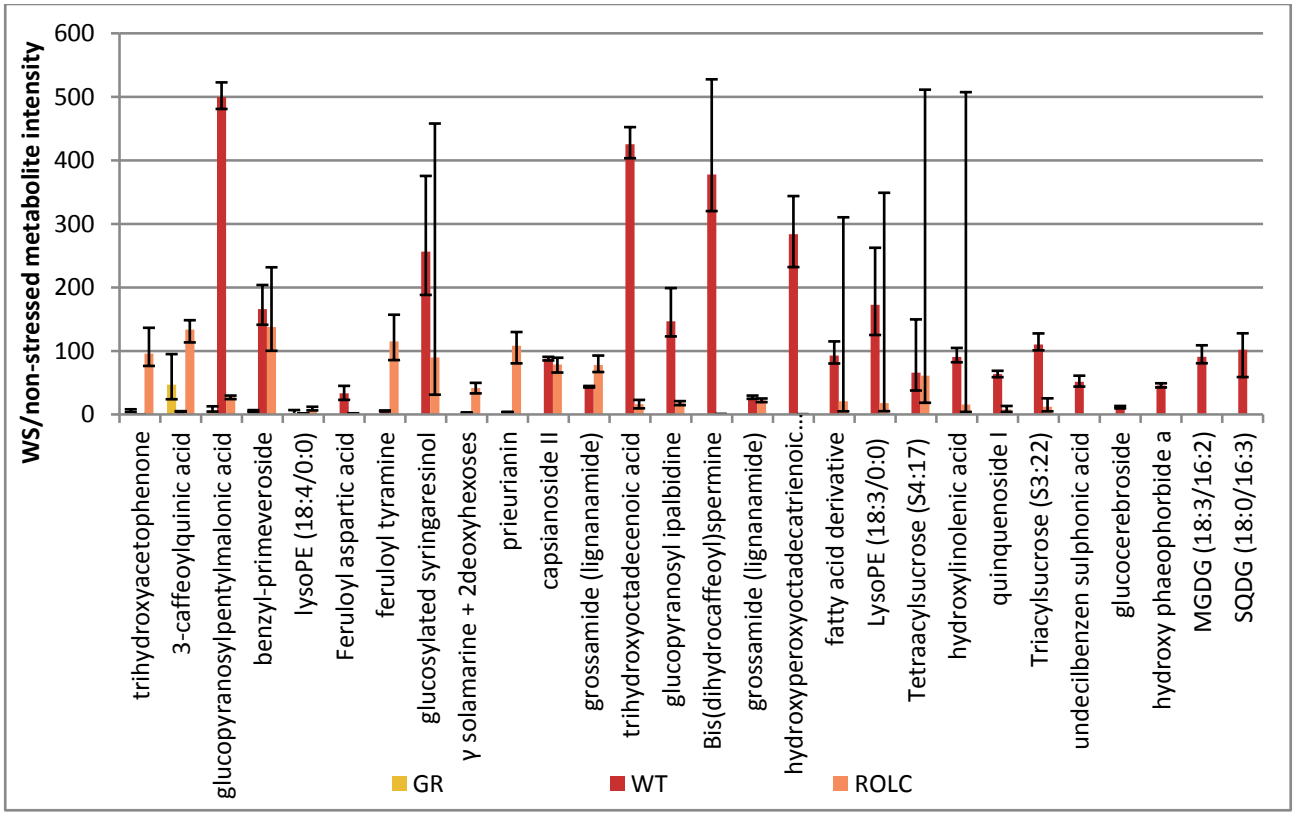

Figure 24 Ratio between WS samples and controls for the metabolites showing statistically significant changes. Error bars represents the maximum and minimum values of the ratios.

2.4 DISCUSSION

2.4.1 SA, SHA, JA validated method

Here we presented a fast, accurate and quantitative methodology for the simultaneous determination of three key molecules, SA, SHA and JA, involved in plant metabolism and stress response. To our, known, no other methodologies which permits the combined quantification of these compounds have been reported. Several analytical procedures have been used for the determination of SHA, which is generally analyzed to be employed in the production of anti-influenza drugs. The analytical techniques include NMR [67], GC-MS [229], spectrophotometry [158], HPLC [230] and HPLC-UV [157, 231]. However these procedures usually require significant amounts of solvents, time and labor; sensitivity and specificity can also limit their application [232]. JA and SA, well studied as plant phytohormones, are usually determined together or in association with other key regulatory molecules or related compounds [151, 162, 182, 233]. Phytohormone determination protocols were traditionally based on enzyme-linked immunosorbent assays (ELISAs) [234] and radioimmunoassay tests [235, 236]; more recently, spectrophotometric [237], and chromatographic [150, 178, 191, 238] methodologies have been developed. The simultaneous determination of multiple phytohormones still remains a challenging task, due to the different physicochemical characteristics of compounds and their typical low concentrations in plants. Our methodology, employing the HPLC-HRMS technique, permits a good separation of compounds, their unambiguous identification and their sensitive quantification; moreover, due to the high mass accuracy measurements, HRMS avoid unreliable quantifications deriving from the presence of interfering molecules, thus permitting a simplification of the sample treatment procedure. In comparison with the GC-MS and LC-MS previously reported methods for SHA [80, 229] and JA and SA [161, 182, 239], our protocol do not require the derivatization step, which is highly time consuming and a possible source of non-systematic errors. Our validated method, showed the highest LOD for SHA, which is a highly polar molecules; the use of an acidic mobile phase, required to increase the retention time, negatively influenced the ionization of this compound, diminishing the method sensitivity. Despite this, the LOD obtained resulted even better that the ones reported in previous studies [67, 229, 240]. SA and JA, which are generally determined simultaneously, showed better LOD, comparable with the one reported by [179] and higher than the one of [161, 162, 232]. SHA demonstrated the highest recovery, while JA and SA showed low recoveries. However, the use of the internal standard quantification method for these compounds prevents their unreliable quantification. The assessed repeatabilities and precisions resulted better for SHA than for SA and JA. This is consistent with the use of two different quantification methods for SHA and JA and SA. Therefore, the low JA and SA accuracies could derive from variations of the basal SA and SHA content in the matrix used for validation or could suggest a slightly different chemical behavior of the internal standard with respect to the analytes. However, the biological variability is known to be generally around 15% and the repeatability obtained for the analyte quantification was included among the variability of each sample set.

The validation of an analytical methodology on biological samples needs to take into account the matrix effect, in addition to an effective estimation of the quality parameters, due to the potential different instrumental responses between standard solutions and samples [69]. In this study we demonstrated that SHA and JA are both affected by *Nicotiana* matrix effect, the former showing exceptional signal suppression, the latter being enhanced. The use of FR prepared in a matrix solution completely solved this problem. The use of an instrumental factor response for quantification is highly recommended, to avoid signal variation effects which could compromise the reproducibility. All these aspects of method validations are not always considered in literature [101, 179, 232, 239] and could conduct to misleading results.

2.4.2 Fatty acid semi-quantitative method

The developed method for the determination of fatty acids includes the extraction and trans-esterification with a MeOH-BF3 solution; this approach, despite trans-esterifying all fatty acids, could prevent the complete extraction of lipids, thus compromising the results. We observed high increase in the fatty acid levels which were confirmed by the changes associated to lipid composition in HS samples. However, we can't exclude that an incorrect evaluation of fatty acids has been performed. Further investigations will be made in order to understand the real method efficiency, performing a complete validation.

2.4.3 Metabolite analysis

A multi-approach analysis has been performed on wild and transgenic *Nicotiana Langsdorfii* plants, subjected to different abiotic stresses, in order to identify the main metabolic changes associated with stress exposition and the effects of genetic modifications on plant responses. The GR modified plants were obtained by insertion of a mouse glucocorticoid receptor which is thought to possess high affinity with the plant brassinosteroids; moreover, the inserted gene possess the constitutive CaMV promoter, which, being continuously active, could permit an higher activation of phytosteroids in GR transgenic plants [151, 152]. The integration of Rol C gene in plants has shown to determine the alteration of many physiological and metabolic aspects; these effects have been related to an increased activity of cytokinins and to the induction of many secondary metabolite pathways [147, 189, 241]. The insertion of GR and Rol C genes do not affected the content of SA and SHA, which showed similar levels in all the controls. The JA levels, however, resulted reduced in Rol C controls; this effect could be due the ability of the enhanced cytokinin activity to suppress the JA levels, which have been assessed during stress conditions [242, 243]. Conversely, GR seems to highly influence the basal levels of fatty acids, which showed a reduction of 87% with respect to the WT controls. This fact could be related to an enhanced activation of brassinosteroids which are known to regulate genes involved in the fatty acid metabolism [244].

With respect to the samples exposed to abiotic stresses, the GR plants do not showed significant increases of phytohormones while in WTCR and Rol C-CR plants, the levels of SHA and JA resulted enhanced. SHA is a key molecule in the biosynthesis of flavonoids and phenolics, antioxidant molecules which production is known to increase during stress conditions [31, 60]; the high levels of SHA therefore indicate an induction of the shikimate pathway, which is consistent with the results obtained by the metabolomic analysis, where the presence of many HCAs derivatives in both WTCR and RolC-CR samples was observed. Also the levels of JA resulted increased in both WTCR and RolC-CR plants. The role of JA has generally been related to biotic stresses [146, 172, 182, 245] but it could also take part to abiotic stress responses [45, 162, 172]. JA is known to induce the synthesis of phenylammonium lyase (PAL), the first enzyme of phenyl propanoid pathway and chalcone synthase (CHS), the key enzymes of flavonoid biosynthesis [246]. The SA levels resulted enhanced only in WTCR samples, while presenting almost comparable levels in RolC-CR and GR-CR plants. The SA is known to be involved in chemical stress response [247, 248]. An interesting thing observed in RolC-CR samples was the reduction of hydroxyl and peroxy fatty acid levels with respect to controls; we suggest that this effect could be related to the RolC gene ability of reduce the production of ROS species during abiotic stresses, thus resulting in a low lipid oxidation [154].

With respect to HS samples, they present the highest changes in the content of phytohormones, with SA significantly increased in GRHS and decreased in ROLCHS; JA showed enhanced levels in RolCHS and WTHS plants. Since increased SA levels have been recognized to determine a positive effect on plant heat stress tolerance [46, 179, 248, 249], the observed diminution of SA in WTHS could derive from a different mechanisms. For instance, Chen et al, (1993) [250] suggested that SA could be converted to a free radical

during the inhibition of the activity of catalase, which could be a key step in the induction of systemic acquired resistance (SAR); however we didn't observed increased lipid peroxide levels, which is a known effect of H_2O_2 production during SAR activation [251]. On the other hand, the SA diminution could be the effect of an interplay between JA and SA, which are known to act synergically and antagonistically during abiotic stresses; their complex interaction, however, has not been well elucidated yet [46, 161, 162, 177, 178, 252, 253]. With respect to JA, its induction has already been related with heat stress exposition [159], possibly caused by the disruption of cell membranes [254]. Heat stress in plants could cause a decrease in chlorophyll content, increased amylolitic activity, disintegration of thylakoid grana; to alleviate these injuries plants produce antioxidant metabolites including enzymes, phenolics, flavonoids, anthocyanins, lignins [155] deriving principally from the shikimate pathway; interestingly, the SHA levels during heat stress resulted induced only in GRHS and WTHS plants, with respect to controls. However, the levels of RolCHS samples resulted comparable with the others. Despite the increased levels of SHA, the heat stressed samples didn't showed increase phenolic levels; we suggested that the enhanced SHA levels could be more related to the production of SA or to the biosynthesis of glykoalkaloids, through the mevalonate pathway. An increase in the glykoalkaloid content was indeed observed in all the HS samples, and especially in the content of solamarines and solasodines. Interestingly, WTHS and RolCHS showed the highest increases in solamarine compounds while GRHS demonstrated enhanced solamargine levels, indicating a differential induction of the pathway (Figure 26).

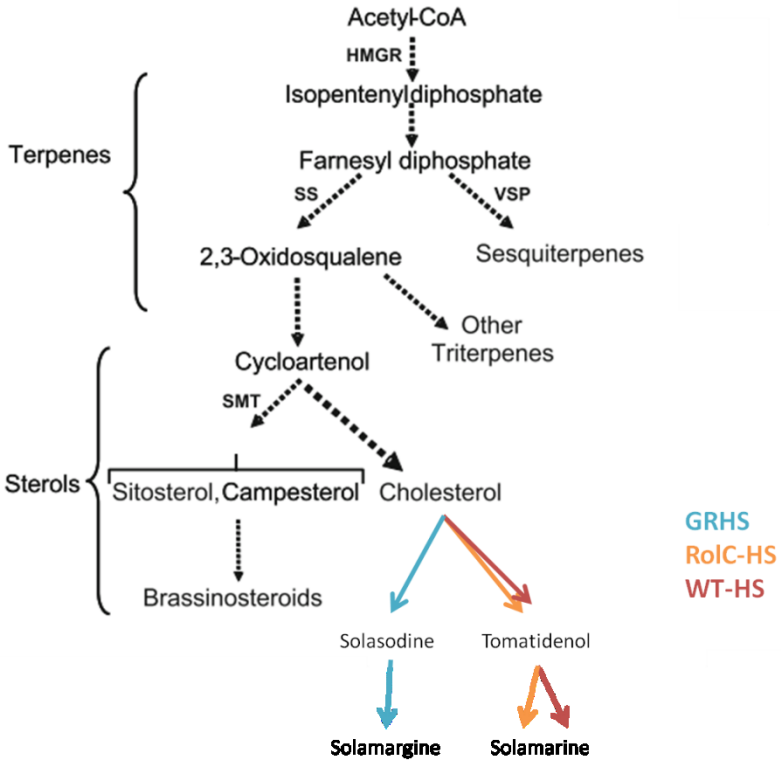

Figure 25 Schematic biosynthesis of the main glykoalkaloids observed in HS samples. Modified from [255].

Increased levels of glykoalkaloids have already been related with intense light and high temperatures in potato and tomato plants [220, 221, 256].

With respect to fatty acids and lipid compounds, the obtained results seem to indicate that the plant response to heat stress leads to an increase in polyunsaturated fatty acids; these compounds, despite the modifications associated with the membrane permeability during the temperature increase, could have a role against this kind of stress. In our study the applied heat stress was short but intensive (2 hours at

50°C); we can thus hypothesize that the increase in unsaturated fatty acids consists more in a first plant stress response than a long-term stress adaptation. Short-term thermal stress studies are not very common in literature; the heat stress applied generally includes a high temperature (35-40 °C) exposition for at least 36 hours and could even considers an acclimatization period. Among the short-term stress studies, the work of Sakhno et al., (2014) [257] showed that the concentrations of fatty acids in plants of *Brassica napus L.* were only slightly affected by the heat treatment at 42°C for 16 h; this finding suggests that the saturation level of compounds is not affected by short term heat stress, unlike the observations obtained from longer thermal stresses. Moreover a study of Gombos et al., (1994) [258] showed that the cyanobacteria Synechocystis PCC6803, exposed to heat stress, tend to decrease the content of dienoic fatty acids while maintaining almost unaltered the content of trienoic fatty acids; the inhibition of the production of these compounds do not seem to determine an advantage against high temperature. Furthermore a study of Zhang et al., (2005) shows how the over expression of two genes, codifying for fatty acid desaturases, determines, in tobacco cell culture and plants, an increase in drought and salt stress tolerance. This could indicate that the enhanced saturation rate of fatty acids in stress conditions reflects more a plant damage than a defensive response. The increased linolenic acid levels, in comparison to their respective saturated, were observed in all HS samples; these increases could be related to an enhanced production of JA. The results obtained from semi-quantitative fatty acid analysis, were confirmed by metabolomic results, in which the major part of detected lipids showed to have a high unsaturation grade. The higher increases in lipid levels were observed for the SQMG/SQDG and MGDG/DGDG groups. These compounds are known to influence the membrane stability and especially the photosynthesis process. One of the major effects of heat stress is indeed the decline of the photosynthetic rate, which is generally attributed to lowered internal CO_2 and to the inhibition of Rubisco. Heat stress particularly suppresses the photochemical efficiency of photosystem PS II by decreasing electron transport and damaging proteins D1 and D2, which represent the heterodimer of the reaction centre of PS II [155]. This effect takes place after a limit temperature, while under this limit the photosynthesis tend to increase. It has been reported that, in tobacco leaves, a temperature of 42 °C for 2 hours determines a reduction in the photosynthesis rate of the 38% [201].The increase of sulfoquinovosyl lipids and DGDG therefore could have a stabilizing function on the photosynthetic system, as part of the plant struggle to maintain unaltered its essential life structures [204]. Sulpholipids are known to accumulate during high/low temperature stress and water deficit [206] with the function of regulation of the catalytic activity of cytochrome oxidase and CF_0-CF_1 ATP synthetase; moreover, both SQDG and DGDG are localized as prosthetic groups at the surface of the native D1/D2 heterodimer and might hold the dimer together; the anionic headgroups of the lipids could have the capacity to bind and conduct protons along the membrane surface, thus providing a more efficient energy transfer between subunits and compensating for the possible inhibition of proton transport under stress conditions [202]. Sato et al. (2003) [203] demonstrated that SQDG protect the PSII system in algal culture of *Chlamydomonas reinhardtii* exposed to heat stress. Among the group of lipids, other compounds showed significant increases in HS samples. The acylsugars could be considered trichome specialized compounds because they are produced and accumulated in these special plant structures; trichomes are also known to store other secondary metabolites, as the glykoalkaloids and the flavonoids, and their role in plant defense against pathogens have been estabilished [192, 209, 212, 259, 260]. In our study, due to the high increase in tri- and tetra-acylsucroses detected in HS plants, we suggested that trichomes could have a role against high temperature stress. Beside secondary metabolite production, trichomes could help the plant heat dispersion or in the regulation of water transpiration [214].

Acylsugars resulted also affected by the water stress, being enhanced particularly in WTWS plants. A similar finding was already reported by Forkner et al., (2000) [261] who however do not registered changes in

trichome number and density between watered and unwatered plants. The other changes observed for WS samples were mainly the increases in some hydroxyl- and peroxy- fatty acids and the simultaneous activation of antioxidant compound production, mainly HCAs derivatives. No significant increases in the levels of JA, SHA and SA were however observed in these samples.

With respect to the comparison of the different responses of the transgenic plants, the GR showed, in all cases, lower increases of the typical stress-derived metabolites, especially in the GRCR samples, where only the HCAs and conjugated amines resulted enhanced. In our study, the RolC plants seem to behave more similarly to WT plants than to GR; SHA and JA levels resulted affected in RolC-WS, RolC-HS and RolC-CR plants, indicating an activation of the plant system response in consequence to the stress perception. On the contrary, GR plants showed significant changes of SHA and SA only in response to heat stress treatment, while the levels of these compounds were never affected by water stress and chemical stress exposition. This important finding strongly indicates that the GR plants could present a higher resistance against water and chemical stress, differently from RolC plants. In a previous study about GR and Rol C modified *Nicotiana Langsdorfii* plants exposed to heavy metal stress, the authors observed an increase in resistance against chromium by the insertion of both these two genetic modifications [153]. In another study, Rol C modified *Rubia Cordifolia* cells has shown to be more resistant to both chilling and heat stress; this finding was mainly related to ROL C expression ability to prevent oxidative burst [154]. GR genetic modification have been scarcely studied in relation to abiotic stresses; another work of Fuoco et al., (2013) [151] showed an hyper-activation of the defense processes in GR plants which, however, negatively interfered with the already high stress-metabolite levels, resulting in the process inhibition. In our study however, we observed higher potential advantages deriving from GR modification which need to be improved and well understand. The behavior of GR plants is particularly interesting in the case of WS stress exposition, where the metabolic profile resulted almost unaffected with respect to controls. Further investigations in order to clearly identify the biological mechanism underlying these findings and their applicability are highly recommended.

2.5 CONCLUSIONS

In this chapter we presented an integration of targeted and untargeted metabolomic approaches with the purpose of identifying the different stress response of wild and genetically modified *N. Langsdorfii* plants. The targeted analysis highlighted the changes associated to a few key metabolites, directly involved in the activation of stress defenses or biomarkers indicating the induction of specific metabolic pathways. The integration of these data with the results of metabolomic analysis permitted a complete evaluation of the plant metabolic status. The effect of GR modification in the control plants was mainly a suppression of the fatty acid biosynthesis, which however seemed to be induced after heat stress exposition. The RolC modification showed a reduction of the content of JA, probably due to an enhanced activation of cytokinins, which can negatively interact with JA. The wild plants showed the highest metabolic changes after stress exposition, demonstrating to be more affected than transgenic plants. The abiotic stress which showed to major affect the metabolic profile of plants was the high temperature exposition. All the heat stressed samples showed an increase mainly of lipid compounds, especially MGDG/DGDG and sulphoquinovosyl derivatives; these compounds are known to protect the photosynthetic system and to be enhanced during increased temperatures [203, 206]. An accumulation of acylsucroses, trichome-specific *Nicotiana* compounds, and glykoalkaloids was also detected after heat and water stress exposition; we suggested that the trichomes, which are known to be involved in pathogens defense, could have a role also in the abiotic stress response[214, 261]. The fatty acid levels resulted enhanced in all HS samples, while the saturation degree was not highly affected. Increased levels of linolenic acid were observed in all HS plants,

probably related to an enhanced production of JA. In all plants exposed to Cr(VI) stress, an induction of the shikimic acid pathway was observed, with increased SHA levels and the presence of many antioxidant compounds, as HCA derivatives and lignins. The effects of chromium stress resulted higher in RolC and WT plants, which showed the presence of hydroxyl- and peroxy- lipids, while the GRCR plants showed changes only in the 4-CQA and HCA-amine conjugates. Moreover, the SHA levels resulted unchanged in GRCR plants. RolC-CR and WT-CR plants also showed the presence of lysolipids, as induction of the release of polyunsaturated fatty acids, which are known precursors of oxylipins [262, 263]. The water stress application mainly resulted in increased levels of antioxidant compounds in WTWS and RolCWS plants, as an effect of the induction of ROS molecule production. The GRWS plants showed an almost unchanged metabolic profile, confirming their higher resistance against abiotic stresses.

3 Metabolomic analysis of *Glycyrrhiza*: characterization of species peculiarities and variety specificities

3.1 INTRODUCTION

Licorice plant belongs to the family of *Fabaceae*, genus *Glycyrrhiza L*., and it is one of the oldest and most popular herbal medicines in the world. The genus *Glycyrrhiza,* which name derives from the ancient Greek *glykos* (sweet) + *rhiza* (root), contains about 30 species widely distributed all over the world. The main commercial sources of licorice are Spain, Iraq, Iran, Turkey, Russia, China and at minor extent the south of Italy. The more significant licorice species are *G. Glabra*, from the Mediterranean region, *G. Uralensis* and *G. Inflata*, from China [71, 264, 265]. The commercially valuable parts of licorice plant are the roots and the rhizomes, which are generally washed, sorted and dried and can be directly consumed; alternatively the roots are crushed by millstones and the pulp is boiled to obtain the extract which is vacuum dried to a dark paste, used to produce confectionery or food flavors [71]. Licorice is economically employed in many industrial manufacturing (cosmetics, food, tobacco, beer) however the most widespread use of licorice is in the pharmaceutical field [266]; a variety of healthy properties have been attributed to licorice extracts, including antioxidant [267], anti-inflammatory [268], antimicrobial [269], hepatoprotective, antispasmodic and expectorant [39, 264]. These effects have been attributed mainly to two classes of compounds: triterpenic saponins and flavonoids [38, 39, 65, 169, 170, 268, 270–272]. More than 50 triterpenic saponins have been identified in *Glycyrrhiza* [55, 109] the most relevant of which is glycyrrhizic acid (GA); GA is an oleanane-type saponin which typically represents the 5-10% of the roots and has been recognized as an efficient sweetening agent, being 50 times sweeter than refined sugar [71, 77]. Among phenolics, the most abundant in licorice are typically liquiritingenin, isoliquiritingenin and their glycosidic derivatives, which represent the ≈ 1% of licorice aqueous extract. The content of bioactive principles in *Glycyrrhiza* is highly dependent from plant species, variety and geographic origin; because changes in the chemical composition of plants are directly related to licorice quality and health effects, a comprehensive metabolite characterization of plants is determinant to ensure the safety and validity of products [111]. A number of studies have been published describing the different chemical composition of *Glycyrrhiza* species, however generally focusing only on a limited number of compounds [77, 94, 273–275]. Licorice, being a traditional medicine of mainly oriental cultures, still suffer a too classical analytical approach to bioactive compound determination [90]. A few recent studies reported more extensive chemical characterizations, despite sometimes lacking in method standardization, identification criteria or in biochemical evaluations [111, 169, 276, 277]. In this context, the metabolomic methodology, especially when combined with pharmacological test, could represent a powerful tool for the chemical investigation of licorice characteristics and the discovery of biological active compounds.

In order to assess the chemical characteristics and peculiarities of different licorice species, we performed the metabolomic study of Chinese *Glycyrrhiza Uralensis* and *Glycyrrhiza Glabra* from Russian (*var. glandulifera)* and Italian origin (*var. typica*). The method was optimized by selection of the more suitable extraction solvent and was applied to 6 different licorice typologies. More than 400 compounds were detected and putatively identified, among which many metabolites already found in licorice roots and some compounds not earlier detected. Specie-specific molecular markers were identified and differences between the two *G. Glabra* varieties were highlighted. The pathway regulation of the two licorice species was evaluated on the basis of the detected metabolites and their relative abundances. To our known, this is the first time in which metabolomic standardized approach is used for licorice characterization.

3.2 MATERIALS AND METHODS

3.2.1 Sample preparation

The *Glycyrrhiza* root samples analyzed for this study were obtained from different sources: the *Glycyrrhiza Glabra* samples, grown in the south region of Russia, were kindly provided by Oxana Astafyeva and the prof. Michael Egorov of the Astrakhan State University, Laboratories of Biotechnologies, with the intermediation of prof. Oreste Piccolo. The *Glycyrrhiza Glabra* samples, collected in Italy (Calabria region), were provided from prof. Oreste Piccolo. The other samples were obtained from commercial sources: the two samples of *Glycyrrhiza Uralensis* were bought in local Chinese herbalists market while the other two *Glycyrrhiza Glabra* roots, of certified Calabria origin, were purchased in two Italian herbalists. Root samples were first cut in small pieces and then grinded and homogenized by using a ball mill (MM 400, Retsch, Verder Scientific, Haan, Germany), equipped with two PTFE vessel and grinder balls; samples were grinded for 6 minutes with a vibration frequency of 20Hz to achieve a final fineness of ≈ 5µm.

3.2.2 Sample treatment procedure

The final selected sample treatment procedure was the same employed for *Nicotiana Langsdorfii* (paragraph 2.2.4.1) metabolomic analysis, with the exception of the extraction solvent. As will be extensively described in paragraph, we selected a solution of ethanol:water 80:20 (v/v) as the most suitable extraction solvent. Moreover, differently from the previously described method, for *Glycyrrhiza* samples we used PTFE filters with 0.45 µm porosity.

We analyzed 6 different typologies of *Glycyrrhiza sp*.: three samples of *Glycyrrhiza Glabra* (*var. typica*) cultivated in Calabria region, one sample of *G. Glabra* (*var. glandulifera*) collected in the region of Astrakhan and two samples of *G. Uralensis* purchased from local markets.

Each licorice typology was analyzed in three independent replicates to verify and assure the reproducibility; each replicate was produced as a representative sample of the available roots of the same typology. Together with the samples, we prepared three mix samples, as a pool of all the kind of samples analyzed, to be used as controls during the instrumental analysis and data processing; these samples were also used for data dependent acquisitions. Together with the sample batch, we analyzed three blank samples , in order to check for eventual contamination deriving from the procedure.

3.2.3 Instrumental method

One of the major objective of metabolomic is the creation of spectral and mass library, both internal and public, in order to simplify the identification of metabolites and create a database which permit a reliable and faster identification. With these main criteria in mind, we decide to use for *Glycyrrhiza* samples the same analytical methodology employed for *Nicotiana Langsdorfii* (paragraph 2.2.4.2), in order to maintain the data comparability and start the creation of an internal library with mass, retention time and all the molecule characteristics.

3.2.4 Data processing

The data processing was performed by means of MetAlign and MSClust, with the previously described criteria (paragraph 1.3.2). The parameters used for MetAlign and MSClust processing are reported in Appendix in Table A 8 and 9 respectively. The list of masses produced as output by MetAlign was partially manually processed, in order to properly prepare it for the MSClust processing. We removed the

metabolites which were not present in at least three samples (the three replicates) and we randomized the noise values. The MSClust parameters were selected manually on the basis of the final number of metabolites produced, of the repetitions and of the significance of metabolites. After MSClust processing, we removed from the final mass list the compounds with a centroid factor lower than 0.85. We then perform the statistical analysis, mainly the principal component analysis, by means of Statistica 8.0 (StatSoft, Inc., 2007). The metabolite intensities before PCA analysis were standardized using range scaling. We selected, through the score plot and the loading plot, the most relevant variable to identify; we also decided to identify the most intense compounds in each typology of licorice, in order to better characterize the plant root metabolic profiles. We selected, through the tools of Microsoft Office Excel (Version 2007), the fifty most abundant metabolites in each three replicates of licorice.

3.3 RESULTS

3.3.1 Selection of the extraction solvent

In order to select the best extraction methodology we tested different solvents which are generally employed for the extraction of bioactive principles; we also tested the solvent mixture previously used for *Nicotiana Langsdorfii* metabolomic analysis. Namely, we tested ethyl acetate, diethyl carbonate, ethanol:H_2O/ 80:20 and methanol:H_2O/75:25 acidified with 0.1 % of Formic acid. We weighted 100 ± 1 mg of licorice roots, we extracted them for 30 minutes extraction by ultrasonic bath, we centrifuged the extracts and we filtered the surnatant which was finally analyzed by HPLC-HRMS.

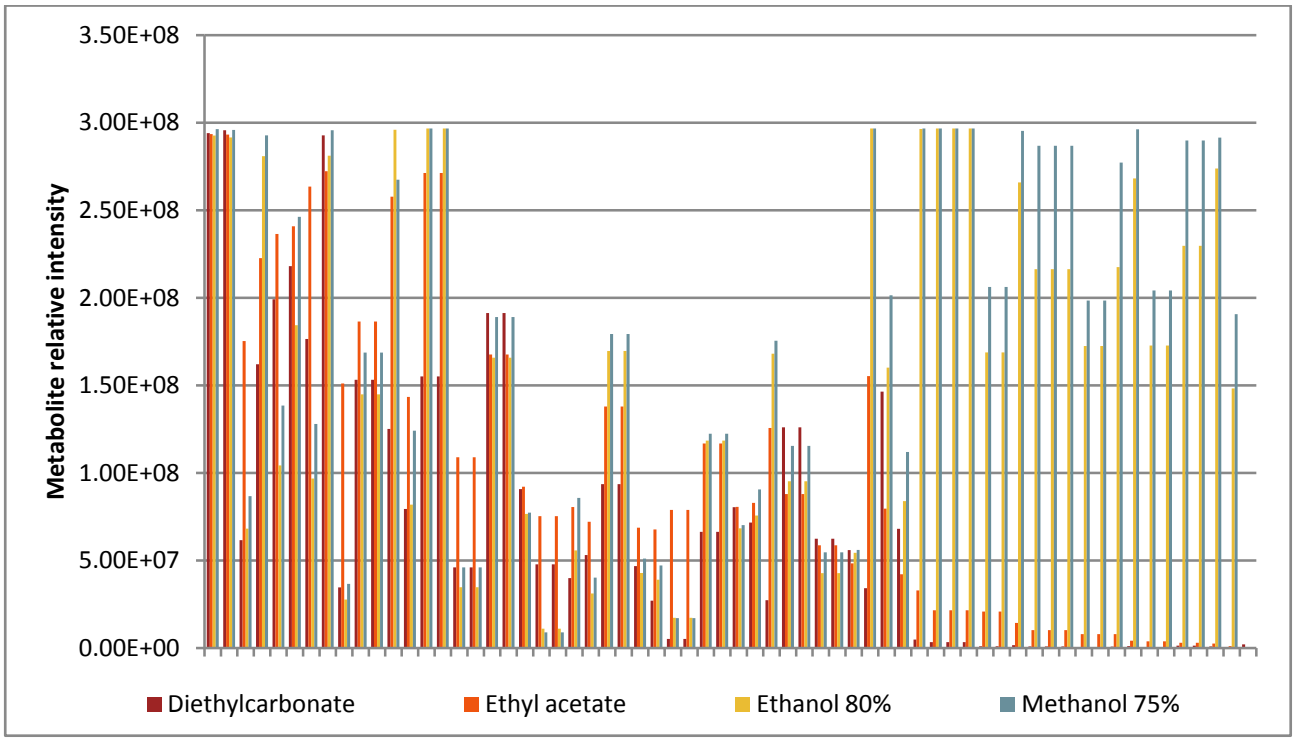

Figure 26 Metabolite signal intensity for the four different extraction solvents used on *Glycyrrhiza Glabra* samples

In Figure 26 are reported the intensities of the main metabolites obtained with the different extraction solvents. All the solutions resulted efficient in the metabolite extraction but, as highlighted in the right part of the figure, ethanol and methanol seemed to provide the widest metabolite extraction range; ethyl acetate and diethyl carbonate were not able to extract a group of metabolites (almost 15 molecular masses) which were thereafter identified as triterpenic saponins. Due to high relevance of this class of

compounds for licorice characterization, we excluded diethyl carbonate and ethyl acetate from the eligible solvents. We checked the extraction efficiency of ethanolic and methanolic solvents for these specific compounds in both *Glycyrrhiza Glabra* from Calabria and Russia. The extraction efficiency of the two solvents resulted comparable for all the identified saponins. To further verify the validity of the tested solvents, we decided to spike the *Glycyrrhiza Glabra* samples with standard solutions of three relevant metabolites (Glabridin, Glycyrrhizin and Ursolic acid), selected as representatives of flavonoids, triterpenic saponins and triterpene acids respectively. The standard compounds were purchased from Sigma Aldrich®(Buchs, Switzerland) and were added to licorice matrices in concentrations of 1 µg/g. The samples were processed as previously explained in this paragraph. The metabolites were identified by comparison (molecular ion, retention time and mass fragments) with the pure standard solutions injected at a concentration level of 1 ng/µL. The chromatographic peaks were integrated manually by means of Xcalibur software 2.1 (Thermo Fisher Scientific, Germany) and corrected on the basis of metabolite concentration already present in the samples. The test showed an extremely higher extraction efficiency of ethanol and methanol solutions for glycyrrhizin and ursolic acid while demonstrating a similar efficiency of all the solvents for glabridin. The results are showed in Figure 27.

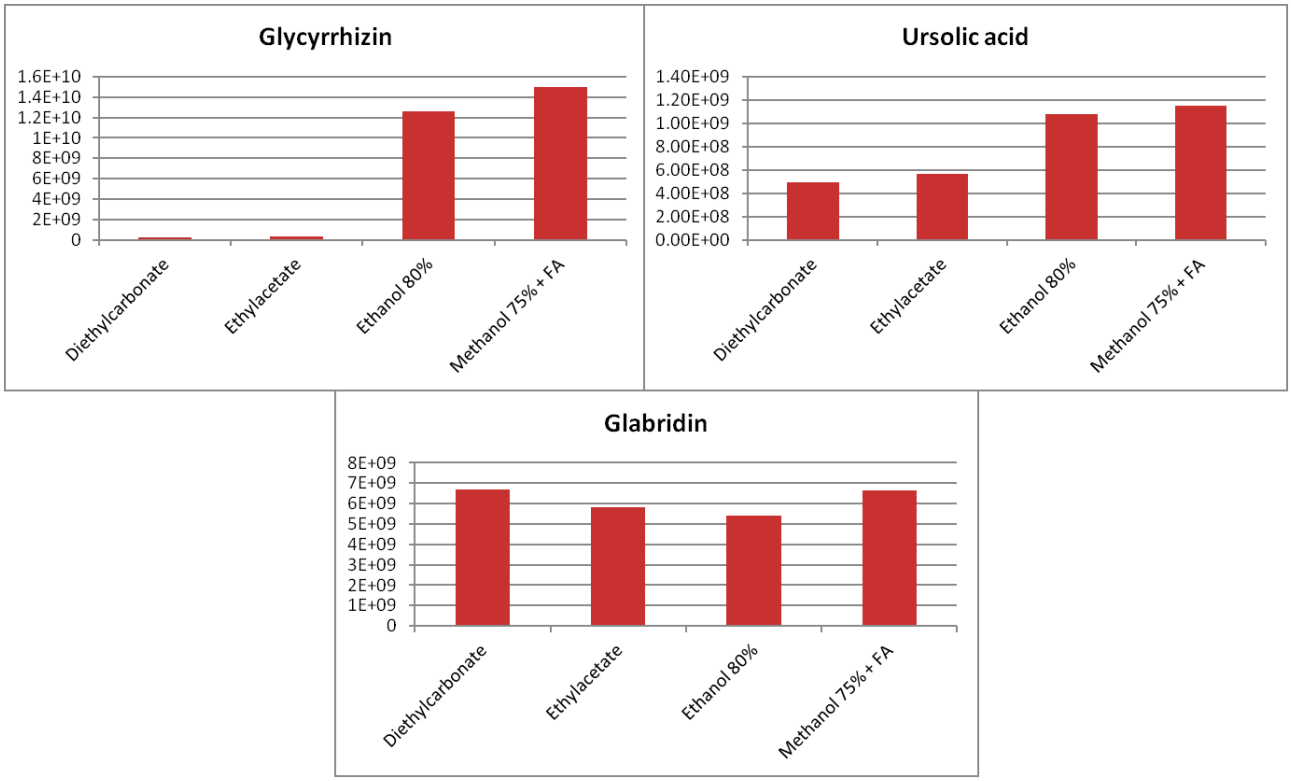

Figure 27 Signal intensities of spiked standard solutions of Glycyrrhizin, ursolic acid and glabridin in *Glycyrrhiza* samples

The extraction recovery of methanol 75% and ethanol 80% resulted similar for the major part of tested metabolites. In this case, we preferred to use ethanol:H_2O /80:20 which is a lower toxic mixture, more compatible for pharmacological and alimentary applications.

3.3.2 Main constituents of *Glycyrrhiza* roots

The analysis of *Glycyrrhiza Glabra* and *Glycyrrhiza Uralensis* samples aim to find out the main metabolites presents in the two species and to identify the principal distinctive molecules which are intimately related to the bioactive properties. After HPLC-LTQ-Orbitrap analysis, the chromatograms were baseline corrected and aligned by means of MetAlign; the ions with low intensity were filtered out; the resulting 14000 ions

were clustered by MSClust analysis in 438 reconstructed metabolites, 281 in negative polarity and 157 in positive polarity. Among these, we were able to putatively annotate 141 metabolites at level 2 while 28 compounds were identified at level 3 [140]; the complete lists of metabolites are reported in appendix (Table A 10 and Table A 11). The principal component analysis was conducted on all the separated metabolites in both polarities and the score plot is reported in Figure 28; in Table 5 is reported the explained variance by the first three factors of the PCA analysis.

Table 5 Variance explained by the PCA performed on *Glycyrrhiza* samples in the first three factors

Factors	Explained variance (%)	Cumulative variance (%)
1	42,78	42,78
2	16,52	59,3
3	14,74	74,04

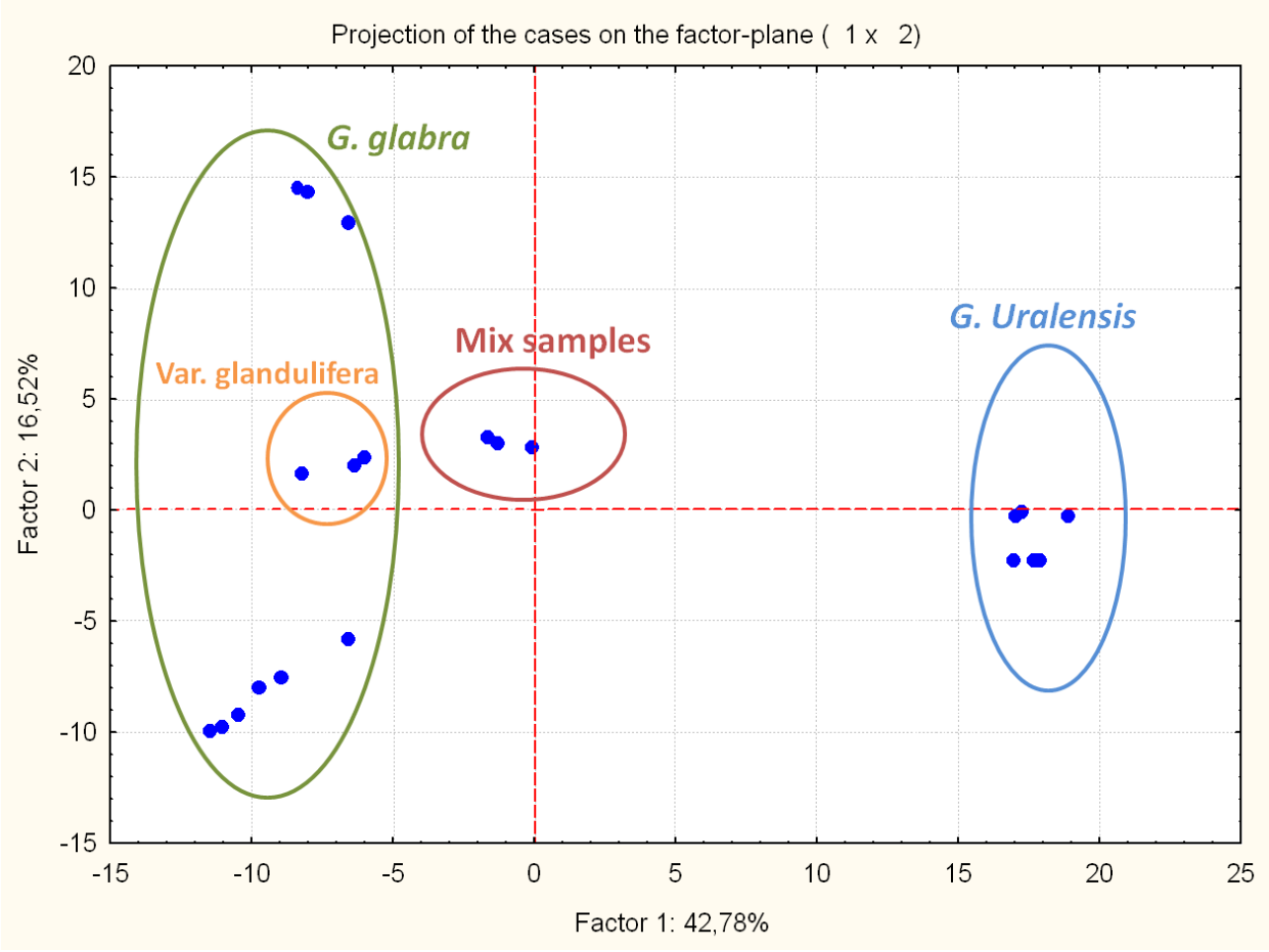

Figure 28 Score plot between the first and second factors of *Glycyrrhiza* samples

The Mix samples prepared as a pool of all the other samples are shown in the center of the plot as analysis and processing controls. As highlighted in the plot, the samples are well separated, as expected, mainly on the basis of *Glycyrrhiza* species; the first factor discriminates between the *G. Uralensis* samples, in the right part of the plot, and the *G. Glabra* samples, in the left part. The second factor is able to discriminate the internal variability, slightly separating the two different samples of *G. Uralensis* species and the four samples of *G. Glabra* (in three replicates each one). Interestingly the Russian *G. Glabra* samples of the variety *glandulifera* seem to be completely included in the variability of *G. Glabra var. typica* samples. The main part of identified compounds belongs to the flavonoid class, including glycosidic flavonoids, isoflavanes, isoflavones and flavanones, followed by triterpenic saponins; we moreover detected a few number of coumarins, chalcones and benzofurans.

The triterpenic saponins, which mainly ionized in negative polarity, showed a diagnostic fragment of 351.05594 corresponding to $[2GluA-H]^-$ or of 497.11371 representing the $[2GluA+Rham-H]^-$ ion, as already suggested [109]. The compounds with the highest signal intensity, detected in both polarities, are reported in Figure 29 and Figure 30.

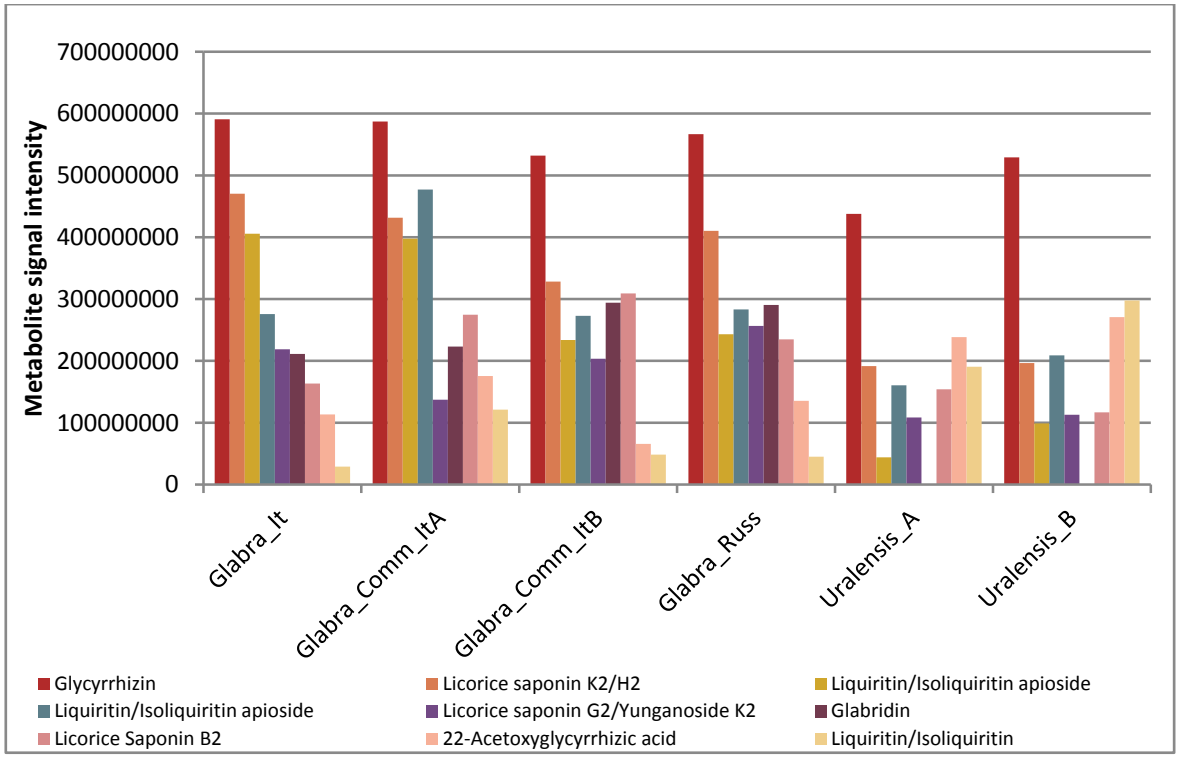

Figure 29 Selected metabolites detected in *Glycyrrhiza* samples in negative polarity

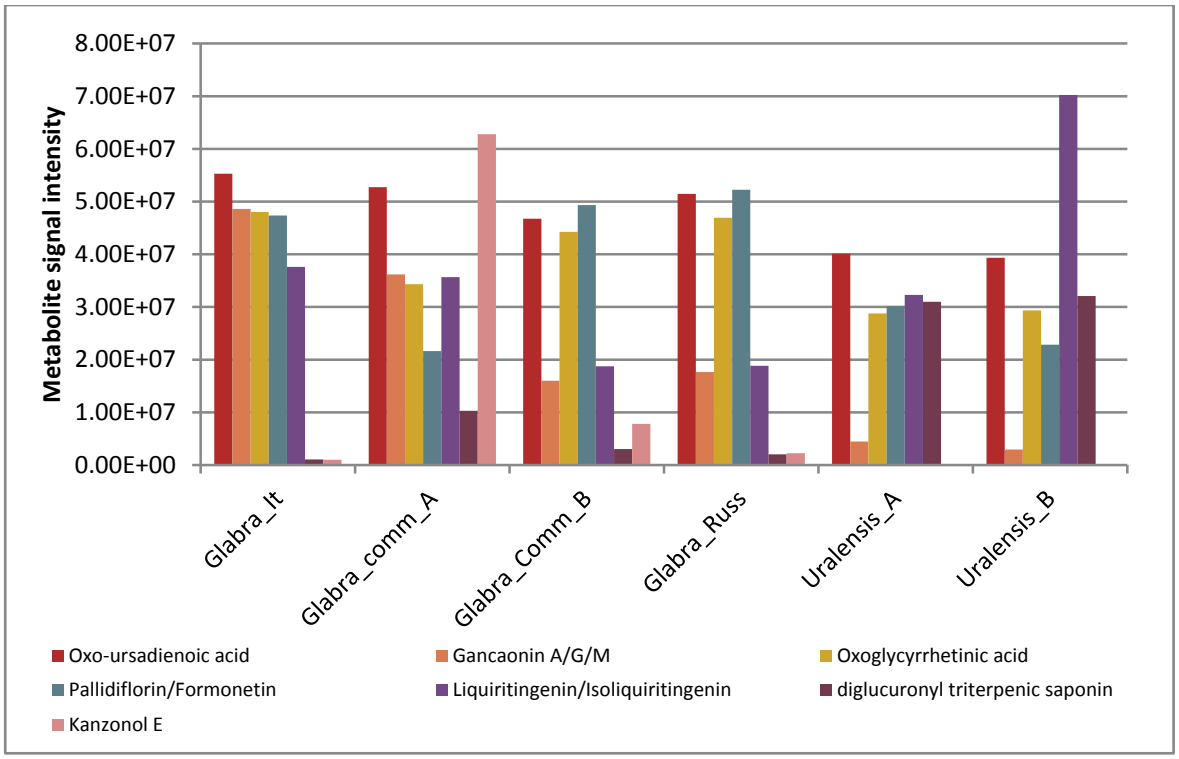

Figure 30 Selected metabolites detected in *Glycyrrhiza* samples in positive polarity

In the negative mode, the most intense molecule was the glycyrrhizin, which showed similar abundances in all the *Glycyrrhiza* samples. This compound, differently from the others, was identified at level 1, by means of a reference standard solution, by comparison of retention time, accurate mass and fragmentation pattern. Glycyrrhizin is the main licorice bioactive principle, generally representing the 5-25 % of *Glycyrrhiza* roots. Beside glycyrrhizin, the other relevant constituents of licorice were the triterpenic saponins of aglycone oleanane-type, K2/H2, G2 and B2. Saponins K2 and H2, as saponin G2 and Yunganoside K2, have the same molecular formula and accurate mass and presents very similar fragmentation spectra; in negative polarity we detected three ions with m/z 821.3961, corresponding to Licorice saponin K2/H2 or isomers, which are very hardly distinguished without standard comparison. The major part of studies maintains the double possible identification. With respect to licorice saponin G2/ Yunganoside K2, the two compounds are generally distinguished on the basis of their relative abundance [55, 278]; in our samples, we detected five isomers with m/z 837.3905, corresponding to these compounds, with different intensities depending from the species; we then preferred to maintain both the name for all the five ions, which nevertheless could be isomers or other unknown molecules. Other compounds with high signal intensity in negative polarity were the glycosidic flavonoids liquiritin/isoliquiritin and liquiritin/isoliquiritin apiosides; liquiritin/isoliquiritin shows a higher intensity in *G. Glabra* than *G.Uralensis* while the apiosidic derivatives exhibit the opposite behavior. Liquiritin/Isoliquiritin and their glycosidic derivatives were identified in negative polarity on the basis of their diagnostic fragment 255.06546, corresponding to $[\text{liquiritigenin-H}]^-$. However, liquiritin/isoliquiritin and the derivatives have same molecular formula, accurate mass and similar MS fragmentation; also in this case the discrimination between the two isomers was not possible, due to the lack of PDA information. The isoflavane glabridin, which is included among the most intense compounds for *G. Glabra*, is present in *G. Uralensis* under the level of noise. Glabridin is a widely recognized marker compound of *Glabra* species [75, 279, 280]. With respect to the metabolites detected in the positive polarity, (Figure 30) among the compounds with highest intensity, there are the oxidized form of the principal triterpenic saponin aglycones (oxo-ursadienoic acid

and oxo-glycyrrhetinic acid) which detection in *Glycyrrhiza* roots is not common; among the other relevant compounds, there is the key *Glycyrrhiza* flavonoid liquiritingenin/isoliquiritingenin, the aglycosidic form of liquiritin/isoliquiritin [278, 281, 282], and some isoflavones (Gancaonin A/M/G, pallidiflorin/formononetin, and Kanzonol E).

3.3.3 Characteristic metabolites for species and variety

From the loading plot between the first and the second factors we selected the discriminating variables (metabolites) for the *G. Uralensis* and *G. Glabra* species, choosing the variables with a first component coefficient lower than -0,5 (from -0,5 to -1) and higher than 0,5 respectively, Moreover we performed the t-test between the *G. Glabra* and *G.Uralensis* samples, selecting the metabolites with a p value<0.05; the compounds deriving from this statistical analysis are shown in Figure 31.

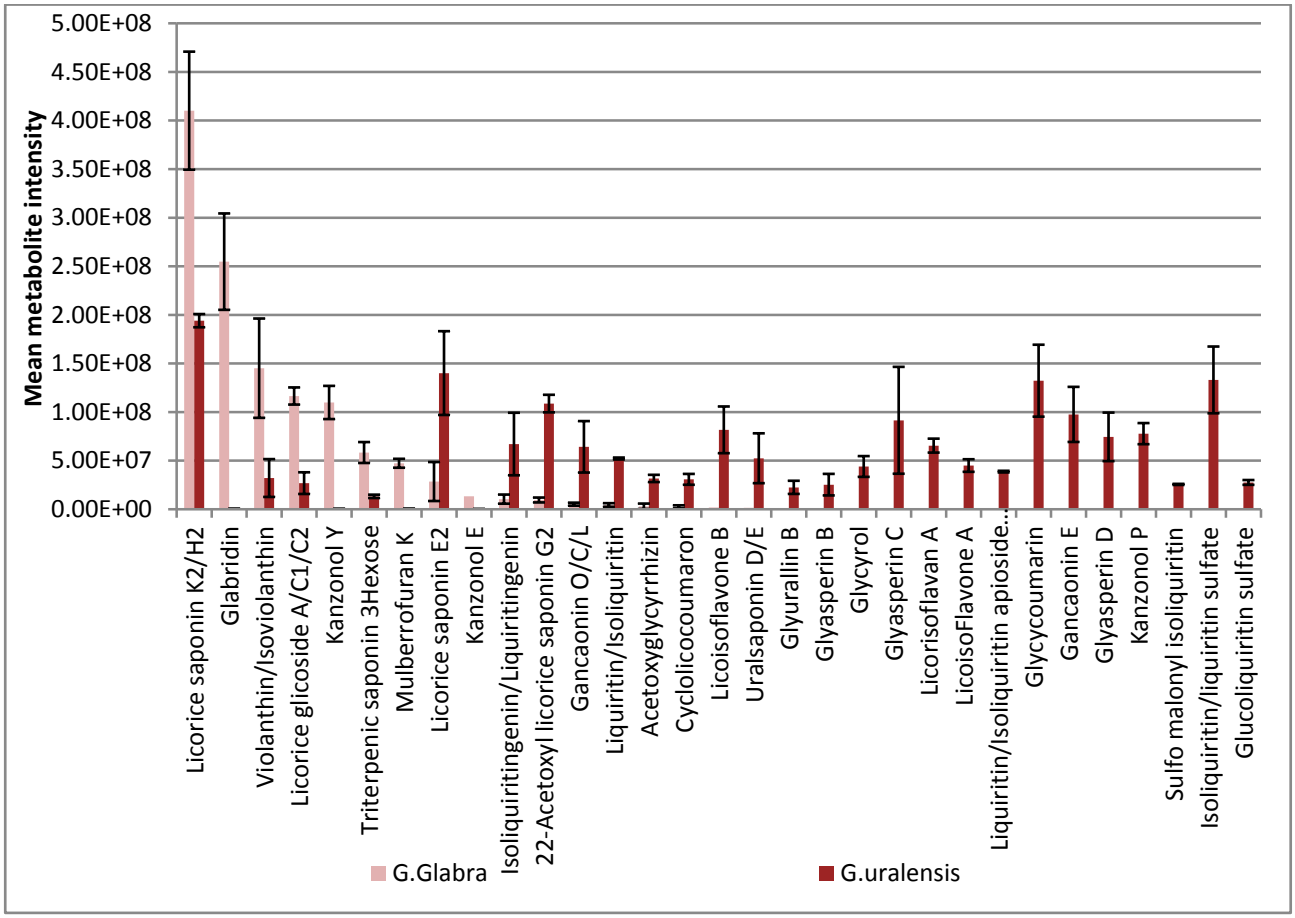

Figure 31 Most relevant metabolites for *G. Uralensis* and *G. Glabra* samples, selected from the PCA loading plot and the t-test. Error bars show the standard deviations

The main differences between *G. Glabra* and *G. Uralensis* seem to be related principally to phenolics compound, more than to triterpenic saponins, as already reported [73, 280]. Among the metabolites appearing typical of *G. Glabra* there are the already cited licorice saponin K2/H2 and glabridin; in addition to these compounds, the glycosidic flavonol violanthin/isoviolanthin, the licorice glycoside A/C1/C2 and the isoprenylated flavonoid Kanzonol Y. Among the relevant compounds for *Glycyrrhiza Glabra*, we also found Mulberrofuran K, which is a benzofuran typically detected in the roots of *Moraceae* plants [283].

One of the most relevant compounds in *Glycyrrhiza Uralensis*, with respect to *Glycyrrhiza Glabra*, is glycycoumarin, which, together with Glyurallin B, is considered a marker metabolite for this specie [169, 264, 268, 280, 284, 285].

Interestingly, several relevant compounds in *G. Uralensis* appear to be flavonoid sulfate molecules, as isoliquiritin/liquiritin, glucoliquiritin, isoliquiritin/liquiritin apiosides and sulfo malonyl isoliquiritin; to our known, their detection in raw licorice roots, haven't been reported before. However it is known that the production of sulfo-flavonoid can occur in a wide typologies of plants as effect of the growth in anaerobic conditions, due to flooding, or as response to high environmental sulfur concentration, with a detoxification role [286, 287]. Relative abundant compounds in *G. Uralensis* resulted also the isoflavanones glyasperin B, C and D, which name derives from *Glycyrrhiza aspera* but known to be presents also in the *Uralensis* species [288, 289]. Other relevant compounds which showed higher abundance intensity in *G. Uralensis* were the pterocarpan and phytoalexin Kanzonol P, Licoisoflavone A and B, Licorisoflavan A. The Uralensis species seems to be characterized also by the presence of licorice saponin E2, acetoxyl licorice saponin G2 and Uralsaponin E/D.

From the principal component analysis performed on the complete set of samples, no significant information come out for the discrimination between the *G. Glabra* (*var. typica*) collected in Italy and the *G. Glabra* collected in Astrakhan (*var. glandulifera*). The identification of the more relevant variables for the two *G. Glabra* varieties has been conducted considering, for the PCA, the subset of *Glycyrrhiza Glabra* samples. The PCA plot, between the first and third factor is reported in Figure 32 and the explained variance in Table 6.

Table 6 Explained variance in the first three factors of the PCA performed on *Glycyrrhiza Glabra* samples

Factors	Explained variance (%)	Cumulative variance (%)
1	38,56	38,56
2	21,81	60,37
3	16,74	77,11

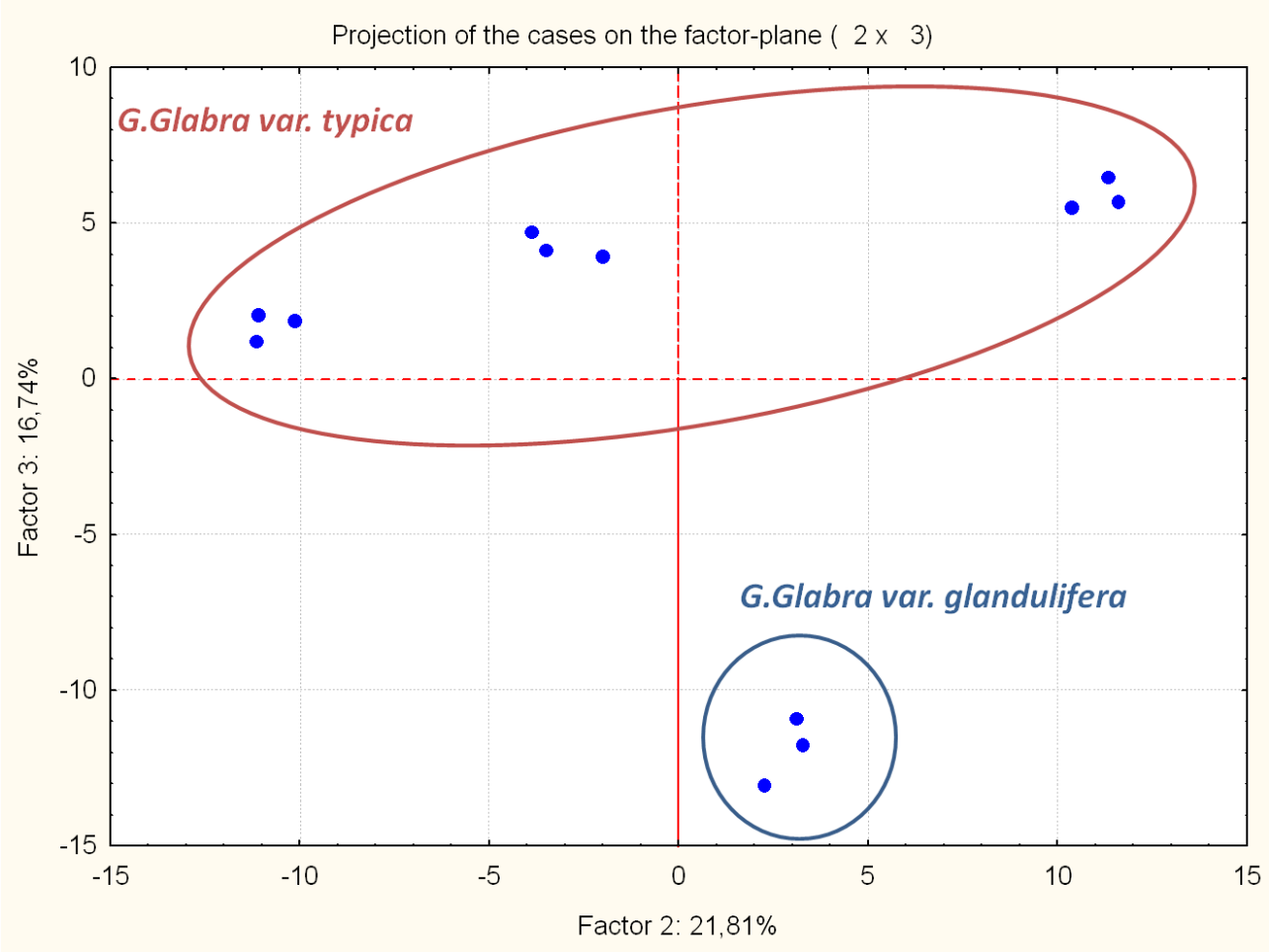

Figure 32 Score plot between the first and third factor of the PCA carried out on *G. Glabra* samples

From the corresponding loading plot between the first and third factor, we selected the variables with a coefficient lower than -0,5 in the third factor, in order to identify the characterizing compound for *G. Glabra* grown in Astrakhan region. Moreover, we performed the t-test between the two licorice varieties and we chose the metabolites with a p value <0.05. The most relevant compounds deriving from statistical analysis are reported in Figure 33. In this case both triterpenic saponins and flavonoids are relevant for the discrimination between the *glandulifera* and *typica* varieties: acetoxyl licorice saponin B2, tri-hexose triterpenic saponin and the triterpenes hydroxyglabrolide and ursolic acid showed higher intensity in the variety *glandulifera* while apioglycyrrhizin and a tetraglycosidic triterpenic saponin resulted more abundant in the variety *typica*. Among the flavonoids, violanthin/isoviolanthin, licorice glycoside E, and formonetin/pallidiflorin showed higher intensity in the *glandulifera* variety.

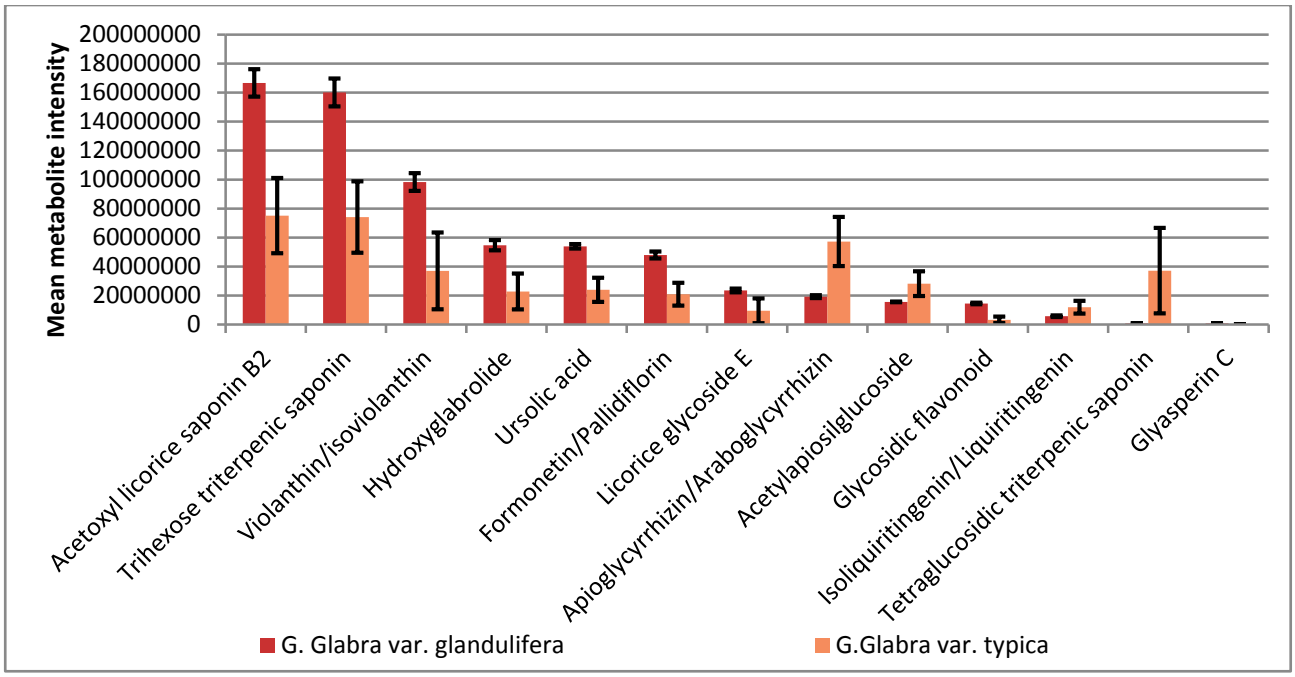

Figure 33 Most relevant compounds discriminating *G. Glabra var. typica and glandulifera*, selected from the PCA loading plot and the t-test. Error bars show the standard deviations

3.4 DISCUSSION

Flavonoids and triterpenic saponins are well known constituents of *Glycyrrhiza* roots; these compounds confer to licorice-derived products the bioactive and organoleptic properties that determine their commercial and pharmacological interest [39, 55, 71, 109, 270, 273, 275, 289–291]. Our analysis showed a metabolic similarity between the *G. Glabra* and *G. Uralensis* species, which are known to be highly genetically related [292, 293]; the glycyrrhizin was confirmed as the most intense metabolite, showing comparable amount in all the samples. Glycyrrhizin and saponin biosynthesis in licorice is an organ-specific process, taking place only in the roots; triterpene synthesis starts from the mevalonate pathway and, through the production of 2-3-oxidosqualene, both the oleanane and ursane-type saponins are synthesized [294]. The main part of detected saponins in our study belongs to the oleanane-type group, which has oleanolic acid or glycyrrhetinic acid as main precursors [295]. The oleanane saponins K2/H2 and C2, both deriving from oleanoic acid, resulted highly related (0.99 with p value<0.001) while not being linked to saponin E2, which has the same biosynthetic origin. No correlation was found among triterpenic saponins deriving neither from glycyrrhetinic acid (glycyrrhizin-G2-J2-B2-A3) nor between these saponins and oxo-glycyrrhetinic acid. This fact could indicate a different regulation, among the species, of the triterpenic substrates utilization, resulting in the independent production of specific compounds. Licorice saponin E2, indeed, was included in the most relevant compounds for *G.Uralensis* while Licorice saponin H2/K2 was characteristic of *G. Glabra*. Licorice saponin B2, while not resulting discriminant between the licorice species, shows a high correlation (0.99 with p value<0.001) with glabrene, glabrone, hydroxyglabrol and Kanzonol Y, which are all specific compounds of the *G. Glabra* species [75, 279, 280]. On the contrary, licorice saponin J2 shows a relation (0.97 with p value<0.001) with glycycoumarin, the molecular marker of *G. Uralensis* species.

With respect to the main licorice flavonoids, no relation was found between liquiritigenin/isoliquiritigenin and its derivatives. Liquiritigenin (flavanone) and isoliquiritigenin (hydroxychalcone) are main intermediates in the flavonoid biosynthetic pathway of *Fabaceae* plants and are produced from the reaction between

malonyl-CoA and p-Coumaryl-CoA. Liquiritigenin and isoliquiritigenin are highly related, being chemically and enzymatically interchangeable by means of a chalcone isomerase; a recent study of Simmler et al., (2014) [94] demonstrated that the ratio of the flavanones to hydroxychalcone isomers is higher in *G. Glabra* than in *G. Uralensis*. Moreover, in accordance with our results, *G. Uralensis* is characterized by a higher content of liquiritin/isoliquiritin while *G. Glabra* shows an enhanced liquiritigenin glycosylation level, especially with apiosyl residues. The regulation of glycotransferase activity in *Glycyrrhiza* has not been well elucidated yet; however, some studies about *G. echinata* cell cultures suggest the presence of both isoflavonoid-specific and substrate-generic enzymes [296, 297]. From liquiritigenin, hydroxyflavanons and methylated flavones, as formonetin, are synthesized; subsequently, the metabolism of formonetin leads to the production of pterocarpans as Kanzonol P. The apigenin biosynthesis, on the contrary, starts from the naringenin chalcone, through a liquiritigenin parallel analogous pathway [298–300]. The presence of a higher content of apigenin derivatives in *G. Glabra* (namely violanthin, apigenin dihexose, hydroxymethylglutaroyl-pentosyl-C-hexosyl-apigenin) could reveal a preferential induction of the naringenin chalcone pathway with respect to the liquiritigenin one. The diverse regulation of the two pathways could be dependent from the different substrate-specificity of chalcone synthases or chalcone isomerases in the two species [298, 301]. *Fabaceae* plants are known to possess enzymes, which are involved in the flavonoid pathway, with different substrate-specificity; the co-existence of these isoenzymes in the same species has been assessed while their activity and regulation is under study; the regulation of chalcone isomerases and synthases is known to be influenced by plant development stage and even by environmental factors [302, 303]. A different regulation of these enzymes has not been studied yet in *Glycyrrhiza* species, therefore further investigations are needed to verify this possibility.

With respect to phenolic compounds, the observed correlations highlight more the existing differences among the sample typologies than common features in biosynthetic pathways regulation. Licorice glycoside B/D1/D2 resulted closely related (0.99 with p value<0.001) to liquiritigenin-glucosyl-apiosyl glucoside, to the coumarone licoagroside B, to Kanzonol Y and the licoflavone A. These compounds, representatives of different classes, showed very similar intensity profiles, with higher relative abundances in *G.Glabra* samples than in *G.Uralensis*, where they resulted almost undetectable. Licorice glycoside E resulted highly correlated (0.99 with p value<0.001) with violanthin/isoviolanthin, highlighting their association with *G. Glabra* variety *glandulifera.*

In *G. Glabra* we detected a group of metabolites which presence seems to be atypical or, at least, not reported before; Kuwanon V, Kuwanon R/Q, Mulberrofuran K are generally detected in the roots of plants of the family *Moraceae* [283]. Kuwanons are prenylated flavonoids while Mulberrofuran K is a furanic flavonoid. *Leguminosae* and *Moraceae* plants are known to be favored sources of both prenylated flavonoids (for instance, Gancaonin O/C/L, glabrene, glyasperins) and furanoflavonoids; these compounds do not showed relevant correlations among them and with other licorice compounds. Only Mulberrofuran K seems to be related to Pongaglabrone (0.97 at a p value<0.001), which is another furanoflavonoid, thus showing a possible biosynthetic relation. The presence of these compounds could be interesting from the pharmacological point of view, because of their bioactivity as HIF-1 inhibitors (Hypoxia-inducible factors) and the effects against inflammatory disease, HIV and cancer [288, 304–306].

With respect to the bioactive potential of licorice species, both *G. Glabra* and *G. Uralensis* showed the presence of molecules with recognized pharmacological properties. In addition to glycyrrhizin, which showed anti-viral, anti-inflammatory, anti-ulcer and anti-hepatotoxic compound [39, 71, 307, 308], the characteristic *G. Glabra* markers glabridin, glabrol and hydroxyglabrol have shown to possess numerous pharmacological properties [39, 267, 269, 271]. With respect to the relevant compounds for *G. Uralensis*,

glycycoumarin have shown to be a potent antioxidant [169, 268] and antibacterial compound [281]; gancaonin E, C/L/O showed high cytotoxic activity in human tumor cells [271]. Licorisoflavan A has also been related to high anti-inflammatory, antioxidative and anti-tumor activities [268, 271, 278, 309] while licoisoflavone A is a recognized antitubercolar compound [269].

3.5 CONCLUSIONS

In this study we reported a comprehensive chemical characterization of two licorice species, *G. Glabra* and *G. Uralensis*, commonly employed for many pharmacological and alimentary purposes. In order to recover the highest number of metabolites, we tested different extraction solvents; an aqueous acidified solution of Methanol (75:15:0.1 v/v/v) and an aqueous solution of ethanol (v/v) resulted comparably the best, being able to extract the widest range of compounds. The other two tested solvents, diethyl carbonate and ethyl acetate, showed low recovery for the class of triterpenic saponins. The ethanolic solution was selected for the final method, being more applicable for pharmacological purposes. The method was specifically tested for the recovery of triterpenes, saponins and flavonoids by means of standard solutions added to the matrix. The analysis permitted to outline a complete metabolomic profile of samples, which resulted in the putative annotation of 141 compounds, the most part of which belonging to flavonoid and triterpenic saponin classes. The principal component analysis clearly separated the two licorice species, highlighting the presence of characterizing metabolites and differences in compound intensities. The PCA performed only on the subset of *G. Glabra* samples also permitted the discrimination between the varieties *typica* and *glandulifera.* We were able to identify the already known molecular markers of *G. Glabra* (glabridin, hydroxyglabrol, glabrol) and *G. Uralensis* (glycycoumarin and glyurallin B); by means of statistical analysis, we selected the most relevant metabolites for each species and for the two *G. Glabra* varieties, which showed statistically significant differences. The most relevant compounds for *G. Glabra* resulted Licorice saponin H2/K2, glabridin, Licorice glycoside A/C1/C2 and Kanzonol Y. We were able to identify some atypical metabolites in *G. Glabra*, which were not previously detected in licorice. Their presence could be relevant for further pharmacological applications. Among the two *G. Glabra* varieties, acetoxyl licorice saponin B2, isoviolanthin/violanthin, hydroxyglabrolide, ursolic acid and formonetin/pallidiflorin resulted most relevant for the Russian *glandulifera*. Araboglycyrrhizin and acetosylapyosylglucoside resulted more intense in the variety *typica*. The *Uralensis* species showed a broad range of specific compounds, mainly flavonoids, and particularly high liquiritin/isoliquiritin intensities. We suggested differential regulation of triterpene and flavonoid pathways in the two species, on the basis of the metabolite intensities and the lack of correlation between biosynthetically linked compounds. More investigations are needed in the field of biosynthesis regulation in *Glycyrrhiza* species, in order to compare our results and clarify the involved processes, even genetically. The two species showed a several number of potential bioactive molecules; however many specific flavonoids, especially for *G. Uralensis*, have not been studied yet; a targeted biological analysis of these compounds could be useful in the understanding of the relation among the identified metabolites and the traditionally known benefic effects of *G. Uralensis* extracts.

This study, providing a wide metabolic profile of *Glycyrrhiza* roots, shows the potential of metabolomic in the characterization of plants and offers an important contribution for the identification of specificity of licorice produced in Italy, Russia and China. In the future, we would like to link this study to toxicological and bioactivity test, in order to improve the comprehension of licorice commercial value and its potential applications.

4 Metabolomic analysis of *Coffea Arabica* and *Canephora* fruits collected at different ripening stages

4.1 INTRODUCTION

Coffee is one of the world's most valuable export commodities, used form the 9th century as a stimulant food [61]. The plant of coffee belongs to the genus of *Coffea*, family of *Rubiaceae*, which comprises a hundred of species spread in Asia, South America and Africa. The commercial production of coffee is based mainly on two species: *Coffea Arabica L.* and *Coffea Canephora Pierre, var. robusta,* which in 2011 accounted for 56% and 44% respectively of the world's production [98]. *Coffea Arabica*, which grow mainly at high altitudes (1200-1300 m a.s.l) in the South of America, is considered the most refined and selected quality, producing a sweet tasting coffee with low caffeine content [99]. On the contrary *Coffea Canephora* gives a stronger, full-bodied coffee, usually with more bitterness and caffeine content than *Arabica* [96]. *Coffea Arabica* is more expensive because of its minor resistance against pathogens and various diseases [310] while C. *Canephora*, being a strongest plant, is cheaper to produce and can be cultivated at low altitudes. The analytical discrimination between *Arabica* and *Robusta* coffee is of great economic interest, in order to assure the commercial quality of products and avoid alimentary frauds concerning illegal adulterations [95, 311]. Caffeine, trigonelline, chlorogenic acids, fatty acids, sugars, putrescine and diterpene alcohols have been assessed as marker compounds for the differentiation between *Arabica* and *Robusta* species [310, 311]. Phenolic and terpenic compounds, which are included in the main bean constituents, are largely responsible for coffee organoleptic properties, as acidity, bitterness and astringency, and for coffee health effects. Phenolics are present predominantly as chlorogenic acids while tannins, lignans and anthocyanins are although present in minor amounts [312]. Coffee diterpenes belong mainly to the kauren and atractyloside families and showed to possess antioxidant, anti-inflammatory and hepatoprotective activities [99, 313, 314] but also cholesterol-rising effects [315].

In order to assure coffee quality, it is crucial to develop appropriate methods to evaluate the presence of these key metabolites. While targeted analysis, focusing on specific metabolite classes, have been extensively used for the characterization of *Coffea* species [310, 316–321] only a few studies employed the untargeted metabolomic approach [95, 311, 322]. Untargeted metabolomic analyses allow the simultaneous detection of many compounds, without need of any preconceived assumption, and help in the comprehension of metabolic processes underlying the production of measured compounds. The biochemical composition of coffee beans is highly dependent from fruit ripening, which involves the development of different tissues and the regulation of metabolic pathways; thus the understanding of biochemical processes during fruit maturation is of critical interest for coffee quality and *Coffea* species characterization.

In this chapter we present a metabolomic study of fruits, collected at different ripening stages and in two harvest seasons, of *C. Arabica* and *C. Canephora*. The metabolomic analysis was performed by means of GC-TOFMS for polar primary metabolite profiling and of LC-LTQ-Orbitrap for secondary metabolite detection. The different parts of the fruit (endosperm, perisperm and pulp) were collected separately, in order to investigate the differential distribution and composition of metabolites in the growing tissues. The analysis permits the separation and identification of more than 100 metabolites, both well known coffee constituents and compounds not previously detected in coffee; clear differences between the species and the tissues considered were revealed. We identified tissue or specie-specific metabolites, highlighting the differential spatial distribution and regulation of the biochemical pathways in the fruits. The ripening

process resulted differently organized in the two species, being independent from the harvest season. The results represent an important contribution to the characterization of *C. Arabica* and *C. Canephora* quality and to the comprehension of the metabolic dynamics in coffee fruit development.

4.2 MATERIALS AND METHODS

The analysis of Coffea samples were completely performed at the Plant Research International (PRI) of the Wageningen University and Research Centre (The Netherland). The samples were provided by Dr. Alan Andrada of the Embrapa Genetic Resources and Biotechnology, Brasília-DF, Brazil.

4.2.1 Sample collection and preparation

The analyzed samples belong to the two species *Coffea arabica* and *C. canephora* and were collected in the savanna region of Planaltina-DF, Brazil. Globally 44 *Coffea* samples were collected and analyzed, 19 for *C. Arabica* and 25 for *C. Canephora*. Samples were picked up during different harvest season (2008-2009 and 2013 for *Coffea Arabica* and 2011-2012 and 2013 for *Coffea Canephora*); the different parts of the fruits were collected separately, when possible, resulting in pulp, perisperm, endosperm, perisperm+endosperm. The samples were collected at different day after flowering (DAF) in order to follow the complete ripening process of the two species: from 90 to 210 DAF for *Coffea Arabica* and from 60 to 300 DAF for *Coffea Canephora*. Before analysis, samples were freeze-dried, milled in small pieces (5 mm) and stored at -20 °C.

4.2.2 Primary metabolites: sample treatment procedure and instrumental analysis

The most polar metabolites (mainly primary metabolites) contained in *Coffea* were analyzed by the GC-TOF technique. The method described have been developed on the basis of the analytical protocol of Lisec et al., (2006) [7]. Briefly, 50 ±0.5 mg of dry material were weighted in a PTFE eppendorf tube and extracted with 1.5 mL of Methanol 75%/H_2O 25% for 30 minutes in ultrasonic bath at 70°C. The internal standard, ribitol, was directly added in the MeOH solution, for a final concentration on samples of 0.5 mg/g. After the extraction, the samples were centrifuged (20.000 rpm for 10 minutes) and 350 µL of surnatant was re-extracted in 350 µL of MilliQ water and 250 µL of Chloroform (centrifugation max speed, 10 min). Before centrifugation each sample was well mixed with Vortex (5 sec); 50 µL of the water phase extract was then transferred in glass vial with insert and dried at 30 °C in speed vac (SPD121P, Thermo Fischer Scientific) overnight. Together with the samples, we prepared 5 blanks, containing only the internal standard and the extraction solvent, to verify the possible contamination deriving from the procedure. Finally the samples were placed in the autosampler of the GC-TOF-MS and derivatized on line with *O*-methylhydroxylamine hydrochloride and *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA). The methoxyamination stabilize the carbonyl moieties, helping in the reduction of the number of derivatives produced from the reducing sugars; the MSTFA is able to efficiently transform the functional groups –OH, -COOH, -SH, -NH into TMS-ethers, TMS-esters, TMS-sulfides and TMS-amine respectively. The analysis were conducted on a GC-TOF-MS system consisting of an Optic 3 high-performance injector (ATAS) and an Agilent 6890 gas chromatograph (Agilent Technologies) coupled to a Pegasus III time-of-flight mass spectrometer (Leco Instruments), equipped with a DB-5-3954 column (Agilent Technologies). A few blank injections were carried out before sample analyses and after 3 sample runs, to check the cleanness of the chromatographical column and the remaining of analytes between one injection and the successive.

Because of was not possible to prepare extra pool samples due to the lack of raw coffee material, we decided to transfer 100 µl of each tube extract in 2 ml Eppendorf tubes to produce 5 extracts mix samples, to be used as controls during instrumental analysis and data processing. The mix samples were run all along the sample sequence, in order to verify the stability of the analytical conditions and of the signal response;

Prior to derivatization, the instrument automatically added to the samples an alkane solution, containing the hydrocarbons from C10 to C34; this step is necessary in order to obtain a sufficient number of landmarks for a proper spectra alignment, and to calculate the retention index from the retention time; in this way, also in case of low metabolite contents, a good number of peaks distributed throughout the chromatogram is obtained.

4.2.3 Data processing

The chromatograms were baseline-corrected in LECO (Leco Corp., St. Joseph, MA, USA) ChromaTOF and further processed by means of MetAlign and the extended version of MSClust (version 2.0.1, Galaxy platform). Moreover, before MSClust processing, the data were elaborated by using MetOT, a Metalign output transformer, used at Wageningen-UR as internal tool; MetOT consents data filtering, missing value replacement, analytical and processing quality control and data preparation for MSClust analysis. The parameters used for MetAlign processing are reported in Table A 12.

The MetOT processing was carried out by selecting the threshold of 3 for the number of samples above the noise for each metabolite; we then set the percentage of randomization of noise at level 90 (that means that the noise is randomized between 90 to 100% of the pre-selected noise value). The output file of MetOT was processed with MSClust Galaxy and included the calculation of retention index, on the basis of alkane. The parameters for MSclust processing are reported in Table A 13.

Finally the resulting files were processed in the NIST library for spectra matching, in order to identify the compounds with spectra and retention index comparison. For each mass peak five hits were considered, namely five possible compounds presents in the selected libraries; the NIST results were manually filtered and controlled, to give the metabolite definitely identification.

4.2.4 Secondary metabolites: sample treatment procedure and instrumentation

The analysis of secondary metabolites in *Coffea* samples has been performed by UHPLC-LTQ Orbitrap technique, following the previously developed protocol of De Vos et al., (2007) [105]. Together with the samples, 6 mix samples pooled from all the different typologies of coffee material were prepared and analyzed with the same procedure. These Mix samples have been used as controls during the instrumental analysis and the data processing, being analyzed all throughout the sample sequence. We also prepared 4 blank samples to be analyzed prior to the samples, to check for contamination. An Accela High Pressure Liquid Chromatography system with a photodiode array (PDA) coupled to an ESI (-)-LTQ-Orbitrap XL Spectrometer (Thermo Fisher Scientific) was used for analysis. A Luna 3 µ C18 column (150x2.00 mm) (Phenomenex, USA) was used to for chromatographical separation, eluted with 0.1% formic acid (phase A) and acetonitrile with 0.1% formic acid (phase B). A linear gradient from 5% to 95% of phase B at a flow rate of 0.19 ml/min was used. The analyses were carried out in negative mode at a resolution of 60000 in full scan modality and with a mass range between m/z 90-1200. The PDA acquisitions were recorded at 240-600nm.

In addition to the full scan spectra we also performed four different scan events, in order to obtain a complete fragmentation pattern of metabolites. We selected the most abundant ion from the full scan (2^{nd} scan event), the most abundant, the second and the third abundant from the second scan event (3^{rd}, 4th and 5^{th} scan events). MSn experiments were carried out at a resolution of 15000 and with normalized collision energy of 35 and an isolation width of 3.

4.2.5 Data processing

The data processing method includes the use of the Metalign, MetOT and MSClust Galaxy software. In Tables A 14 and 15 are reported the parameters used for the elaboration of raw data. The MetOT processing was carried out by selecting the threshold of 3 for the number of samples above the noise and a percentage of randomization of noise at level 90. We report two graphs, supplied by MetOT, which help the quality control of data (Figure 35 and Figure 34).

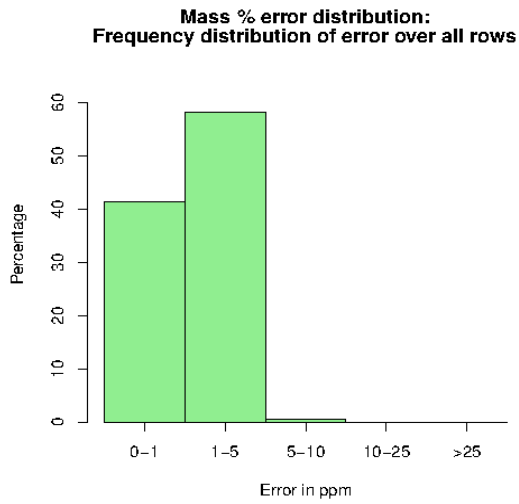

Figure 35 Mass Error distribution in the metabolites

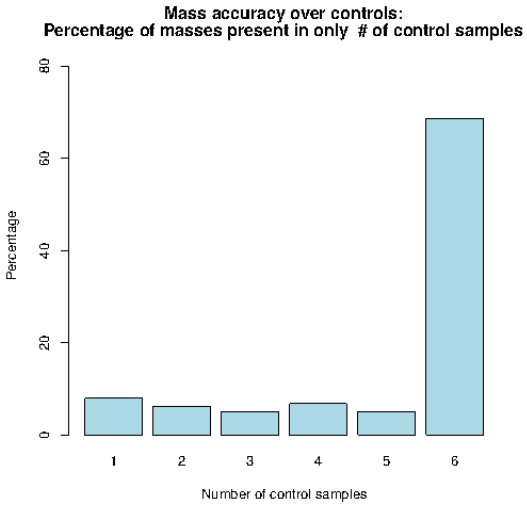

Figure 34 Percentage of masses presents in the control samples (mix)

The mass error distribution represents an indication of the error on the mass measurements, which is generally considered acceptable under the level of 5 ppm. The percentage of masses in the control samples, the mix, is a measure of the repeatability of samples and give indications about the number of peaks (landmarks) used for the alignment. The output file of MetOT was processed with MSClust Galaxy and the final output was used for statistical analysis and metabolite identification.

4.2.6 Statistical analysis

The multivariate data analysis was performed by using Genemaths XT v. 1.6 software (Applied Maths) which is a professional tool for microarray expression analysis; we carried out the principal component analysis and the hierarchical cluster analysis. Data were log2 transformed and normalized prior to statistical analysis. The most relevant metabolites for each *Coffea* specie and each part of the fruit were selected by the analysis of variance (ANOVA), determined by p-values lower than 0.01.

4.3 RESULTS

The metabolomic analysis of *Coffea* samples was carried out in order to highlight the main metabolic differences between the *Arabica* and *Canephora* species and to investigate the fruit ripening process, which determine the metabolite content at maturity and lead to coffee quality. As previously explained, we analyzed different part of the coffee fruits (pulp, endosperm and perisperm) collected at different ripening stages. Globally, with the GC-TOF and LC-MS techniques, we were able to identify 117 molecules, belonging to different primary and secondary metabolites classes. The complete results of LC-MS and GC-TOF analyses are reported in appendix in Table A 16 and Table A 17. The principal component analysis, carried out with all the detected metabolites, highlighted the main relevance of the tissue typology, which determines the clustering of samples. The PCA resulted in pulp, endosperm/perisperm and perisperm

divisions. In Figure 36 is reported the score plot of the first three factors of the PCA and the variance explained is indicated in Table 7.

Table 7 Relative and cumulative explained variance by the first three factors of the PCA performed on *Coffea Arabica* and *Coffea Canephora* samples

Factors	Relative variance (%)	Cumulative variance (%)
1 (x)	27.2	27.2
2 (y)	16.9	44.1
3 (z)	8.3	52.4

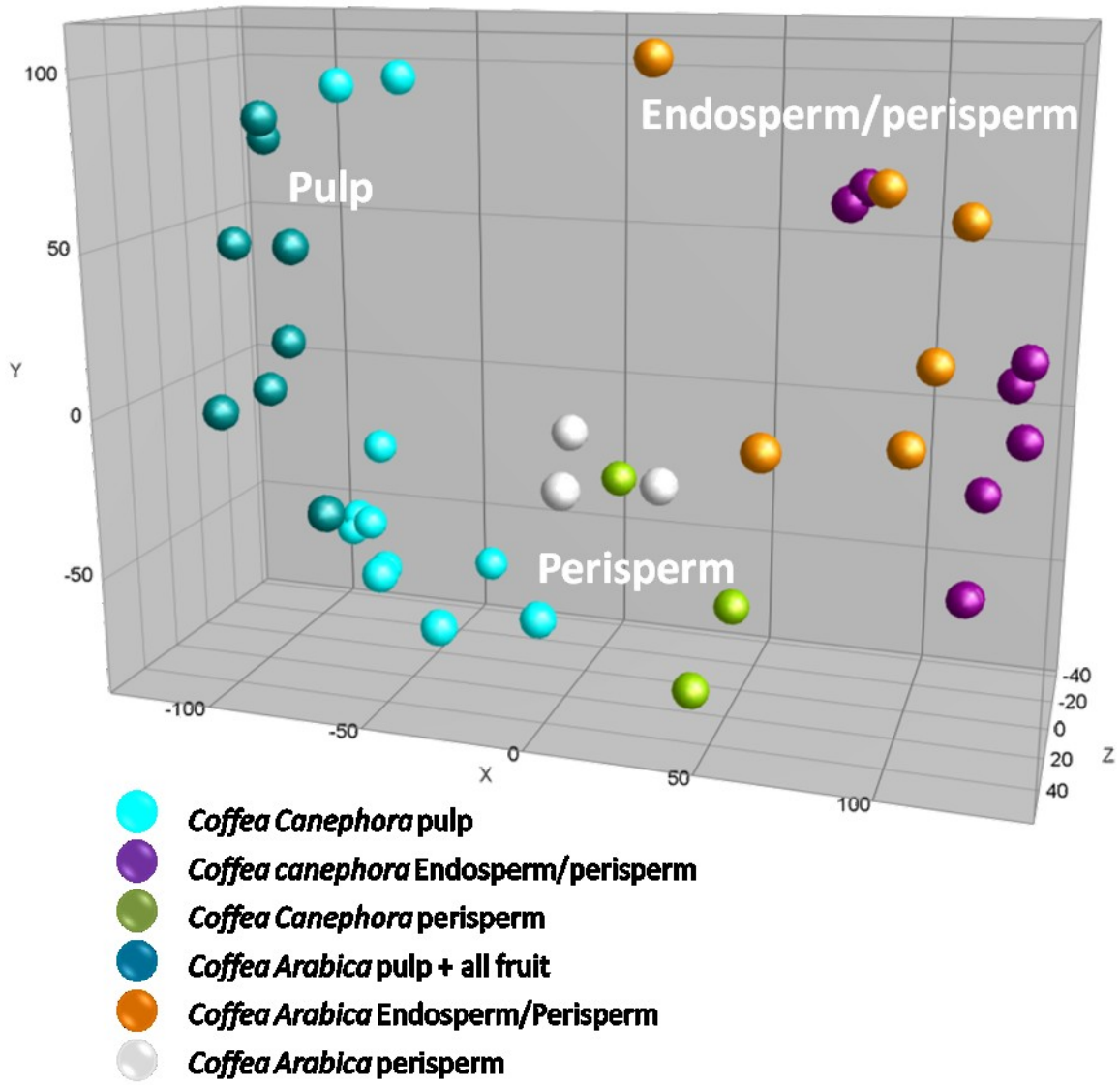

Figure 36 Score plot of the PCA performed on *Coffea* samples by using all the metabolites detected with LC-MS and GC-TOF methods

By using the ANOVA t test (included in the Gene Maths software), we selected the most characteristic metabolites for each subset of samples (pulp, endosperm/perisperm, perisperm), according to a p value<0,05. These statistically significant metabolites were therefore identified.

4.3.1 Main primary metabolites

The GC-TOF analysis and MetAlign preprocessing resulted in the detection of 3498 peaks which were successively filtered and clustered by means of MetOT and MSClust Galaxy workflow. MSClust peak clustering leads to the recognition of 69 compounds, which were identified at level 1 on the basis of retention index and by spectral matching with the NIST internal library. Finally, the metabolites were manually checked and reduced to a final number of 43 compounds.

The identified metabolites belong mainly to the sugar, amino acid, amine and chlorogenic acid classes (Figure 37 and Figure 38). The compounds which showed the higher intensities were the sugars sucrose, glucose and fructose and the quinic acid. The pulp demonstrated the major sugar intensities, especially in *Coffea Arabica* samples. The most abundant sugar in *Coffea* beans is generally sucrose, which constitutes the 2-5% of dry weight in *Canephora* species and 5-8% in *Arabica* species. In our case the sucrose resulted the sugar with higher intensity in both pulp and endosperm/perisperm for *Coffea Canephora* while *Coffea Arabica* showed a higher glucose content in the pulp; interestingly, sucrose showed a similar mean level for both species in the pulp. In addition to fructose and glucose, other relevant sugars for both pulp and endosperm/perisperm were the alcohol sugars myo-inositol and sorbitol. The myo-inositol is generally found in higher concentrations in *Robusta* variety than in *Arabica* species and is known to be involved in many metabolic pathways, being a precursor of inositol phosphates and cell wall polysaccharides, via the myo-inositol oxidation pathway [320]. Sorbitol, which possess the 60% of sweetening power of sucrose, could be synthesized in coffee beans through the conversion from fructose and glucose [323]. With respect to the other primary metabolites, quinic acid resulted present with high intensity in both species and both pulp and endosperm/perisperm. The quinic acid is one of the precursors of CGAs, and it is known to be present in higher concentrations in *Arabica* than in *Canephora* [312, 324].The pulp also contains the organic acids malic and citric acids, which are presents especially in mature seeds and in higher concentrations in *Arabica* than in *Canephora* [323, 325].

The GC-TOF analysis also highlighted the presence of catechin/epicatechin and caffeoylquinic acids, which will be discussed later in the secondary metabolite section.

Other minor metabolites of both pulp and endosperm/perisperm were the ethanolamine and the glycerol, crucial molecules for synthesis of phosphatidylethanolamine (PE), phosphatidylcholine (PC) and glycerolipids respectively. The observed presence of low concentrations of phosphoric acid could derive from the hydrolysis of phytic acid and inositolphosphates [324].

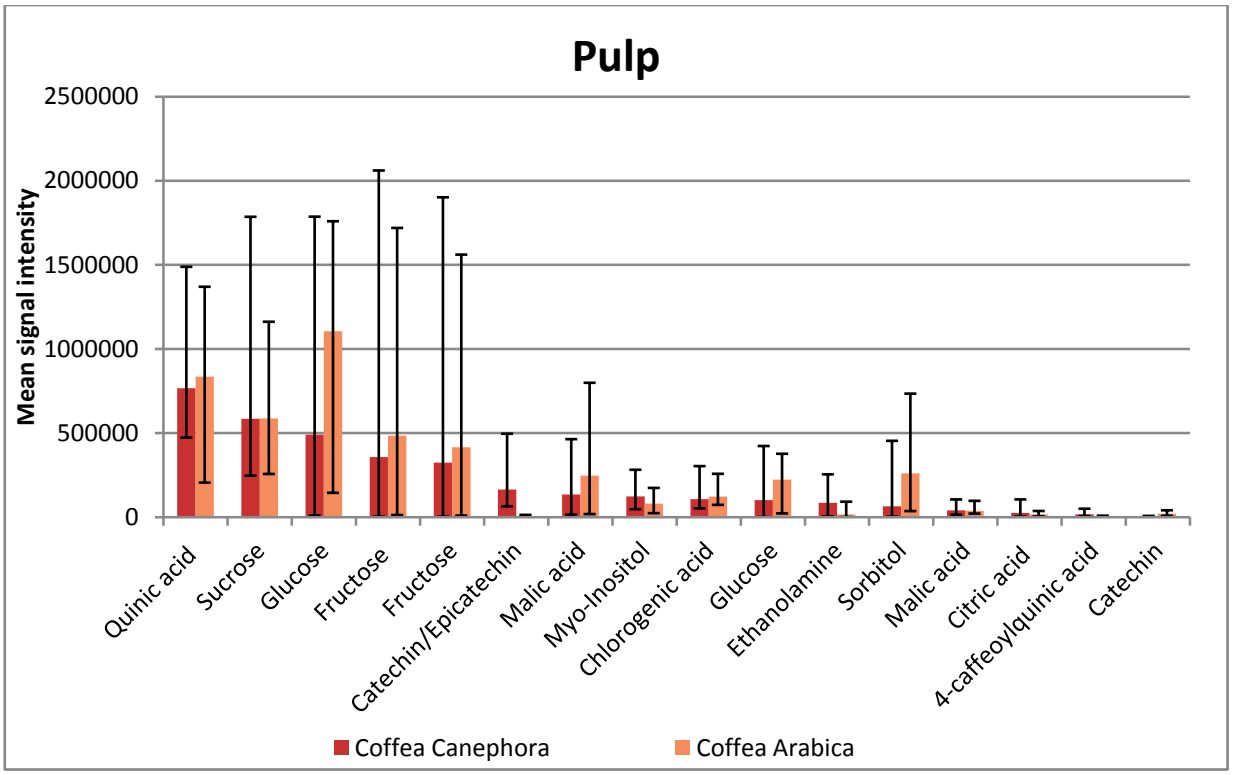

Figure 37 Most relevant metabolites for the pulp, detected by the GC-TOF analysis, with the intensities in *Coffea Canephora* and *Coffea Arabica.* Error bars show the maximum and minimum values

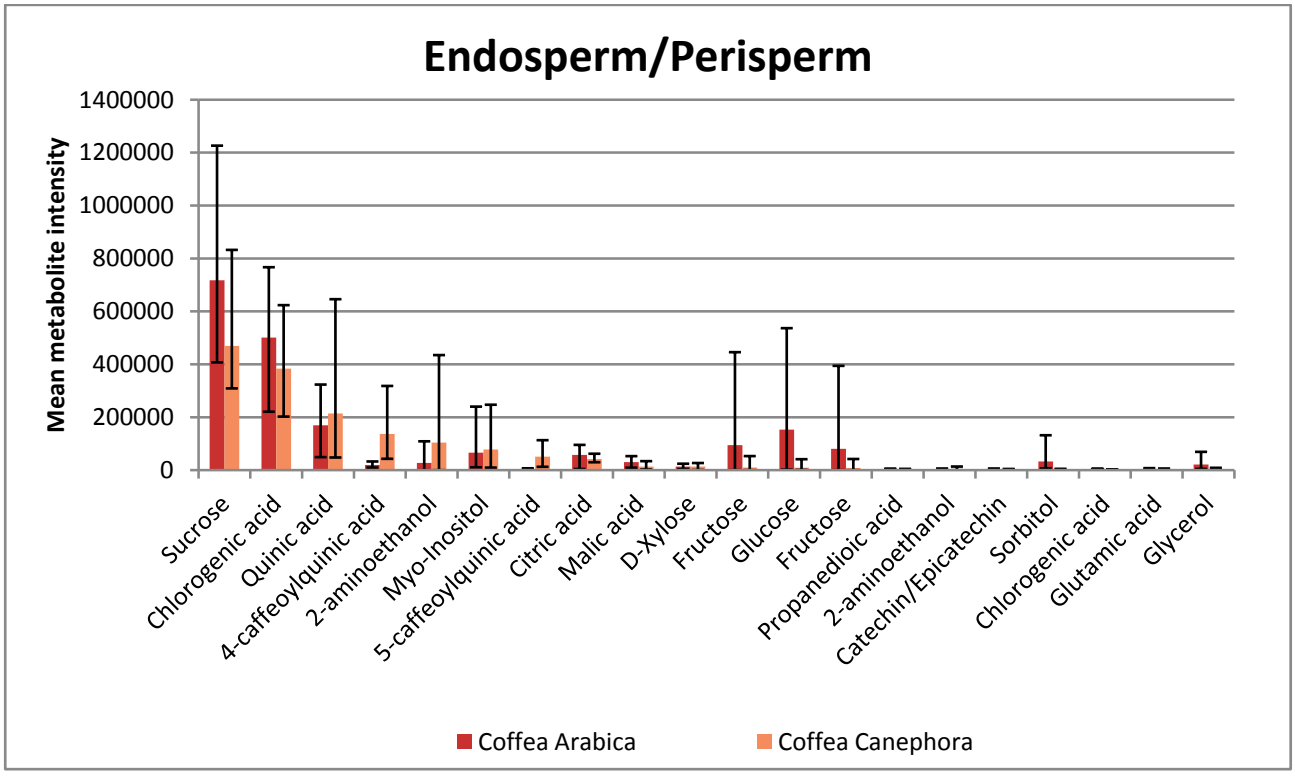

Figure 38 Most relevant metabolites for the endosperm/perisperm samples, detected with GC-TOF analysis, with the intensities in *Coffea Canephora* and *Coffea Arabica.* Error bars show the maximum and minimum values

4.3.2 Secondary metabolites

The LC-MS technique permitted the analysis of medium-polar and low-polar metabolites, which can be mainly classified as secondary metabolites. The MetAlign baseline correction and spectral alignment resulted in a high number of peaks, more than 19000. The data were further processed with MetOT and MSClust, for mass filtering and reduction. Finally we obtained 911 probable metabolite clusters, the most relevant of which were identified; 53 compounds were putatively annotated at level 2 while 14 resulted identified at level 3.

As highlighted in Figure 39 and 40, the major part of relevant metabolites for the pulp are the flavanols monomer, dimers and trimers, namely the catechin, epicatechin and the procyanidins. Both for the pulp and the endosperm/perisperm, higher metabolite intensities were observed in *Coffea Canephora* than in *Coffea Arabica*. The metabolic profile of the two species resulted quite different, showing specific characterizing compound for each part of the fruit. In the pulp of *Coffea Canephora* the most intense metabolites were the procyanidins, especially of B type, and caffeoyl-feruloyl compounds. Catechin showed a high intensity in the *C. Canephora* while *C. Arabica* resulted more characterized by epicatechin. Procyanidins are antioxidant compounds generally abundantly presents in the fruits of many plants, as complex mixtures of dimers, trimers and isomeric forms, together with other polyphenols; the presence of procyanidins could be related to fruit quality, but is also associated to bitterness and astringent taste [41]. In our case, the *Coffea Canephora* showed to contain more procyanidins than *Coffea Arabica*, which is coherent with the well known strong and bitter flavor of this coffee [98, 325]. Generally procyanidins represents the 0.1-1.2% of C*offea* pulp [96]. The presence of higher amount of catechin in *C. Canephora* species and of epicatechin in *C. Arabica* was also reported by previous studies [96, 317, 325]. The pulp of *C. Arabica* resulted characterized by the presence of glycosidic derivatives, in addition to caffeoylquinic acids and some procyanidins. The endosperm/perisperm part of *C.Canephora* fruits showed high intensities of hydroxycinnamic acid derivatives, mainly chlorogenic acids (CGAs), which have been largely studied as the main components of the phenolic fraction of green coffee beans. The identification of CGA isomers was based on the hierarchical classification scheme reported by Clifford et al., (2003) [141] In general, the total CGAs content in green coffee beans vary from 4 to 8.4 % (on dry weight basis) for *Coffea Arabica*, and from 7 to 14.4 % for *Coffea Canephora* [312]. In our samples, the presence of many caffeoyl and feruloyl-quinic acid derivatives was observed, with 4-Caffeoylquinic acid (4-CQA) being the most intense compound in both species, followed by 3-CQA and 5-CQA. Generally the major CQA in *Coffea* is the 5-CQA, with concentrations even five times than the other isomers, followed by 4-CQA and 3-CQA [96, 318, 326, 327]; the higher intensity of 4-CQA, which seems to be atypical, found confirmation in the GC-TOF analysis and NIST metabolite identification (see Figure 38). Interestingly, *Coffea Canephora* also showed the presence of several ferulic derivatives, as feruloylquinic acids (FQA) and feruloylcaffeoylquinic acids (FCQAs), detected in high quantity and in a variety of combinations in coffee beans, especially in *C. Canephora* [33, 316, 326, 327]. *Coffea Canephora* resulted enriched also in some hydroxycinnamic-amino acid conjugates as caffeoyltryptophan, caffeoyltyrosine and coumaroyltyrosine, known to be specie-specific for this kind of coffee [310, 311, 325]. For what concern *Coffea Arabica*, this species showed to contain some particular diterpene glycosides, as carboxyatractyloside I and II (CATR I and II) and Mozambioside, which have been previously reported as characteristic compounds for this species; these group of diterpenes consists of an aglycone with a perhydrophenanthrene structure and a glycoside moiety made up of glucose with sulphate and/or isovalerate [99, 311, 316]; they are structurally similar to the well known diterpenes cafestol and kahwehol, recognized as main *Coffea* constituents [315]. The presence of diterpenes in *C. Canephora* samples resulted quite insignificant, excepting for CATR II.

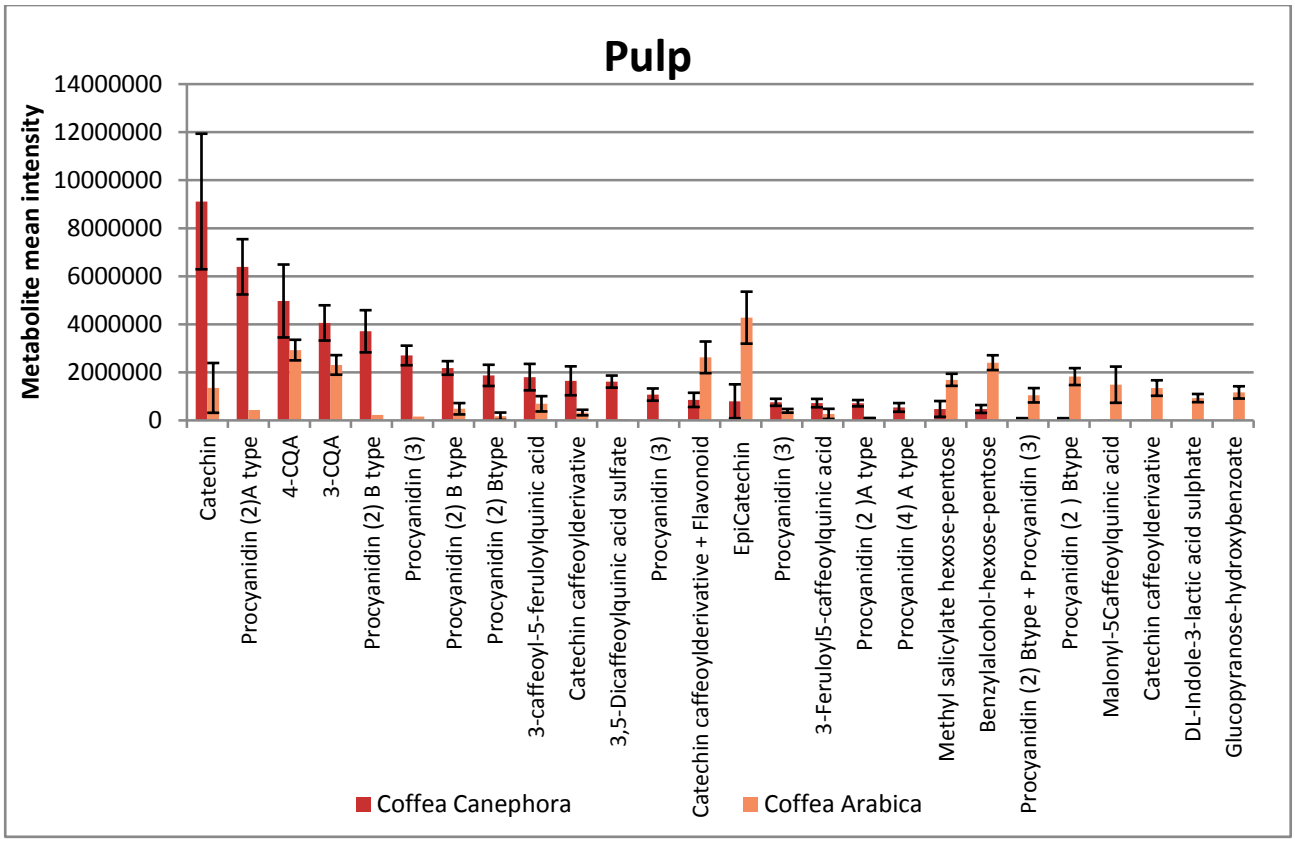

Figure 39 Most relevant metabolites for the pulp, detected with LC-MS analysis, with the intensities in *Coffea Canephora* and *Coffea Arabica.* Error bars show the standard deviations.

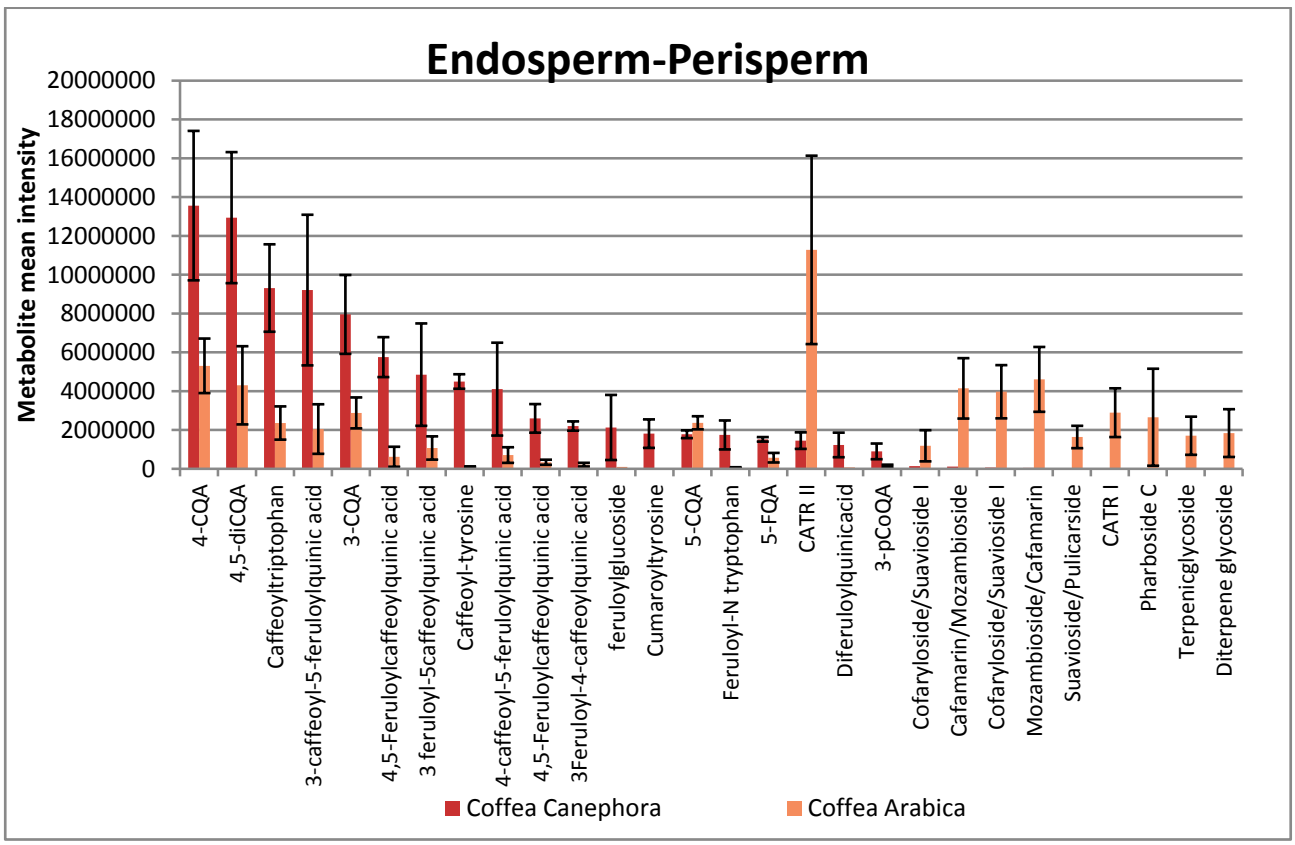

Figure 40 Most relevant metabolites for the endosperm/perisperm samples, detected with LC-MS analysis, with the intensities in *Coffea Canephora* and *Coffea Arabica.* Error bars show the standard deviations.

4.3.1 The perisperm unique metabolic profile

The coffee fruits, in their first developmental stages, are principally constituted by the perisperm and the pericarp. The perisperm is then replaced by the endosperm, remaining as a thin pellicle surrounding the endosperm [328, 329]. For this reason, the perisperm was collected separately from endosperm only in the first stages of ripening (60-150 DAF and 90-180 DAF for *C. Canephora* and *C. Arabica* respectively); the analysis on perisperm samples highlighted a particular metabolic profile, highly different from the one of endosperm/perisperm samples (Figure 41 and Figure 42). The metabolic profile of perisperm samples changes a lot during ripening; here we discuss the mean metabolite intensities, while in the discussion section we will better investigate the maturity process. The perisperm primary metabolite intensities resulted extremely increased in comparison to the endosperm/perisperm samples, being ten times higher; this fact reflects the important storage and biosynthetic role covered by perisperm during fruit maturation. On the contrary, secondary metabolite intensities resulted comparable to endosperm/perisperm samples. The perisperm of *Coffea Arabica* showed higher sugar intensities, especially for sucrose and glucose while *Coffea Canephora* demonstrated lower sugar content but higher intensity of myo-inositol, ethanolamine, quinic, malic and chlorogenic acids. The perisperm, being a transient tissue, is known to be highly enriched also in growth-promoting substances and molecules which function as precursors of structural and storage metabolites. Myo-inositol, for instance, is a precursor of phytic and glucuronic acid, which are relevant compounds in cell-wall polysaccharide biosynthesis; moreover ethanolamine is known to be involved in many growing development processes [60, 184, 330]. Both the molecules, furthermore, represents common headgroups of phospholipids [331]. The presence of hexose free sugars, like glucose, is known to be higher in the first developmental stages, while, when the endosperm starts its formation, they decrease, due to the simultaneous conversion into sucrose [330]. Conversely, the production of glucose and fructose could even take place from the sucrose catabolism. With respect to secondary metabolites, *Coffea Canephora* perisperm showed a poor metabolomic profile, with only CATR II and 3-caffeoyl-5p-coumaroylquinic acid relevant presences. On the contrary, the *Coffea Arabica* perisperm showed a wide variety of diterpenic compounds, mainly presents as glycosides, This finding highlights that, in the early developmental stages of coffee fruits, the diterpene biosynthesis is in full activity and is probably related to the successive biosynthesis of relevant compounds.

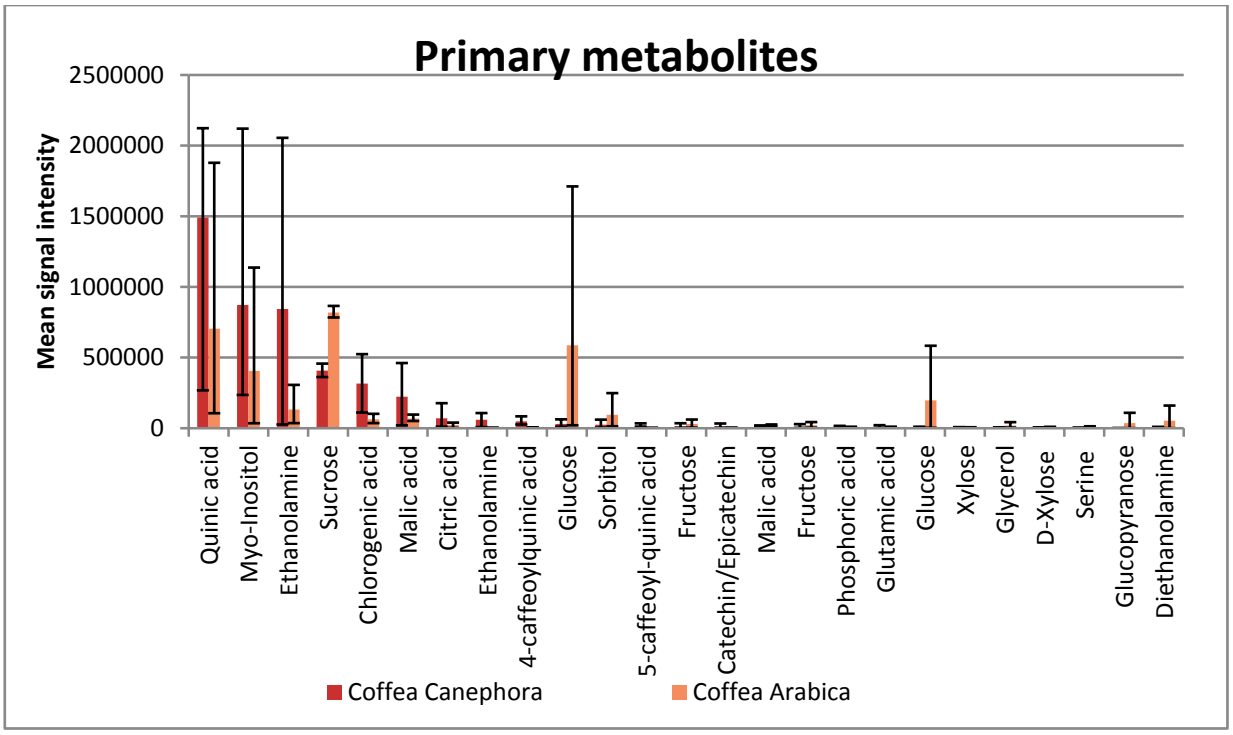

Figure 41 Perisperm mean primary metabolite profile in *Coffea* samples. Error bars show the maximum and minimum values

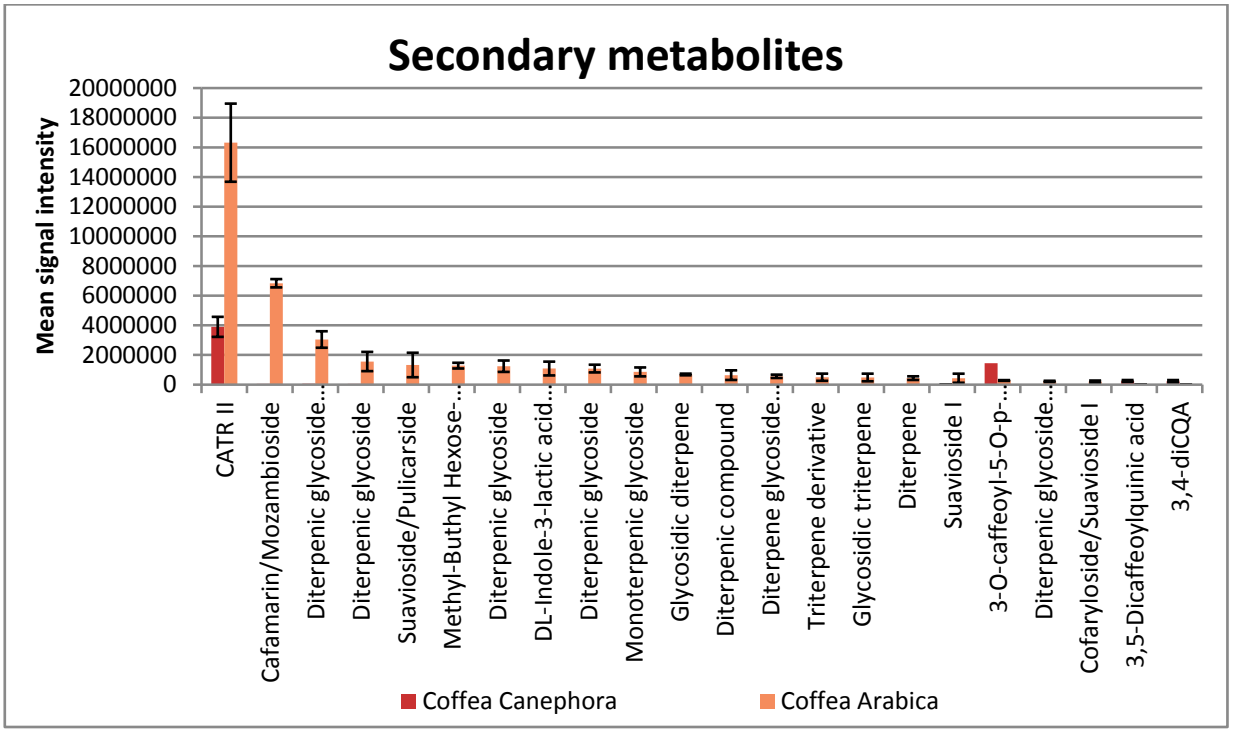

Figure 42 Perisperm mean secondary metabolite profile in *Coffea* samples. Error bars show the standard deviation

4.3.2 Ripening process

The collection of several samples all throughout the ripening process permits the study of the maturation mechanisms of the two species and the evaluation of the major biosynthetic pathways involved. Through principal component analysis, we compared the metabolic profile of the different tissues (pulp and endosperm/perisperm) in the two species at different DAF. In Figure 43 we reported the score plots of the

PCA analyses performed separately on the pulp (A) and the endosperm and perisperm samples (B) of both the species together.

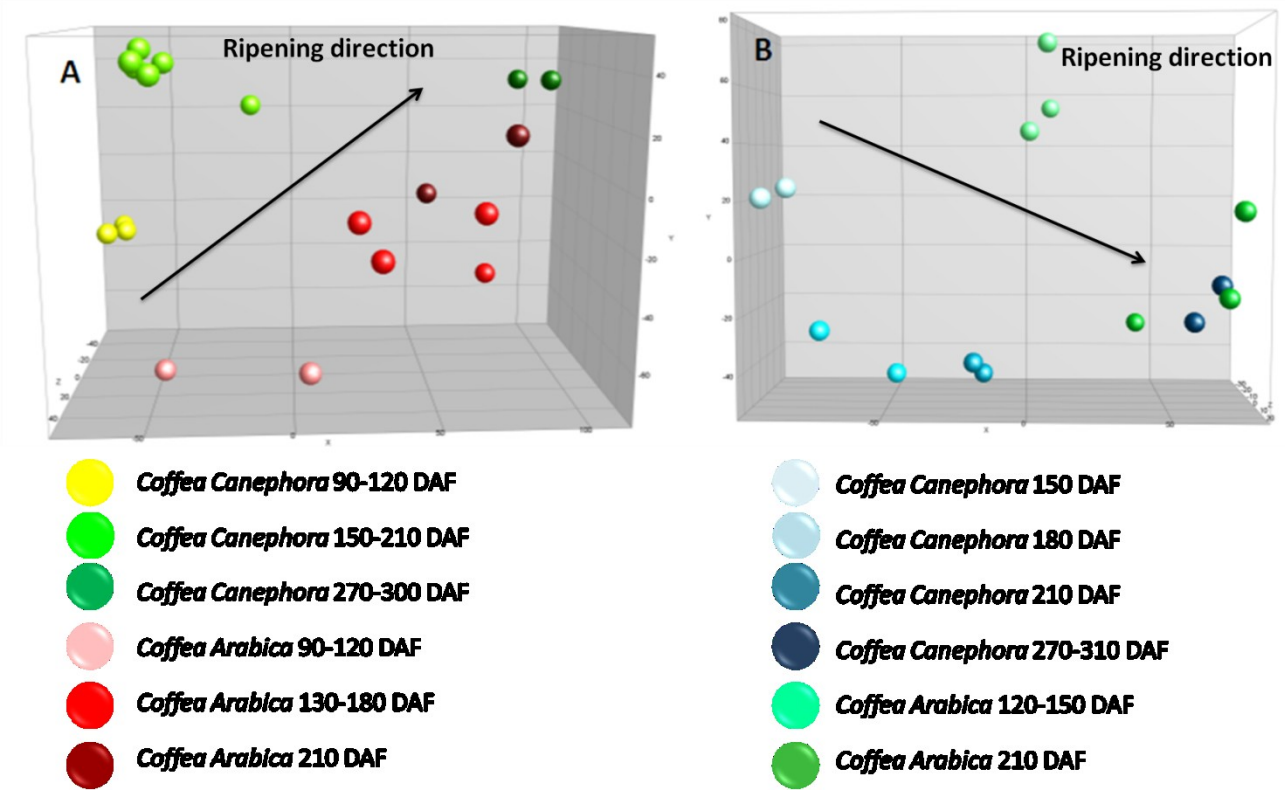

Figure 43 PCA analysis (first three factors) performed separately, with all the variables, on the pulp (A) and the endosperm and perisperm samples (B) of both the species together. Under the PCA plot are reported the figure legends.

Interestingly, no differences were observed among the different harvest seasons, nor in the metabolite intensities and in the ripening process.

As indicated by the arrows, in both the score plots it's possible to identify a ripening direction, which correspond to an increase/decrease of specific metabolites during fruit maturation. Interestingly, the metabolic profile of the last ripening stages of *Coffea Canephora* (270-300 DAF) resulted similar to the one of the 150-180 DAF *Arabica* samples, indicating that there is a difference between the ripening process of the two plants, as already reported [61, 99, 332]. This finding clearly indicates that there are common metabolic mechanisms involved in fruit maturation but which take place at different DAF. In both pulp and endosperm/perisperm, the *C. Canephora* samples seem to dramatically change their metabolic profile from 210 to 270 DAF, abruptly reaching their maturation level; *C. Arabica* shows a more gradual ripening process. The metabolic profile of both the species shows a decrease in almost all the secondary metabolites during ripening, in both pulp and endosperm/perisperm samples; however, this decrease is not necessarily related to a diminution of the amount of metabolites in the fruit while could be a "dilution effect" due to an increase of dimension of the whole fruit. Koshiro et al., (2007) [326], monitoring the concentrations of CGAs during *Coffea Arabica* and *Coffea Canephora* ripening, showed that the absolute content of compounds increase in the seeds and whole fruit while the content (mg/g) tends to decrease during maturation. The decrease seems to be more pronounced in the endosperm/perisperm samples, while, in the pulp, some metabolites showed to increase. In the last two ripening stages of *Coffea Canephora* we observed an increase in the content of epicatechin while, in *Coffea Arabica* pulp, there is an increase in 3-

caffeoyl-5-feruloylquinic acid, 4,5-feruloylcaffeoylquinic acid, 3-caffeoylquinic acid and 4-caffeoylquinic acid.

With respect to the primary metabolites in *Coffea Canephora* and *Arabica* pulp, we observed an increase of the sugar content, especially for sucrose, glucose and fructose. The metabolic profile, as already suggested by the PCA analysis, dramatically change from the 210 to 270 DAF for *Coffea Canephora*, while for *Coffea Arabica* from 180 DAF some changes occurs. The intensities of sucrose, glucose and fructose in mature samples of *Coffea Canephora* and *Coffea Arabica* pulp resulted comparables.

For what concern the endosperm/perisperm samples, during ripening a decrease of concentration for many metabolites is observed in *Coffea Canephora* while in *Coffea Arabica* the main part of metabolites do not show clear changes; however, in both cases, a large increase in sucrose content is observed, which almost double its intensity from the first to the last stage. In the endosperm/perisperm part the final sucrose content in *Coffea Arabica* resulted significantly higher than the one in *Coffea Canephora*.

4.4 Discussion

The analyses conducted on the two *Coffea* species highlighted some interesting differences which could help in the comprehension of the metabolism regulation during coffee fruits development. One of the major observed differences is the wider variety and higher content of hydroxycinnamic derivatives found in *C. Canephora* samples. This finding has been reported in many studies and it is suggested as one of the reason making *C. Arabica* less resistant to pathogens and mechanical stress [96, 318, 322, 326, 333].

The quinic acid levels resulted higher in *C. Canephora* than in *C. Arabica*, in agreement with previous studies [320, 334]. The *C. Canephora* enhanced biosynthesis of quinic acid, which is the main precursor of CGAs, is probably related to the higher CQA production in this species than in *C. Arabica*. The quinic acid level, in the endosperm/perisperm tissue, was high in the early stages while gradually decreasing during ripening. Nevertheless, in the pulp of both species, an oscillation of quinic acid levels was observed, with slightly increased intensities at 180 DAF and 150 DAF in *C. Arabica* and *C. Canephora* respectively. This effect could be related to the transport of quinic acid from the endosperm/perisperm or to the hydrolysis of CQAs into caffeoyl-CoA and quinic acid [334]. The caffeoyl part of chlorogenic acids derives from the phenylalanine, which is converted in coumaroyl/caffeoyl/feruloyl-CoA through the phenylpropanoid pathway; these derivatives are combined with the quinic acid, deriving from the shikimic acid pathway to produce CQAs. Koshiro et al, 2007 [326] demonstrated that the biosynthesis of 5-CQA in *Coffea* fruits starts from the phenylalanine and that the 5-FQA is produced from caffeoyl-CoA through feruloyl-CoA. The biosynthesis of 3- and 4- derivatives is nowadays unclear; their production probably starts from the 5-isomers through not well elucidated reactions; a possible migration of the acyl group has even been suggested [312]. Our data show that the biosynthesis of chlorogenic acids starts early in the developing perisperm and endosperm; the decreased levels of CGAs observed during ripening could be related to an effect of dilution, due to the increased mass of the growing tissues, but also to an effective slight decrease of their biosynthesis. Joet et al., 2009 [330] observed that the absolute content of CGAs could diminish in *Coffea* fruits because of a metabolic re-routing of hydroxycinnamic acids towards lignin biosynthesis. They noticed that the phenylpropanoid pathway could be dissociated into two events during fruit ripening: first, a high amount of CGAs is synthesized; secondly these compounds are catabolized for lignin production, during endosperm hardening. The main enzymes involved in the first steps of the phenylpropanoid pathway don't seem to be active during the lignifications process, indicating that another source of cinnamic monomers is provided.

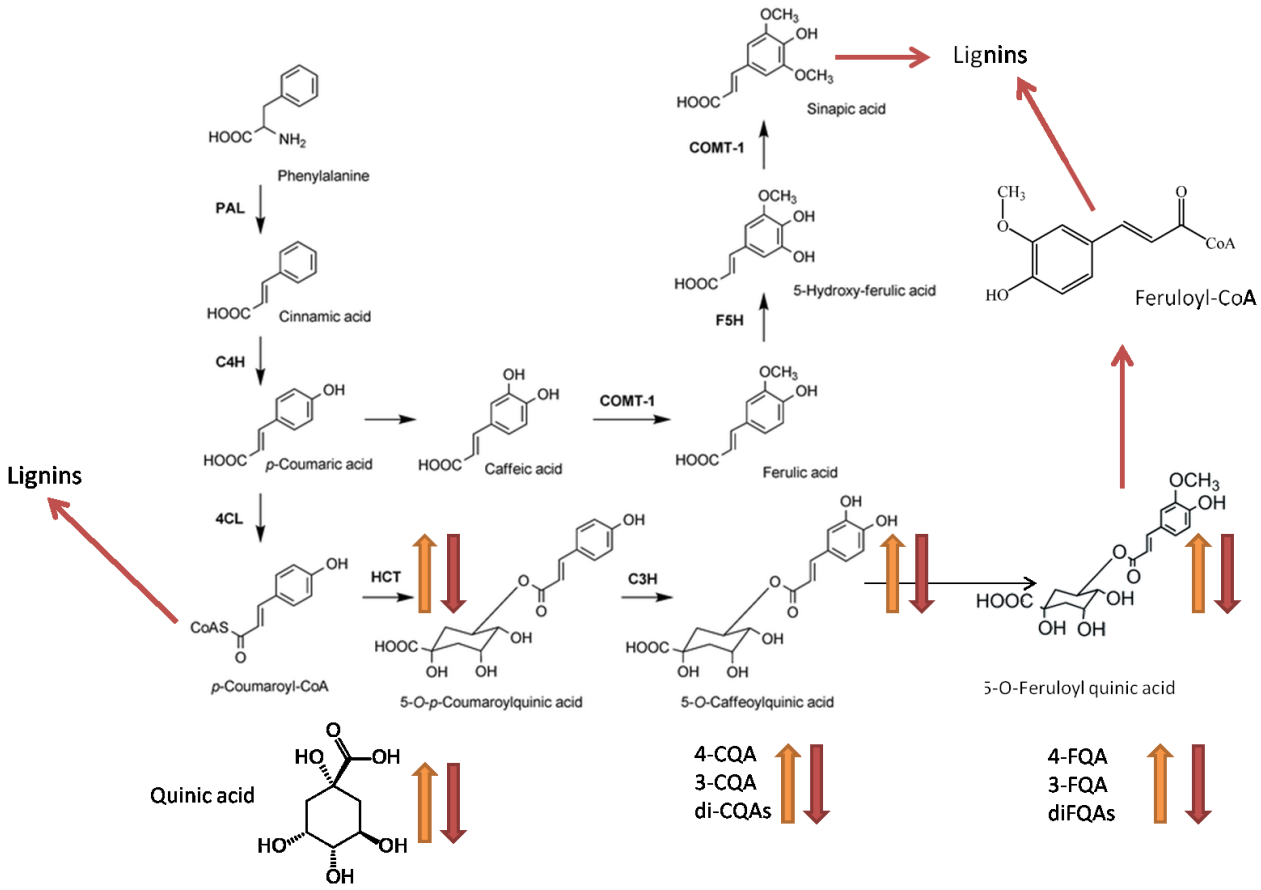

Figure 44 Scheme of main CGAs presents in C. Canephora. Orange arrows indicate the compounds increasing in the early developmental stages while red arrows indicate the changes at maturation level

In our *Coffea* endosperm/perisperm samples, the CQAs and di-CQAs showed a similar profile during ripening: in both *C. Arabica* and *C. Canephora,* an increase of all compounds is observed at 150 DAF, followed by a slight decrease. Previous studies reported that the initial high levels of CGAs in fruits are due to the oxidation of di-CGA compounds, because of the higher oxidase and peroxidase activity in unripe seeds [312]. In our case, however, no inverse correlation was observed among CQAs and di-CQAs. In the pulp, the content of CQAs shows less pronounced variations; in *C. Arabica* 3- and 4-CQAs slightly increases in the last stage while *C. Canephora* presents a small decline of these two compounds, with some variations during the different ripening stages. A study of Salmona et al., 2008 [328] reported than during the last stages of *Coffea* ripening a few genes of the phenylpropanoid pathway are highly expressed; these genes codify for the ferulate 5 hydroxylase, the cinnamoyl CoA reductase and a hydroxyl-cinnamoyl-CoA: shikimate/quinate hydroxycinnamoyl transferase. It is suggested that the expression of the latter gene, which encodes enzymes which esterifed quinic acid with cinnamic acid, could be involved in the biosynthesis or catabolism of CQAs. The expression of ferulate-5-hydoxylase was not explained. The expression of this gene could indicate an enhanced lignin synthesis through the ferulic-sinapic acid way in the last stages. In our case the feruloyl-caffeoyl quinic acid derivatives, mainly presents in the *C. Canephora* endosperm/perisperm, show a decrease during ripening, similar to the other hydroxycinnamic compounds. All CGA compounds could act as substrates for endosperm lignification process; the feruloyl-CoA or the ferulic-sinapic pathway for lignin production is probably more enhanced in *C.Canephora*, due to its higher content in ferulic derivatives [335]. The role of feruloyl compounds in this species could be defensive against pathogens [22]. A scheme of CGAs biosynthesis pathways and lignification process in *C. Canephora* is reported in Figure 44.

The hydroxycinnamic-amino acid conjugates caffeoyl-tryptophan, caffeoyl-tyrosine, coumaroyl-tyrosine, feruloyl-tryptophan, which were found in *C. Canephora* endosperm/perisperm, show almost constant or slight increased levels during the ripening process, demonstrating that their presence is not influenced by the lignifications process. These compounds are recognized as specific markers for the identification of *C. Canephora*, being not detected in the *C. Arabica* [311].

Another relevant class of molecules that we found in *Coffea* samples is the flavon-3-ols catechin/epicatechin and their oligomeric forms. These compounds derive from the phenylpropanoid pathway, through the conversion from flavononols [40].The samples of *C. Arabica* resulted characterized by higher epicatechin levels, while *C. Canephora* showed the presence of catechin as the main flavan-3-ol. Catechin is known to be produced from leucocyanidin through a reaction catalyzed by leucoanthocyanidin-4-reductase (LAR). Conversely, epicatechin derives from cyanidin, by means of anthocyanidin reductase (ANR) [336, 337]. The presence of the ANR pathway has been demonstrated in the leaves of *C. Arabica*. The differential content of catechin and epicatechin clearly indicates a different regulation of LAR and ANR pathways in *C. Canephora* and *C. Arabica* (Figure 45). The oligomerization of both catechin and epicatechin results in the procyanidin constitution.

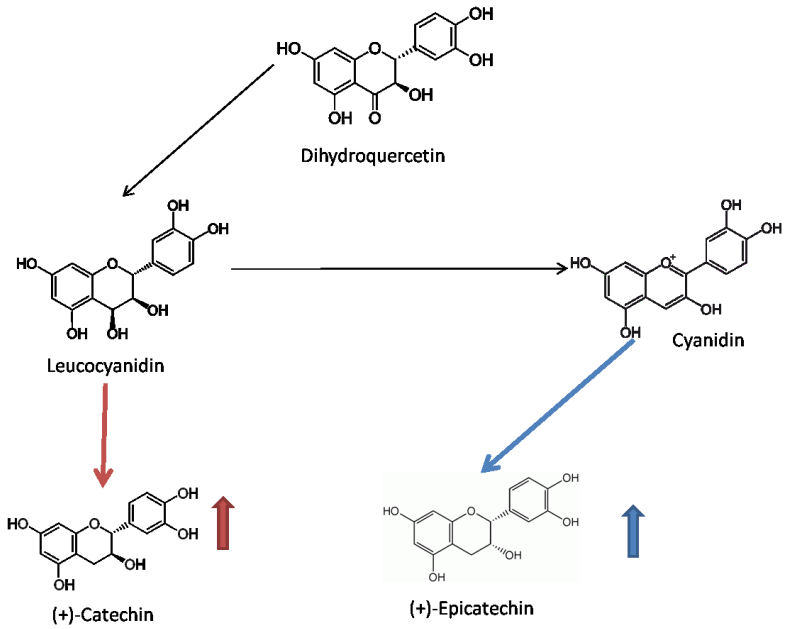

Figure 45 Scheme of catechin and epicatechin biosynthesis in *C. Canephora* (red arrows) and *C. Arabica* (blue arrows).

As already described, in the last stages of both *C.Canephora* and *C. Arabica* pulp, a high increase in sugars is observed. Sucrose in the endosperm/perisperm part of the fruits is present with considerable intensity since the first developing stages, while increasing from 210 DAF for *C. Arabica* and 270 DAF for *C.Canephora* respectively. The metabolism of sucrose resulted different in the pulp: in *C. Arabica* an increase is observed since 180 DAF while *C. Canephora* shows an abrupt augment of sucrose level in the last stage. The increase in sucrose is accompanied by enhanced levels of both glucose and fructose. Rogers et al., 1999 [320] reported low levels of sucrose during endosperm development, in association with a constant decrease of reducing sugars; in accordance with our data, a high increase was observed just before harvest. On the contrary Geromel et al., 2008 [329] reported high and constant levels of sucrose during all the middle ripening of *Coffea racemosa* fruits. Two mechanisms could explain the increase in sucrose levels: firstly sucrose could be directly loaded through sucrose transporters; this hypothesis found confirmation in the enhanced expression of two sucrose transporters genes during the middle developing endosperm phase

[330]. Another mechanism could be the biosynthesis of sucrose directly in the endosperm, starting from the sugar monomers. The biosynthesis could take place in the pulp of *Coffea Arabica*, where the level of glucose decreases until the 150 DAF stage, when the sucrose stars its increase. A previous study [338] identified the major activity of sucrose synthase in coffee fruits at maturity, after 210 DAF. Moreover, the sucrose metabolism in *C. Arabica* seems to be related to the activity of two different enzymes, SUS1, which is mainly active in the early stages, and SUS2, which is probably responsible for the production of sucrose at maturity. SUS1 seems to be involved more in the catabolism of sucrose than in its biosynthesis. Another enzyme deputed to sucrose biosynthesis is sucrose phosphate synthase [328]. Other enzymes as the invertases have even suggested being involved in sucrose metabolism. The sugar trend in *C. Canephora* pulp resulted quite different from *C. Arabica*, with a bimodal distribution of glucose accumulation (90-120 DAF and 270-300 DAF) and a huge increase of sucrose only at 270-300 DAF. In the case of *C. Canephora*, both the external supply of sucrose and the "de novo" biosynthesis seem to be likely. Salmona et al. (2008) [328] indicate the biosynthesis of sucrose as the metabolic point at which the production of cell wall polysaccharides initiates.

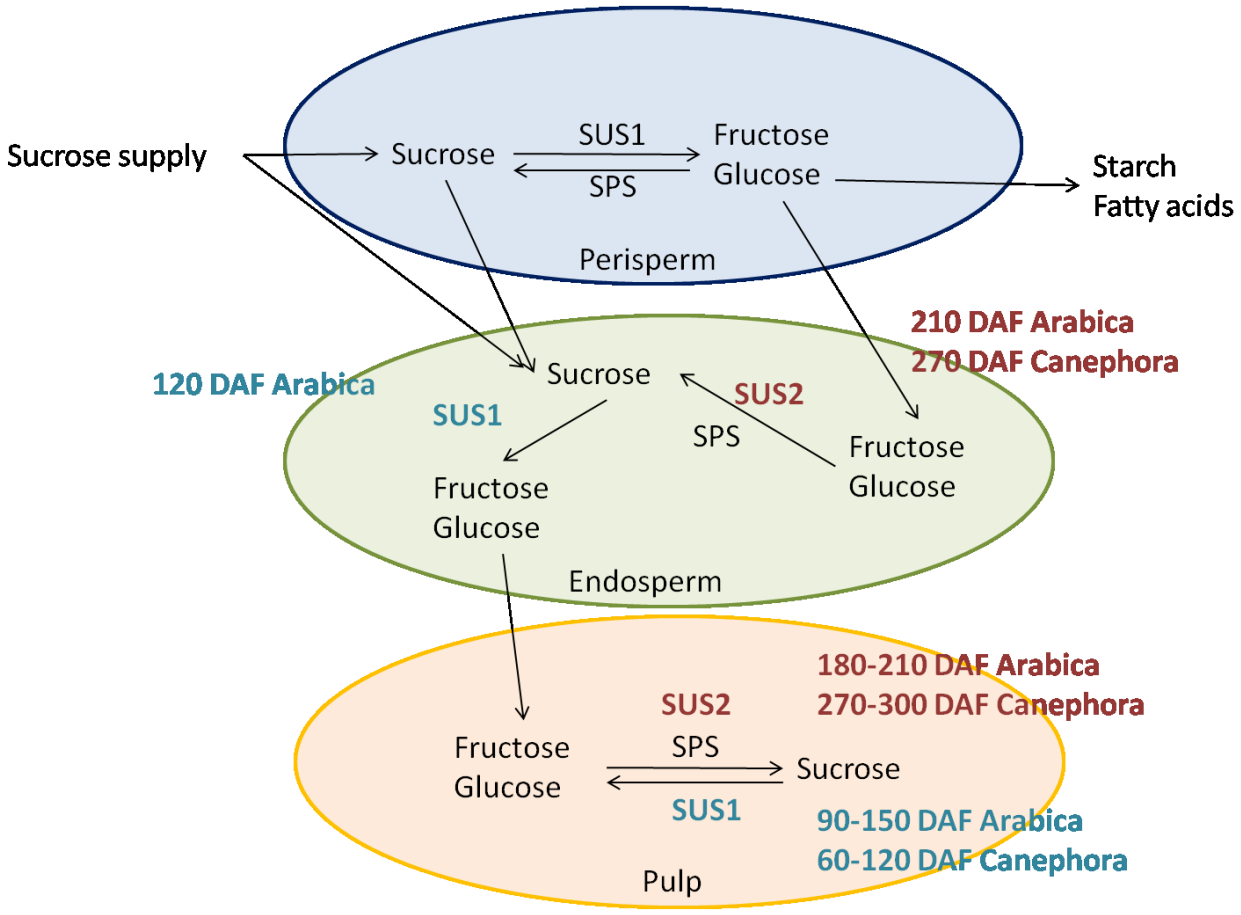

Figure 46 Proposed scheme of main processes involved in sucrose and hexose metabolism in the different part of analyzed coffee fruits. The scheme has been developed in accordance with [330, 332, 338]. SUS1: sucrose synthase 1; SUS2: sucrose synthase2; SPS: sucrose phosphate synthase.

The fructose level shows similar behavior in both pulp species, increasing dramatically together with the increase of sucrose. This could even indicate a role of the sucrose catabolism in the production of fructose in both species. The fairly constant levels of sucrose in the endosperm/perisperm of *C. Arabica*, associated with an abrupt increase of glucose and fructose in the last stage, clearly indicate the existence of sucrose

catabolism activity. The presence of sucrose in the mature coffee beans is particularly important while it has been recognized as one of the main precursor of coffee flavor and aroma [324, 329] and therefore it is crucial in the assessment of coffee quality. The main processes involved in sucrose and hexose metabolism in the different part of coffee fruits are synthesized in Figure 46.

The perisperm is the recognized site in which many important precursors are biosynthesized, to be loaded, stored and transformed in the mature endosperm [328]. The perisperm of *Coffea Arabica* is characterized by high content of glucose and sucrose. This finding is in agreement with previous studies about *C. racemosa* and *C. Arabica* and *C. Canephora* [329, 338]. As already suggested, the increase of hexoses in the early perisperm could be due to an enhanced sucrose catabolism; the sucrose is supplied to the perisperm through the phloem transportation and it is used as hexose source. Glucose and fructose are metabolically activated by phosphatases and used for the biosynthesis of starch and lipids [330]. The *C. Canephora* early perisperm, on the contrary, presents lower sugar amounts and higher myo-inositol, quinic acid and ethanolamine levels. These compounds are main precursors of the secondary metabolites presents in the endosperm. The *C. Arabica* perisperm, since the first stage, presents a more complex secondary metabolite composition, with high CATR II and diterpenic glycoside intensities. The pathways involved in the synthesis of diterpenes seem to be unactive in the *C. Canephora* perisperm, maybe also due to the lower level of available sugars.

The important class of metabolites, diterpene glycosides, was found mainly in the *C. Arabica* species; these compounds are synthesized from the cyclization of geranylgeranylpyrophosphate [339]. The *C. Arabica* species demonstrated a high induction of this pathway with respect to *C. Canephora*. CATR, ATR and their derivatives have been identified in green coffee beans since 1970s and are known inhibitors of mitochondrial oxidative phosphorylation [340, 341]. A recent study of demonstrated that these potentially toxic compounds are fully decomposed during *Coffea* beans roasting [342]. The biosynthesis and role of these compounds have not been well elucidated in *Coffea*, despite their clear and unique specificity for *C. Arabica* species. As for other secondary metabolite classes, more research is needed for the comprehension of the mechanisms which underlie the production and regulation of diterpene glycosides in coffee. The role of feruloyl complexes in *C. Canephora*, the synthesis and regulation of catechin and procyanidins need to be examined in depth in order to understand their effects on coffee quality and properties. In this context, our study provides several data which can contribute to improve the knowledge about Coffea metabolism and ripening process. These data could become even more informative when integrated with genetic and transcriptomic results, in order to confirm and further discuss our findings.

4.5 CONCLUSIONS

The metabolomic analysis of *Coffea Canephora* and *Coffea Arabica* permitted the identification of a wide range of metabolites and consents the characterization of the main peculiarities of the two species. The separated analyses of the different part of coffee fruits resulted highly advantageous in the comprehension of the maturation mechanisms of the plants. The results obtained are in agreement with the major part of the published literature, with the exception of the content of some metabolites (the most important of which, the 4-CQA). The *C. Canephora* showed an enhancement of CGAs content and of feruloyl compound levels, which can be involved in the plant defense mechanisms. These compounds are also known to possess important antioxidant properties and to contribute to the characteristic bitterness and acidity of *Canephora* coffee. *Coffea Arabica* resulted characterized by a wide variety of diterpenic glycosides, some of which already detected in this species, especially presents in the pulp and the perisperm tissues. *Coffea* fruits also showed the presence of flavan-3-ols and procyanidins, in higher number and intensities in *C.*

Canephora than in *C. Arabica*. Interestingly, *C. Arabica* resulted characterized by higher epicatechin levels while *C. Canephora* presents increased catechin amounts. The perisperm tissue showed a unique metabolite pattern, with *C. Canephora* more enriched in precursor compounds as myo-inositol, ethanolamine and quinic acid; on the contrary *C. Arabica* had high sucrose, glucose and fructose levels. The ripening process showed to be slower for *C. Canephora*, reaching maturity at 300 DAF, than for *C. Arabica*, which completed the maturation at 210 DAF. *C. Arabica* maturation process resulted more gradual than the one of *C. Canephora*, which metabolite composition changes abruptly during the last stage. During ripening, the major part of metabolites decreased in both species, while an enhancement of sucrose and hexose sugars was observed. The decreased levels of CQAs were probably related to the higher lignin biosynthesis in the last ripening stages, during endosperm hardening. The observation of metabolite intensities allows hypothesizing the regulation of the major metabolic pathways involved in the ripening process. In this study the application of metabolomic analysis proved high potential in the discrimination of coffee species and in the identification of their peculiarities. The present work shows some interesting biochemical outline of the main coffee metabolites, which directly lead to coffee quality and characteristics; the here presented results could be useful for further coffee metabolic research, especially if associated with genetic and transcriptomic data.

5 General discussion

5.1 METHOD DEVELOPMENT

In this work we presented the development and application of different analytical strategies aimed at the characterization of the plant metabolome. Three different studies have been reported, showing how plant metabolomic could constitute a powerful tool for different purposes. Taking advantage from the use of a HRMS instrument, we were able to detect and characterize a high number of metabolites. For the major part of analyses, we employed the HPLC-HRMS technology, which permits a good separation of multiple compounds, their unambiguous and simultaneous identification and a more sensitive quantification, compared to traditional methods [45, 101, 162, 232, 236, 343, 344]. In comparison to the GC-MS technique, which is confined to the analysis of volatile compounds and could require several sample treatment steps, the LC-MS is free from these limitations and provide high sensitivity, accuracy and reproducibility [108, 232, 236, 343, 344]. The use of high resolution detector coupled to a chromatographic separation, moreover, permits a better identification of compounds based on the retention time and on the high accuracy of the molecular ion (m/z) measurement, thus simplifying the sample treatment procedure and reducing the risk of interfering compounds [72]. Currently, there's a trend toward the use of full scan MS experiments, crucially important to fill in the gap between the conventional target analysis and the metabolomic approach, by which identification of new compounds, data reinterrogation and a more comprehensive metabolome profiling is possible [101]. In this work, the development of targeted methods for the determination of key molecules was necessary, due to their low abundance which prevents their determination without an optimized procedure. A quantitative method for the determination of SA, SHA and JA was developed and validated. This step resulted highly time consuming, because of the different chemical properties of the compounds, the high matrix effect and their different abundance. Moreover, the unavailability of a certified material for the quantification of these metabolites limits the validation step, which required the use of a part of samples. The use of a not-blank matrix with unknown analyte content compromised the reproducibility of the process. The method we proposed is, to our known, the first which combines the determination of these three compounds in a relative short time analysis. The method was finally applied to the samples, resulting informative and adequate to the study purposes. Another semi-quantitative method was then developed to determine the content of fatty acids. In this case, the GC-MS technique was selected as the more suitable, after a failed LC-MS tentatively approach. The method was only partially validated, verifying the signal linearity and the trans-esterification efficiency. The solvent employed for the extraction was probably not- completely adequate, considering the apolarity of many lipids, and could result in a partial recovery of the compounds; however, the results obtained by fatty acid analysis resulted informative and in accordance with the metabolomic results. In the future, we would like to better assess the validity of this method, to validate the results and verify its potential applicability for other studies. A metabolomic analysis method was then created and applied to *Glycyrrhiza* and *N. Langsdorfii* samples. The method was developed following the protocol of De Vos et al., (2007) [105], and firstly tested on *Glycyrrhiza sp.* samples. Despite the use of different plant matrices, the instrumental method was maintained almost unaltered, in order to maintain data comparability and start the creation of an internal metabolite library. The metabolomic method, with an easier and faster procedure, proved highly informative, providing many useful insights in plant biochemistry.

While the validity of results in quantitative methods is assured by the estimation of quality parameters (as recovery and precision) in the metabolomic analysis the main crucial issue is represented by repeatability. Without quantification, the results and their interpretations are based only on the comparison among mass

spectra and chromatograms. In the quantitative plant analysis the main concern is generally the presence of interfering substances or of the matrix effect; these factors can compromise the results even when an internal standard and a FR are used. On the contrary, metabolomics requires avoiding any possible source of variation and of non-systematic errors, from the plant material weighting to the mass detection.

These two different approaches, which have their main origin from different scientific disciplines, have been integrated in this work, showing that their combination could provide interesting results. The developed methods were applied to different plant matrices, showing their versatility for different purposes.

5.2 METABOLOME PROFILING

Three independent studies were presented here, sharing the application of metabolomic approach for the investigation of the plant biochemistry. In the second chapter we proposed the analysis of wild and transgenic *N. Langsdorfii* plants, exposed to different abiotic stresses, in order to evaluate their stress response. In the third chapter we reported the analysis of two different *Glycyrrhiza* species, to identify characteristics and peculiarities in relation to the potential bioactivity. In the fourth chapter we presented a study of *Coffea Arabica* and *Coffea Canephora* fruits, collected at different ripening stages and in two different harvest seasons. The analyses permitted the evaluation of the chemical composition of plants, revealing some interesting findings leading to metabolic pathway regulation depending from species, growth and stress exposition. The shikimate pathway was confirmed as the central plant regulating system, leading to the production of aromatic amino acids, antioxidants (HCAs, flavonoids, CQAs) and phytohormones. The phenyl propanoid pathway, which directly derives from the shikimate way, represented another key regulation point, in relation to lignification rate and antioxidants production. In *Nicotiana langsdorfii* the production of caffeoyl-quinic acids and HCA-amine conjugates was a direct response of chemical stress exposition, especially in WT and RolC plants; the conjugation of HCA with amino acids resulted induced by the application of water stress. In *Coffea*, the presence of caffeoyl-feruloyl- quinic acids, of HCA-amines and HCA-amino acids was a distinguishing trait of *C. Canephora*, with respect to *C. Arabica*. In *C. Canephora,* tyramine and tryptophan, which are the main derivatives of shikimate pathway, were the preferred compound for conjugation. The amines, beside their role in stress response, have been suggested to influence the growth and ripening of coffee fruits, probably in relation to CGA biosynthesis [321]. The phenylpropanoid pathway also showed to influence the production of lignins, as a response of *N. Langsdorfii* plants to water and chemical stress exposition, or in *Coffea* during seed development [34, 332]. Another main regulating system in *N. Langsdorfii* appears the lipid metabolism which, controlling the presence of specific compounds (as SQDG and DGDG), could regulate the membrane stability and the release of free fatty acids for the biosynthesis of other compounds (JA) [203, 206, 262, 263]. The role of fatty acids need to be better investigated in relation to short and long-term heat stress, in order to verify their involvement in signaling mechanisms, in the stability of membranes or in the production of JA [226, 345, 346]. The terpenic pathway showed to be involved in heat stress response, leading to the production of glykoalkaloids, with different inductions among WT, RolC and GR plants. The terpenic pathway also demonstrate to have a prominent role in *Coffea Arabica*, leading to the production of many diterpenic glycosides, which characterize this species [99, 311, 341, 347]; in *Glycyrrhiza* roots, the triterpenes and their derivatives were confirmed as main constituents, and the induction of specific compounds in relation to species and variety was observed [64, 109]. Many sugar-containing compounds were found in all the analyzed matrices; in *N. Langsdorfii* the acyl-sugar accumulation was a direct consequence of heat stress, probably related also to the development of trichomes, specific structures involved in pathogen and insect defense [209, 212, 214]. In *Glycyrrhiza* roots the main part of triterpenes

and flavonoids resulted in their glycosylated form, biologically acting as anti-microbial compounds or nodulating inducers [348]; the glycosylation rate is one of the parameter affecting the bioactivity of flavonoids and the sweetening properties of saponins [37, 291, 349]. In *Coffea,* the metabolism of sugars constituted one of the main differences between the two species and their ripening process, influencing the coffee taste and aroma [324, 329].

6 Concluding remarks

In this thesis we presented different approaches in the studying of the metabolic profiling of plant material. In the three different works, we investigate the metabolome of plants, developing analytical methodologies to fulfill the study purposes. In the second chapter we presented the analysis of wild and transgenic *N. Langsdorfii* plants, exposed to heat, water and chemical stresses. The obtained results suggested a possible role of GR genetic modification in the increase plant resistance to Cr(VI) and water stresses, showing the lower changes in the metabolomic profile and especially in the content of antioxidants and phytohormones. Acyl-sugars and lipids were identified as the main metabolites affected during heat stress, together with increased levels of glykoalkaloids. These findings suggest a role of thricomes as accumulating or defense structures against heat stress. The role of fatty acid need to be further investigated, particularly in relation to their saturation degree as a response to heat stress. In *Nicotiana Langsdorfii* analyses, the integration of quantitative and metabolomic data represented a powerful strategy, permitting the investigation of low-abundant key compounds and of the main components of plant metabolome involved in stress response.

In the last decades a growing interest has been developed in the assurance of food quality, safety and traceability, due to a raised consumer demand and to the even larger globalization of food movement and processing [97]. It is therefore crucial to develop adequate strategies, to fulfill these claims. The food analysis represents today one of the most important application area of analytical chemistry. In this study, through metabolomic analysis, we were able to characterize different species of plants, commonly used worldwide as food supply or for pharmacological applications. In the third chapter metabolomics was applied to different species and varieties of *Glycyrrhiza*, in order to highlight their differences. The main constituent of licorice were confirmed to be glycosidic flavonoids and triterpenic saponins, well known for their potential health benefits [39, 264, 267–269].The chemical composition of plants resulted similar but some specie-specific metabolites were identified. Glabridin resulted specific for *G. Glabra* while Glycycoumarin was present only in *G. Uralensis*. The difference between the two species was mainly attributed to their phenolic profile, with some key compounds being present in higher concentrations in *Uralensis* (liquiritin, isoliquiritin and their sulfate derivatives) and some other in *Glabra* (prenylated flavonoids or chalcones). Some compounds, not previously found in these plants, were even detected, resulting specific for *G. Glabra*. For several of the detected compounds of both species a biological activity has been suggested. However, for many others, no data are available about their potential effects on health. Beside the known and well characterized constituents of licorice, many others are presents, constituting a relevant part of the potential bioactive molecules. For this reason, we suggest the development of specific bioactivity tests for these compounds and, eventually, the evaluation of their potential synergic or antagonistic effects. In the third chapter we presented a study about two *Coffea* species, integrating two different metabolomic approaches: one aiming to the identification of the primary metabolites, by means of GC-TOFMS, and the other aiming to the identification of secondary metabolites, by means of LC-HRMS. This combined approach results in the identification of the main compounds involved in the species characterization and ripening process and permitted to highlight a few differentially regulated pathways. The *C. Arabica* resulted enriched in diterpenes as CATR I, II and Mozambioside, which were not present in *C. Canephora*. Moreover, the content of sugars (sucrose and glucose) resulted higher in *C. Arabica*, especially in the endosperm part of the fruit. *C. Canephora* was on the other hand more enriched in CGAs as caffeoylquinic acids and other ferulic derivatives. We can also observe the ripening process of the plants, highlighting a faster and more gradual ripening for *C. Arabica* with respect to

Canephora, which metabolic profile changes dramatically during the last stages. No relevant differences were observed between the two harvest season considered, strengthening the independence of results from slight difference in climatic conditions.

Concluding, this study showed the potential of the integration of different approaches in analytical chemistry, contributing to the comprehension of plant stress response and suggesting some possible application of the genetic modifications tested. Moreover this work provided useful information about licorice and coffee, demonstrating the potential of metabolomic methodology as a tool in the food characterization and quality assurance. The study represents a good starting point for future works in the field of foodomics and of system biology, highlighting original findings, not reported before, which should be better investigated.

7 Acknowledgements

At the end of this work I would like to thank my tutor, Prof. Gabriele Capodaglio, for give me the possibility to develop my own project and for the scientific support during all the phases of work

I am very grateful to Prof. Robert Hall and Dr. Ric De Vos who host me in Wageningen at the Plant Research International (PRI) and let me the possibility to know and understand metabolomic methodology; thanks are due to them and to the entire metabolomic research group for their kindness during my Dutch staying. All the data about coffee reported here are property of PRI, University of Mato Grosso do Sul (UFMS) and Embrapa Genetic Resources and Biotechnology and their use has been granted to me by Prof. Ric de Vos, Prof. Robert Hall and Dr. Alan Andrada.

All the Coffea samples have been analyzed, produced and processed at the PRI, under the supervision of Dr. Carlos Alexandre Carollo, of the Laboratory of Pharmacognosy of the University du Mato Grosso do Sul, to whom I am very grateful. Thanks are due also to Dr. Alan Andrada from the Embrapa Genetic Resources and Biotechnology, Brazilia-DF, Brazil who collected the samples. The data elaboration and discussion is the fruit of my work at the PRI and the results and findings I reported have to be considered under my responsibility.

I also acknowledge the Department of Biology of the University of Florence for the production of *Nicotiana Langsdorfii* samples which have been analyzed at the University of Venice in the context of the National Interest project (PRIN) “Metodologie analitiche innovative per lo studio della risposta metabolica a stress idrico e stress termico di piante di *Nicotiana langsdorffii* a struttura ormonale geneticamente modificata”, coordinated by Prof. Roger Fuoco of the University of Pisa. I thank prof. Michele Buiatti for his help in data interpretation.

I would like to thank also Prof. Oreste Piccolo who kindly provided the Glycyrrhiza samples from Calabria and help the supply of samples from the State University of Astrakhan, collected by Oxana Astafyeva and prof. Michael Egorov.

I thank Martina Ranaldo, for the fatty acids analysis in *Nicotiana Langsdorfii* and Marta Radaelli for the scientific and human support. I also thank Alice Forlin who during her master degree collaborated in the analysis of *Nicotiana Langsdorfii* samples, especially in the data processing step.

I also thank Prof. Andrea Gambaro, Dr. Roberta Zangrando and Dr. Elena Barbaro who scientifically help and support me during the work development.

Many thanks are due also to Nicola Miotello, who let me to survive to the University burocracy.

Finally I thank my good friends, my family, and my husband for being in my life.

8 Bibliography

1. Villas-Bas SG, Roessner U, Hansen M a. E, et al. (2007) Metabolome analysis: an introduction. doi: 10.1002/0470105518

2. Whitton J (2013) Plant Biodiversity , Overview. In: Levin SA (ed) Encycl. Biodivers., II. Elsevier, pp 56–64

3. National Research Council (1992) Plant biology research and training for the 21st century.

4. Fernie AR (2007) The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. Phytochemistry 68:2861–80. doi: 10.1016/j.phytochem.2007.07.010

5. Cox PA (2013) Pharmacology , Biodiversity and. In: Levin SA (ed) Encycl. Biodivers., II. Elsevier, pp 703–715

6. Bourgaud F, Gravot a., Milesi S, Gontier E (2001) Production of plant secondary metabolites: a historical perspective. Plant Sci 161:839–851. doi: 10.1016/S0168-9452(01)00490-3

7. Lisec J, Schauer N, Kopka J, et al. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc 1:1–10. doi: 10.1038/nprot.2006.59

8. Gibon Y, Rolin D, Deborde C (2012) New Opportunities in Metabolomics and Biochemical Phenotyping for Plant Systems Biology. In: Roessner U (ed) Metabolomics. InTEch, pp 213–240

9. Trenerry VC, Rochfort SJ (2010) Natural Products Research and Metabolomics. In: Mander L, Liu H-W (Ben) (eds) Compr. Nat. Prod. II, Mod. Methods Nat. Prod. Chem. Elsevier, pp 595–628

10. Dunn WB, Erban A, Weber RJM, et al. (2013) Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics 9:44–66. doi: 10.1007/s11306-012-0434-4

11. Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol life Sci 69:3225–43. doi: 10.1007/s00018-012-1091-5

12. Kroymann J (2011) Natural diversity and adaptation in plant secondary metabolism. Curr Opin Plant Biol 14:246–51. doi: 10.1016/j.pbi.2011.03.021

13. Fiehn O (2002) Metabolomics-the link between genotypes and phenotypes. Plant Mol Biol 48:155–71.

14. Moco S, Schneider B, Vervoort J (2009) Plant Micrometabolomics□: The Analysis of Endogenous Metabolites Present in a Plant Cell or Tissue. J Proteome Res 8:1694–1703.

15. Sumner LW, Mendes P, Dixon R a (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817–836. doi: 10.1016/S0031-9422(02)00708-2

16. Allwood JW, De Vos RCH, Moing A, et al. (2011) Plant metabolomics and its potential for systems biology research background concepts, technology, and methodology. In: Jameson D, Verma M, Westerhoff H V. (eds) Methods Enzymol., 1st ed. Elsevier Inc., pp 299–336

17. Skalicka-Wozniak K, Hanos MŁ, Glowniak K (2011) Phytochemistry, Phytopharmacology and the Biological Role of Plant Metabolites. In: Waksmundzka-Hajnos M, Sherma J (eds) High Perform. Liq. Chromatogr. Phytochem. Anal. CRC Press Taylor & Francis group, pp 89–106

18. Hounsome N, Hounsome B, Tomos D, Edwards-Jones G (2008) Plant metabolites and nutritional quality of vegetables. J Food Sci 73:R48–65. doi: 10.1111/j.1750-3841.2008.00716.x

Bibliography

19. Van der Kooy F, Verpoorte R, Marion Meyer JJ (2008) Metabolomic quality control of claimed anti-malarial Artemisia afra herbal remedy and A. afra and A. annua plant extracts. South African J Bot 74:186–189. doi: 10.1016/j.sajb.2007.10.004

20. Mohr H, Schopfer P (1994) Plant Physiology, I. Springer, New York

21. Buchanan BB, Gruissem W, Jones RL (2003) Biochimica e biologia molecolare delle piante. Zanichelli

22. Nuringtyas TR, Choi YH, Verpoorte R, et al. (2012) Differential tissue distribution of metabolites in Jacobaea vulgaris, Jacobaea aquatica and their crosses. Phytochemistry 78:89–97. doi: 10.1016/j.phytochem.2012.03.011

23. Beckles DM, Roessner U (2012) Plant metabolomics: applications and opportunities for agricultural biotechnology. In: Altman A, Hasegawa PM (eds) Plant Biotechnol. Agric., First Edit. Elsevier Inc., pp 67–82

24. Nuutila AM, Oksman-Caldentey K-M (2003) Secondary Metabolism in Plant Cell Cultures. In: Brau T (ed) Encycl. Appl. Plant Sci. Academic Press Elsevier, pp 1388–1395

25. Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, V. W. H. Freeman and Company, New York

26. Heldt H-W, Piechulla B (2011) Sulfate assimilation enables the synthesis of sulfur containing compounds. Plant Biochem. Elsevier Academic Press, pp 323–335

27. Weaver LM, Klaus MH (1997) Dynamics of the shikimate pathway. Trends Plant Sci 1385:346–351.

28. Rawsthorne S (2002) Carbon flux and fatty acid synthesis in plants. Prog Lipid Res 41:182–96.

29. Hildebrand DF, Yu K (2003) Acyl Lipids. In: Brau T (ed) Encycl. Appl. Plant Sci. Academic Press Elsevier, pp 464–477

30. Schmid J, Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. Phytochemistry 39:737–749.

31. Bartwal A, Mall R, Lohani P, et al. (2013) Role of Secondary Metabolites and Brassinosteroids in Plant Defense Against Environmental Stresses. J Plant Growth Regul 32:216–232. doi: 10.1007/s00344-012-9272-x

32. El-Seedi HR, El-Said AM a, Khalifa S a M, et al. (2012) Biosynthesis, natural sources, dietary intake, pharmacokinetic properties, and biological activities of hydroxycinnamic acids. J Agric Food Chem 60:10877–95. doi: 10.1021/jf301807g

33. Jaiswal R, Patras MA, Eravuchira PJ, Kuhnert N (2010) Profile and Characterization of the Chlorogenic Acids in Green Robusta Coffee Beans by LC-MS n□: Identification of Seven New Classes of Compounds. J Agric Food Chem 58:8722–8737. doi: 10.1021/jf1014457

34. Torras-Claveria L, Jáuregui O, Codina C, et al. (2012) Analysis of phenolic compounds by high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry in senescent and water-stressed tobacco. Plant Sci 182:71–78. doi: 10.1016/j.plantsci.2011.02.009

35. Havsteen BH (2002) The biochemistry and medical significance of the flavonoids. Pharmacol Ther 96:67 – 202.

36. Stalikas CD (2007) Extraction, separation and detection methods for phenolic acids and flavonoids. J Sep Sci 30:3268–3295. doi: 10.1002/jssc.200700261

37. Marston A HK (2006) Flavonoids: chemistry, biochemistry, applications, 10th ed. Taylor and Francis group, Boca Raton

Bibliography

38. Cushnie TPT, Lamb AJ (2005) Antimicrobial activity of flavonoids. Int J Antimicrob Agents 26:343–356. doi: 10.1016/j.ijantimicag.2005.09.002

39. Asl MN, Hosseinzadeh H (2008) Review of Pharmacological Effects of Glycyrrhiza sp. and its Bioactive Compounds. Phyther Res 22:709–724. doi: 10.1002/ptr

40. Kong D-Y (2012) Flavonoids. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 169–188

41. Shoji T (2014) Chemical Properties, Bioavailability, and Metabolomics of Fruit Proanthocyanidins. Polyphenols Hum. Heal. Dis. Elsevier Inc., pp 339–351

42. Hu C-Q (2012) Lignans. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 225–245

43. Willför SM, Smeds a I, Holmbom BR (2006) Chromatographic analysis of lignans. J Chromatogr A 1112:64–77. doi: 10.1016/j.chroma.2005.11.054

44. Chen Z-N (2012) Coumarins. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 205–224

45. Pan X, Welti R, Wang X (2008) Simultaneous quantification of major phytohormones and related compounds in crude plant extracts by liquid chromatography-electrospray tandem mass spectrometry. Phytochemistry 69:1773–81. doi: 10.1016/j.phytochem.2008.02.008

46. Kumar D (2014) Salicylic acid signaling in disease resistance. Plant Sci. doi: 10.1016/j.plantsci.2014.04.014

47. Degenhardt J, Gershenzon J (2003) Terpenoids. In: Brau T (ed) Encycl. Appl. Plant Sci. Academic Press Elsevier, pp 500–504

48. Tohge T, de Souza LP, Fernie AR (2014) Genome-enabled plant metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci 966:7–20. doi: 10.1016/j.jchromb.2014.04.003

49. Zhao W-M (2012) Saponins. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 125–145

50. Chen Z-L (2012) Sesquiterpenoids. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 81–100

51. Dinda B, Debnath S, Mohanta BC, Harigaya Y (2010) Naturally Occurring Triterpenoid Saponins. Chem Biodivers 7:2327–2580.

52. Sparg SG, Light ME, Staden J Van (2004) Biological activities and distribution of plant saponins. J Ethnopharmacol 94:219–243. doi: 10.1016/j.jep.2004.05.016

53. Ramirez-Erosa IJ (2008) Studies on triterpene saponins from Saponaria vaccaria seed and their apoptosis-inducing effect on human cancer cell lines.

54. Andrisano V, Bonazzi D, Cavrini V (1995) HPLC analysis of liquorice triterpenoids-applications to the quality control of pharmaceuticals. J Pharm Biomed Anal 13:597–605.

55. Zheng Y, Qi L, Zhou J, Li P (2010) Structural characterization and identification of oleanane- type triterpene saponins in Glycyrrhiza uralensis Fischer by rapid-resolution liquid chromatography coupled with time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 24:3261–3270. doi: 10.1002/rcm

Bibliography

56. Twyman RM, Verpoorte R, Memelink J, Christou P (2003) Alkaloids. In: Brau T (ed) Encycl. Appl. Plant Sci. Academic Press Elsevier, pp 494–500

57. Xu R-S, Ye Y (2012) Alkaloids. In: Xu R, Ye Y, Zhao W (eds) Introd. to Nat. Prod. Chem. CRC Press Taylor & Francis group, pp 55–79

58. O'Connor SE (2010) Alkaloids. Nat. Prod. Struct. Divers. Second. Metab. Organ. Biosynth. Elsevier, pp 977–1007

59. Ritala A, Dong L, Imseng N, et al. (2014) Evaluation of tobacco (Nicotiana tabacum L. cv. Petit Havana SR1) hairy roots for the production of geraniol, the first committed step in terpenoid indole alkaloid pathway. J Biotechnol 176:20–8. doi: 10.1016/j.jbiotec.2014.01.031

60. Fraire-velázquez S, Balderas-hernández VE (2013) Abiotic Stress in Plants and Metabolic Responses. In: Vahdati DK (ed) Abiotic Stress - Plant Responses Appl. Agric. INTECH, pp 25–48

61. Oestereich-Janzen S (2010) Chemistry of Coffee. In: Liu LM and H-W (Ben) (ed) Compr. Nat. Prod. II. Elsevier, pp 1085–1096

62. Steinmann D, Ganzera M (2011) Recent advances on HPLC/MS in medicinal plant analysis. J Pharm Biomed Anal 55:744–757. doi: 10.1016/j.jpba.2010.11.015

63. Funari CS, Eugster PJ, Martel S, et al. (2012) High resolution ultra high pressure liquid chromatography-time-of-flight mass spectrometry dereplication strategy for the metabolite profiling of Brazilian Lippia species. J Chromatogr A 1259:167–78. doi: 10.1016/j.chroma.2012.03.069

64. Wang Y, Yang Y (2007) Simultaneous quantification of flavonoids and triterpenoids in licorice using HPLC. J Cromatography B 850:392–399. doi: 10.1016/j.jchromb.2006.12.032

65. Hennell JR, Lee S, Khoo CS, et al. (2008) The determination of glycyrrhizic acid in Glycyrrhiza uralensis Fisch . ex DC . (Zhi Gan Cao) root and the dried aqueous extract by LC – DAD. J Pharm Biomed Anal 47:494–500. doi: 10.1016/j.jpba.2008.01.037

66. Gaquerel E, Heiling S, Schoettner M, et al. (2010) Development and Validation of a Liquid Chromatography - Electrospray Ionization - Time-of-Flight Mass Spectrometry Method for Induced Changes in Nicotiana attenuata Leaves during Simulated Herbivory. J Agric Food Chem 58:9418–9427. doi: 10.1021/jf1017737

67. Monakhova YB, Schütz B, Schäfer H, et al. (2013) Validation studies for multicomponent quantitative NMR analysis: the example of apple fruit juice. Accredit Qual Assur 19:17–29. doi: 10.1007/s00769-013-1026-3

68. Fiehn O (2008) Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. Trends Anal Chem 27:261–269.

69. Matuszewski BK, Constanzer ML, Chavez-Eng CM (2003) Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem 75:3019–30.

70. Essers AJA, Alink GM, Speijers GJA, et al. (1998) Food plant toxicants and safety Risk assessment and regulation of inherent toxicants in plant foods. Environ Toxicol Pharmacol 5:155–172.

71. Isbrucker RA, Burdock GA (2006) Risk and safety assessment on the consumption of Licorice root (Glycyrrhiza sp .), its extract and powder as a food ingredient , with emphasis on the pharmacology and toxicology of glycyrrhizin. Regul Toxicol Pharmacol 46:167–192. doi: 10.1016/j.yrtph.2006.06.002

72. Wu H, Guo J, Chen S, et al. (2013) Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography – mass spectrometry. J Pharm Biomed Anal 72:267–291. doi: 10.1016/j.jpba.2012.09.004

73. Rauchensteiner F, Matsumura Y, Yamamoto Y, et al. (2005) Analysis and comparison of Radix Glycyrrhizae (licorice) from Europe and China by capillary-zone electrophoresis (CZE). J Pharm Biomed Anal 38:594–600. doi: 10.1016/j.jpba.2005.01.038

74. Yuliana ND, Khatib A, Choi YH, Verpoorte R (2011) Metabolomics for Bioactivity Assessment of Natural Products. Phyther Res 25:157–169.

75. Shanker K, Fatima A, Negi AS, et al. (2007) RP-HPLC Method for the Quantitation of Glabridin in Yashti-madhu (Glycyrrhiza glabra). Chromatographia 65:771–774. doi: 10.1365/s10337-007-0230-x

76. Zhou Y, Yu H, Zhang Y, et al. (2010) Evaluation on intrinsic quality of licorice influenced by environmental factors by using FTIR combined with 2D-IR correlation spectroscopy. J Mol Struct 974:127–131. doi: 10.1016/j.molstruc.2010.03.017

77. Farag MA, Porzel A, Wessjohann LA (2012) Comparative metabolite profiling and fingerprinting of medicinal licorice roots using a multiplex approach of GC – MS , LC – MS and 1D NMR techniques. Phytochemistry 76:60–72. doi: 10.1016/j.phytochem.2011.12.010

78. Lin I, Lee M, Chuang W (2006) Application of LC/MS and ICP/MS for establishing the fingerprint spectrum of the traditional Chinese medicinal preparation Gan-Lu-Yin. J Sep Sci 29:172 – 179. doi: 10.1002/jssc.200500147

79. Wang Y, He S, Cheng X, et al. (2013) UPLC-Q-TOF-MS/MS fingerprinting of Traditional Chinese Formula SiJunZiTang. J Pharm Biomed Anal 80:24–33. doi: 10.1016/j.jpba.2013.02.021

80. Pongsuwan W, Fukusaki E, Bamba T, et al. (2007) Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. J Agric Food Chem 55:231–236. doi: 10.1021/jf062330u

81. Goodacre R, York E V., Heald JK, Scott IM (2003) Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. Phytochemistry 62:859–863. doi: 10.1016/S0031-9422(02)00718-5

82. Zhang X-J, Qiu J-F, Guo L-P, et al. (2013) Discrimination of multi-origin chinese herbal medicines using gas chromatography-mass spectrometry-based fatty acid profiling. Molecules 18:15329–43. doi: 10.3390/molecules181215329

83. Mungur R, Glass ADM, Goodenow DB, Lightfoot DA (2005) Metabolite Fingerprinting in Transgenic Nicotiana tabacum Altered by the Escherichia coli Glutamate dehydrogenase Gene. J Biomed Biotechnol 2:198–214. doi: 10.1155/JBB.2005.198

84. Nakabayashi R, Saito K (2013) Metabolomics for unknown plant metabolites. Anal Bioanal Chem 405:5005–5011. doi: 10.1007/s00216-013-6869-2

85. Moco S, Bino RJ, De Vos RCH, Vervoort J (2007) Metabolomics technologies and metabolite identification. Trends Anal Chem 26:855–866. doi: 10.1016/j.trac.2007.08.003

86. Glauser G, Veyrat N, Rochat B, et al. (2013) Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. J Chromatogr A 1292:151–9. doi: 10.1016/j.chroma.2012.12.009

87. Bedair M, Sumner LW (2008) Current and emerging mass-spectrometry technologies for metabolomics. Trends Anal Chem 27:238–250. doi: 10.1016/j.trac.2008.01.006

88. Agnolet S, Wiese S, Verpoorte R, Staerk D (2012) Comprehensive analysis of commercial willow bark extracts by new technology platform: combined use of metabolomics, high-performance liquid chromatography-solid-phase extraction-nuclear magnetic resonance spectroscopy and high-resolution radical scavengin. J Chromatogr A 1262:130–7. doi: 10.1016/j.chroma.2012.09.013

89. Yin Q, Wang P, Zhang A, et al. (2013) Ultra-performance LC-ESI/quadrupole-TOF MS for rapid analysis of chemical constituents of Shaoyao-Gancao decoction. J Sep Sci 36:1238–46. doi: 10.1002/jssc.201201198

90. Wang X, Sun H, Zhang A, et al. (2011) Potential role of metabolomics approaches in the area of traditional Chinese medicine: as pillars of the bridge between Chinese and Western medicine. J Pharm Biomed Anal 55:859–68. doi: 10.1016/j.jpba.2011.01.042

91. Shyur L-F, Yang N-S (2008) Metabolomics for phytomedicine research and drug development. Curr Opin Chem Biol 12:66–71. doi: 10.1016/j.cbpa.2008.01.032

92. Wahyuni Y, Ballester A, Tikunov Y, et al. (2013) Metabolomics and molecular marker analysis to explore pepper (Capsicum sp .) biodiversity. Metabolomics 9:130–144. doi: 10.1007/s11306-012-0432-6

93. Bernillon S, Biais B, Deborde C, et al. (2013) Metabolomic and elemental profiling of melon fruit quality as affected by genotype and environment. Metabolomics 9:57–77. doi: 10.1007/s11306-012-0429-1

94. Simmler C, Jones T, Anderson JR, et al. (2013) Species-specific Standardisation of Licorice by Metabolomic Profiling of Flavanones and Chalcones. Phytochem Anal n/a–n/a. doi: 10.1002/pca.2472

95. Choi M-Y, Choi W, Park JH, et al. (2010) Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. Food Chem 121:1260–1268. doi: 10.1016/j.foodchem.2010.01.035

96. Mullen W, Nemzer B, Stalmach a, et al. (2013) Polyphenolic and Hydroxycinnamate Contents of Whole Coffee Fruits from China, India, and Mexico. J Agric Food Chem. doi: 10.1021/jf4003126

97. Garcia-Canas V, Simo C, Herrero M, et al. (2012) Present and Future Challenges in Food Analysis: Foodomics. Anal Chem 84:10150–10159.

98. Cagliani LR, Pellegrino G, Giugno G, Consonni R (2013) Quantification of Coffea arabica and Coffea canephora var. robusta in roasted and ground coffee blends. Talanta 106:169–73. doi: 10.1016/j.talanta.2012.12.003

99. Garrett R, Schmidt EM, Pereira LFP, et al. (2013) Discrimination of arabica coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics. LWT - Food Sci Technol 50:496–502. doi: 10.1016/j.lwt.2012.08.016

100. Kueger S, Steinhauser D, Willmitzer L, Giavalisco P (2012) High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. Plant J 70:39–50. doi: 10.1111/j.1365-313X.2012.04902.x

101. Van Meulebroek L, Bussche J Vanden, Steppe K, Vanhaecke L (2012) Ultra-high performance liquid chromatography coupled to high resolution Orbitrap mass spectrometry for metabolomic profiling of the endogenous phytohormonal status of the tomato plant. J Chromatogr A 1260:67–80. doi: 10.1016/j.chroma.2012.08.047

102. Marti G, Erb M, Boccard J, et al. (2013) Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. Plant Cell Environ 36:621–39. doi: 10.1111/pce.12002

103. Pezzopane JRM, Salva TDJG, de Lima VB, Fazuoli LC (2012) Agrometeorological parameters for prediction of the maturation period of Arabica coffee cultivars. Int J Biometeorol 56:843–851. doi: 10.1007/s00484-011-0486-6

104. Choi YH, Kim HK, Linthorst HJM, et al. (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in Nicotiana tabacum leaves. J Nat Prod 69:742–8. doi: 10.1021/np050535b

105. De Vos RCH, Moco S, Lommen A, et al. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protoc 2:778–91. doi: 10.1038/nprot.2007.95

106. Madalinski G, Godat E, Alves S, et al. (2008) Direct Introduction of Biological Samples into a LTQ-Orbitrap Hybrid Mass Spectrometer as a Tool for Fast Metabolome Analysis. Anal Chem 80:3291–3303.

107. Lei Z, Huhman D V, Sumner LW (2011) Mass Spectrometry Strategies in Metabolomics. J Biol Chem 286:25435–25442. doi: 10.1074/jbc.R111.238691

108. Pan X, Wang X (2009) Profiling of plant hormones by mass spectrometry. J Chromatogr B, 877:2806–13. doi: 10.1016/j.jchromb.2009.04.024

109. Ji S, Wang Q, Qiao X, et al. (2014) New triterpene saponins from the roots of Glycyrrhiza yunnanensis and their rapid screening by LC/MS/MS. J Pharm Biomed Anal 90:15–26. doi: 10.1016/j.jpba.2013.11.021

110. Ding L, Huang X, Yang J, et al. (2006) Determination of glycyrrhetic acid in human plasma by LC – ESI – MS. J Pharm Biomed Anal 40:758–762. doi: 10.1016/j.jpba.2005.10.047

111. Montoro P, Maldini M, Russo M, et al. (2011) Metabolic profiling of roots of liquorice (Glycyrrhiza glabra) from different geographical areas by ESI/MS/MS and determination of major metabolites by LC-ESI / MS and LC-ESI / MS / MS. J Pharm Biomed Anal 54:535–544. doi: 10.1016/j.jpba.2010.10.004

112. Van Der Hooft JJJ, Akermi M, Unlu FY, et al. (2012) Structural Annotation and Elucidation of Conjugated Phenolic Compounds in Black, Green, and White Tea Extracts. J Agric Food Chem 60:8841–8850.

113. Capanoglu E, Vos RCH De, Hall RD, et al. (2013) Changes in polyphenol content during production of grape juice concentrate. Food Chem 139:521–526. doi: 10.1016/j.foodchem.2013.01.023

114. Bird SS, Marur VR, Sniatynski MJ, et al. (2011) Lipidomics profiling by High resolution LC-MS and HCD Fragmentation: Focus on Characterization of Mitochondrial Cardiolipins and Monolysocardiolipins. Anal Chem 83:940–949. doi: 10.1021/ac102598u.Lipidomics

115. Van der Hooft JJJ, Vervoort J, Bino RJ, De Vos RCH (2012) Spectral trees as a robust annotation tool in LC – MS based metabolomics. Metabolomics 8:691–703. doi: 10.1007/s11306-011-0363-7

116. Makarov A, Scigelova M (2010) Coupling liquid chromatography to Orbitrap mass spectrometry. J Chromatogr A 1217:3938–3945. doi: 10.1016/j.chroma.2010.02.022

117. Tikunov YM, Laptenok S, Hall RD, et al. (2012) MSClust□: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8:714–718. doi: 10.1007/s11306-011-0368-2

118. Creek DJ, Jankevics A, Burgess KE V., et al. (2012) IDEOM: An Excel interface for analysis of LC-MS based metabolomics data. Metabolomics 8:29–36. doi: 10.1007/s11306-011-0341-0

119. Lommen A, Kools HJ (2012) MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. Metabolomics 8:719–726. doi: 10.1007/s11306-011-0369-1

120. Lommen A (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem 81:3079–86. doi: 10.1021/ac900036d

121. Katajamaa M, Oresic M (2007) Data processing for mass spectrometry-based metabolomics. J Chromatogr A 1158:318–28. doi: 10.1016/j.chroma.2007.04.021

122. Dehaven CD, Evans AM, Dai H, Lawton K a (2010) Organization of GC/MS and LC/MS metabolomics data into chemical libraries. J Cheminform 2:9. doi: 10.1186/1758-2946-2-9

123. Rojas-cherto M, Peironcely JE, Kasper PT, et al. (2012) Metabolite Identification Using Automated Comparison of High-Resolution Multistage Mass Spectral Trees. Anal Chem 84:5524–5534.

124. Wang Y, Griffiths WJ (2008) Mass Spectrometry for Metabolite Identification. In: Griffiths WJ (ed) Metabolomics, Metabonomics Metab. profiling. pp 29–39

125. Hu Q, Noll RJ, Li H, et al. (2005) The Orbitrap□: a new mass spectrometer. J mass Spectrom 40:430–443. doi: 10.1002/jms.856

126. Allwood JW, Goodacre R (2010) An Introduction to Liquid Chromatography – Mass Spectrometry Instrumentation Applied in Plant Metabolomic Analyses. Phytochem Anal 21:33–47. doi: 10.1002/pca.1187

127. Neumann S, Böcker S (2010) Computational mass spectrometry for metabolomics□: Identification of metabolites and small molecules. Anal Bioanal Chem 398:2779–2788. doi: 10.1007/s00216-010-4142-5

128. Bollina V, Kumaraswamy GK, Kushalappa AC, et al. (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against Fusarium head blight. Mol Plant Pathol 11:769–782. doi: 10.1111/J.1364-3703.2010.00643.X

129. Makarov A, Scigelova M Orbitrap Mass analyser-Overview and applications in proteomics. In: Thermo Electr. Corp.

130. Stumpf CL, Goshawk J (2004) The MarkerLynx application manager: informatics for mass spectrometric metabonomic discovery. Waters_Application Note November:1–4.

131. Lopez MF, Kuppusamy R, Sarracino DA, et al. (2011) Mass Spectrometric Discovery and Selective Reaction Monitoring (SRM) of Putative Protein Biomarker Candidates in First Trimester Trisomy 21 Maternal Serum. J Proteome Res 10:133–142.

132. Smith C a, Want EJ, O'Maille G, et al. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–87. doi: 10.1021/ac051437y

133. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. Anal Chem 84:5035–9. doi: 10.1021/ac300698c

134. Katajamaa M, Miettinen J, Oresic M (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22:634–6. doi: 10.1093/bioinformatics/btk039

135. Bunk B, Kucklick M, Jonas R, et al. (2006) MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. Bioinformatics 22:2962–5. doi: 10.1093/bioinformatics/btl526

136. Broeckling CD, Reddy IR, Duran AL, et al. (2006) MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. Anal Chem 78:4334–41. doi: 10.1021/ac0521596

137. Luedemann A, Strassburg K, Erban A, Kopka J (2008) TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics 24:732–7. doi: 10.1093/bioinformatics/btn023

138. Lommen A, van der Kamp HJ, Kools HJ, et al. (2012) metAlignID: a high-throughput software tool set for automated detection of trace level contaminants in comprehensive LECO two-dimensional gas chromatography time-of-flight mass spectrometry data. J Chromatogr A 1263:169–78. doi: 10.1016/j.chroma.2012.09.056

139. Tikunov Y MSClust User Manual.

140. Sumner LW, Amberg A, Barrett D, et al. (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics 3:211–221. doi: 10.1007/s11306-007-0082-2

141. Clifford MN, Johnston KL, Knight S, Kuhnert N (2003) Hierarchical Scheme for LC-MS n Identification of Chlorogenic Acids. J Agric Food Chem 51:2900–2911.

142. Todeschini R Introduzione alla Chemiometria. Edises, Napoli

143. Brereton R (1992) Multivariate pattern recognition in chemometrics illustrated by case studies. Elsevier

144. Intrieri MC, Buiatti M (2001) The horizontal transfer of Agrobacterium rhizogenes genes and the evolution of the genus Nicotiana. Mol Phylogenet Evol 20:100–110. doi: 10.1006/mpev.2001.0927

145. Ichikawa T, Ozeki Y, Syono K (1990) Evidence for the expression of the rol genes of Nicotiana glauca in genetic tumors of N. glauca X N. langsdorffii. Mol Gen Genet 220:177–80.

146. Heidel AJ, Baldwin IT (2004) Microarray analysis of salicylic acid- and jasmonic acid- signalling in responses of Nicotiana attenuata to attack by insects from multiple feeding guilds. Plant, Cell Envirinment 27:1362–1373.

147. Kiselev K V, Kusaykin MI, Dubrovina AS, et al. (2006) The rolC gene induces expression of a pathogenesis-related beta-1,3-glucanase in transformed ginseng cells. Phytochemistry 67:2225–31. doi: 10.1016/j.phytochem.2006.07.019

148. Kiselev K V., Grishchenko O V., Zhuravlev YN (2010) CDPK gene expression in salt tolerant rolB and rolC transformed cell cultures of Panax ginseng. Biol Plant 54:621–630. doi: 10.1007/s10535-010-0112-1

149. Bulgakov V., Tchernoded G., Mischenko N., et al. (2004) The rolB and rolC genes activate synthesis of anthraquinones in Rubia cordifolia cells by mechanism independent of octadecanoid signaling pathway. Plant Sci 166:1069–1075. doi: 10.1016/j.plantsci.2003.12.027

150. Giannarelli S, Muscatello B, Bogani P, et al. (2010) Comparative determination of some phytohormones in wild-type and genetically modified plants by gas chromatography-mass spectrometry and high-performance liquid chromatography-tandem mass spectrometry. Anal Biochem 398:60–8. doi: 10.1016/j.ab.2009.10.038

151. Fuoco R, Bogani P, Capodaglio G, et al. (2013) Response to metal stress of Nicotiana langsdorffii plants wild-type and transgenic for the rat glucocorticoid receptor gene. J Plant Physiol 170:668–675. doi: 10.1016/j.jplph.2012.12.009

152. Bogani P, Calistri E, Muscatello B, et al. (2012) The effect of the integration of the rat glucocorticoid GR receptor on the response of Nicotiana to abiotic stresses. Proc. 56th Ital. Soc. Agric. Genet. Annu. Congr. 16:

153. Del Bubba M, Ancillotti C, Checchini L, et al. (2013) Chromium accumulation and changes in plant growth, selected phenolics and sugars of wild type and genetically modified Nicotiana langsdorffii. J Hazard Mater 262:394–403. doi: 10.1016/j.jhazmat.2013.08.073

154. Bulgakov VP, Aminin DL, Shkryl YN, et al. (2008) Suppression of reactive oxygen species and enhanced stress tolerance in Rubia cordifolia cells expressing the rolC oncogene. Mol Plant Microbe Interact 21:1561–70. doi: 10.1094/MPMI-21-12-1561

155. Lipiec J, Doussan C, Nosalewicz a., Kondracka K (2013) Effect of drought and heat stresses on plant growth and yield: a review. Int Agrophysics 27:463–477. doi: 10.2478/intag-2013-0017

156. Ramakrishna A, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav 6:1720–1731. doi: 10.4161/psb.6.11.17613

157. Pline WA, Wilcut JW, Duke SO, et al. (2002) Tolerance and Accumulation of Shikimic Acid in Response to Glyphosate Applications in Glyphosate-Resistant and Nonglyphosate-Resistant Cotton (Gossypium hirsutum L .). J Agric Food Chem 50:506–512.

158. Beltrano J, Ruscitti M, Arango C, Ronco M (2013) Changes in the accumulation of shikimic acid in mycorrhized Capsicum annuum L . grown with application of glyphosate and phosphorus. Theor Exp plant Physiol 25:125–136.

159. Dong-Mei C, Yun-Shan W, Li-Fang K, et al. (2007) Effect of high temperature on jasmonic acid (JA) concentration and antioxidation of Anthurium andraeanum seedling leaf. Chinese J Eco-Agriculture 05:30031.

160. Xin Z, Zhou X, Pilet P-E (1997) Level changes of jasmonic, abscisic, and indole-3yl-acetic acids in maize under desiccation stress. J Plant Physiol 151:120–124. doi: 10.1016/S0176-1617(97)80047-8

161. Engelberth J, Schmelz EA, Alborn HT, et al. (2003) Simultaneous quantification of jasmonic acid and salicylic acid in plants by vapor-phase extraction and gas chromatography-chemical ionization-mass spectrometry. Anal Biochem 312:242–250.

162. Matsuura H, Aoi Æa, Satou Æc, et al. (2009) Simultaneous UPLC MS / MS analysis of endogenous jasmonic acid , salicylic acid , and their related compounds. Plant Growth Regul 57:293–301. doi: 10.1007/s10725-008-9347-7

163. Creelman R a, Mullet JE (1995) Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. Proc Natl Acad Sci U S A 92:4114–9.

164. Pearcy RW (1978) Effect of Growth Temperature on the Fatty Acid Composition of the Leaf Lipids in Atriplex lentiformis (Torr .) Wats. New Phytol 61:484–486.

165. Murakami Y (2012) Trienoic Fatty Acids and Plant Tolerance of High Temperature. Science (80-). doi: 10.1126/science.287.5452.476

166. Mathieu Y, Barbier-brygoo H, Laurière C (2002) Activation by fatty acids of the production of active oxygen species by tobacco cells. October 40:313–324.

167. Routaboul J-M, Skidmore C, Wallis JG, Browse J (2012) Arabidopsis mutants reveal that short- and long-term thermotolerance have different requirements for trienoic fatty acids. J Exp Bot 63:1435–43. doi: 10.1093/jxb/err381

168. Sakamoto A, Sulpice R, Hou C, et al. (2003) Genetic modification of the fatty acid unsaturation of phosphatidylglycerol in chloroplasts alters the sensitivity of tobacco plants to cold stress. 99–105.

169. Li Y, Chen J, Li Y, et al. (2011) Screening and characterization of natural antioxidants in four Glycyrrhiza species by liquid chromatography coupled with electrospray ionization quadrupole time-of-flight tandem mass spectrometry. J Chromatogr A 1218:8181– 8191. doi: 10.1016/j.chroma.2011.09.030

170. Li W, Koike K, Asada Y, et al. (2002) Flavonoids from Glycyrrhiza pallidiflora hairy root cultures. Phytochemistry 60:351–355.

171. Kreps JA, Wu Y, Chang H, et al. (2002) Transcriptome Changes for Arabidopsis in Response to Salt , Osmotic , and Cold Stress. Plant Physiol 130:2129–2141. doi: 10.1104/pp.008532.with

172. Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63:3523–43. doi: 10.1093/jxb/ers100

173. Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221–7. doi: 10.1093/jxb/erl164

174. Larkindale J, Huang B (2004) Changes of lipid composition and saturation level in leaves and roots for heat-stressed and heat-acclimated creeping bentgrass (Agrostis stolonifera). Environ Exp Bot 51:57–67. doi: 10.1016/S0098-8472(03)00060-1

175. Yordanov I, Velikova V, Tsonev T (2000) Plant reponse to drought acclimation stress. Photosynthetica 38:171–186.

176. Ganesh KS, Baskaran L, Rajasekaran S, et al. (2008) Chromium stress induced alterations in biochemical and enzyme metabolism in aquatic and terrestrial plants. Colloids Surf B Biointerfaces 63:159–63. doi: 10.1016/j.colsurfb.2007.11.016

177. Berger S (2002) Jasmonate-related mutants of Arabidopsis as tools for studying stress signaling. Planta 214:497–504. doi: 10.1007/s00425-001-0688-y

178. Fragnière C, Serrano M, Abou-Mansour E, et al. (2011) Salicylic acid and its location in response to biotic and abiotic stress. FEBS Lett 585:1847–52. doi: 10.1016/j.febslet.2011.04.039

179. Segarra G, Jauregui O, Casanova E, Trillas I (2006) Simultaneous quantitative LC – ESI-MS / MS analyses of salicylic acid and jasmonic acid in crude extracts of Cucumis sativus under biotic stress. Phytochemistry 67:395–401. doi: 10.1016/j.phytochem.2005.11.017

180. Zelaya I a, Anderson J a H, Owen MDK, Landes RD (2011) Evaluation of spectrophotometric and HPLC methods for shikimic acid determination in plants: models in glyphosate-resistant and -susceptible crops. J Agric Food Chem 59:2202–12. doi: 10.1021/jf1043426

181. Bochkov D V, Sysolyatin S V, Kalashnikov AI, Surmacheva I a (2012) Shikimic acid: review of its analytical, isolation, and purification techniques from plant and microbial sources. J Chem Biol 5:5–17. doi: 10.1007/s12154-011-0064-8

182. Schmelz EA, Engelberth J, Alborn HT, et al. (2003) Simultaneous analysis of phytohormones , phytotoxins , and volatile organic compounds in plants. Proc Natl Acad Sci U S A 100:10552–10557.

183. Sánchez-Rodríguez E, Ruiz JM, Ferreres F, Moreno D a (2011) Phenolic metabolism in grafted versus nongrafted cherry tomatoes under the influence of water stress. J Agric Food Chem 59:8839–46. doi: 10.1021/jf201754t

184. Krasavina MS, Burmistrova NA, Raldugina GN (2014) The Role of Carbohydrates in Plant Resistance to Abiotic Stresses. Emerg. Technol. Manag. Crop Stress Toler. - Vol. 1. Elsevier Inc., pp 229–270

185. Yang YY, Chen J, Liu Q, et al. (2012) Comparative proteomic analysis of the thermotolerant plant Portulaca oleracea acclimation to combined high temperature and humidity stress. J Proteome Res 11:3605–23. doi: 10.1021/pr300027a

186. Irdani T, Caroppo S, Ambrogioni L (2003) Response of Nicotiana tabacum plants overexpressing a glucocorticoid receptor to Meloidogyne incognita (Nematoda Tylenchida) infestation. Redia 86:35–38.

187. Bettini P, Michelotti S, Bindi D, et al. (2003) Pleiotropic effect of the insertion of the Agrobacterium rhizogenes rolD gene in tomato (Lycopersicon esculentum Mill.). Theor Appl Genet 107:831–836. doi: 10.1007/s00122-003-1322-0

188. Bulgakov VP (2008) Functions of rol genes in plant secondary metabolism. Biotechnol Adv 26:318–324. doi: 10.1016/j.biotechadv.2008.03.001

189. Palazón J, Cusidó RM, Roig C, Piñol MT (1998) Expression of the rol C gene and nicotine production in transgenic roots and their regenerated plants. Plant Cell Rep 17:384–390. doi: 10.1007/s002990050411

190. Nilsson O, Olsson O (1997) Getting to the root: The role of the Agrobacterium rhizogenes rol genes in the formation of hairy roots. Physiol Plant 100:463–473. doi: 10.1034/j.1399-3054.1997.1000307.x

191. Schmulling T, Fladung M, Grossmann K, Schel J (1993) Hormonal content and sensitivity of transgenic tobacco and potato plants expressing single rol genes of Agrobacterium rhizogenes T-DNA. plant J 3:371–382.

192. Faber A, Cutsem E Van, Morsomme P, Boutry M (2011) Gel-based and gel-free proteomic analysis of Nicotiana tabacum trichomes identifies proteins involved in secondary metabolism and in the (a)biotic stress response. Proteomics 11:440–454. doi: 10.1002/pmic.201000356

193. Pharmaceuticals-canada D, Amore FJ, Cox G V, et al. (1980) Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry. Anal Chem 52:2242–2249.

194. Skoog AD, Holler FJ, Crouch RS (2007) Chimica analitica strumentale, II. Edises, Napoli

195. Weston TR, Derner JD, Murrieta CM, et al. (2008) Comparison of Catalysts for Direct Transesterification of Fatty Acids in Freeze-Dried Forage Samples. Crop Sci 48:1636. doi: 10.2135/cropsci2007.07.0376sc

196. Lu W, Bennett BD, Rabinowitz JD (2008) Analytical strategies for LC-MS-based targeted metabolomics. J Chromatogr B Anal Technol Biomed life Sci 871:236–242. doi: 10.1016/j.jchromb.2008.04.031.Analytical

197. Chu H, Tso TC (1968) Fatty Acid Composition in Tobacco I. Green Tobacco Plants. Plant Phisiology 43:428–433.

198. Koiwai A, Suzuki F, Matsuzaki T, Kawashima N (1983) The fatty acid coposition of seeds and leaves of Nicotiana species. Phytochemistry 22:1409–1412.

199. Rivas-Ubach A, Sardans J, Pérez-Trujillo M, et al. (2012) Strong relationship between elemental stoichiometry and metabolome in plants. Proc Natl Acad Sci U S A 109:4181–6. doi: 10.1073/pnas.1116092109

200. Saidi Y, Peter M, Finka A, et al. (2010) Membrane lipid composition affects plant heat sensing and modulates Ca 2+ -dependent heat shock response. Plant Signal Behav 5:1530–1533. doi: 10.4161/psb.5.12.13163

201. Hasanuzzaman M, Nahar K, Fujita M (2009) Extreme Temperature Responses , Oxidative Stress and Antioxidant Defense in Plants. In: Vahdati K, Leslie C (eds) Abiotic Stress - Plant Responses Appl. Agric. pp 169–205

202. Taran N, Okanenko A, Musienko N (2000) Sulpholipid reflects plant resistance to stress-factor action. Biochem Soc Trans 28:922–924.

203. Sato N, Aoki M, Maru Y, et al. (2003) Involvement of sulfoquinovosyl diacylglycerol in the structural integrity and heat-tolerance of photosystem II. Planta 217:245–51. doi: 10.1007/s00425-003-0992-9

204. Essemine J, Govindachary S, Ammar S, et al. (2012) Enhanced sensitivity of the photosynthetic apparatus to heat stress in digalactosyl-diacylglycerol deficient Arabidopsis. Environ Exp Bot 80:16–26. doi: 10.1016/j.envexpbot.2011.12.022

205. Chen J, Burke JJ, Xin Z, et al. (2006) Characterization of the Arabidopsis thermosensitive mutant atts02 reveals an important role for galactolipids in thermotolerance. Plant Cell Environ 29:1437–1448. doi: 10.1111/j.1365-3040.2006.01527.x

206. Burgos A, Szymanski J, Seiwert B, et al. (2011) Analysis of short-term changes in the Arabidopsis thaliana glycerolipidome in response to temperature and light. Plant J 66:656–68. doi: 10.1111/j.1365-313X.2011.04531.x

207. Süss K-H, Yordanov IT (1986) Biosynthetic cause of in vivo acquired thermotolerance of photosynthetic light reactions and metabolic responses of chloroplasts to heat stress. Plant Physiol 81:192–9.

208. Otsuru M, Yu Y, Mizoi J, et al. (2013) Mitochondrial Phosphatidylethanolamine Level Modulates Cyt c Oxidase Activity to Maintain Respiration Capacity in Arabidopsis thaliana Rosette Leaves. Plant cell Physiol 54:1612–1619. doi: 10.1093/pcp/pct104

209. Kim J, Kang K, Gonzales-vigil E, et al. (2012) Striking natural diversity in glandular trichome acylsugar composition is shaped by variation at the acyltransferase2 Locus in the wild tomato Solanum habrochaites. Plant Physiol 160:1854–1870. doi: 10.1104/pp.112.204735

210. Ghosh B, Westbrook TC, Jones AD (2013) Comparative structural profiling of trichome specialized metabolites in tomato (Solanum lycopersicum) and S . habrochaites : acylsugar profiles revealed by UHPLC / MS and NMR. Metabolomics 10:496–507. doi: 10.1007/s11306-013-0585-y

211. Zhu H, Feng Y, Yang J, et al. (2013) Separation and characterization of sucrose esters from Oriental tobacco leaves using accelerated solvent extraction followed by SPE coupled to HPLC with ion-trap MS detection. J Sep Sci 36:2486–95. doi: 10.1002/jssc.201300294

212. Glas JJ, Schimmel BCJ, Alba JM, et al. (2012) Plant Glandular Trichomes as Targets for Breeding or Engineering of Resistance to Herbivores. Int J Mol Sci 13:17077–17103. doi: 10.3390/ijms131217077

213. Schilmiller A, Shi F, Kim J, et al. (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. plant J 62:391–403. doi: 10.1111/j.1365-313X.2010.04154.x

214. Pérez-estrada LB, Cano-santana Z, Oyama KEN (2000) Variation in leaf trichomes of Wigandia urens : environmental factors and physiological consequences. tree Physiol 20:629–632.

215. Kuc J (1984) Steroid glykoalkaloids and related compounds. Am Potato J 61:123–139.

216. Nema PK, Ramayya N, Duncan E, Niranjan K (2008) Potato glycoalkaloids : formation and strategies for mitigation. J Sci Food Agric 1881:1869–1881. doi: 10.1002/jsfa

217. Udalova Z V, Zinov S V, Vasil IS, Paseshnichenko VA (2004) Correlation between the Structure of Plant Steroids and Their Effects on Phytoparasitic Nematodes. Appl Biochem Microbiol 40:109–113.

218. Bagalwa JM, Voutquenne-nazabadioko L, Sayagh C, Bashwira AS (2010) Evaluation of the biological activity of the molluscicidal fraction of Solanum sisymbriifolium against non target organisms. Fitoterapia 81:767–771. doi: 10.1016/j.fitote.2010.04.003

219. Coria NA, Sarquı JI, Penalosa I, Urzua M (1998) Heat-Induced Damage in Potato (Solanum tuberosum) Tubers : Membrane Stability , Tissue Viability , and Accumulation of Glycoalkaloids. J Agric Food Chem 46:4524–4528.

220. Nitithamyong A, Vonelbe JH, Wheeler RM, Tibbitts TW (1999) Glycoalkaloids in potato tubers grown under controlled environments. Am J Potato Res 76:337–343. doi: 10.1007/BF02910006

221. Dimenstein L, Lisker N, Kedar N, Levy D (1997) Changes in the content of steroidal glycoalkaloids in potato tubers grown in the field and in the greenhouse under different conditions of light , temperature and daylength. Physiol Mol Plant Pathol 50:391–402.

222. Petersson E V, Arif U, Schulzova V, et al. (2013) Glycoalkaloid and Calystegine Levels in Table Potato Cultivars Subjected to Wounding, Light, and Heat Treatments. J Agric Food Chem 61:5893–5902.

223. Mithöfer A, Schulze B, Boland W (2004) Biotic and heavy metal stress response in plants: evidence for common signals. FEBS Lett 566:1–5. doi: 10.1016/j.febslet.2004.04.011

224. Lee S, Suh S, Kim S, et al. (1997) Systemic elevation of phosphatidic acid and lysophospholipid levels in wounded plants. plant J 12:547–556.

225. Narváez-vásquez J, Florin-christensen J, Ryan CA (1999) Positional Specificity of a Phospholipase A Activity Induced by Wounding , Systemin , and Oligosaccharide Elicitors in Tomato Leaves. Plant Cell 11:2249–2260.

226. Ryu SB, Wang X (1998) Increase in free linolenic and linoleic acids associated with phospholipase D-mediated hydrolysis of phospholipids in wounded castor bean leaves. Biochim Biophys Acta - Lipids Lipid Metab 1393:193–202. doi: 10.1016/S0005-2760(98)00048-4

227. Tiburcio AF, Altabella T, Cordeiro A (2003) Polyamines. In: Brau T (ed) Encycl. Appl. Plant Sci. Academic Press Elsevier, pp 1042–1051

228. Van de Mortel J, Schat H, Moerland PD, et al. (2008) Expression differences for genes involved in lignin , glutathione and sulphate metabolism in response to cadmium in Arabidopsis thaliana and the related Zn / Cd-hyperaccumulator Thlaspi caerulescens. Plant Cell Environ 31:301–324. doi: 10.1111/j.1365-3040.2007.01764.x

229. Liu H, Li Q, Zhang Y, Zhou Y (2009) Analysis of (−)-Shikimic Acid in Chinese Star Anise by GC–MS with Selected Ion Monitoring. Chromatographia 69:339–344. doi: 10.1365/s10337-008-0898-6

230. Mardones C, Hitschfeld A, Contreras A, et al. (2005) Comparison of shikimic acid determination by capillary zone electrophoresis with direct and indirect detection with liquid chromatography for varietal differentiation of red wines. J Chromatogr A 1085:285–292. doi: 10.1016/j.chroma.2005.06.022

231. Matallo MB, Almeida SD., Cerdeira AL, et al. (2009) Microwave-assisted solvent extraction and analysis of Shikimic acid from plant tissues. Planta Daninha 27:987–994.

232. Pan X, Welti R, Wang X (2010) Quantitative analysis of major plant hormones in crude plant extracts by high performance liquid chromatography-mass spectrometry. Nat Protoc 5:986–991.

233. Birkemeyer C, Kolasa A, Kopka J (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. J Chromatogr A 993:89–102.

234. Yang J, Zhang J, Wang Z, et al. (2001) Hormonal changes in the grains of rice subjected to water stress during grain filling. Plant Physiol 127:315–23.

235. Weiler EW, Ziegler H (1981) Determination of phytohormones in phloem exudate from tree species by radioimmunoassay. Planta 152:168–170.

236. Durgbanshi A, Arbona V, Pozo O, et al. (2005) Simultaneous determination of multiple phytohormones in plant extracts by liquid chromatography-electrospray tandem mass spectrometry. J Agric Food Chem 53:8437–42. doi: 10.1021/jf050884b

237. Ulger S, Sonmez S, Karkacier M, et al. (2004) Determination of endogenous hormones, sugars and mineral nutrition levels during the induction, initiation and differentiation stage and their effects on flower formation in olive. Plant Growth Regul 42:89–95. doi: 10.1023/B:GROW.0000014897.22172.7d

238. Liu H-T, Li Y-F, Luan T-G, et al. (2007) Simultaneous Determination of Phytohormones in Plant Extracts using SPME and HPLC. Chromatographia 66:515–520. doi: 10.1365/s10337-007-0350-3

239. Kallenbach M, Baldwin IT, Bonaventure G (2009) A rapid and sensitive method for the simultaneous analysis of aliphatic and polar molecules containing free carboxyl groups in plant extracts by LC-MS / MS. Plant Methods 11:5–17. doi: 10.1186/1746-4811-5-17

Bibliography

240. Avula B, Wang Y, Smillie TJ, Khan IA (2009) Determination of Shikimic Acid in Fruits of Illicium Species and Various Other Plant Samples by LC – UV and LC – ESI – MS. Cromatographia 69:307–314. doi: 10.1365/s10337-008-0884-z

241. Palazón J, Cusidó RM, Gonzalo J, et al. (1998) Relation between the amount of rolC gene product and indole alkaloid accumulation in Catharanthus roseus transformed root cultures. J Plant Physiol 153:712–718. doi: 10.1016/S0176-1617(98)80225-3

242. Sano H, Seo S, Koizumi N, et al. (1996) Regulation by Cytokinins of Endogenous Levels of Jasmonic and Salicylic Acids in Mechanically Wounded Tobacco Plants. Plant Cell Physiol 37:762–769. doi: 10.1093/oxfordjournals.pcp.a029011

243. Biondi S, Scoccianti V, Scaramagli S, et al. (2003) Auxin and cytokinin modify methyl jasmonate effects on polyamine metabolism and ethylene biosynthesis in tobacco leaf discs. Plant Sci 165:95–101. doi: 10.1016/S0168-9452(03)00147-X

244. Goda H, Shimada Y, Asami T, et al. (2002) Microarray Analysis of Brassinosteroid-Regulated Genes in Arabidopsis. Plant Phisiology 130:1319–1334. doi: 10.1104/pp.011254.exposure.

245. Wasternack C (2007) Jasmonates: An Update on Biosynthesis , Signal Transduction and Action in Plant Stress Response , Growth and Development. Ann Bot 100:681–697. doi: 10.1093/aob/mcm079

246. Heldt H-W, Piechulla B (2011) Multiple signals regulate the growth and development of plant organs and enable their adaptation to environmental conditions. Plant Biochem., IV. Academic Press Elsevier, pp 451–485

247. Yadav SK (2010) Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. South African J Bot 76:167–179. doi: 10.1016/j.sajb.2009.10.007

248. Horváth E, Szalai G, Janda T, Horvath E (2007) Induction of Abiotic Stress Tolerance by Salicylic Acid Signaling. J Plant Growth Regul 26:290–300. doi: 10.1007/s00344-007-9017-4

249. Guy C, Kaplan F, Kopka J, et al. (2008) Metabolomics of temperature stress. Physiol Plant 132:220–235. doi: 10.1111/j.1399-3054.2007.00999.x

250. Chen Z, Silva H, Klessig D (1993) Active oxygen species in the induction of plant systemic acquired resistance by salicylic acid. Science (80-) 262:1883–1886. doi: 10.1126/science.8266079

251. Anderson MD, Chen Z, Klessig DF (1998) Possible involvement of lipid peroxidation in salicylic acid-mediated induction of PR-1 gene expression. Phytochemistry 47:555–566.

252. Fujita M, Fujita Y, Noutoshi Y, et al. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9:436–42. doi: 10.1016/j.pbi.2006.05.014

253. Depuydt S, Hardtke CS (2011) Hormone signalling crosstalk in plant growth regulation. Curr Biol 21:R365–73. doi: 10.1016/j.cub.2011.03.013

254. Hamilton III EW, Coleman JS (2001) Heat-shock proteins are induced in unstressed leaves of Nicotiana attenuata (Solanaceae) when distant leaves are stressed. Am J Bot 88:950–5.

255. Ginzberg I, Tokuhisa JG, Veilleux RE (2008) Potato Steroidal Glycoalkaloids: Biosynthesis and Genetic Manipulation. Potato Res 52:1–15. doi: 10.1007/s11540-008-9103-4

Bibliography

256. ANDRE CM, SCHAFLEITNER R, GUIGNARD C, et al. (2009) Modification of the Health-Promoting Value of Potato Tubers Field Grown under Drought Stress : Emphasis on Dietary Antioxidant and Glycoalkaloid Contents in Five Native Andean Cultivars (Solanum tuberosum L .). J Agric Food Chem 57:599–609.

257. Sakhno LO, Slyvets MS, Korol NA, et al. (2014) Changes in Fatty Acid Composition in Leaf Lipids of Canola Biotech Plants under Short-time Heat Stress Changes in Fatty Acid Composition in Leaf Lipids of Canola Biotech Plants under Short-time Heat Stress. J Stress Physiol Biochem 10:24–34.

258. Gombos Z, Wada H, Hideg E, Murata N (1994) The Unsaturation of Membrane Lipids Stabilizes Photosynthesis against Heat Stress. Plant Physiol 104:563–567.

259. Schilmiller AL, Last RL, Pichersky E (2008) Harnessing plant trichome biochemistry for the production of useful compounds. Plant J 54:702–11. doi: 10.1111/j.1365-313X.2008.03432.x

260. Guo Z, Wagner GJ (1995) Biosynthesis of cembratrienols in cell-free extracts from trichomes of Nicotiana tabacum. Plant Sci 110:1–10. doi: 10.1016/0168-9452(95)04174-S

261. Forkner RE, Hare DJ (2000) Genetic and environmental variation in acyl glucose ester production and glandular and non-glandular trichome densities in Datura wrightii. J. Chem. Ecol. 26:

262. Schuck S, Kallenbach M, Baldwin IT, Bonaventure G (2014) The Nicotiana attenuata GLA1 lipase controls the accumulation of Phytophthora parasitica-induced oxylipins and defensive secondary metabolites. Plant Cell Environ. doi: 10.1111/pce.12281

263. Cho K, Kim Y, Wi SJ, et al. (2012) Nontargeted Metabolite Pro fi ling in Compatible Pathogen- Inoculated Tobacco (Nicotiana tabacum L. cv. Wisconsin 38) Using UPLC-Q-TOF/MS. J Agric Food Chem 60:11015–11028.

264. Zhang Q, Ye M (2009) Chemical analysis of the Chinese herbal medicine Gan-Cao (licorice). J Cromatography A 1216:1954–1969. doi: 10.1016/j.chroma.2008.07.072

265. Siracusa L, Saija A, Cristani M, et al. (2011) Phytocomplexes from liquorice (Glycyrrhiza glabra L .) leaves — Chemical characterization and evaluation of their antioxidant , anti-genotoxic and anti-inflammatory activity. Fitoterapia 82:546–556. doi: 10.1016/j.fitote.2011.01.009

266. Hayashi H, Sudo H (2009) Economic importance of licorice. Plant Biotechnol 26:101–104.

267. Chin Y-W, Jung H-A, Liu Y, et al. (2007) Anti-oxidant Constituents of the Roots and Stolons of Licorice (Glycyrrhiza glabra). J Agric Food Chem 55:4691–4697.

268. Fu Y, Chen J, Li Y-J, et al. (2013) Antioxidant and anti-inflammatory activities of six flavonoids separated from licorice. Food Chem 141:1063–71. doi: 10.1016/j.foodchem.2013.03.089

269. Gupta VK, Fatima A, Faridi U, et al. (2008) Antimicrobial potential of Glycyrrhiza glabra roots. J Ethnopharmacol 116:377–380. doi: 10.1016/j.jep.2007.11.037

270. Simons R, Vincken J, Mol LAM, et al. (2011) Agonistic and antagonistic estrogens in licorice root (Glycyrrhiza glabra). Anal Bioanal Chem 401:305–313. doi: 10.1007/s00216-011-5061-9

271. Nomura T, Fukai T, Akiyama T (2002) Chemistry of phenolic compounds of licorice (Glycyrrhiza species) and their estrogenic and cytotoxic activities *. Pure Appl Chem 74:1199–1206.

272. Fu B, Liu J, Li H, et al. (2005) The application of macroporous resins in the separation of licorice flavonoids and glycyrrhizic acid. J Cromatography A 1089:18–24. doi: 10.1016/j.chroma.2005.06.051

273. Liao WC, Lin Y, Chang T, Huang W (2012) Identification of two licorice species , Glycyrrhiza uralensis and Glycyrrhiza glabra , based on separation and identification of their bioactive components. Food Chem 132:2188–2193. doi: 10.1016/j.foodchem.2011.12.051

274. Shibano M, Ozaki K, Watanabe H, et al. (2010) Determination of Flavonoids in Licorice Using Acid Hydrolysis and Reversed-Phase HPLC and Evaluation of the Chemical Quality of Cultivated Licorice. Planta Med 76:1–10. doi: 10.1055/s-0029-1240690

275. Tanaka K, Ina A, Hayashi K, Komatsu K (2010) Comparison of chemical constituents in Glycyrrhiza uralensis from various sources using a multivariate statistical approach. J Tradit Med 27:210–216.

276. Xu T, Yang M, Li Y, et al. (2013) An integrated exact mass spectrometric strategy for comprehensive and rapid characterization of phenolic compounds in licorice. Rapid Commun mass Spectrom 27:2297–309. doi: 10.1002/rcm.6696

277. Zhang J, Gao W, Gao Y, et al. (2011) Analysis of influences of spaceflight on chemical constituents in licorice by HPLC- ESI-MS/MS. Acta Physiol Plant 33:2511–2520. doi: 10.1007/s11738-011-0796-7

278. Tan G, Zhu Z, Zhang H, et al. (2010) Analysis of phenolic and triterpenoid compounds in licorice and rat plasma by high-performance liquid chromatography diode-array detection, time-of-flight mass spectrometry and quadrupole ion trap mass spectrometry. Rapid Commun Mass Spectrom 24:209–218. doi: 10.1002/rcm

279. Simons R, Vincken J, Bakx EJ, et al. (2009) A rapid screening method for prenylated flavonoids with ultra-high-performance liquid chromatography/electrospray ionisation mass spectrometry in licorice root extracts. Rapid Commun Mass Spectrom 23:3083–3093. doi: 10.1002/rcm

280. Kondo K, Shiba MS, Nakamura RN, et al. (2007) Constituent Properties of Licorices Derived from Glycyrrhiza uralensis , G . glabra , or G . inflata Identified by Genetic Information. Biol Pharm Bull 30:1271–1277.

281. Hatano T, Shintani Y, Aga Y, et al. (2000) Phenolic Constituents of Licorice . VIII . 1) Structures of Glicophenone and Glicoisoflavanone , and Effects of Licorice Phenolics on Methicillin-Resistant Staphylococcus aureus. Chem Pharm Bull 48:1286–1292.

282. Sung MW, Li PCHL (2004) Chemical analysis of raw , dry-roasted , and honey-roasted licorice by capillary electrophoresis. Electrophoresis 25:3434–3440. doi: 10.1002/elps.200305988

283. Rao AS (1990) Root flavonoids. Bot Rev 56:1–84. doi: 10.1007/BF02858531

284. Zhu S, Sugiyama R, Batkhuu J, et al. (2009) Survey of Glycyrrhizae Radix resources in Mongolia□: chemical assessment of the underground part of Glycyrrhiza uralensis and comparison with Chinese Glycyrrhizea Radix. J Nat Med 63:137–146. doi: 10.1007/s11418-008-0303-7

285. Hayashi H, Inoue K, Ozaki K, Watanabe H (2005) Comparative Analysis of Ten Strains of Glycyrrhiza uralensis Cultivated in Japan. Biol Pharm Bull 28:1113–1116.

286. Barron D, Ibrahim RK (1988) Synthesis of Flavonoid Sulfates . II . The Use of Aryl Sulfatase in the Synthesis of Flavonol-3-sulfates. Z Naturforsch 43c:625–630.

287. Flamini G, Antognoli E, Morelli I (2001) Two flavonoids and other compounds from the aerial parts of Centaurea bracteata from Italy. Phytochemistry 57:559–64.

288. Nomura T, Hano Y, Fukai T (2009) Chemistry and biosynthesis of isoprenylated flavonoids from Japanese mulberry tree. Proc Japan Acad Ser B 85:391–408. doi: 10.2183/pjab.85.391

289. Kwon H, Kim H, Ryu YB, et al. (2010) In vitro anti-rotavirus activity of polyphenol compounds isolated from the roots of Glycyrrhiza uralensis. Bioorg Med Chem 18:7668–7674. doi: 10.1016/j.bmc.2010.07.073

290. Mukhopadhyay M, Panja P (2008) A novel process for extraction of natural sweetener from licorice (Glycyrrhiza glabra) roots. Sep Purif Technol 63:539–545. doi: 10.1016/j.seppur.2008.06.013

291. Kitagawa I, Chen W-Z, Kazuyuki H, et al. (1998) Chemical studies of Chinese Licorice-Roots II. Five new flavonoid constituents from the roots of Glycyrrhiza aspera PALL. Collected in Xinjiang. Chem Pharm Bull 46:1511–1517.

292. Liu T, Lin HM (2011) Preliminary assessment of genetic diversity in cultivated Glycyrrhiza uralensis, G. inflate and G. glabra by chemical fingerprint and inter-simple sequence repeat markers. In: Pan W, Ren J, Li Y (eds) Adv. Mater. Res. pp 1318–1325

293. Yamazaki M, Sato A, Shimomura K, et al. (1994) Genetic Relationships among Glycyrrhiza Plants Determined by RAPD and RFLP Analyses. Biol Pharm Bull 17:1529–1531. doi: 10.1248/bpb.17.1529

294. Seki H, Sawai S, Ohyama K, et al. (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. Plant Cell 23:4112–23. doi: 10.1105/tpc.110.082685

295. Augustin JM, Kuzina V, Andersen SB, Bak S (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72:435–57. doi: 10.1016/j.phytochem.2011.01.015

296. Nagashima S, Inagaki R, Kubo A, et al. (2004) cDNA cloning and expression of isoflavonoid-specific glucosyltransferase from Glycyrrhiza echinata cell-suspension cultures. Planta 218:456–9. doi: 10.1007/s00425-003-1118-0

297. Tian L, Pang Y, Dixon R a. (2007) Biosynthesis and genetic engineering of proanthocyanidins and (iso)flavonoids. Phytochem Rev 7:445–465. doi: 10.1007/s11101-007-9076-y

298. Aoki T, Akashi T, Ayabe S (2000) Flavonoids of Leguminous Plants: Structure, Biological Activity, and Biosynthesis. J Plant Res 113:475–488. doi: 10.1007/PL00013958

299. Nakamura K, Akashi T, Aoki T, et al. (1999) Induction of isoflavonoid and retrochalcone branches of the flavonoid pathway in cultured Glycyrrhiza echinata cells treated with yeast extract. Biosci Biotechnol Biochem J 63:1618–1620.

300. Shimada N, Aoki T, Sato S, et al. (2003) A Cluster of Genes Encodes the Two Types of Chalcone Isomerase Involved in the Biosynthesis of General Flavonoids and Legume-Specific 5-Deoxy (iso) flavonoids in Lotus japonicus. Plant Physiol 131:941–951. doi: 10.1104/pp.004820.al.

301. Kimura Y, Aoki T, Ayabe S (2001) Chalcone Isomerase Isozymes with Different Substrate Specificities towards 6 ¢ -Hydroxy- and 6 ¢ -Deoxychalcones in Cultured Cells of Glycyrrhiza echinata , a Leguminous Plant Producing 5-Deoxyflavonoids. Plant cell Physiol 42:1169–1173.

302. Tunen AJ Van, Koes RE, Spelt E, et al. (1988) Cloning of the two chalcone flavanone isomerase genes from Petunia hybrida□: coordinate , light-regulated and differential expression of flavonoid genes. EMBO J 7:1257–1263.

303. Koes RE, Spelt CE, Mol JN (1989) The chalcone synthase multigene family of Petunia hybrida (V30): differential, light-regulated expression during flower development and UV light induction. Plant Mol Biol 12:213–25. doi: 10.1007/BF00020506

304. Barron D, Ibrahimt RK, Bat LDBV, et al. (1996) Isoprenylated flavonoids--a survey. Phytochemistry 43:921–982.

305. Zafar MS, Muhammad F, Javed I, et al. (2013) White Mulberry (Morus alba): A brief phytochemical and Pharmacological Evaluations Account. Int J Agric Biol 15:612–620.

306. Ferrari F, Delle Monache F (2004) A new phenolic glycoside from Sorocea ilicifolia stem bark. Fitoterapia 75:417–9. doi: 10.1016/j.fitote.2003.12.011

307. Sabbioni C, Ferranti A, Bugamelli F, et al. (2006) Simultaneous HPLC Analysis , with Isocratic Elution , of Glycyrrhizin and Glycyrrhetic Acid in Liquorice Roots. Phytochem Anal 17:25–31. doi: 10.1002/pca.877

308. Sabbioni C, Mandrioli R, Ferranti A, et al. (2005) Separation and analysis of glycyrrhizin , 18 beta-glycyrrhetic acid and 18 alfa-glycyrrhetic acid in liquorice roots by means of capillary zone electrophoresis. J Cromatography A 1081:65–71. doi: 10.1016/j.chroma.2005.03.044

309. Man S, Guo S, Gao W, et al. (2013) Identification of metabolic profiling of cell culture of licorice compared with its native one. Anal Bioanal Chem 405:3321–9. doi: 10.1007/s00216-013-6776-6

310. Casal S, Mendes E, Alves MR, et al. (2004) Free and conjugated biogenic amines in green and roasted coffee beans. J Agric Food Chem 52:6188–92. doi: 10.1021/jf049509u

311. Garrett R, Vaz BG, Hovell AMC, et al. (2012) Arabica and Robusta Coffees: Identification of Major Polar Compounds and Quantification of Blends by Direct-Infusion Electrospray Ionization − Mass Spectrometry. J Agric Food Chem 60:4253–4258.

312. Farah A, Donangelo CM (2006) Phenolic compounds in coffee. Brazilian J Plant Physiol 18:23–36.

313. Bakuradze T, Lang R, Hofmann T, et al. (2010) Antioxidant effectiveness of coffee extracts and selected constituents in cell-free systems and human colon cell lines. Mol Nutr Food Res 54:1734–43. doi: 10.1002/mnfr.201000147

314. Kitzberger CSG, Scholz MBDS, Pereira LFP, et al. (2013) Diterpenes in green and roasted coffee of Coffea arabica cultivars growing in the same edapho-climatic conditions. J Food Compos Anal 30:52–57. doi: 10.1016/j.jfca.2013.01.007

315. Kurzrock T, Speer K (2001) Diterpenes and diterpene esters in coffee. Food Rev Int 17:433–450.

316. Anthony F, Clifford MN, Noirot M (1993) Biochemical diversity in the genus Coffea L.: chlorogenic acids, caffeine and mozambioside contents. Genet Resour Crop Evol 40:61–70. doi: 10.1007/BF00052636

317. Ramirez-Coronel MA, Marnet N, Kolli VSK, et al. (2004) Characterization and estimation of proanthocyanidins and other phenolics in coffee pulp (Coffea arabica) by thiolysis-high-performance liquid chromatography. J Agric Food Chem 52:1344–9. doi: 10.1021/jf035208t

318. Perrone D, Farah A, Donangelo CM, et al. (2008) Comprehensive analysis of major and minor chlorogenic acids and lactones in economically relevant Brazilian coffee cultivars. Food Chem 106:859–867. doi: 10.1016/j.foodchem.2007.06.053

319. Casal S, Alves MR, Mendes E, et al. (2003) Discrimination between arabica and robusta coffee species on the basis of their amino acid enantiomers. J Agric Food Chem 51:6495–6501. doi: 10.1021/jf034354w

320. Rogers WJ, Michaux S, Bastin M, Bucheli P (1999) Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (Coffea canephora) and Arabica (C. arabica) coffees. Plant Sci 149:115–123. doi: 10.1016/S0168-9452(99)00147-8

321. Sridevi V, Giridhar P, Ravishankar GA (2009) Endogenous polyamine profiles in different tissues of Coffea sp., and their levels during the ontogeny of fruits. Acta Physiol Plant 31:757–764. doi: 10.1007/s11738-009-0289-0

322. Wei F, Furihata K, Koda M, et al. (2012) C NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin. J Agric Food Chem 60:10118–10125.

323. Joët T, Laffargue A, Descroix F, et al. (2010) Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. Food Chem 118:693–701. doi: 10.1016/j.foodchem.2009.05.048

324. Bradbury AGW (2001) Chemistry I: Non-Volatile compounds. Coffee Recent. Dev. 69:

325. T. K. L (2012) Edible Medicinal and Non-Medicinal Plants. doi: 10.1007/978-90-481-8661-7

326. Koshiro Y, Jackson MC, Katahira R, et al. (2007) Biosynthesis of chlorogenic acids in growing and ripening fruits of Coffea arabica and Coffea canephora plants. Z Naturforsch C 62:731–42.

327. Alonso-salces RM, Guillou C, Berrueta LA (2009) Liquid chromatography coupled with ultraviolet absorbance detection , electrospray ionization , collision- induced dissociation and tandem mass spectrometry on a triple quadrupole for the on-line characterization of polyphenols and methylxanthines in gree. Rapid Commun mass Spectrom 23:363–383. doi: 10.1002/rcm

328. Salmona J, Dussert S, Descroix F, et al. (2008) Deciphering transcriptional networks that govern Coffea arabica seed development using combined cDNA array and real-time RT-PCR approaches. Plant Mol Biol 66:105–124. doi: 10.1007/s11103-007-9256-6

329. Geromel C, Ferreira LP, Bottcher A, et al. (2008) Sucrose metabolism during fruit development in Coffea racemosa. Ann Appl Biol 152:179–187. doi: 10.1111/j.1744-7348.2007.00199.x

330. Joët T, Laffargue A, Salmona J, et al. (2009) Metabolic pathways in tropical dicotyledonous albuminous seeds: Coffea arabica as a case study. New Phytol 182:146–162. doi: 10.1111/j.1469-8137.2008.02742.x

331. Postle AD (2008) Phospholipid Profiling. In: Griffiths W (ed) Metabolomics, metabonomics Metab. profiling, RCSPublish. pp 116–133

332. De Castro RD, Marraccini P (2006) Cytology , biochemistry and molecular changes during coffee fruit development. Brazilian J Plant Physiol 18:175–199.

333. Clifford MN, Kazi T (1987) The influence of coffee bean maturity on the content of chlorogenic acids, caffeine and trigonelline. Food Chem 26:59–69. doi: 10.1016/0308-8146(87)90167-1

334. Lepelley M, Cheminade G, Tremillon N, et al. (2007) Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in C. canephora. Plant Sci 172:978–996. doi: 10.1016/j.plantsci.2007.02.004

335. Campa C, Noirot M, Bourgeois M, et al. (2003) Genetic mapping of a caffeoyl-coenzyme A 3-O-methyltransferase gene in coffee trees. Impact on chlorogenic acid content. Theor Appl Genet 107:751–6. doi: 10.1007/s00122-003-1310-4

336. Pfeiffer J, Kühnel C, Brandt J, et al. (2006) Biosynthesis of flavan 3-ols by leucoanthocyanidin 4-reductases and anthocyanidin reductases in leaves of grape (Vitis vinifera L.), apple (Malus x domestica Borkh.) and other crops. Plant Physiol Biochem 44:323–34. doi: 10.1016/j.plaphy.2006.06.001

337. Peng Q-Z, Zhu Y, Liu Z, et al. (2012) An integrated approach to demonstrating the ANR pathway of proanthocyanidin biosynthesis in plants. Planta 236:901–18. doi: 10.1007/s00425-012-1670-6

338. Geromel C, Ferreira LP, Guerreiro SMC, et al. (2006) Biochemical and genomic analysis of sucrose metabolism during coffee (Coffea arabica) fruit development. J Exp Bot 57:3243–58. doi: 10.1093/jxb/erl084

339. Garcia PA, Braga de Oliveira A, Batista R (2007) Occurrence, Biological Activities and Synthesis of Kaurane Diterpenes and Their Glycosides. Molecules 12:455–483. doi: 10.1002/chin.200735259

Bibliography

340. Obatomi DK, Bach PH (1998) Biochemistry and Toxicology of the Diterpenoid Glycoside Atractyloside. Food Chem Toxicol 36:335–346.

341. Lang R, Fromme T, Beusch A, et al. (2013) 2-O-β-D-Glucopyranosyl-carboxyatractyligenin from Coffea L. inhibits adenine nucleotide translocase in isolated mitochondria but is quantitatively degraded during coffee roasting. Phytochemistry 93:124–35. doi: 10.1016/j.phytochem.2013.03.022

342. Lang R, Fromme T, Beusch A, et al. (2014) Raw coffee based dietary supplements contain carboxyatractyligenin derivatives inhibiting mitochondrial adenine-nucleotide-translocase. Food Chem Toxicol 70:198–204. doi: 10.1016/j.fct.2014.05.017

343. Chiwocha SDS, Abrams SR, Ambrose SJ, et al. (2003) A method for profiling classes of plant hormones and their metabolites using liquid chromatography-electrospray ionization tandem mass spectrometry: an analysis of hormone regulation of thermodormancy of lettuce (Lactuca sativa L.) seeds. Plant J 35:405–417. doi: 10.1046/j.1365-313X.2003.01800.x

344. Erb M, Glauser G (2010) Family business: multiple members of major phytohormone classes orchestrate plant stress responses. Chemistry (Easton) 16:10280–9. doi: 10.1002/chem.201001219

345. Upchurch RG (2008) Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. Biotechnol Lett 30:967–77. doi: 10.1007/s10529-008-9639-z

346. Kachroo P, Shanklin J, Shah J, et al. (2001) A fatty acid desaturase modulates the activation of defense signaling pathways in plants. Proc Natl Acad Sci U S A 98:9448–53. doi: 10.1073/pnas.151258398

347. Speer K, Kölling-speer I (2006) The lipid fraction of the coffee bean. Brazilian Plant Physiol 18:201–216.

348. Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol 131:878–885. doi: 10.1104/pp.102.017319.Flavonoids

349. Kitagawa I (2002) Licorice root . A natural sweetener and an important ingredient in Chinese medicine *. Pure Appl Chem 74:1189–1198.

350. Cheynier V, Comte G, Davies KM, et al. (2013) Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. Plant Physiol Biochem 72:1–20. doi: 10.1016/j.plaphy.2013.05.009

9 Appendix

Table A 1 MS ESI and optical parameters selected for the analysis of SHA, JA and SA

MS parameters		
Source	**Polarity**	Negative
	Ionization temperature (°C)	365
	Sheath gas flow (µL/min)	41
	Aux gas(µL/min)	14
	Sweep gas(µL/min)	0
	Ion Spray voltage (kV)	4.0
	Capillary temperature (°C)	275
	Capillary Voltage (V)	-20
	Tube Lens (V)	-57.97
Lens	**Multipole RF Amplifier (Vp-p)**	400
	Multipole 00 Offset (V)	4.5
	Lens 0 Voltage (V)	4.5
	Multipole 0 Offset (V)	6
	Lens 1 Voltage (V)	11
	Gate Lens Offset (V)	54
	Multipole 1 Offset (V)	18
	Front Lens (V)	5.5

Appendix

Table A 2 Concentrations of JA, SHA and SA in the *Nicotiana Langsdorfii* samples. Means, relative standard deviations (RSD %), minimum and maximum values for each set

GR				Wild Type				ROLC			
	SA	SHA	JA		SA	SHA	JA		SA	SHA	JA
GR11	7 x10^3	2,9 x10^3	1,8 x10^2	WT1	1,6 x 10^4	3,2 x10^3	7 x10^1	ROLC12	8, x10^3	2,9 x10^3	1 x10^1
GR12	1x 10^4	3,6 x10^3	8 x10^1	WT6	1,0 x 10^4	3,1 x10^3	1 x10^2	ROLC13	4, x10^3	2,9 x10^3	3 x10^1
GR13	8 x10^3	4,3 x10^3	6 x10^1	WT7	1,2 x 10^4	3,7 x10^3	3 x10^1	ROLC14	1,1 x 10^4	4,2 x10^3	3 x10^1
GR14	1 x 10^4	5,5 x10^3	5 x10^1	WT9	6 x10^3	3,5 x10^3	8 x10^1	ROLC15	9, x10^3	3,6 x10^3	2 x10^1
GR15	1,2 x 10^4	2,6 x10^3	1,0 x10^2	WT10	9 x10^3	2,5 x10^3	5 x10^1	ROLC16	1,3 x 10^4	3,2 x10^3	2 x10^1
GR16	1,2 x 10^4	4,0 x10^3	6 x10^1	WT13	1,1 x 10^4	2,8 x10^3	5 x10^1	ROLC17	1,2 x 10^4	3,6 x10^3	4 x10^1
GR17	9 x10^3	3,2 x10^3	2 x10^1	WT14	1,2 x 10^4	4,1 x10^3	3 x10^1	ROLC18	1,7 x 10^4	4,5 x10^3	3 x10^1
GR18	3 x10^3	4,1 x10^3	9 x10^1	WT15	1,6 x 10^4	4,0 x10^3	5 x10^1	ROLC19	8 x10^3	4,5 x10^3	5 x10^1
GR19	6 x10^3	3,6 x10^3	4 x10^1	WT18	5 x10^3	3,1 x10^3	7 x10^1	ROLC20	1 x 10^4	3,7 x10^3	4 x10^1
GR20	7 x10^3	4,9 x10^3	3 x10^1	WT19	1,4 x 10^4	3,7 x10^3	7 x10^1	ROLC21	1 x 10^4	5,4 x10^3	2 x10^1
Mean	8 x 10^3	3,3x 10^3	7 x10^1	**Mean**	1,1 x 10^4	3,4 x 10^3	6 x10^1	**Mean**	1,0 x 10^4	3,3 x 10^3	3 x10^1
RSD%	32	23	66	**RSD%**	34	15,64	34	**RSD%**	33	20	35
Max	12418	5504	184	**Max**	16328	4146	95	**Max**	16998	5432	46
Min	3067	2588	21	**Min**	5000	2516	33	**Min**	4313	2925	13
GR water stress				**Wild water stress**				**ROL C water stress**			

	SA	SHA	JA		SA	SHA	JA		SA	SHA	JA
GRWS1+3	1,4 x 10^4	3,0 x10^3	1,2 x10^2	WTWS3	2,1 x 10^4	<LOD	7 x10^1	ROLCWS12+15	1,2 x 10^4	6,6 x10^3	3 x10^1
GRWS4	1,5 x 10^4	1,8 x10^3	1 x10^2	WTWS4	5 x10^3	3,7 x10^3	4 x10^1	ROLCWS13+14	1,3 x 10^4	4,3 x10^3	7 x10^1
GRWS6	6 x10^3	2,5 x10^3	8 x10^1	WTWS6	8 x10^3	1,5 x10^3	6 x10^1	ROLCWS17+18	7 x10^3	5,3 x10^3	7 x10^1
GRWS8	1,7 x 10^4	<LOD	1 x10^2	WTWS10	9 x10^3	4,4 x10^3	5 x10^1	ROLCWS19	6 x10^3	3,0 x10^3	4 x10^1
GRWS9	1,4 x 10^4	2,2 x10^3	3,3 x10^2	WTWS13+7	7 x10^3	3,1 x10^3	4 x10^1				
GRWS11	6 x10^3	1,7 x10^3	8 x10^1								
Mean	1,2 x 10^4	2,1 x 10^3	1,3 x10^2	**Mean**	1 x 10^4	3,3 x 10^3	5 x10^1	**Mean**	9 x 10^3	4 x 10^3	5 x10^1
RSD%	36	24	70	**RSD%**	64	39	25	**RSD%**	34	32	39
Max	17061	3048	325	**Max**	20910	4435	67	**Max**	12816	6648	69
Min	5739	1727	78	**Min**	4928	1503	36	**Min**	6122	3033	30

GRCR 50 ppm				**Wild Type CR 50 ppm**				**ROLC CR 50 ppm**			
	SA	SHA	JA		SA	SHA	JA		SA	SHA	JA
GRCR1	1,7 x 10^4	2,9 x10^3	1,5 x10^2	WTCR4	2,0 x 10^4	3,8 x10^3	1,2 x10^2	ROLCCR11	7 x10^3	5,2 x10^3	5 x10^1
GRCR2	1,2 x 10^4	5,9 x10^3	9 x10^1	WTCR7	1,2 x 10^4	4,3 x10^3	6 x10^1	ROLCCR12+13	1,4 x 10^4	7,4 x10^3	5 x10^1
GRCR3	1,0 x 10^4	2,4 x10^3	9 x10^1	WTCR8	1,8 x 10^4	9,7 x10^3	7 x10^1	ROLCCR14	9 x10^3	5,0 x10^3	7 x10^1
GRCR4	1 x 10^4	3,9 x10^3	8 x10^1	WTCR9	2,3 x 10^4	7,3 x10^3	1,3 x10^2	ROLCCR15+18	1,2 x 10^4	1,5 x 10^4	6 x10^1
GRCR5	1,4 x 10^4	3,1 x10^3	9 x10^1	WTCR10	2,1 x 10^4	4,7 x10^3	1 x10^2	ROLCCR16	8 x10^3	6,3 x10^3	1,4 x10^2

GRCR6	5 x10^3	4,5 x10^3	1,0 x10^2	WTCR12	2,2 x 10^4	3,4 x10^3	7 x10^1	ROLCCR17+19	8 x10^3	5,7 x10^3	7 x10^1
GRCR7	1 x 10^4	3,7 x10^3	1,1 x10^2	WTCR13	1,8 x 10^4	3,3 x10^3	6 x10^1	ROLCCR20	7 x10^3	1,0 x 10^4	3 x10^1
GRCR8	9 x10^3	4,7 x10^3	9 x10^1	WTCR17	1,6 x 10^4	1,09 x 10^4	8 x10^1				
GRCR9	8 x10^3	4,2 x10^3	8 x10^1	WTCR18	9 x10^3	6,6 x10^3	1,0 x10^2				
GRCR10	2,1 x 10^4	3,7 x10^3	1,1 x10^2	WTCR20	1,7 x 10^4	8,0 x10^3	1,1 x10^2				
GRCR11	1,4 x 10^4	3,8 x10^3	9 x10^1								
Mean	1,2 x 10^4	4,4 x 10^3	1,0 x10^2	**Mean**	1,7 x 10^4	6 x 10^3	9 x10^1	**Mean**	1 x 10^4	7,2 x 10^3	7 x10^1
RSD%	39	24	20	**RSD%**	26	43	28	**RSD%**	30	35	51
Max	21314	5940	153	**Max**	23135	10887	133	**Max**	14189	11539	135
Min	5329	2404	80	**Min**	8710	3320	58	**Min**	6956	4957	31
GR Heat stress				**Wild Type Heat stress**				**ROLC Heat stress**			
	SA	SHA	JA		SA	SHA	JA		SA	SHA	JA
GRHS1	1,1 x 10^4	5,7 x10^3	1 x10^2	WTHS1Tot	6 x10^3	4,5 x10^3	1,3 x10^2	ROLCHS1	1,3 x 10^4	4,3 x10^3	1,8 x10^2
GRHS2S	1,1 x 10^4	5,0 x10^3	1,0 x10^2	WTHS2Aerial	6 x10^3	4,1 x10^3	1,3 x10^2	ROLCHS2	1,4 x 10^4	5,0 x10^3	1,9 x10^2
				WTHS3Aerial	7 x10^3	4,6 x10^3	1,2 x10^2				
Mean	1,1 x 10^4	5,4 x 10^3	1,0 x10^2	**Mean**	6 x 10^3	4,4 x 10^3	1,3 x 10^2	**Mean**	1,3 x 10^4	5 x 10^3	1,8 x 10^2
RSD%	160	469	3	**RSD%**	826	234	4	**RSD%**	696	472	6
Max	11268	5651	101	**Max**	7400	4566	133	**Max**	13543	4981	188

Min	11042	4987	97	**Min**	5855	4122	124	**Min**	12558	4312	179

Table A 3 Total concentrations, mean concentrations (ng/g) and relative standard deviations (RSD %) of fatty acids in samples of *Nicotiana Langsdorfii*

ng/g	palmitoleic acid (16:1)	palmitic acid (16:0)	linolenic acid (18:3)	linoleic acid (18:2)	oleic acid (18:1)	stearic acid (18:0)	arachidonic acid (20:4)	Total content (ng/g)
WT	61	6879	7271	3572	97	638	2.42	18520
RSD %	22	4	62	21	18	2	63	
WTHS	121	8074	19730	10336	365	744	0.72	39371
RSD %	26	5	54	24	28	6	44	
ROL C	71	7243	10025	6826	169	752	1.56	25088
RSD %	32	6	47	27	36	10	26	
ROLCHS	170	8157	27701	11656	260	868	1.44	48814
RSD %	28	5	54	24	28	6	44	
GR	5	612	1015	647	16	75	0.31	2371
RSD %	28	5	54	24	29	6	44	
GRHS	60	6367	21630	8037	260	732	0.43	37087
RSD %	29	5	54	24	27	6	40	

Table A 4 Parameters used for MetAlign data processing of Nicotiana Langsdorfii samples in negative and positive polarities

	Negative	Positive
Mass resolution	50000	50000
Retention begin (scan)	560	549
Retention end (scan)	3197	2830
Maximum amplitude	150000000	400000000
Peak slope factor (x noise)	1	2
Peak threshold factor (x noise)	2	3
Peak threshold (abs value)	500000	200000
Average peak width at half height (scan)	10	10
Max shift (begin)	10	10
Max shift (end)	10	10
Maximum shift (per 100 scans)	35	25
Min. factor (x noise)	4;3	4;3
Min. n° of masses	2;2	2;2

Table A 5 Parameters used for MSClust processing of *Nicotiana Langsdorfii* samples in both negative and positive polarities

Parameters	
efficient peaks	100
peak width	10
peak width margin softness	2
correlation threshold	0.9
correlation threshold margin softness	0.02
PD reduction	0.8
PD reduction softness	0.01
Stop criterion	2

Table A 6 Complete table of metabolites identified in positive polarity in *Nicotiana Langsdorfii* samples, with: retention time, accurate mass measured, chemical formula, Error on mass measure, name of compounds and fragments. The diagnostic fragments used for identification are highlighted. The identifications at level 3 are indicated with an asterisk.

Retention time (min)	Accurate Mass	Chemical Formula	Error (ppm)	Molecule	Fragments
9,06	163,1219	C10H14N2	-6,4	Nicotine	***132.08073***; 106.06503; 130.06526; 120.08078
9,89	179,1167	C10H14N2O	-6,7	Oxi-nicotine	***132.08057***; 161.10715; 106.06503; 130.06526; 120.08078
10,42	284,0972	C10H13N5O5	-6,2	Guanosine	152,05666
11,91	322,2106	C17H27N3O3	-5,9	Feruloyl spermidine	***305.18561***; 248.12772; ***234.11206***; 160.07513; ***177.05434***; 155.11765
12,38	268,1025	C10H13N5O4	-5,7	Adenosine	136,06099
12,38	251,1376	C13H18N2O3	-5,5	Caffeoylputrescine	***234.11153;*** 163.03830; 114.09075; ***98.05950***
12,92	374,1062	C15H19NO10	-5,3	Dihydroxymethoxybenzoaxinone glucoside	***198.07489***; 180.06445; 124.03867; ***106.02806;*** 66.14638
13,43	323,1583	C16H22N2O5	-5,8	Benzyloxicarbonyl, acetyl, methyl ornithine ester	***200.12810***; 154.12228

13,70	266,1372	C14H19NO4	-5,8	Benzyloxicarbonyl-leucine	***207.06474; 163.03876***; 154.10298
14,08	265,15320	C14H20N2O3	-5,5	Feruloyl putrescine	***248.12729; 177.05392***; 152.06992; ***145.02782***; 114.09075
14,25	470,2257	C25H31N3O6	-6,2	Dicaffeoylspermidine	453.2001; 399.15326; ***382.12698; 381.14301; 364.11642***; 356.14771; 279.06406
14,88	474,2571	C25H35N3O6	-5,7	Bis-dihydrocaffeoyl spermidine	457.23346; ***236.12798***; ***222.11227; 165.05458***; 123.04401
15,32	484,2414	C26H33N3O6	-5,8	Caffeoyl feruloyl spermidine	467.21609; ***413.16925; 396.14343; 395.15878***; 370.16351; ***218.11676***; 192.10144
15,46	355,1006	C16H18O9	-5	3-caffeoylquinic acid	193.04849; ***163.03813;*** 145.02768
15,51	325,1377	C15H20N2O6	-5,4	Leucine β xantine	-
15,62	472,2419	C25H33N3O6	-4,9	Dicaffeoyl spermidine monohydrate	455.21646; 310.21140; ***293.18491***; 236.12732; ***234.11162***; ***220.09596; 163.03831***
15,96	488,2727	C26H37N3O6	-5,7	Dihydrocaffeoyl feruloyl spermidine	471.24704; ***250.14264***; 236.12689; ***222.11137; 165.05379***; ***137.05939***; 123.04344; 100.07502
16,04	355,127	C15H22N4O2S2	3,7	Allithiamine	***232.0957***; 163.03818

16,24	470,2261	C25H31N3O6	-5,2	Dicaffeoylspermidine	453.20026; ***308.19534; 291.16891; 234.11145***; 220.09570; ***163.03813***
16,51	738,4383	C39H63NO12	-5,4	δ-solamarine	***720.42792; 702.41833***; 574.36957; 556.35999; 394.30771
16,72	486,2573	C26H35N3O6	-5,3	Caffeoyl feruloyl spermidine monohydrate	469.23172; 310.21124; ***293.18481***; 248.12711; 239.13766; ***222.11151; 165.05365***; 123.04331
16,83	738,4383	C39H63NO12		δ-solamarine	***720.42792; 702.41833***; 574.36957; 556.35999; 394.30771
17,20	484,2414	C26H33N3O6		Caffeoylferuloylspermidine	467.21555; 322.21091; 308.19534; ***305.18420; 291.16867***; 237.12192; ***220.09557***; 177.05356; ***163.03784***; 155.11690
17,31	369,1159	C17H20O9	-5,7	Feruloylquinic acid	177,05342
17,59	500,27281	C27H37N3O6	-5,4	Diferuloyl spermidine monohydrate	483.24591; 350.20551; 324.22653; ***307.19992***; 253.15306; 248.12666; ***236.12669; 179.06912***; 177.05345; 137.05878
17,74	291,0959	C14H14N2O5	-5,8	Malonyltriptophan	***273.08527***; 245.09061

17,83	1048,563	C51H85NO21	-5,2	Dihydrosolasuaveoline	***1030.55273***; 1010.52734; 902.50659; ***884.49573***; 756.44946; 738.43768; 430.33069
17,90	868,5009	C45H73NO15	-5,1	β-solamarine	***850.49109***; ***704.43292***; ***558.37524***; 396.32376; 253.19334
18,28	884,496	C45H73NO16	-4,7	α-solamarine	***866.49042; 848.48169***; 818.43225; 720.43225; ***702.42157***; 412.32159; 394.31088; ***376.29977***; 251.17940
18,28	M+1028.53845	C33H51NO8 + 3 C6H10O4	-5.4	Glucopyranosyl petisidine +3 deoxihexose	***1012.54205; 994.53400***; 882.48145; ***864.46887***; 736.42383; 718.41199; 410.30292; 394.30890
18,33	738,4388	C39H63NO12	-4.7	δ-solamarine	***720.42792; 702.41785***; 574.36963; 412.31863; 394.30832; 251.17798
18,58	868,5016	C45H73NO15	-4,3	Solamargine	***850.49182***; 722.44464; ***704.43298***; ***558.37561***; 396.32388; 378.31375;
18,81	371,2045	C19H30O7	-5,3	Megastigmenine-diol	***209.15236***; 191.14183; 133.10028
18,93	1028.53845 (590 + 3*146)	C33H51NO8 + 3 C6H10O4	-5.4	Glucopyranosyl petisidine +3 deoxihexose	***1010.52826***; 882.48169; ***864.46906***; 736.42426; ***718.41205***; 590.36664; 428.31448; 410.30298
19,04	720.42841 (738 - 18)	C39H63NO12 - H2O	-4.6	δ-solamarine - H2O	***702.41827***; 574.37122; ***556.35986***; 412.31915; 394.30875; ***376.29828***

19,36	576,3872	C33H53NO7	-4	Glucopyranosyl solasodine	***558.37671; 396.32352***; 378.31329; 271.20395; 253.19400; 197.13144; ***157.10023***
19,60	1014.56030 (868+146)	C45H73NO15 + C6H10O4	-3,1	β-solamarine	996.54950; ***868.50317***; ***722.44531***; ***850.49054***; 704.43335; 576.38672; 414.33530; 396.32404; 378.31357
19,66	868,5028	C45H73NO15	-2,9	β-solamarine	850.49194; 722.44501; 704.43445; 576.38733; 414.33493; 396.32422; 378.31396
19,73	576,3875	C33H53NO7	-3.7	Glucopyranosyl solasodine	***558.37695; 396.32388***; 378.31363; 271.20416; ***253.19409***; 211.14709; 197.13150; 171.11606; ***157.10030***; 145.10036
20,01	1014.5592 (868 + 146)	C45H73NO15 + C6H10O4	-4	β-solamarine	***996.54828***; 868.50146; ***850.48993***; 722.44275; ***704.43298***; 576.38660; 396.32382; 378.31339
20,25	722,4441	C39H63NO11	-4.6	γ-solamarina	***704.43500; 558.37555***; 396.32382; 378.31360; 253.19383; ***157.10016***
20,44	762,4385	$C39H65NO12Na^{+}$	-1.9	Licoperoside D	***702.41846; 556.35980***; 394.30798; 376.29755
20,66	576,3864	C33H53NO7	-5.3	Glucopyranosyl solasodine	***558.37610; 396.32330***; 378.31302; 271.20374; 253.19368; 157.10049

20,90	1177,595	C57H92O25	-4,3	Asterlingulatoside D	1031.53223; 869.48462; ***723.42828***; 577.37091; 415.31870;
20,98	868,5017	C45H73NO15	-4.2	Solamargine	***850.49164; 704.43268; 558.37537***; 396.32376; 378.31354
21,12	848.47461 (738 - 2*18 + 146)	C39H59NO10 - 2 H2O + C6H10O4	-8.6	δ-solamarine + deoxihexose	***702.41852***; ***556.36108; 394.30881***; 376.29828
21,39	702.41772 (556 + 146)	C33H49NO6 + C6H10O4	-4,1	Triacetylspirosolene diol + deoxihexose	***556.36072***; 394.30853; 376.29813; 148.11116
21,54	1030,552	C51H83NO20	-5,7	Hiacinthoside	***1012.54315***; 994.52521; 884.49646; ***866.48535***; 738.43848; 720.42847; 582.53406; 412.31766; 394.31018;
25,84	318,2987	C18H39NO3	-4.9	Phytosphingosine	***300.28806; 282.27768***; 270.27759; ***264.2670***3; 252.26694
26,24	625,2514	C36H36N2O8	-4.9	Grossamide	***462.18912***; 351.08478; 336.06091; ***325.10574***; ***323.08948***; 307.09503
26,25	440.29819 (422 + 18)	C24H39NO5 + H2O	-5,4	Hidroxicassaine	***422.28781; 404.27734***; 386.26678; 312.17874
26,90	617,2747	$C27H45O14Na^+$	-5,4	Tetraacylsucrose (S4:14)	***455.22473***; 413.21387; 367.17252; ***353.15628***

27,00	442.31393 (424 + 18)	C24H41NO5 + H2O	-5,4	Dihydrohydroxycassaine	***424.30484***; ***406.29462***; 388.28482; 312.18188
27,16	699,352	$C33H56O14Na^+$	-6,1	Gingerglycolipid A	***537.30365; 375.25073***
27,59	353,2667	C21H36O4	-5.5	MAG (18:3/0:0)	335.25851; 279.23215; ***261.22159***; 243.21097
27,78	631,2903	$C28H47O14Na^+$	-5,3	Tetraacylsucrose (S4:15)	***469.24084; 367.17264***; 353.15707
27,94	476,2745	C23H42NO7P	-5,7	LysoPE (18:3/0:0)	***458.26587; 415.22375***; 335.25751; 304.26422; 261.22104
28,27	659,2849	$C29H48O15Na^+$	-5.5	Tetraacylsucrose (S4:17)	557.21924; ***455.22427***; 367.17212; 353.15634; ***251.08800***
28,42	701,36810	$C33H58O14Na^+$	-5,4	Tetraacylsucrose(S4:20)	***539.31738***; 449.21768; ***377.26486***
28,90	451.31430 (434+ NH3)	$C25H39NO5 + NH4^+$	-5,2	Arachidonoyl glutamate	434.28815; ***322.27267; 305.24606***; 173.09111
29,25	673,3008	$C30H49O15Na^+$	-5	Tetraacylsucrose (S4:18)	571.23712; ***469.24173***; 451.23108; ***367.17343; 265.10498***
29,49	573,30090	C29H44O9	-9,1	Ramnosyl sarmentogenine	**375,25055**
30,01	659,321	$C30H51O14Na^+$	-5,9	Tetraacylsucrose (S4:17)	***497.27191***; 409.21927; ***395.20389***; 353.15692;
30,87	277,2147	C18H28O2	-5.4	Octadecatetraenoic acid	259.20514; ***135.11642***
31,65	991,5548	$C51H84O17Na^+$	-5,3	DGDG (OH-18:3/OH-18:3)	829.50507; 713.33459; ***697.33966***; 681.34497; ***551.28094***; ***535.28625***; ***519.29126***; 405.13669

31,84	701,3319	$C_{32}H_{53}O_{15}Na^+$	-5,2	Tetraacylsucrose (S4:20)	585.25085; ***497.27100***; 409.21851; 395.20300; ***381.18735***; 353.15604
32,08	687,3523	$C_{32}H_{55}O_{14}Na^+$	-5,8	Tetraacylsucrose (S4:19)	***525.30493; 423.23654***; 409.22098; 381.18942; 367.17398; 353.15826
32,22	465,2586	$C_{27}H_{38}O_5Na^+$	-5,5	Muzagenine	***447.25092***; 311.25824
32,37	989,5391	$C_{51}H_{82}O_{17}Na^+$	-5.4	DGDG (2OH-36:4)	827.49103; 711.31909; 693.30872; ***681.34552; 549.26611; 519.29260***
32,53	715,3479	$C_{33}H_{56}O_{15}Na^+$	-4,5	Tetraacylsucrose (S4:21)	613.28265; 571.23541; ***511.28723***; 493.27679; 409.21927; ***367.17239; 265.10434***
32,93	295,225	$C_{18}H_{30}O_3$	-5.8	Hydroxyoctadecatrienoic acid	***277.21481***; 259.20459
33,02	801,4717	$C_{43}H_{70}O_{12}Na^+$	-5,3	MGDG (18:2/18:2)	535.28595; 517.27393; ***507.25479***; 491.25977; 489.24472; 465.20566; ***437.17697***; 425.21228; 409.18188; 397.18262;
33,37	729,363	$C_{34}H_{58}O_{15}Na^+$	-5.2	Tetraacylsucrose (S4:22)	613.28387; ***525.30389***; ***409.21967***; ***381.18848; 265.10416***
33,55	435,2482	$C_{21}H_{39}O_7P$	-5.5	PG (18:2/0:0)	***417.24094***; 337.27420; 263.23724

33,61	729,363	$C_{34}H_{58}O_{15}Na^+$	-5.2	Tetraacylsucrose(S4:22)	613.28455; ***525.30438; 409.21930***; ***381.18811***; 367.17242; 353.15616; 227.05237
34,31	829,5026	$C_{45}H_{74}O_{12}Na^+$	-5,6	MGDG (OH-18:3/OH-18:3)	551.28357; ***535.28851;*** 519.29364
34,52	975,5604	$C_{51}H_{84}O_{16}Na^+$	-4,9	α-galactosyl-DGDG (OH-18:3/18:3)	813.51489; 697.34253; ***681.34802***; ***535.28937; 519.29449***; 405.13812; 347.09601
34,52	743,3787	$C_{35}H_{60}O_{15}Na^+$	-5	Tetraacylsucrose (S4:23)	641.31396; 599.26715; ***539.31879***; ***437.25040***; ***395.20386***; 367.17249
35,09	757,394	$C_{36}H_{62}O_{15}Na^+$	-5,4	Tetraacylsucrose (S4:24)	613.28320; ***553.33466***; 535.32391; ***409.21970***; 265.10452
35,27	977,5757	$C_{51}H_{86}O_{16}Na^+$	-5.2	α-galactosyl-DGDG (OH-36:5)	815.52942; 699.35712; ***681.34711; 537.30341; 521.30872; 519.29340***
35,35	743.41479 (581 + 162)	$C_{32}H_{53}O_9$ + $C_6H_{10}O_5$	-3.9	Glucoconvallosaponin A	***581.36615; 437.25064***
35,44	757,3943	$C_{36}H_{62}O_{15}Na^+$	-5	Tetraacylsucrose (S4:24)	641.31445; 613.28351; ***553.33490; 437.25049; 409.21988***; 391.24573; 227.05208
36,03	953,5759	$C_{49}H_{86}O_{16}Na^+$	-5,2	DGDG (OH-34:3)	791.52814; 697.34088; ***659.36176; 535.28589; 497.30890***; 405.13702; 347.09500

36,27	785,4769	$C_{43}H_{70}O_{11}Na^+$	-5,3	MGDG (3OH-36:4)	675.40558; 519.29138; ***507.25510***; 491.26031; ***397.18207***; 243.08165
36,76	591,4949	$C_{37}H_{66}O_5$	-5.7	DAG (x:i/w:z)	***573.48840***; 335.25824; 313.27383; 261.22141
36,83	785,4246	$C_{38}H_{66}O_{15}Na^+$	-6,1	Tetraacylsucrose (S4:26)	***581.36627***; 437.25089; 409.21973
37,00	787,4921	$C_{43}H_{72}O_{11}Na^+$	-5,8	MGDG (3OH-36:3)	642.31750; ***582.36932***; 535.28796; 493.27731; ***437.25067***; 413.17575
37,10	256,26190	$C_{16}H_{33}NO$	-6.2	Palmitic amide	116.10642; ***102.09078***; 88.07507
37,18	785,4247	$C_{38}H_{66}O_{15}Na^+$	-5,9	Tetraacylsucrose (S4:26)	641.31268; ***581.36505***; 437.24976; 409.21866
37,43	609,2679	$C_{34}H_{40}O_{10}$	-2,5	Genkwanin H	***591.25818***; 559.23212; ***531.23737***; 515.24243; 503.24210; ***487.24728***; 475.24750; 459.25241
37,61	613,4799	$C_{39}H_{64}O_5$	-4,5	DAG (x:i/w:z)	***595.47083***; 577.45898; 539.44348; ***521.43445; 503.42255***
37,96	935,5649	$C_{49}H_{84}O_{15}Na^+$	-5,7	DGDG (34:4)	773.51801; ***681.34711***; 657.34717; ***519.29333***; 405.13779
38,16	282,2774	$C_{18}H_{35}NO$	-6	Octadecenoic amide	***265.25296***; 247.24242; 163.14822; 191.17979; 177.16417; ***135.11691***
38,31	714,5486	$C_{40}H_{75}NO_9$	-4	Glucocerebroside	***696.54144***; 534.48846

38,58	583,41150	C40H54O3	-5,3	Luteine derivative*	***565.40088; 547.39075***; 491.34915; 484.36868; 375.26514; 228.12947
39,23	937,5805	$C_{49}H_{86}O_{15}Na^+$	-5,7	DGDG (18:3/16:0)	775.52979; ***681.34314***; 659.35876; ***519.29034; 497.30603***; 405.13504; 347.09332,
39,36	607,2518	C35H34N4O6	-5,5	Pheophorbide b	579.25751; ***547.23151***; ***519.23688; 475.21048***; ***447.21609;*** 419.22049
39,58	769,4821	$C_{43}H_{70}O_{10}Na^+$	-5.2	MGDG (2OH-36:4)	519.29028; ***491.25891***; 243.08253
39,79	951,5958	$C_{50}H_{88}O_{15}Na^+$	-6	DGDG (35:3)	789.54523; ***681.34265***; 673.37408; ***519.29028; 511.32123***; 405.13461; 347.09341
40,00	593,2731	C35H36N4O5	-4.7	Pheophorbide a	***533.25256***; 547.26672; 515.24078; 492.24884; 473.23077; ***461.23120***; 433.23682
40,44	771,4976	$C_{43}H_{72}O_{10}Na^+$	-5,5	MGDG (2OH-36:3)	519.29059; ***493.27499***; 243.08241
40,57	593,2728	C35H36N4O5		Pheophorbide a	565.28229; ***533.25641***; 492.25165; 476.25827; ***461.23480***; 433.24011
40,95	797,5133	$C_{45}H_{74}O_{10}Na^+$	-5,1	MGDG (18:3/18:3)	***519.29279***; 241.06812
41,32	773,5127	$C_{43}H_{74}O_{10}Na^+$	-6,1	MGDG (2OH-36:2)	521.30872; ***493.27747***

41,79	799,5281	$C45H76O10Na^+$	-6,2	MGDG (36:5)	***519.29199***; 243.08334
42,13	651,4553	C34H67O9P	-6,5	PG (P-16:0/12:0)	541.38708; 373.23566; ***317.20938***; 299.19913
42,67	775,52820	$C43H76O10Na^+$	-6.3	MGDG (18:3/16:0)	756.55518; 521.31018; ***495.29343***
43,12	607,288	C36H38N4O5	-5,7	Methyl pheophorbide a	***547.26886***; 519.27417; 515.24237; 506.26602; 473.23138; ***461.23218***; 433.23761;

Table A 7 Complete table of metabolites identified in negative polarity in *Nicotiana Langsdorfii* samples, with: retention time, accurate mass measured, chemical formula, Error on mass measure, name of compounds and fragments. The diagnostic fragments used for identification are highlighted. The identifications at level 3 are indicated with an asterisk.

Retention time (min)	Accurate Mass	Chemical Formula	Mass error (ppm)	Molecule	Fragments
10,62	240,05103	C10H11NO6	-1,4	Dihydroxybenzoyl-serine	222.04071; 196.06151; ***124.04066***; 115.00393; 66.09714
11,85	399,13217	C15H27O10S	-2.2	Propyl -glucopyranosyl-2-tio-glucopyranoside	380.14566; 354.13031; 353.14194; ; 284.12451; 183.97498; 171.10554; ***136.99147***
12,41	312.09470 (M + AF -H), 266.08911 (M-H)	C10H13O4N5	-1,3	Adenosine	134,04735
12,47	411,17651	C19H28N2O8	-1.9	Glucopyranosyl-caffeoyl putrescine	393.16574; 375.15530; 351.15536; ***321.14481***; 291.13443; 249.06131
13,47	329,08722	C14H18O9	-1.8	Glucosyl-trihydroxyacetophenone	311.16058; ***167.03500***; 152.01155; 123.04528
13,74	353,08701	C16H18O9	-2,2	3-caffeoylquinic acid	335.08783; 191.05603; 179.03496; 135.04527
14,70	447.1499 (M + AF - H) 401.14990 (M-H)	C18H26O10	-2,3	Benzylalcohol hexose-pentose	401,14438
14,89	472,24429	C23H41NO7P	-5.7	LysoPE (18:4/0:0)	454.23398; 362.20770; 350.20789; ***308.19742***; 266.18662; ***186.16087***; 137.06073; 121.02952; ***240.17012***; 198.16086

15,06	241,11925	C11H18N2O4	-0.5	Dihydroxy-undecadienediamide	***197.12946***; 169.13428; 167.03519; ***141.10327***; 130.08760; 127.05133; 82.03002
15,46	353,0875	C16H18O9	-1.2	4-caffeoylquinic acid	191.05608; 179.03500; ***173.04559***; 155.03514; 135.04530
15,67	399,09286	C17H20O11 (C16H18O9+CH2O2)	-1.1	Linderofruticoside A	353.08759(-COO-); ***191.03491***; 176.01167(-C6H9O6); 148.01671(-C7H9O7)
15,90	371,0984	c16h20o10	-0,6	Dihydroferulic acid 4-O-glucuronide	249,0615
16,01	353,11353	C15H22N4O2S2	alto, tipo 6 (0 nel DD)	Allithiamine	309.12375 (-COO); 191.05594
16,22	210,07716	C10H13NO4	-0.1	Enicoflavine*	124,03988
16,24	468,21326	C25H31N3O6	-2.4	Dicaffeoyl spermidine	426.20053; ***332.15869***; 306.17999; ***289.15320***; 264.16956; 247.14346; 261.12228; ***135.04430***
16,60	308,07693	C14H15NO7	-2.1	Feruloyl aspartic acid	290.06567; 264.08655; 246.07622; ***193.04977***; 149.06044; 132.02995
16,90	338,08759	C15H17NO8	-1.7	Glucuronic-hydroxy-methoxyindole	294.09717(-COO); 276.08667(-H2O); ***223.06046***; 191.05566; 132.03009; 177.04289; 173.04495; 127.03952
17,30	367,10263	C17H20O9	-1,4	5-O-feruloylquinic acid	191.0559, 193.04993

17,35	461,23831	C22H38O10	-1,9	Anatolioside	***415.23178***; 340.11765
18,18	417.21228 (371 + 46)	C19H32O7 + CH2O2	-1,7	Glucosylated hydroxy-megastigmenone	***371.20676***; 179.05602; 161.0453
18,72	474,17612	C18H29NO4 + C6H10O5	-1,7	Feruloyl tyramine	**312,12332**
18,91	1072.53076 (1026 + 46)	C50H83N3O15P2 + CH2O2	4	Diacyl glicerol cytidin diphosphate (18:1/20:4)	***1026.52396***; 617.22766; 393.13867
18,94	579,20667	C28H36O13	-2.3	Glucosylated siringaresinol	**417,15439**
19,43	1058.55261 (M+FA - H) 1012.5484	C51H83O19N		γSolamarine +2 Fucoses	866.4870(-FUCOSE), 720.42975 (-FUCOSE); 574.37286
19,63	1220.6062 (M+FA-H) 1174.60034 (M-H)	C57O24H93N		γSolamarine +2Fucoses+ Hexose	1174.54565 (-HEXOSE); 866.4870(-FUCOSE), 720.42975 (-FUCOSE); 574.37286
19,75	187,0977	C9H1604	0.6	Nonadienoic acid	**125.09743** (-coo -H2O)
20,46	716.33606 (M+2AF-2H) 670.83099 (M-2H)	C45 H76 O18	-2	Torvoside a/b + 3 fucoses	693.83350, 670.83099, 597.29938, 524.27124, 659.80945, 451.24249; 1341 massa vera. 903 massa bassa base (451)
20,49	243.1235 (225+ 18)	C12H18O4 + H2O	-1,2	Hydroxyjasmonic acid	***225.11322;*** 181.12350
20,57	761,30255	C38H50O16	-0,1	Prieurianin	743.29108; ***717.31183***; ***689.28094***; 459.17590; ***301.11838***; ***257.12888;*** 242.10593
20,87	1129.52783 (1083 + 46)	C50H84O25 +	-0,5	Capsianoside II*	1083,51929

		CH2O2			
20,90	642.29901 (M-2H+2AF) 1239.60083 (M-H+AF) 1193.59558 (M-H)	C57H94O26	-1,7	Steroidal saponin + 3 fucoses*	1047.53418, 901.47650, 883.46655, 755.41895, 1031.53931, **575.5632**, **737.40863**, 431.31625; 619.29742 (M-2H-AF), 596.29443 (M-2H)
20,97	912.49481 (M +AF H) 866.48859 (M-H)	C45H73NO15	-1,6	β-solamarine	866.48859; 720.43011; 585.23651; 574.37231; 501,21021; 351.12759; **291.10745**
21,55	773.43140(611+hexose)	C39H66O15	-2,3	Steroidal saponin*	755.42047; **737.41077**; 611.37817; 593.36810; **575.35803**; 431.31580; 449.32541; 413.30457
21,93	649.30646*2 (M + 2AF - 2H) 1207.60669 (M-H)	C58H96O26	-3	Hoduloside VII + 2 Fucoses	1207.60669, 1061.55078, 915.49347, 603.29974, 626.30273, 876.61737, 489.21439
22,38	312,12326	C18H19NO4	-2,8	Feruloyltyramine	297.10043;270.11353; 253.08638; ***178.0511; 135.04544***
22,75	1023,46295	C47H76O24	-2,4	Nicotianoside VI	1005.32623; 979.47272; 937.46674
23,03	1113.53027 (921 + 146 + 46)	C44H74O20 + C6H10O4 + CH2O2	-1.9	Liciumoside IV	1067.52466; ***921.46832***; 775.40839; 757.39813; 629.35254; 611.34100
23,17	967.473755 (M + AF)921.46729(M -H)	C44H74O20	-2,9	Lyciumoside VI	921,46729

23,25	623,23798	C36H36N2O8	-3	Grossamide	***460.17490***; 446.15912; 445.15161; 350.13870; 283.09668; 268.07346
23,33	327,21680	C18H32O5	-2.7	Auxin a	***171.10280***, 309.20679(-h2o), 291.19638 (-h2o); 281.21234; 273.18607 (-h2o); 247.20618 ; 229.14456; 227.12840, 193.15941; 171.10280; 165.12828
23,70	1007,46826	C47H76O23	-2.2	Nicotianoside IV	963.4783(-COO); 921.4688; 903.4562; 775.4089 (LYCIUMOSIDE); 757.3984; 741.4038
23,97	641.24884 (623 + 18)	C36H36N2O8 + H2O	-2,5	Grossamide	***623.23816***; 608.21423; ***591.21106***; 551.25293; 489.20187; 460.17471; 432.17993; 328.1181
24,14	329,23248	C18H34O5	-2,4	Trihydroxyoctadecenoic acid	311.22232; 293.21201; ***229.14452***; ***211.13396***; 209.11800; 193.12325; ***183.13876***; 171.10280; 167.14400; 155.14412; 125.09725
24,29	611.25403 (M + AF - H) 565.24841 (M-H)	C25H42O14	-2,9	Tryacylsucrose S3:13 (C4, C4, C5)	565.24872; 523.23822; 481.19128, 467.17551, 425.16483, 505.22867
25,49	597.27478 (M + AF - H) 551.26849 (M - H)	C25H44O13	-3,5	Triacylsucrose S3:12 (C4, C4, C4)	467.21133; 425.16455; 407.15427

25,89	575.30646 (529 + 46)	C28H42N4O6 + CH2O2	-3.8	Bis(dihydrocaffeoyl)spermine	***529.29962; 253.09192***
26,00	359,07623	C18H16O8	-2,8	Trimethoxytrihydroxyflavone	***344.05246; 329.02905;*** 323.1853; 314.00558; 305.17490; ***301.03415***; 286.01089
26,26	623,23792	C36H36N2O8	-3.2	Grossamide	***486.15308***; ***471.13031***; 460.17419; 445.15137; 427.14050; 352.08105; ***297.11187***; 283.09644; ***282.08862***
26,51	593,26208	C27H46O12S	-3,37	SQMG C18:4	551.26563; 524.19019; 482.18076; 471.15204; 427.12585; 385.07922; **299.04291**; 275.20078; **225.00667**; **206.99649**; 164.98605; 152.98637; 125.02438; 66.09728
26,70	309,20636	C18H30O4	-2.5	Hydroperoxyoctadecatrienoic acid	***291.19531***; ***273.18503***; 251.16400; 247.20576; ***239.16443***; 229.1953; 221.15401; 209.11765; 195.13841; 183.10219; ***179. 14381***; 171.10225; ***155.10736***; 139.11263
26,89	639.28528 (M + FA - H) 593.27863 (M-H)	C27H46O14	-3,3	Triacylsucrose S3:15	509.22150, 491.21133, 551.26825, 407.15417, 467.21136, 449.20059, 425.16464

27,09	307,19058	C18H28O4	-2.9	Fatty acid derivative*	***289.18008***; 235.13370; 209.11818; 185.11813; 121.06603
27,09	721.36371 (M + AF - H) 675.35736 (M-H)	C33H56O14	-2,6	Digalactopyranosyl lineoylglicerol (C18:3)	397.13382, 415.14404, 179.05539, 374.50363, 125.02427, 161.04524, 179.05574, 253.09203, 277.09192, 287.07620, 323.09720
27,45	667.27979 (621 + 46)	C28H46O15 + CH2O2	-3.1	Tetraacylsucrose (S4:16)	621.27307; ***579.26300***; 537.21637; 523.23798;509.22083; ***495.20554***; 491.21097; 477.19528; 449.16418
27,61	721.36316 (675 + 46)	C33H56O14 + CH2O2	-2.8	LysoDGDG (18:3/0:0)	***675.35738***; 415.14404; **397.13361**; 374.50363; 323.09720; 305.08667; 287.07620; 253.09203; 235.08170
27,85	653.30066 (M + AF - H) 607.29388 (M-H)	C28H48O14	-2,6	Triacylsucrose 3:16	607.2939; 523.2369; 509.2212; **397.1334**; 235.0813; 565.28387; 481.22638; 425.16470
27,95	474,26108	C23H42NO7P	-3,3	LysoPE(18:3)/(0:0)	277.21631; 275.20081; 259.20599; 233.22694, 275.20081
28,34	265,18015	C16H26O3	-2,6	Hydroxy hexadecatrienoic acid (16:3)	247.16942; 207.13831; 203.17998; 221.19031

28,42	723.37903 (677 + 46)	C33H58O14 + CH2O2	-2.4	LysoDGDG (18:2/0:0)	***677.37360***; 415.14444; ***397.13382***; 379.12329; 323.09695; 305.08682; 287.07611; 253.09200; ***235.08176***; 221.06638; 179.05536; 161.04517; 125.02429
28,65	681.29541 (635 + 46)	C30H50O17 + CH2O2	-3	Tetraacylsucrose (S4:17)	635.28906, 593.27844, 551.23199, 509.23199, 431.21826, 509.22141, 467.17490, 491.21085, 453.15906, 407.15411, 523.23694, 477.19525, 575.26898
28,78	667.31647 (621 + 46)	C29H50O14 + CH2O2	-2.7	Triacylsucrose (S3:17)	621.30884; 537.25201; 523.23651
28,84	699.37903 (653 + 46)	C31H58O14 + CH2O2	-2.4	LysoDGDG (16:0/0:0)	***653.37390***; ***415.14478***; ***397.13416***; 375.12375; 361.11288; 323.09735; 305.08701; 287.07635; 253.09215; ***235.08189***; 179.05550; 161.04526; 143.03499; 125.02431
29,21	593,27142	C27H47O12P	-3,66	Lysophosphatidylinositol (18:3)	277.21609, 315.04715, 413.20789, 467.21106, 233.22659, 259.20563, 152.99541, 241.01076, 223.00040

29,31	695.31091 (649 + 46)	C30H50O15 + CH2O2	-2,8	Tetraacylsucrose (S4:18)	607.29370, **565.24725**, 523.23682, 505.22678, 491.21121, 551.23138, 481.19012, 467.17474, 449.16388, 421.16721
29,50	513.3043+AF	C27H46O9	-3	MGDG C18:3	513.3043, 277.2161, 253.09204, 233.22699, 259.20615, 161.04543
30,16	577,26709	C27H46O11S	-3,925	SQMGC18:3	225.00650, 299.04269, 277.21603 (linolenico), 243.01688, 206.99605, 164.98593, 125.02431
30,55	571,28693	C25H49O12P	-3.4	Phosphatidylinositol (16:0/0:0)	391.22360; 315.04718; ***255.23184***; 241.01115; 237.22194, 223.00066; ***152.99557***
30,57	709.3266 (663 + 46)	C31H51O15 + CH2O2	-2,7	Tetraacylsucrose (S4:19)	663.32013; ***621.30994***, **565.24786**, 523.23755, 579.26343, 537.25232, 505.22681, 439.17984, 425.16412, 407.15372, 421.16937, 323.09695, 603.29980
30,69	709,32684	C31H51O15 + HCOOH	-2,7	Tetraacylsucrose S4:19	621.30994, **565.24786**, 523.23755, 579.26.343, 537.25232, 505.22681, 439.17984, 425.16412, 407.15372, 421.16937, 323.09695, 603.29980

30,87	293,21124	C18H30O3		Hydroxylinolenic acid	275.20090; 211.13359; 265.21677; 195.13881; 177.12810; 66.09716
31,01	505,25571	C24H43O9P	-2,9	PG(18:3/0:0)	413.21021; 397.80786; 325.58813; **277.21759**; 275.20197; 265.53857; 259.20706; 233.22783; 179.18103;
31,22	579,28278	C27H48O11S	-3	SQMG	299.0428; 279.2319; 225.0067; 206.9964; 164.9860; 125.0244; 80.9653
31,58	555,28336	C25H48O11S	-2	SQMG	***299.04404***; ***225.00754; 206.99701***; 164.98657; 152.98669; 125.02486
31,67	1013,56891	C52H86O19	-0,1	Quinquenoside I	***967.56110***; 949.54913; 691.35284
31,74	723.34375 (677.33856 + AF)	C32H54O15 + AF	-0,8	Tetraacylsucrose (S4:20)	677.33539; 635.32489; 579.26276; **565.28613**; 537.25262; 551.27045; 509.22366; 449.20279; 439.18210; 411.14996, 341.06018
31,86	295,22757	C18H32O3	-1	Hydroxylinoleic acid	277.2169; 195.1390; 171.1027; 179.1442; 66.0975
32,12	481,25586	C22H43O9P	-2.8	PG (16:1/0:0)	389.20984; 279.58063; 253.21712; ***245.04312; 152.99605***; 66.09751

32,66	737.35791 (691.35138 + AF)	C33H55O15 + AF	-2	Tetraacylsucrose S4:21	**691.35138**; 649.34094; **607.29443**; **565.25153;** 523.23730; 551.26825; 505.22705, 439.18027, 407.19396, 323.09732, 341.10712, 305.08646, 425.16446, 491.21143, 409.89743
32,75	483,27142	C22H45O9P	-2.9	PG (16:0/0:0)	***255.23193***; 245.04231; ***152.9955***; 237.22159; 66.09727
32,91	741,38098	C47H83O13P	-3.0	Phosphatidylinositol	723.36719; 651.34570; 579.32745; 571.28553, 553.27551; 552.92188; 497.24915; 495.17130; ***485.14029***; ***391.22348***; 323.08847; 297.03677; ***255.23170***; 241.01080, 223.00017; ***187.09712***;
32,95	723.37891 (677.37201 + AF)	C33H48O14 + AF	-2,2	Triacylsucrose (S3:21)	551.26801, 593.31476, 467.21100, 449.20056, 575.30426, 533.25775, 407.15356, 425.16403, 323.09668,
33,26	749.35803 (M + AF - H) 703.35095 (M - H)	C34H56O15	-3	Acylsucrose*	703.35095, 661.34094, 607.29407, 565.28314, 619.29370, 481.19086
33,34	433,23444	C21H39O7P	-3.7	PG (18:2/0:0)	291.70773; 277.53760; ***152.99544***; 171.00587

33,55	737.39459 (691 + 46)	C34H60O14 + CH2O2	-2.6	Triacylsucrose (S3:22)	691.38696; ***565.28320***, 439.17969
33,62	751.37347 (705.36676 + AF)	C34H58O15 + AF	-2,8	Tetraacylsucrose S4:22	663.35651, 607.29419, 621.30945, 565.28253, 537.25287, 547.27234, 481.18988, 523.23676, 579.29883, 439.17981, 593.31451, 467.21103, 421.16916, 323.09641
33,68	737.39441 (691.38696 + AF)	C34H60O14 + AF	-3	Triacylsucrose S3:22	593.3143; **565.2831**; 551.2676; 537.2521; 467.2105
33,78	311,16769	C17H28O3S	-3	Undecilbenzen sulphonic acid	247.20596; ***183.01166***; 170.00407; ***119.05018***; 66.09720
34,54	997.57172 (675 + 146 + 176)	C33H56O14+ C6H10O4 + C6H8O6	-4.9	LysoDGDG (18:3/0:0)	***951.564***; 933.55334; ***675.35638***; 415.14310; 397.13235
34,66	765.38928 (719.38269 + AF)	C35H60O15 + AF	-2,7	Tetraacylsucrose (S4:23)	677.37256, 728.52789, 593.28180, 551.26843, 579.29944, 523.23730, 425.16415, 467.21054, 449.20099, 407.15384, 323.09601

35,06	774.53522 (728 + 46)	C40H76NO8P + CH2O2	7.9	PE	***728.52777;*** 710.51593; 665.22388; 646.95917; 572.37799; 548.46503; ***530.45465***; 374.30432; 324.25198
35,66	779,4046	C36H62O15 + AF	-3	Tetraacylsucrose S4:24	733.3983; 691.3877; 607.2943; **565.2839**; 593.3131; 547.2737; 481.1909; 467.1803; 439.1803: 421.1694, 323.09686, 509.22147
36,51	847,48505	C43H76O14S	-3.2	SQDG	827.46033; 801.47552; 789.44208; 709.37982; ***591.24481***; ***537.27106***; 533.19763; 453.13940; 286.96988; 283.0410; ***225.00621***; 193.89102
36,58	953.54498(907+AF)	C49H80O15	-2,7	DGDG (36:4)	657.34583, 397.13306, 415.14340, 277.21594, 657.34540, 629.31464, 379.12247, 675.35596, 249.18475
37,19	807.43567 (M + AF) 761.43048 (M-H)	C38H66O15 +AF	-3,1	Pentaacylsucrose (S5:25[4])	761.43048; 719.41907; 635.32556; 593.31500; 509.22134; **565.28375**; 467.21115; 439.18015; 421.16977
37,47	744.52448 (M - H +AF) (698.51740 + AF)	C39H73NO9+AF	-3,7	Lucyobroside	698.51733; 507.27069; 463.20764; 327.21631; 279.23178

37,53	807.43567 (M + AF) 761.43048 (M-H)	C38H66O15 +AF	-3,1	Pentaacylsucrose (S5:25[4])	719.41907, 635.32556, 593.31500, 509.22134, 565.28375, 439.18015, 467.21115, 421.16977, 323.09564
37,63	981.57678 (935 + 46)	C51H84O15 + CH2O2	-4.2	DGDG (18:3/18:3)	***935.57153***; 675.35699; 657.34729; 415.14401; 397.13367; 379.12335; 361.11218; 277.21606
38,11	835.51904 (M+ FA - H) 789.51251 (M - H)	C45H74O11	-3,7	MGDG(18:3)	513.30475, 277.21625, 293.21103, 253.09210
38,34	712,53491	C40H75NO9	-2,8	Glucocerebroside	550.48181; 532.47119; 514.46134; 502.46124; 463.45108; 340.28458; 324.25333; 312.25330; 306.24296; 278.24823; 271.22693; 270.24283; 225.22185
38,62	758.54010 (712 + 46)	C40H75NO9 + CH2O2	-3	Glucocerebroside	***712.53333***; 521.24963; 293.21118
38,65	607,25409	C35H36N4O6	-3.5	Hydroxy phaeophorbide a	***575.22754***; 563.26422; 531.23779; 519.27435; 504.25192; ***487.24820***; 475.24838; 459.54971; 405.36691; 384.17078

39,99	791.49506 (M+FA-H) 745.48956 (M-H)	C43H70O10	-0,4	MGDG (C18:3/16:3)	513.30682, 485.27536, 277.21235, 249.18608, 233.22693, 259.20593, 253.09314
40,48	793.50854 (747 + 46)	C43H72O10 + CH2O2	-3.4	MGDG (18:3/16:2)	***747.50214***; 513.30475; 277.21625; 251.20067
41,25	819.52417 (M + AF -H), 773.51764 (M-H)	C45H74O12+AF	-3,2	MGDG(18:3/18:3)	773.51764; 758.53070; 513.30444;**295.22644; 277.21606**;
41,48	787,46521	C41H72O12S	-2,8	SQDG (16:3/18:0)	537.27130, 531.22510, 225.00679, 206.99623, 283.04871, 164.98605
41,69	837,48047	C45H74O12S	-3,6	SQDG (C:18/C:18)	559.25586, 537.27173, 579.28174, 225.00673, 206.99609, 283.04831, 243.01701
41,92	665,41504	C37H63O8P	-5.9	PG (x:i/w:z)	***633.38818***; 615.37842; 589.39874; 354.08438; ***337.08163***; 311.10233; 310.09473; 309.08713; 296.07974; 295.07147; 293.09256; 282.0657; ***267.11337***; 188.03520

42,25	821.53943 (M + AF - H) 775.53333 (M -H)	C45H76O10	-3,2	MGDG(18:2)/(18:3)	775.53333; 515.32019; 513.30481; 279.23175; 277.21631
42,57	831,50012	C43H77O13P	-3,1	PI(C18:3, C16:0)	553.27600; 798.46417; 391.22357; 255.23169; 413.20779; 571.28613; 297.03677; 223.60058; 197.69548; 277.21637; 152.99562; 181.87497
42,82	815,49609	C43H76O12S	-2.9	SQDG (18:0/16:3)	577.26746; ***559.25604***; 555.28198; 537.27216***;*** 396.93982; 297.58200; 277.21628; 255.23151; ***225.00676***
43,35	815,49567	C43H76O12S	-2,7	SQDG (18:3, 16)	559.25604, 537.27216, 225.00674, 283.04855, 206.99649, 457.10858
46,68	951.50964 (M + AF - H) 905.50446 (M - H)	C55H70N4O6Mg	-3,1	Chlorophyll b	626.20087; 597.85999; 555.18665; 540.16290; 495.16629; 444.1101

Table A 8 Parameters used for MetAlign data processing of Glycyrrhiza samples in negative and positive polarities

Parameters	Negative	Positive
Mass resolution	50000	50000
Retention begin (scan)	439	436
Retention end (scan)	3154	3232
Maximum amplitude	600000000	100000000
Peak slope factor (x noise)	2	3
Peak threshold factor (x noise)	3	3
Peak threshold (abs value)	400000	200000
Average peak width at half height (scan)	13	10
Max shift (begin)	10	10
Max shift (end)	10	10
Maximum shift (per 100 scans)	25	25
Min. factor (x noise)	4;3	4;3
Min. n° of masses	2;2	2;2

Table A 9 Parameters used in MSClust data processing of *Glycyrrhiza* samples

Parameters	Negative	Positive
efficient peaks	50	100
peak width	10	10
peak width margin softness	2	2
correlation threshold	0.99	0.97
correlation threshold margin softness	0.02	0.02
PD reduction	0.85	0.88
PD reduction softness	0.01	0.01
Stop criterion	2	2

Table A 10 Complete table of metabolites identified in negative polarity in *Glycyrrhiza* samples, with: retention time, accurate mass measured, chemical formula, Error on mass measure, name of compounds and fragments. The compounds identified at level 3 (21 compounds) are marked with an asterisk.

Retention time (min)	Accurate Mass	Compounds	Molecular formula	Mass error (ppm)	Fragments
13,53	661,14349	Glucoliquiritin sulfate	C27H34O17S	-2,1	417.11768, 579.1705335, 255.06561, 459.12817, 135.00871, 119.05028
13,66	409,04404	Sulfoquinovose sugar derivative*	C14H18O12S	-2,4	241.00206, 222.99135, 138.97064, 96.96023, 180.98085, 164.98616, 150.97052
13,68	661,14349	Glucoliquiritin sulfate	C27H34O17S	-2	417.11768, 579.17035, 255.06561, 135.00871, 119.05028
15,09	499,09082	Liquiritin sulfate	C21H24O12S	-2,4	255,06552
15,30	631,13293	Liquiritin/Isoliquiritin apioside sulfate	C26H32O16S	-1,7	549.15991, 255.06544, 429.1017581, 417.11710
15,40	431,11905	Licoagroside B	C18H24O12	-1,69	329.08731, 369.11853, 203.05606, 243.08731, 125.02466, 179.03474, 269.06631, 225.07626, 207.06563
15,68	499,09116	Isoliquiritin cluster sulfate	C21H23O12S	-2,3	255.06596, 297.07590, 135.00870, 119.05021, 91.01904
16,07	711,21387	Liquiritigenin-apiosyl-glucoside	C32H40O18	-2,8	549.16040, 255.06567, 417.11823, 297.07608
16,29	289,0715	Catechin/Epicatechin	C15H14O6	-3,3	
16,32	593,15033	Apigenin dihexose flavonoid	C27H30O15	-3,4	473.10809, 503.11835, 575.13953, 353.06613, 383.07660, 413.08649, 325.07068, 297.07605
16,43	624,1748	Deoxydiglycosidic flavon sulfate*			542.20184, 577.15485, 415.10223, 255.06570, 297.07571, 253.04990
16,80	717,13324	Sulfo malonyl liquiritingenin apiofuranoside glucoside	C29H34O19S	-2,8	699.12268, 673.14276, 631.13226, 591.17047, 549.16052, 531.14911, 471.11279, 255.06532
17,15	563,14001	Apigenin glycoside (hexose-pentose)	C26H28O14	-2,5	443.09805, 473.10849, 545.12964, 383.07709, 353.06659, 413.08658, 325.07071, 297.07602

17,23	585,09125	Sulfo malonyl isoliquiritin	C24H26O15S	-2,5	541.10095, 459.12814, 499.09058, 417.11774, 255.06546, 297.07645
17,45	737,19275	Apigenin-di-Hexose Hydroxymethylglutaric acid	C33H38O19	-1,9	473.10818, 503.11838, 593.15015, 353.06628, 383.07626, 455.09735, 413.08640, 443.09714
17,74	674.24445 (591.17035 + 83.07428)	Acetylapiosilglucoside*	C28H31O14	-2,6	591.17035, 549.15967, 531.14990, 255.06575, 542.20258, 297.07623, 417.11758, 399.10684, 513.13837, 135.00864, 119.05024
17,86	577,15546	Violanthin/Isoviolanthin	C27H30O14	-2,163	503.11862, 457.11362, 487.12390, 559.14502, 383.07614, 353.06552, 413.08661, 469.11285, 439.10202
18,16	707,18231	Hydroxymethylglutaroyl-pentosyl-C-hexosyl-apigenin	C32H36O18	-1,5	443.09763, 383.07626, 473.10809, 563.13995, 645.18188, 353.06625, 503.11758, 455.09692, 413.08646
18,26	549,16083	Liquiritin/Isoliquiritin apioside	C26H30O13	-1,8	255.06590, 297.07632, 417.11832, 429.11859, 254.05786, 269.08121, 135.0874, 119.05040
18,46	549,16071	Liquiritin/Isoliquiritin apioside	C26H30O13	-1,865	255.0659; 429.11832; 135.00874, 119.05040 (da 429); 297.07632; 417.1183; 279.06534, 269.08121 (da 297)
18,72	417,11847	Liquiritin	C21H22O9	-1,66	255.06592, 329.10245, 373.09241, 135.00874, 119.05033
19,29	565,15564	Glycosidic flavanon*	C26H30O14	-2,4	271.06067, 433.11337, 521.16522, 227.07126, 151.00362, 177.01913
19,34	591,17047	Liquiritingenin acetylapiofuranosyl glucopyranoside	C28H32O14	-2,1	549.16003, 531.15002, 255.06592, 417.11829, 297.07565
19,46	591,17145	Liquiritingenin acetylapiofuranosyl glucopyranoside	C28H32O14	-2,1	549.16010, 531.15015, 255.06596, 399.10785, 417.11737, 297.07574
19,70	591,17145	Liquiritigenin-acetylapiofuranoside-glucoside	C28H32O14	-2,2	549.16010, 531.15015, 255.06596, 399.10699, 471.11371, 417.11737, 297.07574, 513.13861, 135.00873, 119.05013
19,72	561,16095	Glycosidic flavon*	C27H30O13	-1,7	267.06601, 252.04280, 309.07642,

					281.08112, 251.07059, 223.05267, 208.05267, 195.04489, 132.02171
20,10	459,12936	6-Acetylliquiritin	C23H24O10	-2,1	255.06604, 417.11868, 135.00864, 119.05029, 91.01903
20,21	549,16095	Liquiritin/Isoliquiritin apioside	C26H30O13	-2,302	255.0659; 429.11832; 135.00874, 119.05040 (da 429); 297.07632; 417.1183; 279.06534, 269.08121 (da 297)
20,60		Triterpenic saponin 3Hexose (Glu-Glu-Rham*)	C48H74O20	-1,3	951.45599, 907.46771, 793.43530, 549.15936, 497.11359, 419.13297, 351.05573, 193.03473, 289.05548
20,73	417,11874	Isoliquiritin/NeoLiquiritin	C21H22O9	-1,66	255.06593, 297.07651, 135.00871, 119.05037, 91.01910
21,02	669,18207	Glycosidic flavonoid*	C33H34O15	-1	531.14996, 549.15973, 255.06587, 399.10773, 297.07584, 513.13849
21,21	695,19757	Licorice glycoside B/D1/D2	C35H36O15	-0,8	549.16010, 531.15070, 255.06551, 429.10165, 399.10703, 513.13837
21,35	725,20825	Glycosidic licorice A/C1/C2	C36H38O16	-1,7	531.15002, 549.15973, 255.06587, 429.10159, 297.07587, 417.11758, 399.10721, 369.09659, 135.00876, 119.05022
21,48	999,44415	Licorice saponin G2 +Hexose	C48H72O22	-1,4	837.38831, 819.37830, 661.35748, 351.05579, 643.34698, 775.38708
21,63	823,4111	Licorice saponin J2	C42H64O16	-1,6	351.05560, 647.37793, 761.40906, 805.39862; ms2:193.03479, 289.05542, 175.02431
21,73	267,06613	Formonetin	c16h12o4	-1,3	252,04291
21,82	999,4436	Licorice saponin E2 +glucose	C48H72O22	-0,6	837.38885, 819.37842, 351.05579, 497.11331, 661.35742, 643.34650
21,82	837,39075	Licorice saponin G2	C42H62O17	-1,3	351.05591, 661.35760, 775.38873, 819.37854, 193.03493, 289.05551, 175.02451, 113.02439, 333.04507
22,00	285,07654	Licochalcone B	C16H14O5	-1,28	270.05310, 253.05080, 191.03514, 150.03255, 177.01909, 225.05536, 209.06078

22,05	459,12915	6-Acetylliquiritin	C23H24O10	-1,7	297.07632, 255.06587, 269.08130, 135.00873, 119.05031
22,09	853.38580 (M -H) 426.18912 (2M -H)	Hydroxyglabrolidic diglycosidic triterpene*	C42H62O18	-1,1	501.32153, 457.33044, 371.25790, 323.23694, 439.32016, 351.05655, 193.03491, 175.02440, 113.02439, 289.05554, 659.34338
22,14	1101,5116	Tetra glucosydic triterpenic saponin*	C53H82O24	-0,6	1083.49768, 925.47723, 351.05569, 907.46533, 967.48779, 1039.50574, 193.03491, 289.05545, 333.04544, 175.02438, 113.02428
22,31	983,44916	Licorice saponin A3/L3	C48H72O21	-1,1	821.39471, 645.36328, 351.05609, 803.38593, 627.35138, 759.39325, 785.37189, 193.03491, 175.02437, 289.05548
22,47	895,39642	22-Acetoxyl licorice saponin G2	C44H64O19	-1,3	351.05594, 877.38434, 719.36316, 284.06296, 193.03499, 175.02451, 289.05557, 289.05557, 333.04504, 307.06613, 659.34068
22,49	695,19745	Licorice glycoside B/D1/D2	C35H36O15	-1,3	531.14972, 549.15955, 255.06534, 297.07568, 399.10687, 417.11716
22,77	969,4682	Triterpenic saponin 3Hexoses*	C48H74O20	-0,8	951.45709, 793.43555, 907.46747, 497.11368, 351.05585, 775.42163, 569.38153, 193.03488, 289.05545, 333.04492, 307.06561, 175.02446, 113.02434
22,79	1025,45935	Triterpenic saponin 3Hexose (Glu-Glu-Rham)*	C50H74O22	-1,3	497.11334, 1007.44641, 435.11288, 339.09198, 321.08160
22,90	255,06636	Isoliquiritingenin/Liquiritingenin	C15H12O4	-0,8	135.00893, 119.05054
23,03	669,18207	Diglycosidic flavonoid*	C33H34O15	-0,6	531.14948, 255.06543, 417.11758, 513.13867
23,15	837,3907	Licorice saponin G2/Yunganoside K2	C42H61O17		351.0558, 661.3576, 819.3782, 289.0558; MS2: 175.0245, 193.0349, 289.0556, 775.3889
23,27	835,37506	Hydroxylicorice saponin E2	C42H60O17	-0,555	351.05582, 659.34180, 817.36292, 333.04510, 289.05548, 193.03494,

					175.02454, 113.02435, 307.06720
23,31	1011,47974	Licorice saponin D3	C50H76O21	-1,1	993.46716, 949.47827, 689.41962, 497.11353, 435.11313, 339.09241, 321.08157, 417.10156
23,44	879,40161	22-Acetoxyglycyrrhizic acid	c44h64o18	-1,2	351.05579, 861.38855, 703.36749, 643.34674, 193.03497, 175.02452, 289.05557, 333.04498, 131.03500
23,69	965,43805	Triglycosidic saponin (Glu-Glu-Rha)*	C48H70O20	-0,6	497.11365, 435.11322, 643.34723, 321.08200, 947.42603, 339.09195, 417.10251
23,79	879,40155	Acetoxyglycyrrhizin	c44h64o18	-0,9	351.05582, 703.36774, 817.39911, 861.38867, 193.03491, 175.02448, 289.05530
23,91	865,4218	Acetoxyl licorice saponin B2	C44H66O17	-1,3	351.0557, 689.3880, 803.4199, 847.9043;ms2:193.0348, 289.0555, 175.0244, 113.0242, 333.0450
24,07	837,39032	Licorice saponine G2/Yunganoside K2	C42H62O17	-1,3	351.05582, 661.35748, 819.37823, 289.05563, 333.04501, 193.03499, 175.02451, 113.02439, 485.32513, 513.29395, 643.34570
24,12	807,41693	Licorice saponin B2	C42H64O15	-0,38	351.05530, 631.38251, 789.40332, 745.41327, 193.03496, 289.05557, 175.02452, 113.02439, 631.38165
24,21	967,45386	YunganosideJ-GluA-GluA-Rhamn	C48H72O20	-0,9	949.44116, 905.45093, 645.36267, 497.11334, 435.11307, 435.11276, 339.09198, 321.08157
24,28	819,38031	Licorice saponin E2	C42H60O16	-1,4	351.05582, 643.34668, 757.37811, 801.36792, 625.33691, 289.05554, 333.04514, 193.03493, 175.02448, 113.02436, 131.03490
24,35	985,46423	Yunganoside G1	C48H74O21	-1,2	497.11371, 967.45258, 663.37329, 645.36169, 487.34064, 407.11795, 949.44147, 435.11288, 339.09201, 321.08163, 923.46185, 879.43427, 851.44043, 803.37775

24,54	837.39105 839.40637	Licorice saponin G2/Yunganoside G1	C42H64O17 (+2H)	-1,6	351.05597, 663.37286, 821.39386, 486.32837, 556.33710, 289.05563, 193.03499, 175.02451, 307.06592, 333.04501
24,58	837,3905	Licorice saponinG2/Yunganoside K2	C42H62O17	-1,7	351.05563, 661.35724, 819.37793; MS2: 289.05551, 193.03481, 175.02443, 333.04504, 661.35645, 775.38739, 485.32544
24,72	839,40637	Licorice saponin G2/Yunganoside G1	C42H64O17	-1,3	661.35834, 351.05621, 775.38953, 819.37964, 193.03502, 289.05569, 333.04532, 175.02458, 113.02444
24,88	821,3947	Glycyrrhizin	C42H62O16	-1,813	351.05609; 645.36316; 803.38440; MS2: 193.03491, 289.05554, 175.02448, 113.0243
25,11	807,41815	Licorice saponin B2	C42H64O15	-1,2	627.35205, 745.41437, 537.35699, 351.05582, 469.33075, 425.34067, 583.36182, 493.36682, 289.05566, 193.03502, 175.02457
25,14	807,41663	Licorice Saponin B2	C42H64O15	-1,219	351.05576, 789.40411; MS2: 289.05539, 193.03481, 175.02437, 745.41339, 631.38464
25,29	969,46851	Triterpenic saponin 3Hexose (GluA-GluA-Rham)*	C48H74O20	-1,6	951.45715, 497.11365, 889.45465, 435.11273, 339.09195, 321.08154, 907.46252, 933.44464, 629.36591, 471.34613
25,33	823,41083	Licorice saponin J2	C42H64O16	-1,6	351.05603, 803.38440, 759.39471, 687.37292, 645.36304, 193.03497, 289.05566, 333.04526, 175.02457
25,49	821,39587	Licorice saponin K2/H2	C42H62O16	-1,9	351.05585, 645.36267, 803.38397, 759.39313, 627.34900, 333.04514, 289.05557, 193.03494, 175.02449, 131.03491
25,71	939,49554	Glucosydic flavonoid trihexose(GLU-GLU-RHA)*	C48H76O18	-1,2	497.11365, 921.48370, 617.40430, 435.11331, 339.09244
25,76	821,39594	Licorice saponin K2/H2	C42H62O16	-1,2	351.05582, 289.05573, 645.36243,

					803.38367, 627.35144, 583.36157, 193.03490, 175.02446, 113.02435, 307.06616, 333.04496
25,85	255,0663	Isoliquiritingenin	C15H12O4	-1,1	135.00899, 119.05063
25,95	821,39612	Licorice saponin K2/H2	C42H62O16	-0,99	351.05597, 645.36292, 803.38416, 469.33066, 509.32550, 539.33563, 289.05563, 193.03497, 175.02457, 113.02444
26,12	823,41132	Licorice saponin J2	C42H64O16	-1,5	351.05585, 647.37830, 761.40906, 805.39893; ms2:193.03482, 289.05545, 175.02437, 333.04510
26,36	777,40607	Apioglycyrrhizin	C41H62O14	-0,9	627.35339, 715.40546, 537.35822, 469.33182, 583.36182, 565.35144, 425.34097, 493.36697
26,77	805,40112	Licorice saponin C2	C42H62O15	-1	787.38721, 743.39789, 629.36646, 351.05521, 453.33563, 333.044480, 289.05539, 307.06595, 193.03484, 175.02444, 113.02433
26,84	369,13409	Glyasperin B	C21H22O6	-2	229.08678, 191.03494, 139.04021, 176.01126
26,99	807,41718	Licorice saponin B2	C42H64O15	-1,7	351.05569, 631.38300, 789.40399, 745.41461, 289.05551, 193.03487, 175.02454, 113.02435, 307.06665, 333.04514
27,61	805,39935	Licorice saponin C2	C42H62O15	-3	787.38721, 743.39789, 629.36646, 351.05521, 453.33563, 333.044480, 289.05539, 307.06595, 193.03484, 175.02444, 113.02433
27,81	367,11761	Glycycoumarin	C21H20O6	-3,7	309.03946, 297.03964, 284.03201, 352.09415, 253.05016, 269.04495, 281.04477, 265.04999, 201.01909
27,84	793,4361	Diglucosydic saponin*	C42H66O14	-2,3	351.05533, 617.40314, 731.43427, 775.42377, 193.03487, 289.05545, 175.02441, 113.02430, 333.04504
28,08	807,41528	Licorice saponin B2	C42H64O15	-2,8	351.05530, 631.38251, 789.40332,

					745.41327, 193.03496, 289.05557, 175.02452, 113.02439, 631.38165
28,15	255,15396	Glyasperin C	C21H24O5	-3,4	323.12781, 337.14340, 233.11786, 254.05785, 286.08395, 207.10228, 254.05791, 279.13837, 268.07352, 237.12813, 192.07877, 137.02431, 124.01655
28,26	339,12314	Flavon*	c20h20o5	-3,6	167.03517, 229.08714, 321.11295; MS3: 211.07607, 201.09180, 183.08133, 135.04518; 149.02422, 293.11749, 249.12775
28,55	353,10223	Gancaonin O/C/L	C20H18O6	-3,3	338.11469, 321.11209, 310.12003, 295.06012, 283.06024, 270.05270, 266.05759, 323.09146
28,63	323,13821	Licoagrochalcone A	C20H20O4	-2,9	135.00873, 187.11258, 132.05803
28,67	353,13812	Licochalcone D	C21H22O5	-3,7	338.11520, 321.11255, 295.06058, 283.06073, 270.05298, 266.05759, 255.06554, 239.07079, 270.05261, 323.09149, 310.11996
28,88	321,1123	Glabrene	C20H18O4	-2,4	306.08896, 266.05792, 199.07629, 175.07645, 145.02956, 261.09109
28,91	339,12268	Cyclolicocoumarone	C20H20O5	-3,9	324.09912, 310.11566, 241.04588, 267.06052, 292.10562, 281.04449, 269.04465, 256.03702
29,04	325,1069	ω-HydroxyMoracin N	C19H18O5	-3,8	215.07085, 175.07622, 109.02956, 149.02435
29,34	353,10196	Gancaonin O/C/L	C20H18O6	-3,3	125.02442, 337.10281, 191.07117, 297.07120, 147.08144, 244.06902, 282.04813, 293.11298, 320.09991, 309.10736
29,36	335,09155	Glabrone	C20H16O5	-2,8	307.09650, 320.16793,0291.10176, 161.02426, 236.04749, 252.04247, 263.03391, 279.10175, 289.08633
29,62	323,128	Glabridin	C20H20O4	-3,04	135.04515; 201.09177; 213.09171; 147.04512; 121.02959; 175.07631.

29,88	353,10191	LicoIsoFlavone A	C20H18O6	-3,3	285.11230, 177.09167, 216.04240, 133.10219, 122.03725, 109.02953
29,96	321,1124	Licoflavone A/ Kanzonol D/B	C20H18O4	-3	306.08844, 277.12256, 227.03452, 293.11740, 291.06534, 196.05260, 262.09924, 250.06285, 234.06813
30,10	367,11755	Flavonoid*	c21h20o6	-3,9	349.10709, 321.11237, 305.11713, 331.09622, 163.03983, 177.05536, 189.05531, 227.07068
30,31	365,10208	Glycyrol	C21H18O6	-3,346	350.07867, 307.02390, 295.02432
30,37	407,18567	Flavonoid*	C25H28O5	-2,8	379.19061, 363.19614; ms2: 189.09183, 294.12534, 307.13318, 320.14075, 310.12039, 177.09196, 201.09186, 217.08679, 380.19415, 335.20093
30,56	353,10184	Gancaonin O/C/L	C20H18O6	-3,28	298.04721, 269.04483, 325.10712, 231.06589, 241.04991, 225.05498, 297.11215, 282.05240, 254.02112, 231.06555, 175.07587
30,72	425,19556	Kushenol T	C25H30O6	-2,5	235.09711, 217.08673, 177.09193, 407.18542
30,83	351,08636	Licoisoflavone B	C20H16O6	-3	283.09665, 307.09656, 265.08643, 241.08659, 239.10719, 175.03981, 255.10196, 235.03952, 221.09669, 211.03951
30,91	661,24261	Kuwanon R/Q	C40H38O9	-2,5	337.10681, 349.10657, 483.14288, 551.20355, 643.23016, 455.14801, 319.09634, 467.14853, 533.19391, 305.11719, 281.11768, 263.10782, 239.07059, 213.05508, 189.05544
31,13	379,19092	Gancaonin U	c24h28o4	-3,4	309.11258, 292.1101, 323.12817, 335.20087, 264.11517, 241.05038;ms2: 254.05801, 267.10208, 281.11774, 294.08914
31,31	369,16959	Glyasperin D	C22H26O5	-2,8	337.14331, 221.11786, 245.11763, 135.04507, 147.04501, 279.06546,

					267.06564, 322.12000, 119.05021, 151.03995, 163.03983, 206.09441, 191.07124
31,51	409,202	Kanzonol Y	C25H30O5	-3,4	235.0974; 391.1911; MS2: 205.0868; 207.1023; 189.0920; 123.0453; 217.0868; 177.0920; 32..1209; MS3: 161.0240
31,76	367,15488	Kanzonol P	C22H24O5	-1	352.13126, 337.10800, 321.07651, 309.07663, 297.07684, 293.11816
31,96	409,2017	Kanzonol Y	C25H30O5	-0,58	235.09770; 391.19144; MS2: 217.08720; 177.09239; 213.0920. MS3: 174.0325; 189.09235; 161.02480
32,13	391,19135	Glabrol	C25H28O4	-0,5	203.07114, 187.11275, 159.08157; MS3: 132.05829
32,28	405,1705	Licocoumarin A	c25h26o5	-0,38	293.0452, 336.0998, 361.1803, 349.1076 377.17529 307.06076 281.04538; ms2: 249.05544, 265.05038; 293.04514, 281.04526; 306.12555, 292.11005
32,55	643,23352	Mulberrofuran K	C40H36O8	-0,3	455.14862, 467.14825, 533.19482, 319.09662, 307.09662, 335.12781, 267.06583, 279.06567, 253.05022, 241.05031, 333.11258
32,68	393,2071	Kanzonol X	C25H30O4	-0,4	215.1075, 203.1076, 189.0920, 177.0921, 324.1362, 349.2167, 375.1959; MS2: 160.0530, 200.0842, 122.0375, 109.0297, 133.1024, 281.0815, 255.0660, 309.1127
32,68	423,18112	Gancaonin E	C25H28O6	-0,3	229.08675, 193.08684, 124.01678, 149.09734, 174.03217, 230.09027
32,90	643,23328	Mulberrofuran K	c40h36o8	-0,7	455.14926, 467.14822, 335.12836, 291.13852, 320.10458, 267.06595, 279.06583, 253.05046, 241.05040, 333.11267

33,67	421,16443	Glyurallin B		-2,8	193.08644, 149.09715, 124.01662, 137.02448, 151.07632
33,95	407,1853	Hydroxyglabrol	C25H28O5	-2,2	203.0711; 229.0864; 159.0813; 221.0813, 174.03188, 146.03732
34,25	391,19034	Hispaglabridin A	C25H28O4	-2,9	347.20004, 322.11975, 215.10698, 203.10709, 177.09164, 189.09152, 134.03729, 148.05286, 121.02957, 109.02948, 133.10226
34,65	657,21124	Glabridin + SophoraCoumestan A	C40H34O9	-2,3	333.07562, 323.12836, 305.11734, 253.12288, 213.09158, 201.09167, 175.07619, 135.04504, 147.04500, 318.05219, 303.02917, 279.02936, 121.02946, 109.02951
35,18	657,21136	Furanic flavonoid+ 2 fucoses*	C40H34O9	-2,9	455.14832, 481.12711, 639.20038, 305.04465, 293.04465, 267.02917, 279.06534, 253.04990, 241.04999, 463.11694, 451.11694
35,43	437,23206	Licorisoflavan A	C27H34O5	-2,6	405.20587, 422.20816, 229.08646, 215.10725, 177.09184, 203.10735, 347.12735, 390.18198, 335.12744, 200.08403, 160.05293, 197.09692, 171.11795, 148.05283, 134.03714, 159.11790
35,70	407,18515	Prenylated flavanon*	C25H28O5	-2,4	245.08139, 229.08669, 185.09697, 201.09181, 189.01920, 130.04231
36,63	455,35156	Ursolic acid	C30H48O3	-3,6	407.32910, 439.31964
37,77	935,57153	DGDG 18:3/18:3	C51H84O15	-2,6	675.35730, 657.34729, 379.12314, 397.13345, 277.21628, 415.14432, 565.29895
38,26	712,53668	Sphingogalactolipid C (14:2)/(C20:0)	C40H75NO9	-0,9	550.48395, 532.47333, 271.22803, 502.46289, 312.25446, 270.24341, 340.28555, 324.25430, 278.24890, 225.22247, 306.24384, 463.45331, 339.88483
38,41	983.59485 (M-H +AF)	Galactolipid C18:2 C24:4	C51H86O15	-3	659.36444, 675.36090, 657.34930,

	937.58978 (M-H)				379.12311, 397.13449
38,76	617,38293	Cumaroyl chorasolic acid	C39H54O6	-2,3	587.33527, 571.37665, 437.34177, 543.34637
39,20	959.59216 (M + AF - H), 913.58624 (M -H)	DGDG 16:0/18:3	C49H86O15	-3,5	635.36261, 657.34521, 379.12329, 397.13358, 277.21616
39,27	985.60785 (M-H +AF) 939.60083 (M-H)	DGDG 18:2/18:2	C51H88O15	-3,6	659.36188, 849.51001, 677.37286, 379.12296, 397.13342, 279.23160
40,08	961.60840 (M-H +AF) 915.60199 (M-H)	DGDG 18:2/16:0	C49H88O15	-2,3	635.36285, 659.36285, 677.37354, 379.12302, 397.13318, 279.23175, 567.23175
43,75	817,51184	SQDG C (18:2/16:0)	C43H78O12S	-2,88	561.27148, 537.27185

Table A 11 Complete table of metabolites identified in positive polarity in *Glycyrrhiza* samples, excluding the ones already identified in negative mode, with: retention time, accurate mass measured, chemical formula, Error on mass measure, name of compounds and fragments. The compounds identified at level 3 (7 compounds) are marked with an asterisk.

Retention time (min)	Accurate Mass	Compounds	Chemical Formula	Mass Error (ppm)	Fragments
16,2475	352,13678	methylacetyl benzaminodeoxy furanoside*	C17H21NO7	-5,9	308.14688, 263.12585, 273.11020, 290.13678, 237.07397, 217.08426, 189.08952, 189.08984, 199.07404
17,29633	565,15228	Apigenin dihexose	C26H28O14	-4,58	547.14099, 529.13092, 499.12061, 445.11032
18,61302	723,21082	Hydroxy-methylglutaroyl isoviolanthin	C33H38O18	-4,5	705.19983, 687.18939, 669.17914, 651.16882, 619.16370, 625.15326, 601.15308, 525.13721, 459.12686, 441.11642, 423.10599, 40509543, 325.06949
19,05676					420.18356, 540.24048, 258.13168, 240.12123, 278.61023, 144.10065, 288.14233, 180.10054
19,13604	579,16785	Violanthin/Isoviolanthin	C27H30O14	-4,4	433.11017, 271.05832, 278.61084
20,18466	269,07928	Pallidiflorin/Formonetin	C16H12O4	-5,7	254.05528, 237.05254, 213.08922, 136.01440, 118.04032, 226.06090
20,70869	257,07941	Liquiritingenin/Isoliquiritingenin	C15H12O4	-5,5	239.06874, 147.04291, 163.03772, 211.07399, 137.02225, 119.04819, 117.03256
22,64743	841,41949	Yunganoside G2	C42H64O17	-4,27	453.33487, 471.34515, 629.36664, 317.04907, 647.37714
23,12401	469,3291	Oxoglycyrrhetinic acid	c30h44o4	-4,5	451.31848, 433.30817, 405.31339, 175.14714, 415.29739, 373.28702, 334.26379, 217.15761, 147.11606, 133.10043,

					119.08485, 105.06920
23,34632	841,41766	Yunganoside G2	C42H64O17	-5,3	453.33331, 471.34354, 629.36407, 647.37421, 611.35199, 435.32315, 409.34280
24,33165	883,42902	Triterpenic saponin dihexoses*	C44H66O18	-3,7	671.37598, 611.35516, 565.34979, 495.34424, 453.33414, 435.32376, 407.32376, 407.32883, 689.38611, 707.40002, 513.35498, 471.34424, 395.29187, 189.16254, 201.16232, 389.31738
25,47415	453,33423	Oxo-ursadienoic acid	C30H44O3	-4,7	435.32355, 407.32877, 389.31815, 417.31299, 371.27094, 365.19354
25,69631	269,07928	Pallidiflorin/Formonetin	C16H12O4	-5,8	254.05598, 237.05336, 213.08986, 107.04832, 118.04053, 136.01466, 181.06377, 226.06129, 198.06650
26,23671	957,50568	Yunganoside A	C48H76O19	-5,2	811.44806, 599.39484, 423.36237, 617.40509, 581.38416, 405.35162, 793.43781
26,34797	453,33688	Oxo-ursadienoic acid	C30H44O3	-5,4	435.362629, 407.33124, 389.32062, 241.19518, 189.16377, 389.32037, 227.18007
26,71384	825,42633	Licorice saponin J2	C42H64O16	-4	613.37323, 631.38385, 567.36761, 455.35141, 437.34122, 409.34631, 353.07104, 189.16333, 391.33475
27,07975	825,4267	Licorice saponin J2	C42H64O16	-4	613.37329, 631.38361, 595.36261, 455.35147, 437.34125, 409.34595, 397.30945, 391.33511, 189.16354, 201.16347, 243.21028, 269.22650, 343.26218

27,22292	321,11148	Flavon*	C20H16O4	-5,5	303.10068, 279.10095, 265.12173, 251.11093, 223.11093, 147.040370, 247.11134, 275.10620, 288.07755
27,57281	437,33957	Triterpenic saponin*	C30H44O2	-5,7	391.33405, 189.16277, 243.20938, 299.19904, 285.18356, 215.17816, 201.16266, 335.27191, 295.24084, 267.24084
27,97042	825,42682	Licorice saponin J2	C42H64O16	-5,2	613.37329, 455.35159, 437.34122, 391.33594
28,14537	301,10513	Flavonoid/Coumarin*	C17H16O5	-6,3	289.10675, 235.13239, 221.11681, 123.04367, 165.05441, 137.05948, 153.05443, 191.10645
28,43171	647,3775	Monoglucuronylglycyrrhetinic acid	C36H54O10	-4,8	453,33478
28,73396	353,10001	LicoisoflavoneB	C20H16O6	-5,5	335.09045, 307.09576, 297.07501, 227.06961, 153.01764, 320.06693
29,24293	353,1277	LicoisoflavoneB	C20H16O6	-5,8	297.07498, 165.01796, 269.08057, 183.02855, 213.09082, 241.08563
29,60878	189,09044	Trigoforin	C12H12O2	-6,1	147.07980, 171.07974, 161.09538, 119.08507, 143.08543, 128.06181
30,21305	283,05829	Coumarin/parano Flavon*	C16H10O5	-6,2	255.06546, 165.01833, 121.02850, 229.04980
30,53122	299,05551	Isotrifoliol	C16H10O6	-6,4	271.06070, 243.06583, 225.05524, 197.06018, 165.01852, 121.02861
30,67438	389,17258	Kanzonol E	C25H24O74	-5,5	333.11206, 374.15118, 305.11719, 318.08859, 291.06500, 279.06500, 263.10663, 235.07521, 315.10120, 344.05219,

					331.08087, 197.00763
31,04027	353,13626	Gancaonin A/G/M, Glyurallin A	C21H20O5	-5,8	335.12668, 325.14240, 297.07480, 229.08504, 307.13174, 211.07471, 177.05383, 153.05389
31,29471	337,1424	Licoagrochalcone B	C21H20O4	-5,4	281.07974, 253.08597, 225.09102, 197.09608, 167.03389, 118.04123
31,70834	337,10638	Glabrone	C20H16O5	-5,5	319.09555, 263.10593, 201.05411, 245.09554, 263.10599, 291.10031, 301.08502
31,89924	647,26276	Kuwanon V	C40H38O8	-4,8	591.19989, 445.16354, 321.11130, 469.16342, 537.22571, 267.06439, 293.11618, 303.10074, 275.10580, 211.07437, 573.18890, 535.13715, 481.16324, 455.14816, 319.09567
34,25258	645,245	Mulberrofuran K	C40H36O8	-5,2	587.16986, 535.21136, 523.21136, 337.14328, 321.11185, 527.14600, 463.11633, 559.17236
34,4753	351,12241	Glycyrrhizol B	C21H18O5	-7	295.05978, 177.01819, 211.07509, 239.06989, 267.06488
35,6684	409,19888	Glyinflanin A	C25H28O5	-5	205.08583, 189.09103, 353.13870, 341.13852, 231.10167
38,38857	615,49603	Diacylglycerol*	C39H66O5	-6,6	597.48535, 523.44977, 505.38660, 541.46063, 486.52112
40,26511	1171,7033	Ganglioside A1	C60H102O20N2	-6,4	1011.63269, 879.59058, 827.40051, 1025.64771, 865.57489

Table A 12 MetAlign parameters used for *Coffea* GC-TOF data processing

Parameters	Values
Mass resolution	Nominal data Mass bin: 0.65 (-0.35-0.65)
Retention begin (scan)	70
Retention end (scan)	30000
Maximum amplitude	20000000
Peak slope factor (x noise)	2
Peak threshold factor (x noise)	3
Peak threshold (abs value)	200
Average peak width at half height (scan)	35
Max shift (begin)	25
Max shift (end)	25
Maximum shift (per 100 scans)	35
Min. factor (x noise)	7;3
Min. n° of masses	5;2

Table A 13 Parameters used in MSClust processing of *Coffea* GC-TOF data

Parameters	Values
efficient peaks	100
peak width	7
peak width confidence	80
correlation threshold	0.80
correlation threshold confidence	98
PD reduction	0.8
PD reduction softness	0.01
Stop criterion	2
MIC membership treshold	0.8

Table A 14 MetAlign parameters used for *Coffea* LC-MS data processing

Parameter	Values
Mass resolution	60000
Retention begin (scan)	20
Retention end (scan)	2000
Maximum amplitude	50000000
Peak slope factor (x noise)	3
Peak threshold factor (x noise)	4
Peak threshold (abs value)	6000
Average peak width at half height (scan)	15
Max shift (begin)	35
Max shift (end)	35
Maximum shift (per 100 scans)	35
Min. factor (x noise)	8;4
Min. n° of masses	3;2

Table A 15 Parameters used in MSClust processing of *Coffea* LC-MS data

Parameters	Values
efficient peaks	200
peak width	5
peak width confidence	80
correlation threshold	0.80
correlation threshold confidence	90
PD reduction	0.8
PD reduction softness	0.01
Stop criterion	2
MIC membership treshold	0.8

Table A 16 Complete table of metabolites identified by LC-MS in negative polarity in *Coffea sp*. samples, with: retention time, accurate mass measured, chemical formula, Error on mass measure, name of compounds and fragments. The compounds identified at level 3 (14 compounds) are marked with an asterisk

Retention time (min)	Accurate Mass	Compound	Formula	Fragments
5,22	865,19989	Procyanidin (3)	C45 H38 O18	695.14105, 575.12042, 811.84143, 287.05707, 525.87933
7,00	543,2664	Glycosidic diterpene*	C26H40O12	
8,24	353.08823 707.18433 dimer 191.05643 fragment	3-CQA	C16H18O9	Base peak 191.05656, 179.03545 (50%), 135.0455
9,00	283.03995 567.08728 203.08307	Amino acid*	C11 H12 N2 O5 S	283.04019: 222.02339, 203.08292, 238.01807, 142.06650
9,53	577,13629	Procyanidin (2) B type	C30H26012	Fragments:425.08807, 289.07187, 407.07730, 245.04546, 451.10379. MS3: from 289.07187:245.08224, 205.05092;
10,42	577.13641 1155.2799 dimer	Procyanidin (2) B type	C30H26012	425.08807, 407.07736, 289.07184, 451.10379, 559.12476
10,76	337,09308	3-pCoQA	C16 H18 O8	163.04037
11,20	289.0723 579.15240 dimer	Catechin	C15 H14 O6	245.08228, 205.05099
12,45	353.08887 707.18561 dimer 1061.28247 173.04622 fragment	4-CQA	C16H18O9	Base peak 173.04597, 179.0352 (60%), 191.05643 (30%)
12,82	349.18802 395.19357	Monoterpenic glycoside*	C16 H30 O8 C16 H30 O8 + FA	

12,82	367,10379	5-FQA	C17 H20 O9	353.08185; MS3: 191.05621
13,07	577.13629 1155.28015	Procyanidin (2) Btype	C30H26O12	425.08795, 407.07721
13,19	575.12103 1151.24854	Procyanidin (2)A type	C30 H24 O12 DIMER	
13,78	355,10367	feruloylglucoside	C16 H20 O9	193.05090 BP, 175.04027, 217.05089, 235.06151 (minor)
13,92	447.1516 401.1462	Benzylalcohol-hexose-pentose	C18H26O10 +FA C18H26O10	Fragment 269.10315 (benzyl alcohol hexose)
14,13	299.0777 599.16266	Glucopyranose-hydroxybenzoate	C13H16O8 DIMER	152.01161, 225.04117; MS3: 108.02187 from 152.01
14,54	353,0878	5-CQA	C16 H18 O9	191.05623 BASE PEAK , 179.03502 (<5%)
14,74	289.0722 579.15240 dimer	EpiCatechin	C15H14O6	
14,95	439.08905 395.09924 233.06683	Malonyl-5Caffeoylquinic acid	C19H20O12	395.09814 (-co2) 353.08847 (5-caffeoylQA), 233.06711(-co2 glucoside) MS3: from353.08: 191.05623 base peak, 179.03502; from 233.06: 173.04564

15,02	863,18396	Procyanidin (3)	C45H36O18	711.13525, 573.10394, 411.07202, 289.071. MS3 from 711: 693.12488, 559.08838, 525.88397
16,19	863.18506 911.20575	Procyanidin (3)	C45 H36 O18 C45 H38 O18 + FA	From 863.18: 711.13501, 573.10370, 289.07169, 411.07220
16,33	427.18225 381.17673	Methyl-Buthyl Hexose-pentose	C16H30O10 +FA C16H30O10	
16,97	491,1416	Methyl salicylate hexose-pentose	C19H26O12 + FA	445.1353 (-COOH2), 293.08777, 413.10870
17,42	577,1366	Procyanidin (2) Btype	C30H26O12	425.08795, 289.07196, 525.85150
18,56	559,27722	Cofaryloside/Suavioside I	C26H42O10 +FA	351.21762 Ms3: from351.22 257.1914, 269.1914, 287.2021;
19,20	1151,2489	Procyanidin (4) A type	C60 H48 O24	
19,43	575.12097 621.12640	Procyanidin (2)A type	C30 H24 O12 C30 H24 O12 + FA	289.07227, 449.08789, 423.07227
20,70	525.23492 1051.47778	CATR II	C26 H38 O11 DIMER	481.24442; MS3: 301.18109
21,69	451.1041 609.14746	Catechin caffeoylderivative + Flavonoid	C24 H20 O9 C27 H30 O16	From 609: 301.035558, 271.02518, 343.046611, 525.87329; MS3 from 301: 178.99884, 271.02518, 151.00395

21,81	342.09872 206.04628 685.20514 dimer 1028.31152 trimer	Caffeoyl-tyrosine	C18 H17 O6 N	206.04628, MS3: 163.04025, 119.05049
22,31	284,02402	DL-Indole-3-lactic acid sulphate	C11 H11 O6 N S	222.02336 (C10H9O3NS), 204.06693 (C11H10O3N), 210.02333 (C9H8O3NS); MS3 from222.02: 142.06654 (-SO3) From204.06: 186.056227 (-H2O), 158.06136 (C10H8ON), 116.05083 (C8H6N); from210.02333:131.48 (-SO3)
22,79	553.23029 507.22528	Mozambioside/Cafamarin	C26H36O10 +FA	345.1711 . MS3: 346.1749, 328.1633, 179.0564
23,27	559.27655 513.2700	Cofaryloside/Suavioside I	C26H42O10 +FA	351.21777
23,59	559.27814 513.27252	Cofaryloside/Suavioside I	C26H42O10 +FA	
23,79	391.17709 347.18677	Diterpenic compound*	C20 H25 O5 +FA C20 H25 O5	
24,19	553.23010 507.22519	Cafamarin/Mozambioside	C26H36O10 +FA	MS3: 346.1749, 328.1633, 179.0564. All the fragments fit.
25,11	559,27795	Diterpene glycoside derivative*	C27 H44 O12	
25,17	515.12097 1031.24939	4,5-diCQA	C25 H24 O12	353.08813, 335.07758 (10%), MS3: 173.04576

25,63	451,10413	Catechin caffeoylderivative	C24 H20 O9	Fragments: 341.06702, 299.05652, 189.01964; from 341.06702:217.01463, 297.07745, 177.01962; from 299.05652 189.01953, 137.02460, 255.06610
26,72	595,07727	3,5-Dicaffeoylquinic acid sulfate	C25 H24 O12	433.04468, 353.08829. MS3: 271.01312, 253.00217, 415.03247, 535.191
27,18	515,11969	3,4-diCQA	C25H24O12	353.08884. 335,07800. MS3: 173.04576 BP, 179.03516, 191.05635
27,59	529,2674	Diterpene glycoside*	C26H42O11	Fragments: 321.20740, 483.26035
28,66	541,26544	Diterpene glycoside derivative*	C27 H42 O11	
29,12	559.27612 513.27081	Suavioside I	C26 H42 O10 + FA C26 H42 O10	351.21753
29,16	515,1209	3,5-Dicaffeoylquinic acid	C25 H24 O12	353.08801 MS3: 191.05634 BP, 179.03511
29,74	529,13574	3Feruloyl-4-caffeoylquinic acid	C26 H26 O12	
29,89	499,12625	3-O-caffeoyl-5-O-p-Coumaroylquinic acid	C25H24O11	353.08780, 337.09280. MS3: from 353 : 191.05634 BP, 179.03532 (50%). From 337: 191.05647
30,42	529,1355	4,5-Feruloylcaffeoylquinic acid	C26 H26 O12	367.10355, 335.07758; MS3 from 367.10355: 193.05093, 173.04572
30,49	771,34509	CATR I	C37 H56 O17	727.35486

30,77	451.10413 903.21613	Catechin caffeoylderivative*	C24 H20 O9 DIMER	341.06668, 299.05533; MS3 from 341.06: 217.01448
31,06	529,13574	3 feruloyl-5caffeoylquinic acid	C26H26O12	367.10361, 335.22308; MS3 193.05093, 173.04572
31,20	705.33557 659.33051	Suavioside/Pulicarside	C32 H52 O14 + FA	
31,43	529.13605 1059.27942 dimer	3-caffeoyl-5-feruloylquinic acid	C26 H26 O12	353.08844, 367.10376; MS3 from353.10: 191.05635, 179.03551
31,72	541,26733	Diterpene glycoside derivative*	C27H42O11	
32,79	529,13586	4,5-Feruloylcaffeoylquinic acid	C26 H26 O12	367.10355, 335.07758; MS3 from 367.10355: 193.05093, 173.04572
33,09	659,32861	Terpenicglycoside*	C35 H48 O12	
33,30	529,13678	4-caffeoyl-5-feruloylquinic acid	C26 H26 O12	353.08810, 367.10352; MS3: 173.04587, 179.03528
33,89	326,10382	Cumaroyltyrosine	C18 H17 N O5	190.05119, 135.04536, 147.04539
34,05	365.11539 731.23846 1097.36072	Caffeoyltriptophan	C20 H17 O5 N2 DIMER TRIMER	229.06227. MS3: 186.05634, 185.07230, 100.00423
34,53	865,3891	Triterpenic glycoside*	C43 H62 O18	
35,40	543,28302	Diterpenic glycoside*	C27H44011	335.22311 (Kauranoic acid residue) , 497.27628
35,95	743,29504	Triterpene derivative*	C38 H48 O15	697.28650, 590.94499, 340.27167
36,60	543.28229 497.27625	Pharboside C	C27 H44 O11 C26H41O9 + FA	335.22308 (Kauranoic acid residue)
36,73	543,15155	Diferuloylquinicacid	C27 H28 O12	

37,86	749.34045 703.33276	Diterpenic glycoside*	C38 H54 O15 + FA C37H52O13	351.21768 BP, 703.33276
38,04	683,27258	Diterpene*	C36 H44 O13	353.08878
39,66	379,13028	Feruloyl-N tryptophan	C21 H20 N2 O5	229.06216, 186.05644, 335.13995

Table A 17 Complete table of metabolites identified by GC-TOF in negative polarity in *Coffea sp*. samples, with: retention index, rentention index-retention index expectedname, chemical formula, error on spectra matching.

Retention Index	Retention Index- Retention Index expected	Name	Chemical formula	Relative error on spectra matching %
0	1038	2-amino-Ethanol	C2NH7O	-1
1205	75	Propanedioic acid	C3H4O4	1
0	1246	2-amino-Ethanol	C2NH7O	5
1283	33	Phosphoric acid	H3PO4	22
1282	32	Glycerol	C3H8O3	7
1293	15	Proline	C5H9NO2	6
1368	33	Serine	C3H7NO3	9
1480	103	Malic acid	C4H6O5	-
0	1465	Malic acid	C4H6O5	-
0	1498	Aspartic acid	C4H8N2O3	-
0	1516	3,3-dimethoxy-Propanenitrile	C5H9NO2	-42
0	1534	2,3-Dihydroxybutanoic acid	C4H8O4	-8
1612	11	Glutamic acid	C5H9NO4	-
0	1654	L-Asparagine	C4H8N2O3	6
0	1784	Galactofuranoside	C20H48O6Si4	16
0	1792	Arabinofuranose	C17H42O5Si4	3
1830	32	Fructose	C6H12O6	-

1808	5	Citric acid	C6H8O7	-
1842	0	Quinic acid	C7H12O6	-
1856	3	Fructose	C6H12O6	-23
1865	2	Fructose	C6H12O6	-23
1881	1	Glucose	C6H12O6	-24
1900	0	Glucose	C6H14O6	-25
0	1909	Cadaverine/Putrescine	C5H14N2/C4N2H12	3
1920	9	Sorbitol	C6H14O6	-23
1920	3	Sorbitol	C6H14O6	-23
1959	5	Glucopyranose	C6H12O6	-
2077	2	Myo-Inositol	C6H12O6	12
0	2225	Diethanolamine	C4O2H11N	-31
1078	1164	Acetic acid	C2H2O3	-54
0	2358	Xylose	C5H10O5	-47
0	2611	Sucrose	C12H22O11	-5
2865	34	Catechin/Epicatechin	C15O6H14	-
2865	13	Catechin/Epicatechin	C15O6H14	-
0	2872	D-Xylose	C5H10O5	-36
0	2895	Galactopyranose	C6H12O6	-25
0	2940	D-Xylose	C5H10O5	-38

0	2950	Chlorogenic acid	C16H18O9	-
0	2955	Mannose	C6H12O6	-39
3097	30	Chlorogenic acid	C16H18O9	-
3154	33	4-caffeoylquinic acid	C16H18O9	-
3177	35	5-caffeoylquinic acid	C16H18O9	-