UNIVERSITÁ DEGLI STUDI DI CATANIA

Dipartimento di Scienze Biologiche, Geologiche e Ambientali

PhD In

Scienze Geologiche, Biologiche e Ambientali

Coordinatore: Prof.ssa A. Di Stefano

Giusi D'Amante

Study of bioactive compounds in *Silybum marianum* L. and characterization of genes involved in the biosynthesis of terpenes

PhD thesis

Tutor: Prof. S.A. Raccuia

Co-tutor: Prof.ssa M.V. Brundo

XXXII Cycle

AA.AA. 2016/2019

Preface

This work has been carried out in collaboration with:

CNR AGROFOOD Department – Institute for Mediterranean Agriculture and Forest Systems of Catania.

Bioscience Team, University of Wageningen and Research, Netherlands.

Gutta cavat lapidem non vi, sed saepe cadendo

Cadendo la goccia scava la pietra, non per la sua forza, ma per la sua costanza.

The drop hollows out the stone not by strenght, but by constant falling.

Lucrezio

Ai miei genitori ed a tutti i miei amici, che sempre hanno creduto in me.

To my parents and all my friends, who always believed in me.

Abstract

Plants have been used by man since ancient times. In a first moment, people used vegetables for their nutritive purposes but after the discovery of therapeutic properties, different human communities used plants for illness cure and health progress. Egyptian papyruses displayed that coriander and castor oil were used for curative purposes, cosmetics and preservatives (Vinatoru, 2001). During Greek and Roman epoch, a thousand of medicinal usages of natural flora was described by some scholars specifically Hippocrates, Theophrastus, Celsus, Dioscorides and many others (Paulsen, 2010). The usage of herbal plants in the ancient time shows the history of bioactive molecules, even though our ancestors didn't know about these compounds (Azmir *et al*., 2013).

Milk thistle (*Silybum marianum* L.) is an officinal herb, famous for silymarin, a pharmaceutical compound present in its fruits (achenes) (Cappelletti and Caniato, 1984; Andrzejewska *et al*., 2011). The bioactive compounds of milk thistle exhibit functional roles in plant metabolism and nutraceutical effects on human health (Lucini *et al*., 2016). Terpenes, other bioactive compounds, are also present in this plant. Terpenes are known to be important bioactive compounds in different members of the *Asteraceae* family, e.g. artichoke (MacLeod *et al*., 1982; Eljounaidi *et al*., 2014; Shakeri and Ahmadian, 2014), chicory (Cankar *et al*., 2011; Fan *et al*., 2017), sunflower (Adams and TeBeest, 2017) etc. Despite the important roles of terpenes in the Asteraceae family, the biosynthetic pathway responsible for terpenes biosynthesis in milk thistle is unknown and few articles are present in the literature concerning the content of terpenes in the tissues of this plant.

For this reason, in this research, ecophysiological and metabolomic studies were carried out to investigate the metabolomic profile of different silymarin constituents (silybin, silychristin and silydianin) in various milk thistle tissues. Metabolomic analyses were also integrated with bioinformatic and genetic methodologies to examine the terpenes metabolomic profile in different milk thistle tissues and to characterize the genes involved in the biosynthesis of these bioactive compounds.

Seed germination physiology was investigated and a correlation between afterripening and percentage of germination was detected.

Metabolomic profile of silymarin constituents and terpenes were carried out in flowers from different development stages, leaves, stems and roots of milk thistle. Silymarin constituents were detected by liquid chromatography-mass spectrometry (LC-MS) in flowers at the third development stage. The silymarin constituents revealed a different distribution in the milk thistle chemotypes analyzed, showing a high concentration in flowers at the third development stage.

Terpenes were detected by gas chromatography-quadrupole mass spectrometry (GC-qMS) of the headspace sampled by solid-phase microextraction (SPME). A different distribution of sesquiterpenes, monoterpenes and terpenes derivates was found in the different milk thistle tissues. The D-Limonene was the most detected terpenes in milk thistle tissues. β-elemene was found mainly in the roots.

The biosynthetic pathway responsible for sesquiterpenes biosynthesis in milk thistle (*Silybum marianum* L. Gaernt.) is unknown, for this reason in this study, bioinformatic, metabolomic, biochemical and molecular methodologies were integrated to investigate the candidate genes involved in the biosynthesis of these volatile compounds. Five SmTPS genes were isolated and two were found to be functional: SmTps1 was a multi-product enzyme, catalyzing the formation of amorpha 4,11-diene (small peak), β-Curcumene, β-Sesquiphellandrene and Zingiberene, while SmTPS2 was a single-product enzyme catalyzing the formation of Germacrene A.

This study provides a molecular basis for the production of volatile terpenes in milk thistle tissues. In our study, for the first time the presence of terpenes was detected not only in the seed extracts, but also in other milk thistle tissues and we also report for the first time the isolation and expression of key genes involved in sesquiterpenes biosynthesis in this species.

This study is very important both for the scientific community and for possible pharmacological applications. In fact, apart from their importance in plant physiology and ecology, volatile terpenoids are also utilized as natural flavor and aroma compounds and have beneficial effect on humans as health promoting compounds

(Wagner and Elmadfa, 2003; Nagegowda, 2010). Indeed, terpenes are also known for their anti-inflammation, anti-carcinogenesis and neuroprotection effects (Cho *et al*., 2017; Kiyama, 2017). Therefore, these results could lead to further uses of milk thistle not only for therapeutic purposes, but also as insecticide.

Further studies are needed to understand better: the expression of silymarin in the different chemotypes; the functions of terpenes in the different milk thistle tissues and the expression of the genes involved in the biosynthesis of terpenes in this species.

Riassunto

L'uomo ha utilizzato le piante fin dall'antichità. In un primo momento per scopi nutritivi e successivamente, dopo la scoperta delle loro proprietà terapeutiche, anche per la cura di malattie. Il coriandolo e l'olio di ricino, per esempio, sono stati utilizzati dagli egiziani per scopi terapeutici, come cosmetici ed anche come conservanti (Vinatoru, 2001). Durante l'epoca greca e romana l'uso terapeutico delle piante è stato descritto da *Hippocrates, Theophrastus, Celsus, Dioscorides* ed altri autori (Paulsen, 2010). L'utilizzo delle piante medicinali, fin dai tempi antichi, rispecchia la storia dei composti bioattivi, anche se la loro esistenza non era ancora nota ai nostri antenati (Azmir *et al*., 2013).

Il cardo mariano (*Silybum marianum* L.) è una pianta officinale, nota principalmente per la presenza nei suoi frutti (acheni) della silimarina, una mistura di composti con proprietà farmaceutiche (Cappelletti and Caniato, 1984; Andrzejewska *et al*., 2011). I composti bioattivi del cardo mariano mostrano importanti funzioni nel metabolismo della pianta ed anche effetti nutraceutici sulla salute dell'uomo (Lucini *et al*., 2016). Altri composti bioattivi presenti in questa pianta sono i terpeni, già individuati in differenti piante appartenenti alla famiglia delle Asteraceae, fra cui il carciofo (MacLeod *et al*., 1982; Eljounaidi *et al*., 2014; Shakeri and Ahmadian, 2014), la cicoria (Cankar *et al*., 2011; Fan *et al*., 2017), il girasole etc. (Adams and TeBeest, 2017). Nonostante l'importante ruolo dei terpeni nelle Asteraceae, pochi articoli riguardanti il contenuto terpenico nei tessuti del cardo mariano sono presenti in letteratura e ad oggi, nessun dato è noto sul *pathway* di biosintesi di tali composti nella pianta. Per tale motivo, in questa ricerca, studi ecofisiologici e metabolomici sono stati effettuati per analizzare il profilo metabolico dei diversi costituenti della silimarina (silibina, silicristina e silidianina) nei differenti tessuti di cardo mariano. Per esaminare il profilo metabolico dei terpeni nei tessuti di tale pianta ed i geni coinvolti nella loro biosintesi, analisi metaboliche sono state integrate con metodologie sia di bioinformatica che di genetica.

Per quanto concerne lo studio fisiologico sulla germinazione dei semi, è stata individuata una correlazione tra la dormienza dei semi e la percentuale di germinazione.

È stato identificato il profilo metabolico della silimarina e dei terpeni in tre differenti stadi di sviluppo delle infiorescenze, foglie, steli e radici di cardo mariano. I costituenti della silimarina sono stati individuati, mediante l'utilizzo della cromatografia liquida accoppiata alla spettrometria di massa (LC-MS), principalmente nelle infiorescenze al terzo stadio di sviluppo. È stata osservata una differente distribuzione dei costituenti della silimarina nei tre chemiotipi presi in esame, con un'elevata concentrazione identificata nelle infiorescenze al terzo stadio di sviluppo.

Il profilo metabolico dei terpeni è stato analizzato mediante un tipo di gas-cromatografia accoppiata alla spettrometria di massa specifica per i composti volatili (*gas chromatography quadrupole mass spectometry* - GCqMS - *of the headspace sampled by solid-phase micro extraction* - SPME). Una differente distribuzione dei sesquiterpeni, monoterpeni e terpeni derivati è stata osservata nei differenti tessuti di cardo mariano presi in esame. Il D-Limonene è stato il composto principalmente identificato nei tessuti del cardo mariano, mentre la β-elemene è stata trovata principalmente nelle radici.

Come detto sopra, il *pathway* di biosintesi dei sesquiterpeni nel cardo mariano non è ancora noto, per questo motivo, metodologie bioinformatiche, metabolomiche e molecolari sono state sfruttate per analizzare i geni coinvolti nella biosintesi di questi composti volatili. Cinque geni *SmTPS* sono stati isolati, due dei quali erano funzionali: SmTPS1, un enzima capace di produrre più di un composto, ha prodotto amorpha 4,11-diene, β-Curcumene, β-Sesquiphellandrene e Zingiberene; mentre SmTPS2 è stato capace di produrre un solo composto, la Germacrene A.

Questo studio fornisce delle basi molecolari sulla produzione dei terpeni nei tessuti di cardo mariano. In questa ricerca, per la prima volta, è stata individuata la presenza dei terpeni non solo in estratti ottenuti dai semi, ma anche in altri tessuti di cardo mariano e sono stati anche identificati ed espressi, per la prima volta, alcuni dei geni coinvolti nella biosintesi dei sesquiterpeni di questa importante specie.

Tale studio assume una grande rilevanza oltre che per la comunità scientifica anche per le applicazioni farmacologiche. Infatti, i terpeni hanno rilevanti ruoli non solo per la fisiologia ed ecologia della pianta, ma sono anche utilizzati come composti aromatici ed hanno effetti benefici sulla salute dell'uomo (Wagner and Elmadfa, 2003; Nagegowda, 2010), grazie alle loro proprietà anti-infiammatorie, anti-carcinogeniche e per gli effetti neuroprotettivi (Cho *et al*., 2017; Kiyama, 2017). Per cui, i risultati ottenuti potrebbero aprire nuovi orizzonti nella ricerca scientifica per eventuali ed ulteriori utilizzi del cardo mariano non solo a scopo terapeutico, ma anche come insetticida.

Ulteriori studi sono necessari per comprendere meglio: l'espressione della silimarina nei diversi chemiotipi; le funzioni dei terpeni nei differenti tessuti del cardo mariano e l'espressione dei geni coinvolti nella biosintesi dei terpeni in questa specie.

Index

1. Introduction

1.1. Bioactive compounds in plants

Bioactive compounds in plants are secondary metabolites having pharmacological or toxicological effects in human and animals. Different chemical compounds are produced by every biological system, from one cell bacterium to million cell plants for their survival.

These compounds can be divided into two wide categories: primary metabolites, which are chemical substances involved in growing and development (such as carbohydrates, amino acids, proteins and lipids) and secondary metabolites, compounds (other than primary metabolites) which help plants to rise their survival capacity and adaptation to environmental changes (Harborne, 1993). Plants can live without secondary metabolites, as shown by Kadereit *et al*. (2014) from which they have been removed by breeding. Though nutrients produce pharmacological or toxicological effects when consumed at high dosages (e.g. vitamins and minerals), these metabolites in plants are usually not considered bioactive compound. Thus, the typical bioactive compounds in vegetations are synthesized as secondary metabolites. Some of which have effect on biological systems, for this reason they are considered as bioactive (Harborne, 1993; Bernhoft, 2010; Azmir *et al*., 2013; Wink, 2016).

In different words, secondary compounds are often produced in a phase of successive to growth, have no function in growth (though they may have survival function), are synthesized by certain restricted taxonomic groups of microorganisms, have uncommon chemical structures, and are often produced as mixtures of tightly correlated members of a chemical family (Martin and Demain, 1978). The production of secondary metabolites in different species is mostly selected during evolution in correlation to the needs of that species. For example, synthesis of aroma by floral species to attract insect for their pollination and fertilization, and synthesis of toxic chemical produced against pathogens, herbivores and for suppressing the growth of nearby plants (Dudareva and Pichersky, 2000).

The usage of herbal plants in the ancient time shows the history of bioactive molecules, even though our ancestors didn’t know about these compounds. In fact, the plants have been used by man since ancient times. In a first moment, people used flora for their nutritive purposes but after the discovery of therapeutic properties, different human communities used plants for illness cure and health progress.

Egyptian papyruses displayed that coriander and castor oil were used for curative purposes, cosmetics and preservatives (Vinatoru, 2001). During Greek and Roman epoch, a thousand of medicinal usages of natural flora was described by some scholars specifically Hippocrates, Theophrastus, Celsus, Dioscorides and many others (Paulsen, 2010). Romanians are known for their use of therapeutic plants too (Azmir *et al*., 2013). *Silybum marianum* L. (milk thistle) extracts have also been used as traditional herbal remedies for almost 2000 years. These extracts are still widely used to protect the liver against toxins and to control chronic liver illnesses. Furthermore, recent experimental and clinical studies suggest that milk thistle extracts also have anticancer, antidiabetic, and cardioprotective effects (Tamayo and Diamond, 2007).

1.1.1. Classification, biosynthesis and storage of bioactive compounds

Classification of bioactive compounds in diverse categories is still inconsistent rather it is related to the different type of classification used. For example, biosynthetic classifications (based on the description of biosynthetic pathways) will not match the purpose of pharmacological classification (Azmir *et al*., 2013).

Based on their biosynthetic origins, plant bioactive compounds are divided into three main groups:

- terpenes and terpenoids (approximately 25,000 types);
- alkaloids (approximately 12,000 types);
- phenolic compounds (approximately 8,000 types).

These compounds belong to different families, each of which has peculiar structural features deriving from biosynthetic pathway (Fig. 1).

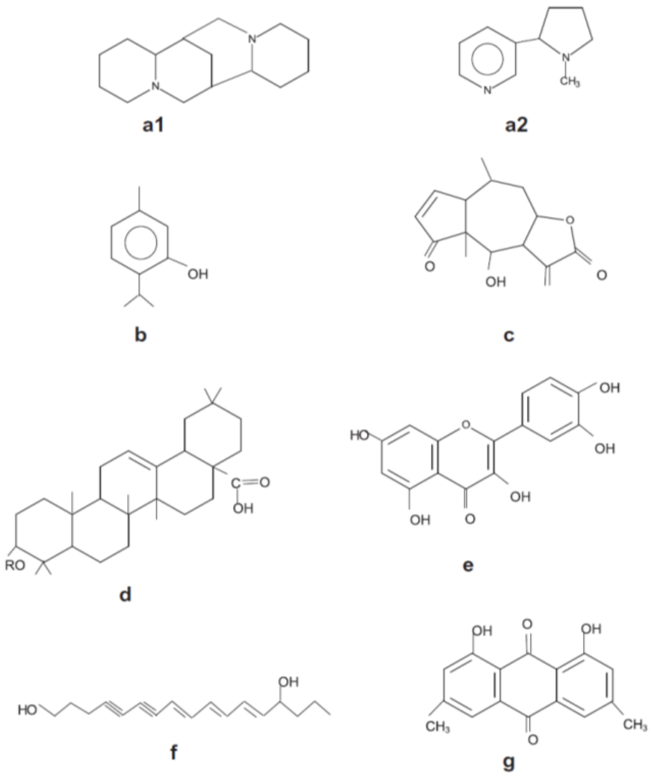


Figure 1. General structure of various categories of plant bioactive compounds: alkaloids (a1 and a2), monoterpenes (b), sesquiterpenes (c), triterpenes, saponins, steroid (d), flavonoids (e), polyacetylenes (f), polyketides (g) (Wink, 2003; Azmir *et al*., 2013).

There are four main pathways for synthesis of secondary metabolites or bioactive compounds: Shikimic acid pathway, malonic acid pathway, Mevalonic acid pathway and non-mevalonate (MEP) pathway (Fig. 2). Alkaloids are synthesized by aromatic amino acids (derived from shikimic acid pathway) and by aliphatic amino acids (originated from tricarboxylic acid cycle). **Phenolic compounds** are produced through **shikimic acid pathway** and **malonic acid pathway**. While **terpenes** are synthesized through **mevalonic acid pathway** and **MEP pathway** (Fig. 2) (Croteau *et al*., 2000; Tiaz and Zeiger, 2006; Azmir *et al*., 2013).

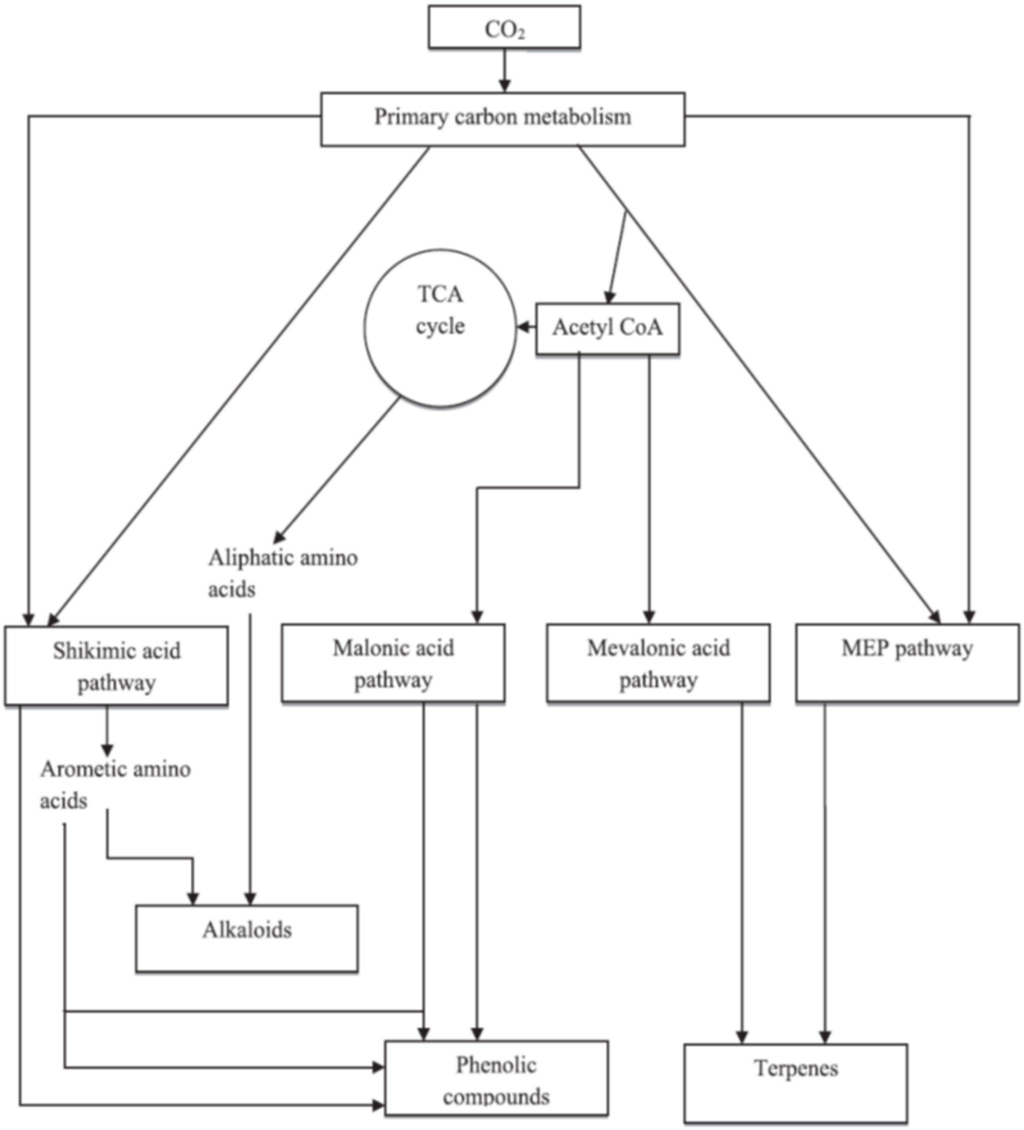


Figure 2. A simplified view of pathways for production of three major groups of plant bioactive compounds (adapted from Tiaz and Zeiger, 2006).

Furthermore, primary and secondary metabolites cannot be distinguished on the basis of precursor molecules, chemical structures, or biosynthetic origins. For example, both primary and secondary metabolites are found between the diterpenes (C_{20}) and triterpenes (C_{30}). Equally, the essential amino acid proline is classified as a primary metabolite, whereas the C_6 analog pipecolic acid is considered an alkaloid and thus a natural product. In the absence of a valid division based on either structure or biochemistry, we use the functional definition, with primary products involved in nutrition and essential metabolic processes of the plant, and natural (secondary) products influencing ecological interactions between the plant and its environment (Croteau *et al*., 2000).

Secondary metabolites are not only synthesized but they need to be stored to perform their ecological functions. Most of these compounds may also disturb the metabolism of the plant producing them, probably for this reason the site of synthesis of secondary metabolite is not always the site of its accumulation (Boller and Wiemken, 1986; Wink, 1997; Kutchan, 2005). The cytoplasm is generally the site for the biosynthesis of the secondary metabolites. However, other biosynthetic sites are present in plants cells: hydroxylations via cytochrome oxidases are performed at the smooth endoplasmic reticulum; while other biosynthetic pathways or parts of them can take place in chloroplasts (some terpenoids via the pyruvate-glyceraldehyde pathway; a few alkaloids, such as quinolizidine and steroid alkaloids), mitochondria (some amines, the alkaloid coniine), or vesicles (some isoquinoline alkaloids) (Kutchan, 2005; Zenk and Jünger, 2007; Wink, 2010b; Krauss and Nies, 2014; Wink, 2016).

The large central **vacuole** of most plant cells is not a site for biosynthesis but rather for the accumulation of defense, signal, and storage compounds (Boller and Wiemken, 1986; Wink, 1997; Kutchan, 2005; Wink, 2016). The vacuole is a storage compartment for many of the hydrophilic secondary metabolites, such as alkaloids, cyanogenic glucosides, glucosinolates, betalains, flavonoids, anthocyanins, saponins, cardiac glycosides, anthraquinone glycosides, NPAAs (Non-Proteinogenic Amino Acids), and organic acids. Instead the Lipophilic metabolites (especially terpenoids, phenylpropanoids, polyacetylenes) are accumulated in dead cells or compartments, such as resin ducts, oil cells, glandular scales, trichomes, or on the cuticle. Some Papaveraceae (poppy family), Asteraceae (aster family), Euphorbiaceae (spurge family), and Apocynaceae (dog bane family) have latificers (latex ducts) containing milky juice full of toxic alkaloids, sesquiterpenes, or diterpenes (Wink, 2008a, b, 2010b; Krauss and Nies, 2014; Wink, 2016).

Vacuoles often hold high concentrations of these compounds. To achieve the vacuole, the mostly polar and hydrophilic secondary metabolites have to cross the tonoplast (a biomembrane around the vacuole) against a concentration gradient. A few compounds can diffuse across the tonoplast and are later trapped in the vacuole as charged molecules (e.g., nicotine and some other alkaloids) (therefore called ion-trap mechanism). In most other cases, active transporters are necessary, such as energy-dependent ABC transporters or proton-dependent antiporters (Wink, 1997; Kutchan, 2005; Yazaki, 2006; Rea, 2007; Wink, 2016). ABC transporters are present in cancer cells, parasites, or drug-resistant bacteria where they pump out any lipophilic substance that has entered a cell via free diffusion (Wink *et al*., 2012, 2016). Genomic researches have revealed that some ABC transporter genes are also present in plants and that several of them apparently lead the import of polar secondary metabolites into the vacuole (Yazaki, 2006; Rea, 2007; Wink, 2016).

As mentioned above, the organ of biosynthesis of secondary metabolite is not always the organ of its accumulation. In fact, several of these metabolites, such as nicotine or tropane alkaloids, are produced in the root but are stored in all other plant tissues. For instance, lupines and other correlated legumes (Fabaceae) make quinolizidine alkaloids in photosynthetically active tissues (especially leaves) but store them in other tissues, mostly seeds (Wink, 1992, 2013, 2016). Plants always produce and store mixtures of secondary metabolites, which are generally synthesized by differing pathways. For example, an alkaloid producing plant often makes phenolics and terpenoids at the same time. The composition of such mixtures differs among organs and developmental stages: **the secondary metabolites profile usually differs between the roots, leaves, flowers, and seeds.** In addition, **secondary metabolites patterns diversify among individual plants and populations in nature**, this apparent variability is perhaps important as a strategy against adaptation of herbivores and microbial pathogens (Wink, 2008b, 2016). Moreover, some of these compounds undergo a regular turnover. Because several of these metabolites may be inactivated by a spontaneous racemization or polymerization process, their degradation and continuous new synthesis always ensure the presence of their active form.

Furthermore, plants can actively react when defied by an herbivore or microbe, by either activating preexisting bioactive compounds or by stimulating the synthesis of existing or new forms of secondary metabolites (called phytoalexins). In fact, various secondary metabolites are stored as inactive prodrugs which are activated in emergency by hydrolysis with glucosidases or esterases which come in contact with their substrate afterward tissue decompartmentation (Wink, 2003, 2008a, 2010b; Krauss and Nies, 2014, Wink, 2016). Production, transport, and storage of these compounds are energetically expensive, for the synthesis of precursors and subsequently the secondary metabolites themselves (in the form of ATP and/or NADPH), but also for the production of proteins, genes and morphological structures that are involved with bioactive compounds formation (Wink, 2010b, 2016).

1.1.2. Function of bioactive compounds

1.1.2.1. Function of bioactive compounds in plants

Plants cannot run away from enemies and they don't have a potent immune system with antibodies against infection from bacteria, fungi, and viruses. Therefore, similar to other sessile or slowly moving organisms (such as nudibranchs, jelly fish, corals, sponges) or toxic organisms (such as many amphibians), vegetables began early in their evolution to synthetize and store a variety of toxic, deterring or repellent secondary metabolites (Levin, 1976; Wink, 1988; Harborne, 1993; Wink, 1993, 2003; Hartmann, 2007; Wink, 2010a). In fact, through the use of deterrent, pungent, bitter, or toxic compounds, plants can protect themselves against most herbivores (approximately 60% of animals are herbivores), such as arthropods, worms, and vertebrates, but also against other plants, competing for light, water, and nutrients (called allelopathy). In some cases, vegetables have developed structural specialties to protect themselves from other organisms. They include stinging hairs of nettles (containing formic acid, acetylcholine, and histamine) which hurt after contact and cause an inflammation; spines, thick cell walls, leaf hairs, and inert bark. After an infection or wound, some biosynthetic pathways are activated in which jasmonic acid (made from α-linolenic acid) and salicylic acid are important signaling molecules. These compounds regulate the expression of genes of secondary metabolites or proteinase inhibitors. In addition, the volatile ethylene also induces the expression of defense genes. After the release of the elicitors (derived from cell walls of bacteria and fungi), the plant triggers its defense mechanisms and transport receptors on their surfaces that can identify these elicitors (Harborne, 1993; Kadereit *et al*., 2014; Krauss and Nies, 2014). The structures of these compounds have been modified during evolution, enabling plants to defend themselves against different species of herbivores and microbes (Wink, 1988, 2008a, b). Some secondary metabolites can interact with a multitude of molecular targets in animals and microbes. Numerous secondary metabolites attack more than a single target (multi-target agents) present in both herbivore (such as many **phenolics, terpenoids,** and some **alkaloids**) and microbes (Wink, 2007b, 2008b). On the other hand, some compounds seem to have been selected only as antimicrobials (some saponins, antimicrobial peptides) and contribute to the innate immune system of plants. Therefore, after an infection, plants also produce enzymes which degrade bacterial or fungal cell walls and stabilize their walls through incorporation of lignin (Kadereit *et al*., 2014). Although defense is the main function of secondary metabolites for plants, many vegetables sometimes use these compounds to **attract animals for pollination** (especially insects, sometimes small rodents, fruit bats, hummingbirds, sunbirds) or seed dispersal (several frugivorous birds, fruit bats, primates). The attraction is allowed by the color of some secondary metabolites in flowers (anthocyanins, chalcones, aurones, betalains, and carotenoids) and/or **volatile and aromatic terpenoids**, alcohols, aldehydes, ketones, amines, and phenylpropanoids. These compounds serve as signal compounds from a distance (pollinators should only be attracted to flowers, and not eat them) and as a deterrent when an animal would like to get nectar, oils, or pollen from the flower (Fig. 3).

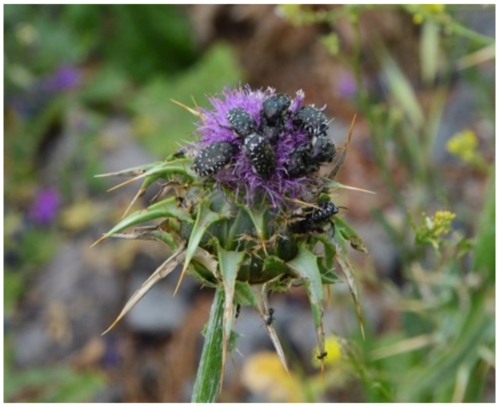


Figure 3. Milk thistle flower attacked by insects.

Fruits also store secondary metabolites and are consequently often unpalatable or even toxic when they are not mature. Upon maturation, they present attractive colors (anthocyanins, carotenoids, aurones), perfumes (many volatile mono- and sesquiterpenes), and sweet sugars (glucose) (Harborne, 1993; Wink, 2010b; Kadereit *et al*., 2014). Plants also attract frugivores, which eat the pulp, but leave the seeds intact. The secondary metabolites (especially flavonoids) also serve as signal substances in the interaction between symbiotic nitrogen-fixing rhizobial bacteria and plants (Harborne, 1993; Krauss and Nies, 2014).

Numerous bioactive compounds are stored in epidermal cells, hair cells, or the bark. This makes sense, considering the role as defense compounds, the first tissues that come into contact with intruders (Roberts and Wink, 1998; Wink 1988, 2003).

As discussed above, these compounds (being biologically active) may also interfere with the metabolism of the plants producing them. Vegetables have probably adopted some strategies to protect themselves against their own toxins, such as: their synthesis as inactive prodrugs; the presence of their affect targets only in animals (e.g. alkaloids which modulate neuroreceptors) (Roberts and Wink, 1998; Wink *et al*., 1998; Wink, 2000, 2007b) and the sequestration of secondary metabolites in cellular compartments, by which secondary metabolites can't interfere with the cell's metabolism (Wink *et al*., 2016).

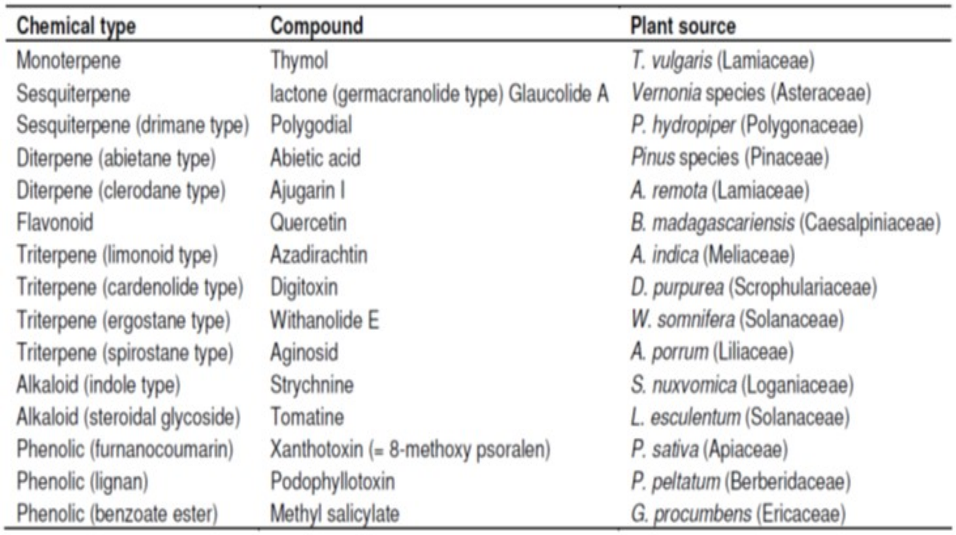
1.1.2.2. Bioactive compounds in pharmacology and phytotherapy

As discussed above, plants synthetize different types of bioactive compounds (Wink and Schimmer, 2010), whose bioactivities were discovered by humans a long time ago. Toxic alkaloids and cardiac glycosides have been employed as arrow poisons for hunting and war. Plants were also used to poison wild animals, rivals, or enemies (Mann, 1992; Wink, 1988; Wink and Van Wyk, 2008). Aromatic secondary metabolites (**mono**-and **sesquiterpenes**, phenylpropanoids) were used as perfumes, colored compounds for staining of clothes or skin. Neurotoxins (specially alkaloids and amines) were (and still are) largely selected as stimulants, intoxicants, and hallucinogens. For thousands of years humans have used plants and plant extracts to treat infections, inflammations, pains and diseases (Wink, 2000; Wink and Van Wyk, 2008). Several isolated bioactive compounds, such as morphine, colchicine, vinblastine, paclitaxel, galanthamine, huperzine, or emetine, are employed in medicine as registered drugs (Schmeller and Wink, 1998; Van Wyk and Wink, 2004). In addition, traditional medicines and phytotherapy use extracts from medicinal plants that contain complex mixtures of **phenolics**, **terpenoids**, saponins, and polysaccharides. These extracts recognize a multitude of targets and are consequently prescribed to treat more different illnesses (Wink, 2008, 2015). From a perspective of pharmacology, it is evident that secondary metabolites of plants, but also of microbes and sessile marine organisms (Proksch and Ebel, 1998), represent a source of potential active agents (Wink, 2007a, b). For this reason, it is important to preserve the diversity of plants and animals to allow future generations to find new possible drugs for medical and other applications (Wink, 2016).

1.1.2.3. Bioactive compounds in plants as deterrents against insect pests

Plants were utilized against biting insects by the ancient Greeks and are still used by vast number of people today as biopesticides. The use of plant and plant-derived products to control pests in the developing world is well known and before the discovery of synthetic pesticides, plant or plant-based products were the only pest-managing agents available to farmers around the world (Owen, 2004). Biopesticides are a group of natural and slow-acting agents that are usually safer to humans with minimal residual effects to the environment than conventional pesticides. Biopesticides can be biochemical or microbial. Biochemical pesticides comprise plant-derived pesticides that can interfere with the growth, feeding or reproduction of pests or insect pheromones used for mating disruption, monitoring or attract-and-kill strategies. The plants with this activity are very interesting as potential sources of natural insect control agents considering the increasing number of insects, showing resistance against chemical insecticides (Jacobson, 1975, 1989; Adeyemi, 2010). Secondary metabolites present in plants have an important role as defense, which inhibits reproduction and other processes in insects (Rattan, 2010). Plant-derived insecticides comprise natural insecticides, deterrents or repellents that belong to various groups of chemicals such as alkaloids, rotenoids and pyrethrins (Adeyemi, 2010). These are substances that decrease consumption (feeding) by an insect (the terms anti-feedant and feeding deterrent are used synonymously). They are behavior modifying substances that deter feeding, through a direct action on peripheral sensilla (taste organs) in insects (Isman, 2002). This definition excludes chemicals that suppress feeding by acting on the central nervous system (following ingestion and absorption) or compounds that have sub-lethal toxicity to the insect. Some of these insect anti-feedants are triterpenoids (based on a 30-carbon skeleton). Specially well studied in this regard are the limonoids from the neem (*Acalypha indica*) and chinaberry (*Melia azedarach*) trees and azadirachtin, toosendanin and limonin from Citrus species. Other anti-feedant triterpenoids comprise cardenolides, steroidal saponins and withanolides. Several types of diterpenes (based on a 20-carbon skeleton) have also antifeedants functions, including the clerodanes and the abietanes. **Sesquiterpenes** (15-carbon skeleton) with potent anti-feedant action include the drimanes, e.g. drimane polygodial from foliage of the water pepper, *Polygonum hydropiper* and the sesquiterpene lactones. **Monoterpenes** (based on a 10-carbon skeleton) which are major constituents of many plant “essential oils” deter insect feeding too. Among the plant phenolics, the furanocoumarins and the neolignans are antifeedants compounds. Furthermore, alkaloids with anti-feedant action on insects include certain indoles and the solanaceous glycol alkaloids. Specific examples of well documented anti-feedants from plants are showed in Table 1 (Adeyemi, 2010).

Table 1. Some examples of potent insect anti-feedants isolated from terrestrial plants (Adeyemi, 2010).



Chemical type	Compound	Plant source
Monoterpene	Thymol	*T. vulgaris* (Lamiaceae)
Sesquiterpene	lactone (germacranolide type) Glaucolide A	*Vernonia* species (Asteraceae)
Sesquiterpene (drimane type)	Polygodial	*P. hydropiper* (Polygonaceae)
Diterpene (abietane type)	Abietic acid	*Pinus* species (Pinaceae)
Diterpene (clerodane type)	Ajugarin I	*A. remota* (Lamiaceae)
Flavonoid	Quercetin	*B. madagascariensis* (Caesalpiniaceae)
Triterpene (limonoid type)	Azadirachtin	*A. indica* (Meliaceae)
Triterpene (cardenolide type)	Digitoxin	*D. purpurea* (Scrophulariaceae)
Triterpene (ergostane type)	Withanolide E	*W. somnifera* (Solanaceae)
Triterpene (spirostane type)	Aginosid	*A. porrum* (Liliaceae)
Alkaloid (indole type)	Strychnine	*S. nuxvomica* (Loganiaceae)
Alkaloid (steroidal glycoside)	Tomatine	*L. esculentum* (Solanaceae)
Phenolic (furnanocoumarin)	Xanthotoxin (= 8-methoxy psoralen)	*P. sativa* (Apiaceae)
Phenolic (lignan)	Podophyllotoxin	*P. peltatum* (Berberidaceae)
Phenolic (benzoate ester)	Methyl salicylate	*G. procumbens* (Ericaceae)

1.1.3. Terpenes

Terpenes consist of a group of 30,000 chemicals, with an important role as constituents of flavors, antifeedants and pheromones (Breitmaier, 2006; Kiyama, 2017). A number of modified products and derivatives, collectively called **terpenoids**, have also been characterized, and comprise steroids/sterols, saponins and meroterpenes (Cho *et al*., 2017; Kiyama, 2017).

Terpenoids are included in the **V**olatile **O**rganic **C**ompounds group (**VOC**s). In fact, a relatively wide group of plant secondary metabolites consists of volatile organic compounds, lipophilic liquids with low molecular weight and high vapor pressure at ambient temperatures. Physical properties of these compounds allow them to freely cross cellular membranes and to reach the surrounding environment (Pichersky *et al*., 2006). VOCs have been found in 90 different plant families belonging to both angio- and gymnosperms (Knudsen *et al*., 2006). Biosynthesis of VOCs depends on the availability of carbon, nitrogen and sulfur as well as energy furnished by primary metabolism, indicating the high grade of connectivity between primary and secondary metabolism. Therefore, biosynthesis of several different VOCs branches off from only some primary metabolic pathways. Based on their biosynthetic origin, all VOCs are classified into several groups, including terpenoids (as mentioned above), phenylpropanoids/benzenoids, fatty acid derivatives and amino acid derivatives in addition to a few species-/genus-specific compounds not represented in those main classes (Dudareva *et al*., 2013).

Volatile terpenoids, represented by principally isoprene, monoterpenes and sesquiterpenes, constitute the largest class of plant volatile compounds. They have important roles in direct and indirect plant defense against herbivores and pathogens, in reproduction by attraction of pollinators and seed disseminators, and in plant thermotolerance (Dudareva *et al*., 2006). Apart from their importance in plant physiology and ecology, volatile terpenoids are also utilized as natural flavor and aroma compounds and have beneficial effect on humans as health promoting compounds (Nagegowda, 2010; Wagner and Elmadfa, 2003). In fact, terpenes are also known for their anti-inflammation, anti-carcinogenesis and neuroprotection effects (Cho *et al*., 2017; Kiyama, 2017).

1.1.3.1. Biosynthetic pathway of terpenes

Terpenes have vast chemical structural diversity that is generated by several terpenoid metabolic pathways as well as the specialized cell types involved in their biosynthesis (Zulak *et al*., 2010; Cho *et al*., 2017). Terpenes are produced from the isoprene unit (C_5) and, according to the number of units, they are classified as hemi- (C_5), **mono-** (C_{10}), **sesqui**- (C_{15}), di- (C_{20}), sester- (C_{25}), tri- (C_{30}), tetra- (C_{40}), or polyterpenes ($(C_5)_n$; $n > 8$) (Fig. 4) (Cho *et al*., 2017; Kiyama, 2017).

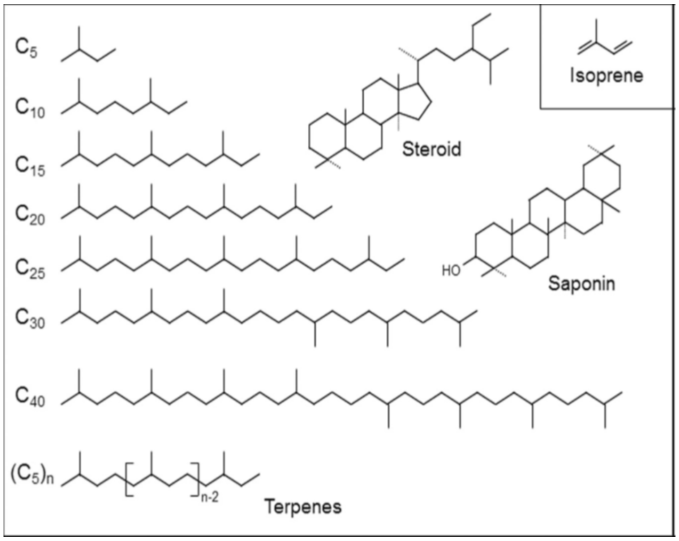


Figure 4. Terpenes, steroid and saponin (triterpene class) (Kiyama, 2017).

In plants, the two independent, compartmentally separated pathways responsible for the synthesis of these C_5-isoprene building units are: the mevalonic acid (MVA) and methylerythritol phosphate (MEP) pathway (Fig. 5). The MVA pathway produce volatile sesquiterpenes (C_{15}), while the MEP pathway provides precursors to volatile hemiterpenes (C_5), monoterpenes (C_{10}) and diterpenes (C_{20}). The **MEP pathway** is localized exclusively in plastids, in fact only in these organelles are present the corresponding enzymes, based on experimental evidence and predictions of their subcellular localization (Hsieh *et al*., 2008). By contrast, subcellular localization of the MVA pathway is not as clear. Historically, this pathway was referred to as being cytosolic; in any case, new evidence suggests that the MVA pathway is localized between the cytosol, endoplasmic reticulum and peroxisomes (Simkin *et al*., 2011; Pulido *et al*., 2012). The **MVA pathway** consists of six enzymatic reactions and is initiated by a condensation of three molecules of acetyl-CoA to 3-hydroxy-3-methylglutaryl-CoA, which undergoes reduction to MVA followed by two subsequent phosphorylations and a decarboxylation/elimination step with synthesis of IPP as the final product (Fig. 5) (Lange *et al*., 2000). Up to now, it is still uncertain which subcellular pool of acetyl-CoA is employed for terpenoid biosynthesis, as acetyl-CoA cannot readily cross membranes, and pools is localized in chloroplasts, peroxisomes, mitochondria, cytosol and nucleus (Oliver *et al*., 2009). The **MEP pathway** involves seven enzymatic steps and starts with the condensation of D-glyceraldehyde 3-phosphate (GAP) and pyruvate (Pyr) to produce 1-deoxy-D-xylulose 5-phosphate, which is then subjected to isomerization/reduction with synthesis of the pathway's characteristic intermediate, MEP (Fig. 5).

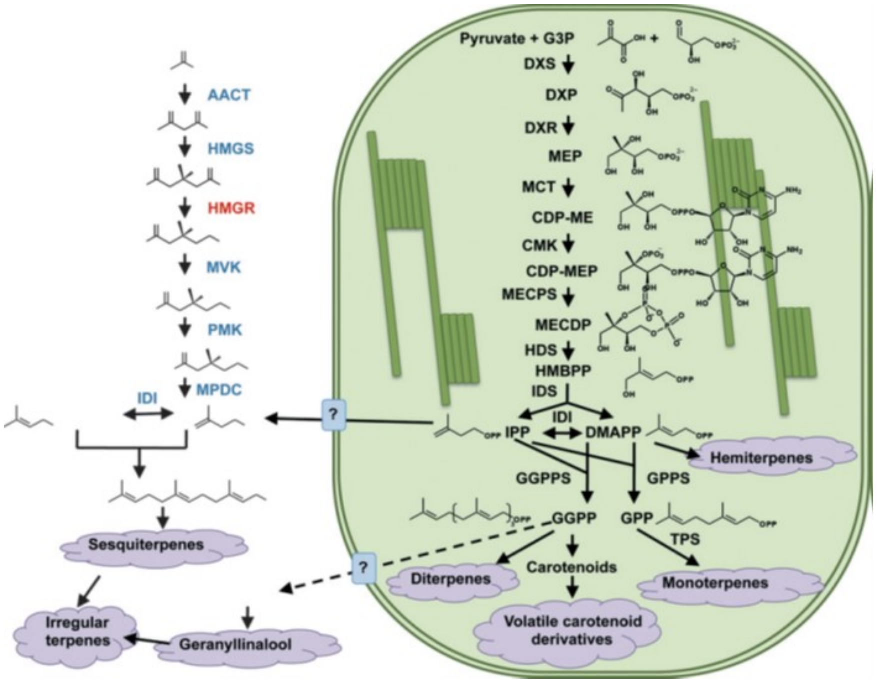


Figure 5. Biosynthetic pathways and their compartmentalization leading to volatile terpenoids in plants. AACT, acetoacetyl-CoA thiolase; AcAc-CoA, acetoacetyl-CoA; CDP-ME,4-(cytidine 50-diphospho)-2-C-methyl-D-erythritol; CDP-ME2P, 4-(cytidine 50-diphospho)-2-C-methyl-D-erythritol phosphate; CMK, CDP-ME kinase; DMAPP, dimethylallyl diphosphate; DOXP, 1-deoxy-D-xylulose 5-phosphate; DXR, DOXP reductoisomerase; DXS, DOXP synthase; FDS, farnesyl diphosphate synthase; FPP, farnesyl diphosphate; GA-3P, glyceraldehyde-3-phosphate; GDS, geranyl diphosphate synthase; GGDS, geranyl geranyl diphosphate synthase; GGPP, geranyl geranyl diphosphate; GPP, geranyl diphosphate; HDR, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HDS, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HMBPP, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; IDI, isopentenyl diphosphate isomerase; IPP, isopentenyl diphosphate; ISPS, isoprene synthase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; ME-2,4cPP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; MVD, mevalonate diphosphate decarboxylase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; TPS, terpene synthase (Dudareva *et al*., 2013).

After five consecutive steps MEP is converted to IPP and DMAPP. The MEP pathway relies on primary metabolism for the provision of Pyr and GAP, with the second derived from both glycolysis and the pentose phosphate pathway (PPP). To date, the origin of Pyr in the chloroplasts is not fully understood, as plastids have low activities of the main glycolytic enzymes, phosphoglycerate mutase and enolase (Andriotis *et al*., 2010; Joyard *et al*., 2010; Bayer *et al*., 2011), and might not be able to support high Pyr request for isoprenoid biosynthesis (Furumoto *et al*., 2011). Both **IPP** and **DMAPP** are substrates for short-chain prenyltransferases, which produce prenyl diphosphate precursors, geranyl diphosphate (**GPP**), farnesyl diphosphate (**FPP**) and geranylgeranyl diphosphate (GGPP), for a wide family of **terpene synthases/cyclases** (**TPSs**) (Fig. 5) (Cane, 1999; Wise and Croteau, 1999). While the MVA pathway synthesis only IPP, the MEP pathway produces both IPP and DMAPP at a 6:1 ratio (Rohdich *et al*., 2003). Thus, both pathways depend on isopentenyl diphosphate isomerase (IDI), which reversibly converts IPP to DMAPP (Nakamura *et al*., 2001) and checks the balance between them. IPP, DMAPP, and short prenyl diphosphates (GPP and FPP) simplify the metabolic crosstalk between the compartmentally divided MVA and MEP pathways by acting as linking metabolites (Nabeta *et al*., 1997; Adam *et al*., 1999; Hemmerlin *et al*., 2003; Wu *et al*., 2006; Orlova *et al*., 2009). Transporting of these compounds across the inner envelope membrane of plastids is allowed by an unidentified metabolite transporter (Bick and Lange, 2003; Flϋgge and Gao, 2005). Such connectivity of the isoprenoid biosynthetic pathways allows the MEP pathway, often with a higher carbon flux than the MVA way, to sustain biosynthesis of terpenoids in the cytosol (Laule *et al*., 2003; Dudareva *et al*., 2005; Ward *et al*., 2011). The contribution of these pathways to terpenoid biosynthesis is species- and/or organ-specific: for example, in snapdragon flowers, the MEP pathway synthetized precursors for cytosolic sesquiterpene production (Dudareva *et al*., 2005), in carrot (leaves and roots) sesquiterpenes are resulted from both MEP and MVA pathways (Hampel *et al*., 2005). In spite of the presence of IPP and DMAPP in numerous compartments, biosynthesis of prenyl diphosphate intermediates is compartment-specific and depends on subcellular localization of corresponding short-chain prenyltransferases. In the cytosol, the sequential head-to-tail condensation of two IPP molecules with one molecule of DMAPP catalyzed by FPP synthase produces **FPP**, the precursor of volatile **sesquiterpenes**. In plastids GPP and GGPP synthases are liable for the head-to-tail condensation of one DMAPP molecule with one or three IPP molecules to give rise **GPP** and GGPP, respectively, the corresponding precursors of **mono** and diterpenes. The great diversity of volatile terpenoids in plants is due to the action of **terpene synthases (TPSs)** (Fig. 5), **many of which have the typical ability to synthesize multiple products from a single prenyl diphosphate substrate** (Degenhardt *et al*., 2009). Indeed, in Arabidopsis two sesquiterpene synthases (TPS21 and TPS11) are involved in the biosynthesis of almost all 20 sesquiterpenes found in the floral volatile mixture (Tholl *et al*., 2005). Moreover, numerous TPSs accept more than one substrate (Tholl, 2006; Bleeker *et al*., 2011), which increases the diversity of produced terpenoids by directing bifunctional enzymes to diverse compartments with a varying range of available substrates (Aharoni *et al*., 2004; Nagegowda *et al*., 2008; Huang *et al*., 2012; Gutensohn *et al*., 2013). However, it is still unknown if this is a general capacity of TPSs. To date, the **TPS gene family** consists of more than 100 members characterized from various plant species, with about one-third identified in flowers and fruits. This gene family has been classified into **seven subfamilies** (designated TPS-a through TPS-g) based on sequence affinity, functional assessment, and gene architecture (Bohlmann *et al*., 1998; Aubourg *et al*., 2002). The TPS-a clade was further divided into a dicot-specific subclade and a monocot-specific subclade, this clade is composed of angiosperm-specific sesquiterpene synthases. TPS-b clade, containing the RRX8W motif in the N-terminal region for monoterpene cyclization, which is normally found in angiosperm-specific monoterpene synthases (Hyatt *et al*., 2007; Chen *et al*., 2011). Regarding the TPS-g clade, previous studies have showed that a prominent feature is the prevalence of acyclic products, because of the lack of an RRX8W motif in this clade (Dudareva *et al*., 2003).

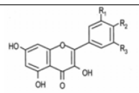
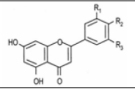
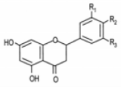
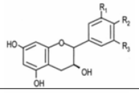
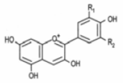
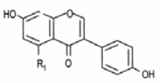
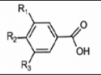
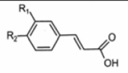
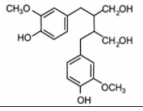
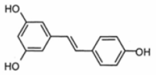
TPS proteins of the TPS-g subfamily identified from grapevine were shown to synthesize acyclic monoterpenes, sesquiterpenes, and diterpenes specifically (Martin *et al*., 2010). Furthermore, terpenoid diversity is further increased by other enzymes that are able to modify the TPS products via hydroxylation, dehydrogenation, acylation, or other reactions, thus enhancing their volatility and changing their olfactory properties (Dudareva *et al*., 2004). Floras also produce irregular volatile terpenoids with carbon skeletons ranging from C_8 to C_{18}, which are synthetized from carotenoids via three step modifications, including an initial dioxygenase cleavage followed by enzymatic transformation and acid-catalyzed conversion to volatile compounds (Fig. 5) (Winterhalter and Rouseff, 2001; Dudareva *et al*., 2013).

1.1.4. Phenols

Phenols are very important in plant physiology for their role in pigmentation, flavour, growth, reproduction and resistance to pathogens and predators. Several thousand phenols with at least one aromatic ring (phenolic ring) bearing hydroxyl groups have been identified. The phenols can be divided in the following subgroups based on the structural components (Tab. 2): phenolic acids, flavonoids, stilbenes and lignans.

Most phenols have antioxidant activity and are the most abundant antioxidants in our diet. The main dietary sources of phenols are fruits and beverages (fruit juice, wine, tea, coffee and chocolate) and, to a lesser extent vegetables, dry legumes and cereals. Certain phenols are found in different plants, while others are specific to particular plants. In most cases, foods comprehend complex mixtures of phenols. Phenol content of plants is related to environmental factors such as soil type, sun exposure, temperature and rain fall. The degree of ripeness also affects the concentration and proportion of phenol, where phenolic acid concentration reduces during ripening, whereas anthocyanin concentrations increase. **Flavonoids** can be divided into 6 major subclasses: flavones, flavonols, flavanones, flavanols (cathechins and proanthocyanidins), anthocyanins, and isoflavones. Flavonoids provide flavour and colour to fruits and vegetables (Astley *et al*., 2002; Manach *et al*., 2004, 2005; Williamson *et al*., 2005; Blomhoff, 2010). Flavonols are the most extensively represented flavonoids in food. Fruits often contain 5-10 different flavonol glycosides (glycosylated form of flavonols). These compounds are accumulated in the skin and leaves of the fruit because their biosynthesis is stimulated by light. Thus, the flavonols concentration of fruits and vegetables depends on the exposure to light. **Phenolic acids** can be classified into two subclasses: hydroxybenzoic acids and hydroxycinnamic acids. Hydroxybenzoic acids are identified only in a few plants eaten by humans such as certain berries and onions, whereas hydroxycinnamic acids are more common and are found in flour, coffee, aubergine, blueberry, kiwi and other fruits and vegetables (Bernhoft *et al*., 2010).

Table 2. Phenols (Bernhoft *et al*., 2010).

Phenolic class **Flavonoids**	**Chemical structure**		**Dietary sources**
Flavonols		Quercetin Kaempherol	Cherry tomato, onion, broccoli, tea, red wine, berries
Flavones		Apigenin Luteolin	Cereals, parsley, celery
Flavanones		Hesperetin Naringenin	Citrus fruits
Flavanols		Catechins	Chocolate, beans, apricot, tea, red wine, cherry, apple
Anthocyanidins		Cyanidin	Aubergine, berries, red wine, red cabbage
Isoflavones		Daidzein Genistein	Soy products, peas
Phenolic acid			
Hydroxybenzoic acid		Gallic acid	Berries Onion
Hydroxycinnamic acid		Caffeic acid Ferulic acid	Blueberry, kiwi, cherry, plum, apple, grains
Lignans		Matairesinol	Linseed, lentils, cereals, garlic
Stilbenes		Resveratrol	Wine, grapes, blueberries

1.2. Milk thistle (*Silybum marianum* L.)

1.2.1. Agronomical characteristics of milk thistle

Silybum marianum L. Gaertn, also known as milk thistle (Fig. 6), is a member of the Asteraceae family. It is a therapeutic herb with a 2000-year history of use (Abenavoli *et al*., 2018).

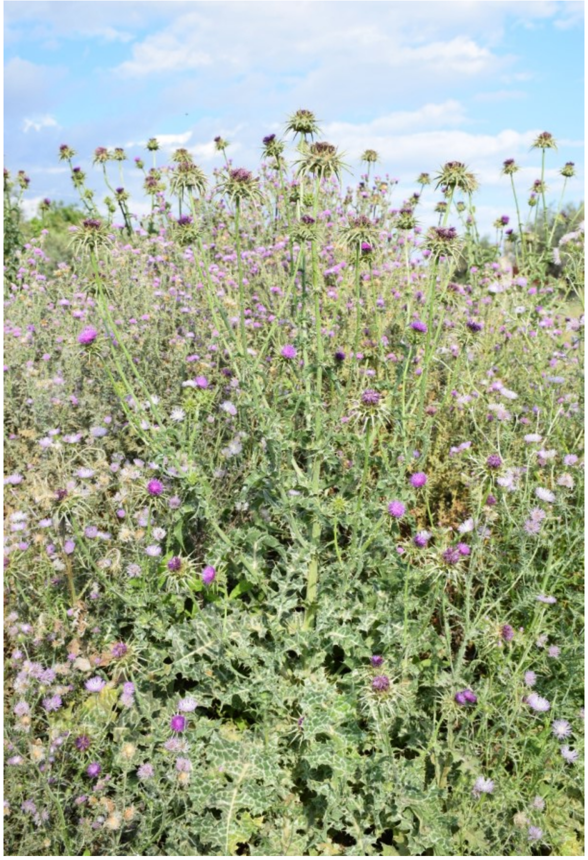


Figure 6. Milk thistle plant.

'Silybum' is the name Dioscorides gave to edible thistles and 'marianum' derives from the legend that the white veins running through the plant's leaves (Fig. 7) were caused by a drop of the Virgin Mary's milk. While looking for a place to nurse the infant Jesus when leaving Egypt, Mary could only find a shelter in an arbor formed from the thorny leaves of the milk thistle (Morazzoni and Bombardelli, 1995). According to this story the folk belief that this herb was good for nursing mothers was born. Other names that have been attributed to this plant include Marian thistle, Mary thistle, St Mary's thistle, Our Lady's thistle, Holy thistle, sow thistle, Blessed Virgin thistle, Christ's crown, Venue thistle, heal thistle, variegated thistle and wild artichoke (Abenavoli *et al*., 2010).

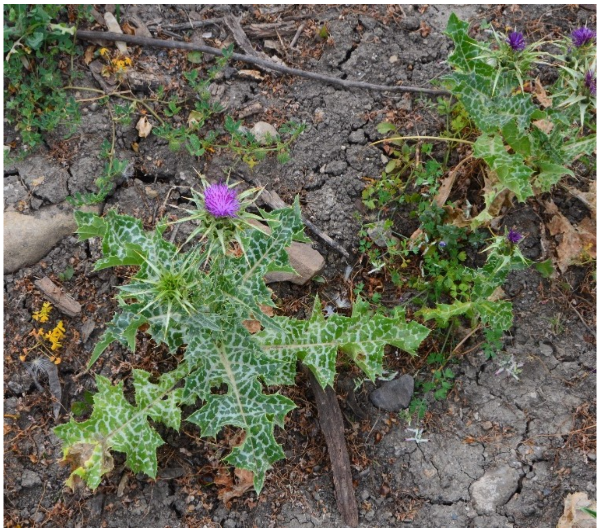


Figure 7. Flowers and leaves with white veins of milk thistle.

Milk thistle is native to the Mediterranean area, and it is naturalized in Central Europe, North and South America, Australia, Asia, and New Zealand (Morazzoni and Bombardelli, 1995; Martin *et al*., 2000) (Fig. 8). Milk thistle has no vernalization requirement and can be usually classified as an annual species even though it can be biennial (Young *et al*., 1978; Groves and Kaye, 1989; Martinelli *et al*., 2016). This plant has an erect stem (that can reach a height of 150-200 cm) and white taproot. Leaves are big, green-coloured, and white-veined, with strong spiny edges. The inflorescences are solitary large purple heads situated at the apex of the stem, or at the primary and secondary branches. The fruits are black, hard-coat achenes, with an elaiosome, an outgrowth on fruit that is due to the large number of oil-storing cells and is attractive to ants and thus aids seeds dispersal. The duration of the biological cycle as well as the main phenological parameters (main stem initiation, onset of flowering, and achene ripening) were strictly influenced by the climate: flowering occurred in Sicily between mid-April and mid-May. Though, the number of days between flowering and achene ripening was similar for all the environments (between 40 and 46 days) (Gresta *et al*., 2006).

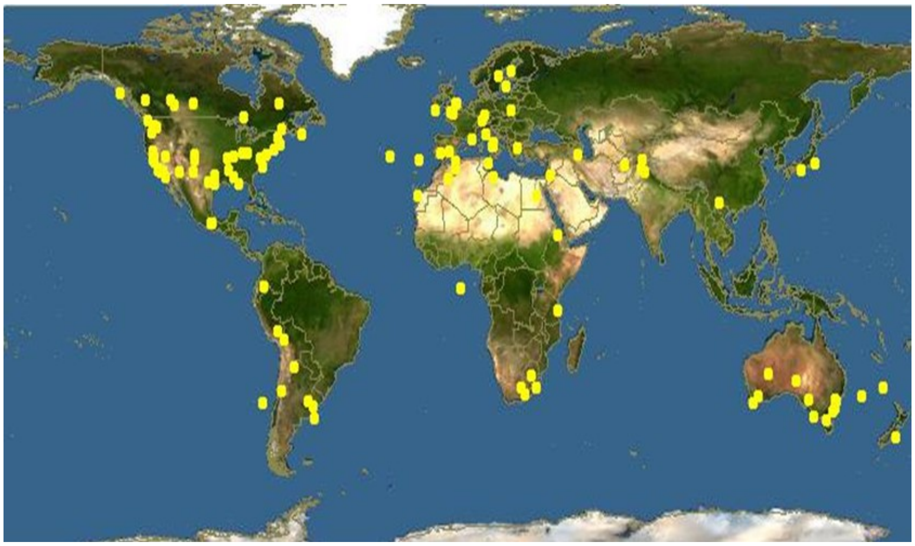


Figure 8. Distribution of milk thistle in the world (https://www.discoverlife.org/mp/20m?kind=Silybum+marianum&guide=Wildflowers&cl=US/CA/Monterey/Hastings_Reserve).

Although many studies have been dedicated to the chemical composition of milk thistle, very little is known about its agronomic characteristics. This lack of information has prevented a wide diffusion of the species, such that in Italy it is cultivated in very restricted areas with a total seed yield of approximately 120 quintals (Vender, 2001; Gresta *et al*., 2006; Raccuia and De Mastro, 2019).

1.2.2. Uses of milk thistle

Milk thistle has been used for centuries in medicine, principally to treat kidney, spleen, liver, and gallbladder illnesses (Schadewaldt, 1969; Flora *et al*., 1998). The Roman naturalist and natural philosopher Pliny the Elder (23–79 AD) wrote that mixing the juice of this plant with honey was indicated to "carry off the bile". The Greek physician, pharmacologist, and botanist Dioscorides (i.e., the author of De Materia Medica) suggested it as tea against serpent bites. The medical use of this plant was reported in the Middle Age in the Saxons records to ward off snakes and to treat the infectious disease contracted after being bitten by a rabid animal. St. Hildegard vòon Bingen

(1098-1179) recommended the herb root and leaves to treat swelling and erysipelas. Subsequently, in the 16th century, two famous English herbalists, John Gerard and Nicholas Culpeper, suggested the use of milk thistle against all melancholy diseases and to cure fever, respectively. This plant was also popular in the German medical tradition and several scientists, including Johannes Gottfried Rademacher (1772–1850), recommended it to treat liver diseases. In the USA, milk thistle is know because of naturopathic medical tradition of the Native Americans as well as of the Eclectic movement, a group of practitioners that suggested milk thistle for varicose veins, menstrual problems, and congestion of the spleen, kidney, and liver in the first half of 19th century. Actually, milk thistle is among the top-selling herbal dietary supplements in the USA (Andrew and Izzo, 2017; Abenavoli *et al*., 2018). Nowadays, this plant is also used as anti-diabetic, hepatoprotective, hypocholesterolaemic, anti-hypertensive, anti-inflammatory, anti-cancer, and as an anti-oxidant. The seeds of milk thistle are also used as an anti-spasmodic, neuroprotective, anti-viral, immunomodulant, cardioprotective, demulcent and anti-haemorrhagic. The plant is also used as a galactagogue and in the treatment of uterine disorders (Kumar *et al*., 2011).

1.3. Principal compounds in milk thistle

1.3.1. Silymarin

The pharmaceutical compound of milk thistle is derived from its fruits (achenes), which contain silymarin in their dry pericarp and seed coat (Cappelletti and Caniato, 1984; Andrzejewska *et al*., 2011). Silymarin is a milk thistle seeds dry extract containing mainly flavonolignans (about 70%-80% w/w) as well as polymeric and oxidized polyphenolic compounds consisting of a mixture of flavonoids. **Flavonolignans**, which have been first found in the seeds of milk thistle, are a relatively small subclass of compounds, where the **flavonoid part** of the molecule is fused with a **lignan**. The principal silymarin flavonolignas are silybin (A and B), isosilybin (A and B), silydianin, and silychristin (Kvasnicka *et al*., 2003) (Fig. 9).

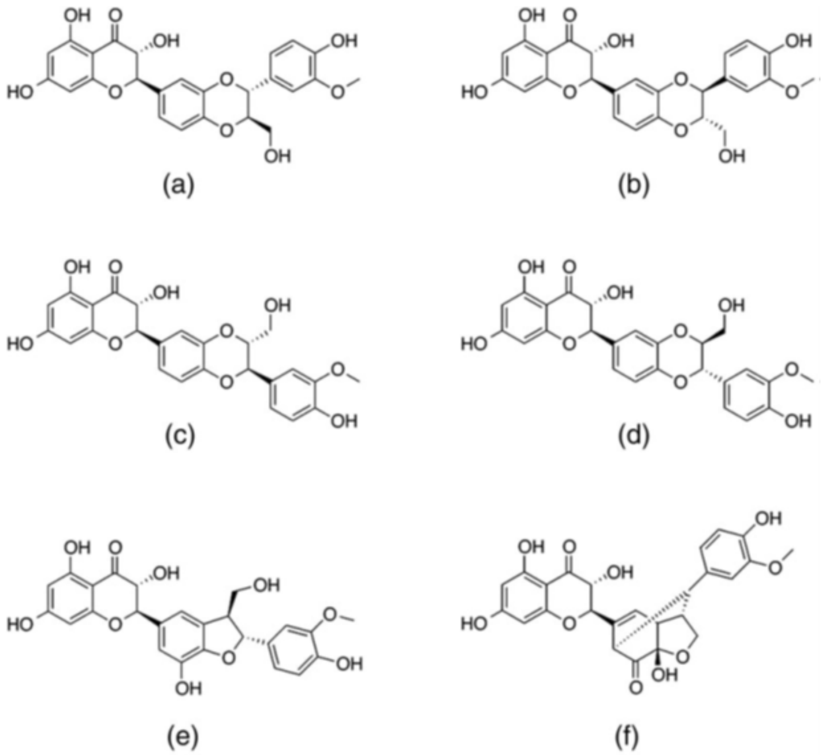


Figure 9. Chemical structures of principal flavolignans contained in silymarin: (a) silybin A, (b) silybin B, (c) isosilybin A, (d) isosilybin B, (e) silychristin, and (f) silydianin (Abenavoli *et al*., 2018).

While many researches describe analytical separation and quantification of silymarin components in the extract in various plant parts, seasons, geographic locations, etc. (Poppe and Petersen, 2016), no comparison of detail flavonolignan profiles in various silymarin preparations is available to date (Chambers *et al*., 2017). Silybin is, from a quantitative point of view, the principal component of silymarin. It has a molecular formula of $C_{25}H_{22}O_{10}$ and a molecular weight of 482.441 (Bijak, 2017). In nature, silybin occurs in the form of two diastereoisomers, namely, silybin A and silybin B, which are present in a roughly quasi-equimolar ratio. Silybin has a low solubility in water and in polar solvents, and it is insoluble in apolar solvents (Biedermann *et al*., 2014). Silybin is a small, extremely functionalized molecule with alternating carbocycles and heterocycles and, due to its structure, it is quite resistant to reduction but oxidizes easily to 2,3-dehydrosilybin. It is stable under acidic conditions and becomes unstable in the presence of Lewis acids or in basic conditions since strong bases or heating can disrupt its structure. In neutral aqueous solutions, silybin behaves as a weak acid (Gu *et al*., 2000). The biosynthesis pathway of silybin is not fully understood. Biomimetic reactions indicated that silybin can be synthesized from coniferyl alcohol and taxifolin by the action of peroxidase (Fig. 10). Five candidate genes for the peroxidase are involved in silybin production, among which Ascorbate Peroxidase 1 showed a good activity as well as the ability to synthesize silybin (Lv *et al.*, 2017).

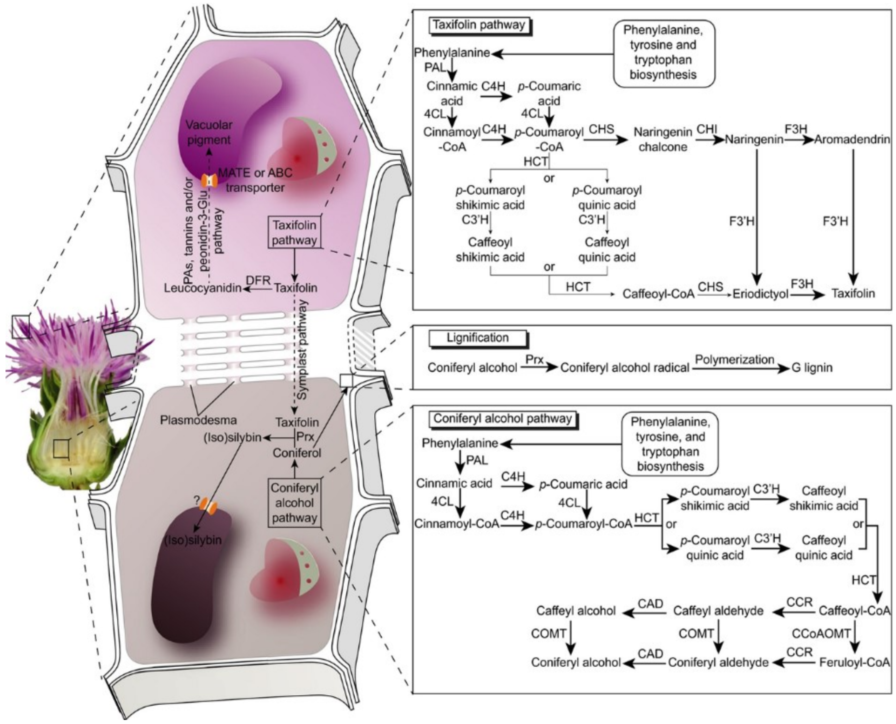


Figure 10. Proposed model for silybin biosynthesis in milk thistle. Silybin is synthesized from coniferyl alcohol and taxifolin by ascorbate peroxidase in the seed coat cell. The accumulation of taxifolin in the seed coat is mainly transported from the flower. The expression profile indicates that the flower principally utilizes the p-coumaroyl-CoA to naringenin pathway for taxifolin synthesis, while the p-coumaroyl-CoA to caffeoyl-CoA pathway is mainly used for coniferyl alcohol synthesis (Lv *et al*., 2017).

Most of the studies have been dedicated to flavonolignans giving less attention to minor components. This has led to problems in determining the exact **composition of silymarin**, **which can vary depending on the processing, variety of the plant, soil composition, and climatic conditions during the plant growth**. Silymarin contains also a mixture of undefined polyphenolic compounds, often mentioned as “polymeric fraction” (Bijak, 2017). Nonetheless, the oil fraction, which contains linoleic, oleic, and palmitic acids, sterols, tocopherol (vitamin E), and phospholipids, has been not fully investigated (Chambers *et al*., 2017; Abenavoli *et al*., 2018). In fact, mature milk thistle seeds also contain, on a dry weight basis, more than 25% lipids of which 42% and 37% are linolenic and oleic acids, respectively (Hamid *et al*., 1983; Carrier *et al*., 2002).

Furthermore, silybin has a wide range of pharmacologic properties such as inhibition of cell proliferation, cell cycle progression, and induction of apoptosis in various cell lines including fibroblasts and breast cancer cells (Ebrahimnezhad *et al*., 2013; Mahmdi *et al*., 2016). This compound has also been described as useful for intervention of hormone refractory human prostate cancer (Zi *et al*., 1999). The mixture of silybin and silycristin has been demonstrated to reduce the nephrotoxic defects of chemical induced injury (Sonnenbichler *et al*., 1999; Carrier *et al*. 2002). Silymarin effects have also been indicated in various diseases of different organs such as prostate, lungs, CNS (Central Nervous System), kidneys, pancreas, and skin (Gazak *et al*., 2007; Kumar *et al*., 2011). It has been also discovered that silymarin may be helpful in slowing down the progression of neurodegeneration in focal cerebral ischemia (El Sherif *et al*., 2013). It has been also demonstrated that silymarin treatment was associated with a reduction of insulin resistance and a significant decrease in fasting insulin levels, suggesting an improvement of the activity of endogenous and exogenous insulin (Cacciapuoti *et al*., 2013; Mahmdi *et al*., 2016).

1.3.2. Terpenes

Various previous studies have demonstrated the biological activities of the milk thistle extracts. However, there are not works, which studied the terpenes composition of milk thistle tissues, but only researches concerning the identification of some of these compounds in seeds, essential oils and whole plant.

Ahmed *et al*. have isolated in 2006 marianine, a new lanostane-type triterpene (1) and marianosides A (2) and B (3), two new triterpenoidal glucosides, respectively (Fig. 11).

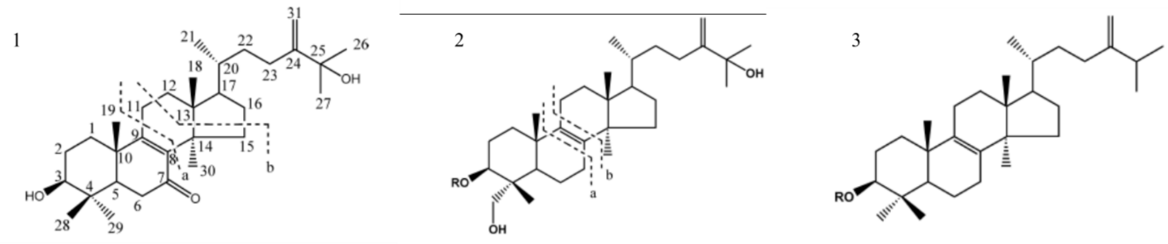


Figure 11. Chemical structure of marianine (**1**) and marianosides A (**2**) and B (**3**) (Ahmed *et al*., 2006).

The same scientists in 2007 have also identified two new pentacyclic triterpenes named as silymins A (1) and B (2), respectively (Fig 12) (Ahmed *et al*., 2007).

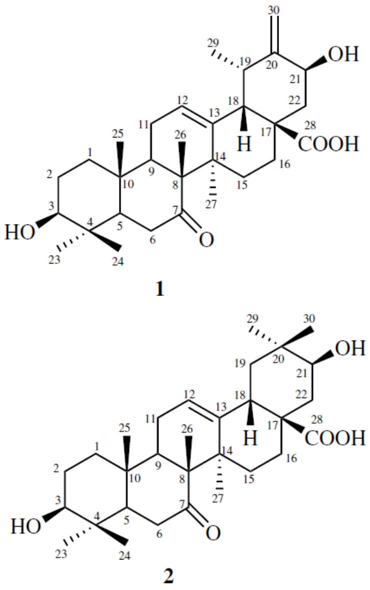


Figure 12. Chemical structure of silymins A (1) and B (2) (Ahmed *et al*., 2007).

Mhamdi *et al*. (2016) have discovered that the essential oils from milk thistle seeds are particularly rich on mono and sesquiterpene hydrocarbons, which could have specific usages in pharmaceutical industry. The main compounds identified by Mhamdi *et al*. (2016) were: γ-cadinene (49.8%), α-Sesquiterpene hydrocarbons (55.4%), monoterpene hydrocarbons (33.29%) and oxygenated monoterpenes presented in minor proportions (5.41%). Ning-Bo *et al*. (2017) have isolated from the seeds of milk thistle four terpenoids including: one new sesquiterpenoid (1), one new monoterpenoid (2) and its glycoside (3), along with a known monoterpenoid (4) (Fig. 13).

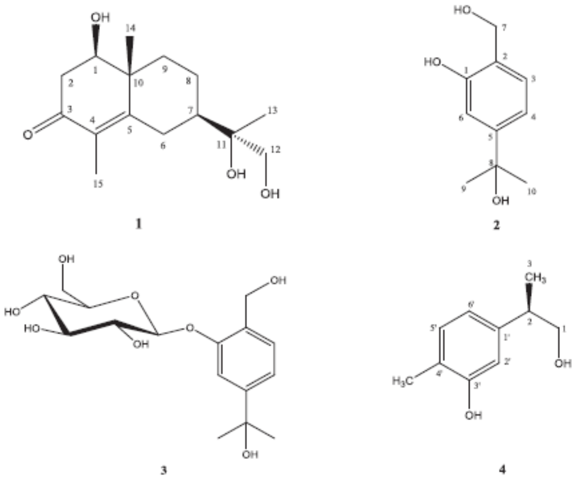


Figure 13. Chemical structure of a new sesquiterpenoid (1), a new monoterpenoid (2) and its glycoside (3) and a known monoterpenoid (4) (Ning-Bo *et al*., 2017).

Furthermore, the results of this study suggested that the bioactive terpenoids may be useful for the development of anti-inflammatory agents and for other related disorders (Ning Bo *et al*., 2017).

In conclusion, essential oils and milk thistle extracts are of great interest in food, cosmetic and pharmaceutical industries. The biological compounds present in these products, can be use as natural additives emerged from a growing tendency to substitute synthetic preservatives by natural ones (Mhamdi *et al*., 2016).

2. Aim of work

Milk thistle has been studied as rich source of phytochemical compounds, mainly for silymarin, a pharmaceutical compound presents in its fruits (achenes).

The bioactive compounds of milk thistle exhibit functional roles in plant metabolism and nutraceutical effects on human health (Lucini *et al*., 2016). In fact, as mentioned above, silybin has a wide range of pharmacologic properties (Ebrahimnezhad *et al*., 2013; Mahmdi *et al*., 2016). This compound has also been described as useful for intervention of hormone refractory human prostate cancer (Zi *et al*., 1999). Silymarin effects have also been indicated in various diseases of different organs such as prostate, lungs, CNS (Central Nervous System), kidneys, pancreas, and skin (Gazak *et al*., 2007; Kumar *et al*., 2011). It has been also discovered that silymarin may be helpful in slowing down the progression of neurodegeneration in focal cerebral ischemia (El Sherif *et al*., 2013).

Terpenes are also known to be important bioactive compounds in different members of the Asteraceae family, e.g. artichoke (MacLeod *et al*., 1982; Shakeri and Ahmadian, 2014; Eljounaidi *et al*., 2014), chicory (Cankar *et al*., 2011; Fan *et al*., 2017), sunflower (Adams and TeBeest, 2017) etc. In fact, as mentioned above, the terpenes play important roles in plant interactions, plant defenses, other environmental stresses and are also used to attract the insects of pollination (Chen *et al*., 2011; Sing *et al*., 2015). Despite the important roles of terpenes in the Asteraceae family, the pathway responsible for terpenes biosynthesis in this plant is unknown and few articles are present in the literature concerning the content of terpenes in milk thistle tissues.

For this reason, in this research, ecophysiological and metabolomic studies were carried out to investigate the metabolomic profiling of different silymarin constituents (silybin, silychristin and silydianin) in various milk thistle tissues.

Metabolomic analyses were also integrated with bioinformatic and genetic methodologies to examine the terpenes metabolomic profiling in different milk thistle tissues and to characterize the genes involved in the biosynthesis of these bioactive compounds.

To achieve these goals, the following approaches were carried out:

1. Ecophysiological studies of seed germination;
2. Evaluation of silymarin metabolomic profiling in milk thistle leaves, stems, roots and flowers from three different development stages;
3. Evaluation of terpenes metabolomic profiling in milk thistle leaves, stems, roots and flowers from three different development stages;
4. Characterization of candidate *SmTPS* (*S. marianum* Terpene Synthases) genes, integrated with sequences and phylogenetic analysis of these genes and their respective encoded enzyme (SmTPS proteins);
5. Comparison between RT-PCR (reverse transcription polymerase chain reaction) and on-line RNASeq data.

3. Materials and methods

Since the important role of milk thistle as therapeutic plant, ecophysiological and metabolomic methodologies were carried out to investigate the metabolomic profile of different silymarin constituents (silybin, silychristin and silydianin). Metabolomic methods were also integrated with bioinformatic and genetic methodologies to examine the terpenes metabolomic profiling in milk thistle tissues and to characterize the genes involved in the biosynthesis of these bioactive compounds.

3.1. Plant materials

For the different trials of this study, we used both field-grown and laboratory-grown conditions plants in correlation to milk thistle's biological cycle. Field-grown milk thistle's plant materials of local genotype (37°32'03.0''N14°54'15.7''E, altitude of 130 m, East of Sicily, Italy) were harvested during the second half of April 2018. Flowers from three diverse development stages (Tab. 3 and Fig. 14), stems, leaves and roots were collected from three different plants (called A, B and C), cut off and rapidly frozen in liquid nitrogen and stored at -80 °C. These plant tissues were then couriered on dry ice to Wageningen University and Research (UR).

The heads of milk thistle of local genotype (37°53'42.53''N14°57'16.50''E, altitude of 905 m, Randazzo, East of Sicily, Italy) were harvested at the beginning of July 2017. Achenes were collected either by shaking or threshing the heads then sieving. All seeds were stored in paper bags at a constant temperature of 20 °C in the dark until used. The specimen was authenticated by the Laboratory of CNR-ISAFOM (National Research Council-Institute for Agricultural and Forest Systems in the Mediterranean), Catania, Italy. All the tissues were ground in a fine power with liquid nitrogen using an electric grinder (IKA A11 basic) to generate a stock sample which was subsequently stored at -80 °C until chemical and genetic analysis.

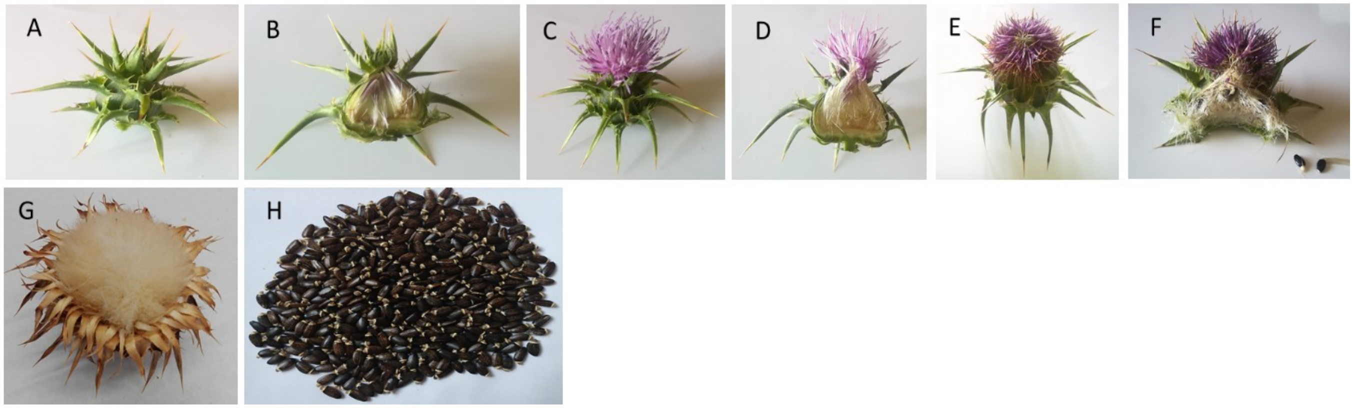


Figure 14. Developmental stages of milk thistle flower heads and seeds used in this study. **A** and **B** Early flowering or stage 1; **C** and **D** Mid-flowering or stage 2; **E** and **F** Late flowering or stage 3; **G** Head with achenes or stage 4; **H** ripe seeds.

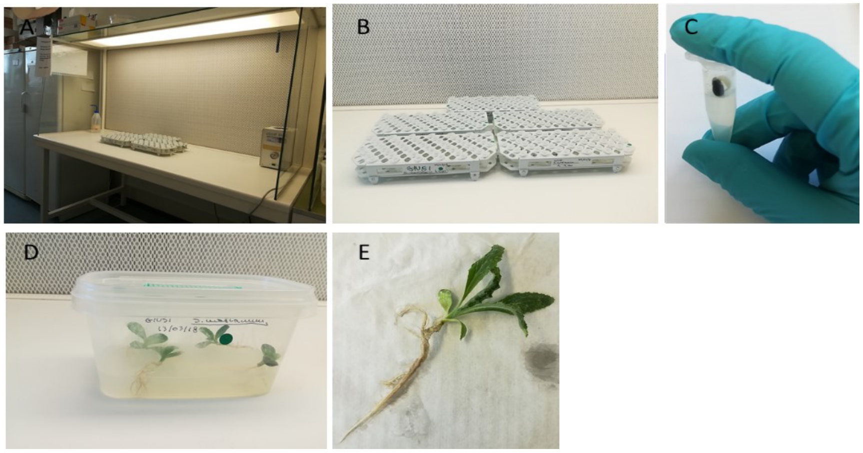
Table 3. Developmental stages of milk thistle flower heads used in this study.

Flowering stage	**Capitulum features**
Early flowering or stage 1	Closed capitulum buds
Mid-flowering or stage 2	Florets open half way to centre of disc
Late flowering or stage 3	Florets drying off with ripe seeds
Head with achenes or stage 4	Capitulum dry, with ripe seeds ready for release

3.2. Laboratory growth conditions

3.2.1. Germination of milk thistle seeds

The seeds were surface sterilized under biological hood (Telstar CLF) (fig. 15A) by immersion in 0.5% NaClO solution for 1 minute and washed three times with sterile distilled water. Sterilized seeds were then placed in 2 mL sterile tubes (fig. 15B) containing Murashige and Skoog (MS) medium (2.026 g/L) solidified with 8 g/L micro agar and were germinated in growth chambers with a photoperiod of 16 hours and with the temperature set at 25 °C. The counting of germinated seeds was done regularly after 24 hours and the appearance of 2 mm or more of radicle was considered as germination (fig. 15C). Germination test was ended after 14 days. Germination percentage was evaluated by counting the numbers of normal seedlings at the end of standard germination test. After 10 days the plants with cotyledons were transferred in plastic pots containing 20 g/L sucrose and 1/2 MS medium solidified with 8 g/L micro agar (fig. 15D) and were grown in growth chambers with a photoperiod of 16 hours and with the temperature set at 25 °C. After about a month leaves and roots in three biological replicas were collected, washed with distilled water (fig. 15E) and rapidly frozen in liquid nitrogen and stored at -80 °C until genetic analysis. MS medium, micro agar and sucrose were purchased from Duchefa Biochemie (Haarlem, The Netherlands).



Figures 15. A Biological hood **B** Tubes with MS, Agar and seeds **C** Germinated seed **D** Plot with MS, agar, sucrose and plants (10 days after seeding) with cotyledons and leaves **E** Plant collected after one month from sowing.

3.3. Metabolomic profiling of silymarin in milk thistle tissues

In order to evaluate the different localization of silymarin in various milk thistle tissues (leaves, stems, roots and flowers from three different development stages), three biological replicate plants were collected in Sicily and couriered on dry ice to Wageningen University and Research (UR). Each tissue was processed and stored (as described in Paragraph 3.1.) before performing metabolomic analysis.

Semi-polar compounds were profiled by liquid chromatography-mass spectrometry (LC-MS) analysis using Q Exactive Plus Orbitrap LC-MS/MS (Thermo Fisher Scientific) (figures 16-18). This analytical platform is designed for use in a wide range of both metabolomics and proteomics analyses. With its high sensitivity, fast scan speed and multiplexing capabilities at high mass resolution, the Q Exactive Orbitrap FTMS mass spectrometer is an outstanding detector for fast and high-throughput separation and mass peak annotation techniques required for metabolomics and proteomics analyses. The Q Exactive is unique in its ability for fast ionization mode switching at high mass resolution, providing alternating positive/negative scans of accurate mass ions during fast chromatographic separation. In addition, the high resolution MS/MS data, generated by using the quadrupole (Q) option with high collision dissociation (HCD), enables metabolite and peptide identification and quantitation of even more compounds with greater confidence. Attached to a Dionex UltiMate 3000 U-HPLC system (Dionex, Sunnyvale, CA, USA) coupled with Q Exactive^{plus}-Orbitrap FTMS mass spectrometer (Thermo Fisher Scientific) for fast chromatographic separation, a photodiode array detector and the online (+/-) switching capability at high mass resolution allows the most comprehensive metabolomics profiling of compounds present in complex samples and saves considerable time during experiments in which screening in both ionization modes is necessary or desired.

Compounds or peptides in crude or more purified extracts are firstly separated by LC and then ionized at the source of the Q Exactive Orbitrap FTMS. The S-lens at the source filters the ions from non-charged compounds and impurities. Subsequently, the Quadrupole can be activated to filter for only one specific ion of interest up to a wide range of ions that are transferred to the C-trap. Here ions can be sent to the Orbitrap mass analyzer with or without high collision fragmentation in the HCD cell, depending upon the users demands. Within the Orbitrap, the m/z values of the entering ions are accurately determined at high mass resolution, based on Fourier transformation of the mass-dependent ion oscillation frequency. The high mass resolution of all ions in combination with indicative MS/MS fragments ensures sensitive and accurate detection and quantification of a large number of target molecules present in the extracts. In the untargeted mode, fast scanning at both a wide m/z range and a high mass resolution enables the detection and relative quantification of hundreds to thousands of molecules (https://www.wur.nl/en/product/Q-ExactivePlus-Orbitrap-LC-MSMS.htm).

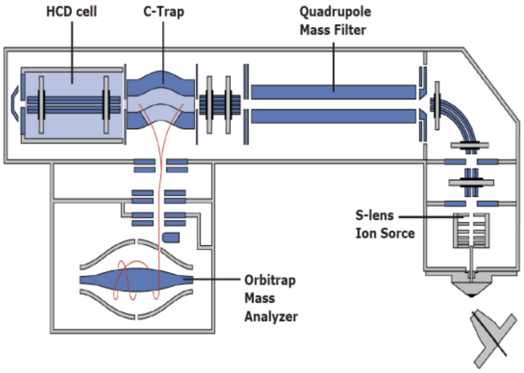


Figure 16. Schematic representation of LC-MS with Q Exactive Plus Orbitrap LC-MS/MS (https://www.wur.nl/en/product/Q-ExactivePlus-Orbitrap-LC-MSMS.htm).

3.3.1. Chemicals used for LC-MS analysis

The following standards were used for LC-MS analysis: silybin, silychristin and silydianin purchased from HWI ANALYTIK GmbH (Rülzheim, Germany). Methanol (MeOH) absolute were purchased from Biosolve (Dieuze, France), while formic acid (FA) from Sigma Aldrich (St. Louis, MO, USA).

3.3.2. Metabolites extraction and analysis of silymarin by LC-MS

For LC-MS analysis, 300 mg of ground frozen tissue for each biological replica were extracted with 700 µL of 100% methanol (MeOH) and 0.13% formic acid (FA) (Fig. 17). Afterwards samples were mixed by vortexing (15 seconds). Extracts were sonicated for 15 minutes and centrifuged at 14,000 rpm for 15 minutes.

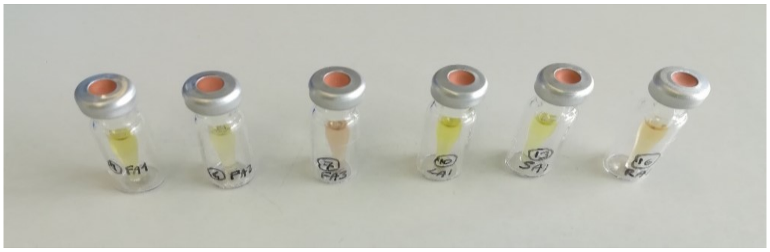


Figure 17. Vials with ground tissues dissolved in MeOH (100%) and FA (0.13%) solution.

The supernatant was analyzed by accurate mass LC-MS as described previously (De Vos *et al*., 2007). An UltiMate 3000 U-HPLC system (Dionex, Sunnyvale, CA, USA) was used to create a 45 min linear gradient of 5–75% acetonitrile in 0.1% FA in water at a flow rate of 0.19 ml min^{-1}. Of each extract, 5 µl was injected and compounds were separated on a Luna C18 column (2.0 x 150 mm, 3 µm; Phenomenex) at 40 °C. A Q Exactive^{plus}-Orbitrap FTMS mass spectrometer (Thermo Fisher Scientific), operating at a resolution of 35 000 with scan-to-scan switching between negative and positive electrospray ionization (ESI) mode both over the *m/z* range 95–1350, was used to detect eluting compounds. Raw data files of both + and – ESI mode LC-MS analysis were subsequently processed in an untargeted manner using the dedicated Metalign-MSClust workflow described previously (van Treuren *et al.*, 2018), including unbiased peak picking, alignment and assembling of mass signals likely derived from the same metabolite.

For the identification and quantification of silymarin constituents, three commercial standards were injected in a known concentration (5 µg mL^{-1} in acidified aqueous methanol at a final concentration of 70% methanol and 0.13% formic acid) and the peak areas of the respective compounds were identified in the study samples were calculated using the formula:

F = (extraction volume/g sample) x [(5µg x area peak sample)/ (area peak standard)] µg g^{-1} FW (fresh weight).

These standards (see Paragraph 3.3.1.) were compared for retention time and accurate mass with LC peaks of silybin, silychristin and silydianin. A single LC-MS in the negative ionization mode was used to analyze the silymarin constituents, allowing a mass deviation of 5 ppm. Visualization of the data was performed using Xcalibur 2.1 software (Thermo). QualBrowser was used for identified compounds in the different tissues.

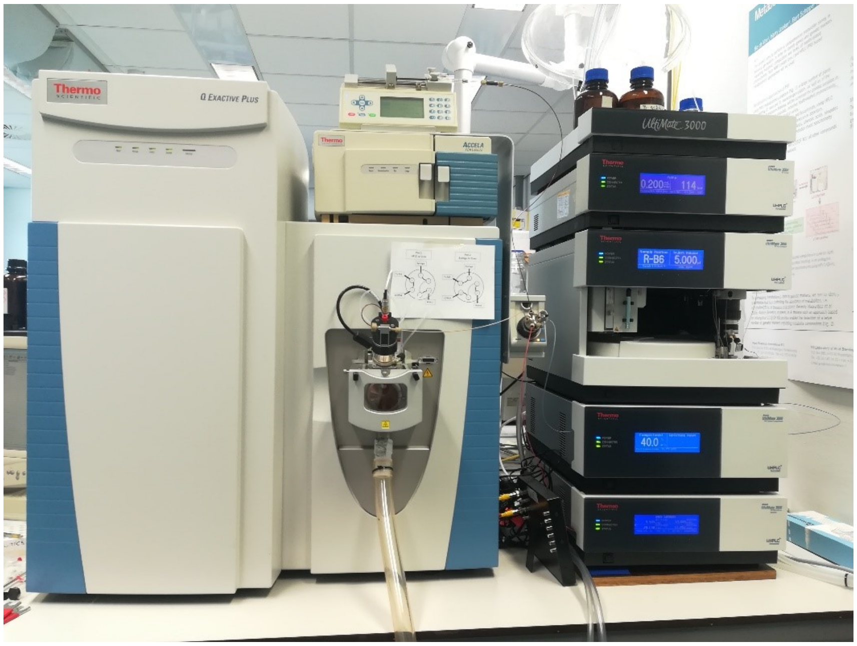


Figure 18. Photo of the used UltiMate 3000 U-HPLC system (Dionex) coupled with Q Exactiveplus-Orbitrap FTMS mass spectrometer (Thermo Fisher Scientific).

3.4. Metabolomic profiling of terpenes in milk thistle tissues

In order to detect as many compounds as possible in a given sample, an untargeted metabolomics workflow was followed, using GC-MS for the profiling of the volatile organic compounds (VOCs).

Volatile organic compounds were analyzed using gas chromatography-quadrupole mass spectrometry (GC-qMS) by sampling the headspace sampled using solid-phase micro extraction (SPME) (figures 19-21) at the laboratory of Wageningen University and Research (The Netherlands). Solid-phase microextraction (SPME) is a technique, which decreases the drag of sample preparation and thereby, reduces the analysis time. Extraction and handling the SPME device are simple (Pawliszyn, 2002) and are like handling a syringe (Figure 19). The basic principle involved in SPME is to expose a precoated surface to the sample matrix of interest (Camarasu, 2000). Once equilibrium is reached between the sample and the headspace, volatiles from the headspace are absorbed on the coating of the fiber. After equilibrium is reached, further exposure of the fiber does not increase the amount of compound extracted. Hence, by SPME, sample extraction and pre-concentration processes could be attained in one single step (Davoli *et al*., 2003; Tuduri *et al*., 2003; Balasubramanian *et al*., 2011). The volatiles are usually thermally desorbed and introduced to the GC-MS for separation and detection.

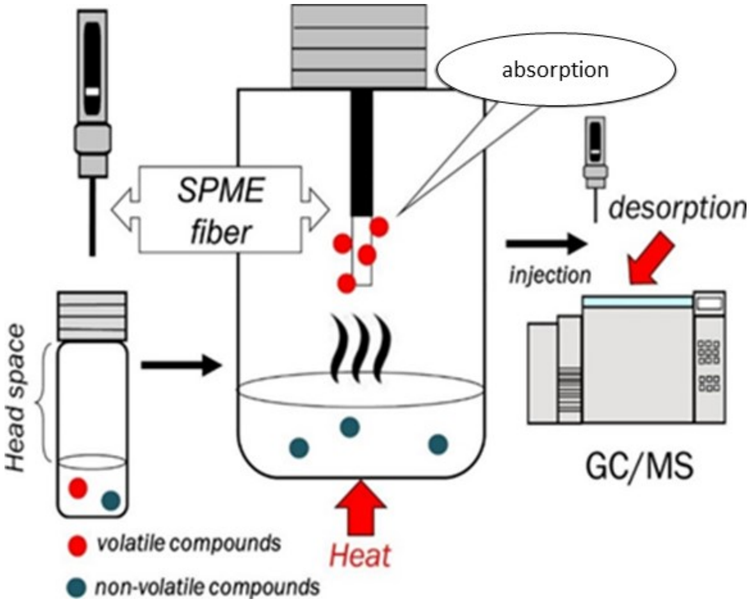


Figure 19. Schematic representation of GC-MS with HS-SPME method (adapted from Saito *et al*., 2019).

3.4.1. Chemicals used for GC-MS analysis

The following standards were used for GC-MS analysis: β-elemene, nootkatone and valencene where purchased from Isobionics (RD Geleen, The Netherlands); farnesal, indole, α and β-pinene, limonene were provided by Sigma Aldrich (St. Louis, MO, USA); ginger oil and sabinene were purchased from Carl Roth GmbH + Co. KG (Karlsruhe, Germany); linalool oxide was provided by Fluka Chemika (Buchs, Switzerland). EDTA (Ethylenediaminetetraacetic acid) and $CaCl_2$ (calcium chloride) were purchased from Sigma Aldrich (St. Louis, MO, USA).

3.4.2. Metabolites extraction and analysis of terpenes by SPME-GC-MS

For GC-MS analysis, 500 mg of frozen ground tissue for each biological replicate was dissolved with 1 mL of EDTA (50 mM) and $CaCl_2$ (5M) (Fig. 20). EDTA solution was chosen for its effectiveness compared to a number of alternative buffers tested in a previous study (Tikunov *et al*., 2005). $CaCl_2$ was added to stop enzyme activity and to drive the volatiles into the headspace (Bezman *et al*., 2003; Verdonk *et al*., 2003). Afterwards samples were thoroughly vortexed.

Figure 20. Vials with ground tissues dissolved in EDTA (50 mM) and $CaCl_2$ (5 M) solution.

Headspace volatiles were collected by solid phase microextraction (SPME) using a *50/30 um DVB/Carboxen/PDMS* (Supelco, Bellefonte, USA) similarly as described previously (Cordovez *et al*., 2015). The volatile compounds were thermally desorbed at 250 °C by inserting the fibre for 2 minutes into the GC injection port (Agilent GC7890A). The released compounds were transferred onto the analytical column (*ZB-5ms*, 30 m × 0.25 mm ID, 1 µm film thickness) in splitless mode. The temperature program ran from 45 °C (2-minutes hold) and rose 5 °C min^{-1} to 250 °C (5-minutes hold). The column effluent was ionised by electron impact at 70 eV (Agilent MSD 5978C). Mass scanning was done from m/z 33 to 550 with a scan time of 2.8 scans s^{-1}. The chromatography and mass spectral data were converted and evaluated using Xcalibur 2.1 software (Thermo). The QualBrowser software module was used to inspect the mass spectra and identify volatile compounds in the different tissues. Volatile compounds were identified based on comparison of the obtained mass spectra and retention times with those of authentic reference standards (10 µg mL^{-1}) and a ginger oil with known composition and concentration (50 µg mL^{-1}). Other compounds were putatively identified by matching mass spectra with those of commercial

(NIST14) and in-house libraries and by comparing the retention indices with those published in literature. Linear retention indices were calculated based on a series of alkanes (C8-C22).

GC–MS raw data were processed using MetAlign software (Lommen, 2009) to extract and align the mass signals (s/n ≥ 3). The software tool MSClust was used to remove signal redundancy per metabolite and to reconstruct compound mass spectra as previously described (Tikunov *et al.*, 2012).

For the quantification of specific terpenes, ten commercial standards were injected in a known concentration (10 μg mL^{-1} pentane) and the peak areas of the respective compounds were identified (α-pinene, β-elemene, limonene and valencene) in the study samples were calculated using the formula:

F = (extraction volume/g sample) x [(10μg x area peak sample)/ (area peak standard)] μg g^{-1} FW (fresh weight).

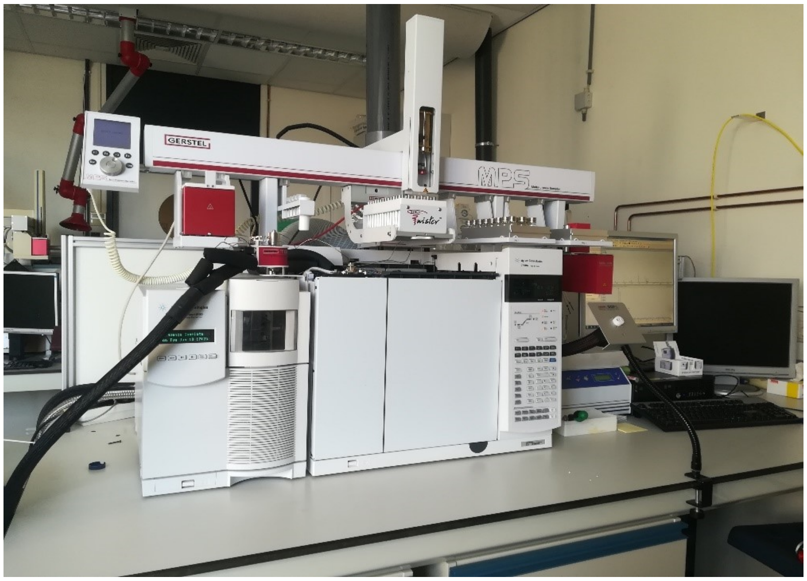


Figure 21. Photo of the used gas chromatography mass spectrometry system (Agilent) coupled to a Gerstel Multi Purpose Sampler (MPS) with headspace solid-phase microextraction (HS-SPME).

3.5. Functional genetics

In order to characterize the genes involved in the biosynthesis of terpenes in milk thistle, flowers from three diverse development stages, leaves, stems and roots from plants grown in field conditions and leaves and roots from plant grown in growth chambers were collected, stored and processed as described above (Paragraph 3.1.).

3.5.1. RNA extraction

Three different methods (described below) were used for the RNA extraction. The best method was selected in correlation to the different tissues to have a good quality of RNA. All the tissues were ground in a fine power with liquid nitrogen using an electric grinder (IKA A11 basic) or in a mortar with a pestle based on hardness of the tissue.

1) For some samples RNA was extracted using the protocol described in the RNeasy Plant Mini Kit (Quiagen) according to the manufacturer's instruction with minor modifications, as follows:

1. Add 10 µl β-mercaptoethanol (β-ME) to 1 mL *buffer RLT* before use. Disrupt a maximum of 100 mg plant material with liquid nitrogen using a mortar and pestle or an electric grinder. Decant tissue powder and liquid nitrogen into RNase-free, liquid-nitrogen–cooled, 2 mL microcentrifuge tube. Allow the liquid nitrogen to evaporate, but do not allow the tissue to thaw. Proceed immediately to step 2.
2. Add 450 µL *buffer RLT* (lysis buffer) to a maximum of 100 mg tissue powder. Vortex vigorously.
3. Transfer the lysate to a *QIAshredder spin column* (lilac) placed in a 2 mL collection tube. Centrifuge for 2 minutes at full speed. Transfer the supernatant of the flow-through to a new microcentrifuge tube without disturbing the cell-debris pellet.
4. Add 500 µL of ethanol (96–100%) to the cleared lysate and mix immediately by pipetting. Do not centrifuge. Proceed immediately to step 5.
5. Transfer the sample (usually 650 µL), with any precipitate, to a *RNeasy Mini spin column* (pink) in a 2 mL collection tube. Close the lid, and centrifuge for 15 seconds at ≥8000 x g (≥10,000 rpm) at 4 °C. Discard the flow-through.

6. Add 700 µL *buffer RW1* to the *RNeasy spin column*. Close the lid, and centrifuge for 15 seconds at ≥8000 x g at 4 °C. Discard the flow-through.
7. Add 500 µL *buffer RPE* to the *RNeasy spin column*. Close the lid, and centrifuge for 15 seconds at ≥8000 x g. Discard the flow-through.
8. Add 500 µL *buffer RPE* to the *RNeasy spin column*. Close the lid, and centrifuge for 2 minutes at ≥8000 x g.
9. Place the *RNeasy spin column* in a new 1.5 mL *collection tube*. Add 30 µL RNase-free water directly to the *spin column membrane*. Close the lid, and centrifuge for 1 minute at ≥8000 x g to elute the RNA.

2) For the samples for which a good RNA was not obtained with the first method, RNA was extracted using the protocol described in the Spectrum^{TM} Plant Total RNA Kit (Sigma-Aldrich, Life Science) according to the manufacturer's instruction with minor modifications, as follows:

1. Pipette 500 µL of the Lysis Solution/2-ME (2-mercaptoethanol) mixture to 100 mg of tissue powder and vortex immediately and vigorously for at least 30 seconds. Incubate the sample at 56 °C for 3–5 minutes (do not vortex the sample during or after the heat incubation).
2. Centrifuge the sample at maximum speed for 3 minutes (14,000-16,000 x g) to pellet cellular debris.
3. Pipette the lysate supernatant into a *filtration column* (blue retainer ring) seated in a 2-mL *collection tube* by positioning the pipette tip at the bottom of the tube but away from the pellet. Close the cap and centrifuge at maximum speed for 1 minute to remove residual debris. Save the clarified flow-through lysate.
4. Pipette 500 µL of *binding solution* into the clarified lysate and mix immediately and thoroughly by pipetting at least 5 times or vortex briefly. Do not centrifuge. Pipette 700 µL of the mixture into a *binding column* (red retainer ring) seated in a 2-mL *collection tube*. Close the cap and centrifuge at maximum speed for 1 minute to bind RNA. Decant the flow-through liquid and tap the *collection tube* (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the *collection tube*

and pipette the remaining mixture to the column and repeat the centrifugation and decanting steps. Continue to Step 5.
5. Pipette 500 µL of *wash solution 1* into the column. Close the cap and centrifuge at maximum speed for 1 minute. Decant the flow-through liquid and tap the *collection tube* (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the *collection tube*.
6. Pipette 500 µL of the diluted *wash solution 2* into the column. Close the cap and centrifuge at maximum speed for 30 seconds. Discard the flow-through liquid and tap the *collection tube* (upside down) briefly on a clean absorbent paper to drain the residual liquid. Return the column to the *collection tube* (this step was repeated twice).
7. Centrifuge the column at maximum speed for 1 minute to dry. Carefully remove the column-tube assembly from the centrifuge to avoid splashing the residue flow-through liquid on the dried column.
8. Transfer the column to a new, clean 2 mL *collection tube*. Pipette 30 µL of *elution solution* directly onto the center of the binding matrix inside the column. Close the cap and let the tube sit for 1 minute. Centrifuge at maximum speed for 1 minute to elute. Purified RNA is now in the flow-through eluate and ready for immediate use or storage at -20 °C (short term) or -80 °C (long term).

3) For the samples for which a good RNA was not obtained with the first and second method, RNA was extracted using the protocol described by Chang *et al.* (1993) with minor modifications, as follows:
1. About 2 g of powdered tissue were transferred to a 50 mL tube containing 15 mL of pre-heated (65 °C) *RNA extraction buffer* (2% CTAB or hexadecyltrimethylammonium bromide, 2% PVP k30 or polyvinylpyrrolidone, 100 mM Tris-HCl (pH 8.0) or trishyroxymethylaminomethane, 25 mM EDTA or ethylenediaminetetraacetic acid, 2.0 M NaCl, 0.5 g/L spermidine). This solution was autoclaved (121 °C for 20 minutes) and 2% β-mercaptoethanol (β-ME) was added directly to each tube. Afterwards samples were mixed by vortexing (max speed, 30 seconds). Immediately 15 mL of CIA ($CHCl_3$/IAA or chloroform:isoamyl alcohol 24:1) was added, after samples were mixed

vigorously by vortexing for 15 seconds. Tubes were centrifuged at 10,000 x g for 15 minutes at 15 °C. The supernatant was transferred to a new tube and the organic phase and the interphase were discarded (the CIA extraction was repeated twice). The upper aqueous phase was transferred to a fresh tube and ¼ volume of 10 M LiCl was added. Tubes was inverted gently to mix the solution and incubated at 4 °C overnight. Samples were centrifuged at 10,000 rpm for 20 minutes at 4 °C. The supernatant was decanted and 500 µL of pre-heated (60 °C) SSTE (1.0 M NaCl, 0.5% SDS or Sodium Dodecyl Sulphate, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, or ethylenediaminetetraacetic acid) was added to the pellet and it was dissolved by gently pipetting up and down. An equal volume (500 µL) of CIA was added and, after vortexing, the mixture was centrifuged at 10,000 rpm for 10 minutes at room temperature. The aqueous phase was transferred to a new tube and 1 mL of 96% ethanol was added in each tube. The tubes were inverted and stored at -80 °C for one hour. Tubes were centrifuged at 10,000 rpm for 20 minutes at 4 °C, the supernatant was removed (this step was repeated twice). The pellet was washed with 250 µL of 70% ethanol, spun down for 10 minutes and dried for 10 minutes at room temperature. The RNA pellet was resuspended in 50 µL of autoclaved distilled water.

For all the methods described above, the purity and concentration of RNA were assessed using a Nanodrop spectrophotometer (Isogen, Life Science) and agarose gel electrophoresis.

3.5.2. DNase treatment and synthesis of cDNA

RNA was treated with DNase1 (Invitrogen) as follows: 1 µg of RNA, 1 µL of Dnase I reaction buffer (10X), 1 µL Dnase I, at the end MQ water was added up to 10 µL of total volume. The samples were incubated for 15 minutes at room temperature and afterwards 1 µL of EDTA (25 mM) was added. Then the samples were incubated at 65 °C for 10 minutes. After the DNase treatment, 1.0 µg of total RNA extracted from the different tissues of milk thistle was used to synthesize first-strand cDNA using Superscript^{TM} II Reverse Transcriptase (Invitrogen) in 20 µL total volume, as follows:

10 μL of reaction 1 (DNase treatment: 10 μL reaction + 1μL EDTA), 1 μL oligodT (500 μg/mL) and 1 μL of dNTPs (10 mM each) were mixed and incubated in thermocycler at 65 ºC for 5 minutes. Afterwards the tubes were positioned on ice and then centrifuged to spin down. Then 4 μL *first-strand buffer* (5X) and 2 μL of DTT (0.1 M) were added and the total was mixed and incubated at 42 ºC for 2 minutes. 1 μL of *Super Script II RT* was added and mixed by pipetting gently up and down and incubated at 42 ºC for 5 minutes and at the end at 70 ºC for 15 minutes, to synthesize cDNA.

3.5.3. Bioinformatic analysis

Homologous genes involved in the terpene biosynthetic pathway expressed in milk thistle shoot and flowers were screened in a previously reported transcriptome database, 1kp transcriptome project (SRA, ERS1829720) (Naim *et al*., 2014). Terpenic protein sequences from different Asteraceae were screened in milk thistle's genome by using tBlastn. All nucleotide sequences were downloaded from GenBank (http://www.ncbi.nlm.nih.gov). Exonerate software (Slater and Birney, 2005) was used to find predicted TPSs in the genome assembly of milk thistle. The nucleotide sequence, open reading frame (ORF) and deduced amino acid sequence were analyzed with DNASTAR software (Lasergene, USA) and sequence comparison was conducted through database search using BLAST tool (NCBI, http://www.ncbi.nlm.nih.gov). The best hits were taken as candidate genes.

Kallisto was used to quantify the expression of each TPS gene in each of the RNAseq datasets from different milk thistle tissues (Bray *et al*., 2016). Matplotlib was used to make the expression rank picture. The number on the scale bar represents the rank of that gene in that tissue. For example, a gene ranked 3 means that it is the third most expressed in that tissue, while the black genes (rank 0) are highest expressed (black means most expressed and white is the least expressed in the tissue).

SmTPS proteins from milk thistle were submitted to Clustal Omega (Sievers *et al*., 2011) to perform multiple sequence alignment. This alignment was visualized and annotated with TPS motifs using JalView (Waterhouse *et al*., 2009). Conserved regions such as RRx(8)W, RxR, DDxxD, and NSE/DTE motifs were highlighted with different colors (Supplementary Fig. S1). The amino acid sequences of SmTPS proteins were analyzed with Modeller and ChloroP and Predotar to predict their three-dimensional structures (Eswar *et al*., 2007) and subcellular locations (Emanuelsson *et al*., 1999; Small *et al*., 2004), respectively (Fig. 31, Tables 8 and 9). For phylogenetic analysis, the full-length amino acid sequences of SmTPS proteins and their homologs in other plant species (7 DiTPS, 9 MonoTPS, rest sesquiTPS) were aligned using Clustal Omega (Sievers *et al*., 2011) with the Pfam (Finn *et al*., 2013) domains Terpene_synth (Pfam ID: PF01397) and Terpene_synth C (Pfam ID: PF03936) as guides for the alignment. The tree was constructed using the ete3 (Huerta-Cepas *et al*., 2016) and visualized using iTOL (Ivica *et al*., 2006).

3.5.4. Designing of primers and amplification

Specific forward and reverse primers were designed (Tables 4 and 5), using DNASTAR software (Lasergene, USA), to amplify the full-length cDNA sequences (and vector) with appropriate overlaps for cloning (Tab. 5.). PCR products of appropriate length were cloned into the pACYCDuet-1 vector (Novagen) and then transformed into *Escherichia coli* (*E. coli*) DH5α competent cells before sequencing, as described in the following paragraphs.

The full-length genes were amplified from milk thistle cDNA using:

1. High Fidelity DNA Phusion polymerase (Finnzymes). The PCR (Polymerase chain reaction) conditions were as follows: start denaturation at 98 °C for 45 seconds, 30 cycles with denaturation at 98 °C for 10 seconds, annealing at 55 °C for 20 seconds (Tab. 4), extension at 72 °C for 2 minutes and final extension 72 °C for 5 minutes.

2. Q5® High-Fidelity DNA polymerase (NEB). The PCR conditions were as follows: start denaturation at 98 °C for 30 seconds, 30 cycles with denaturation at 98 °C for 10 seconds, annealing for 20 seconds at different genes temperature (Tab. 5), extension at 72 °C for 2 minutes and final extension at 72 °C for 2 minutes.

The amplified fragments were detected by agarose gel electrophoresis and ethidium bromide or SYBR® Safe (Midori) were used as intercalates.

The PCR products were purified by Zymoclean^{TM} Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instruction with minor modifications.

Table 4. Primer designing to isolate complete CDS.

Gene name	Primer name	Primer sequence (5'-3')	Ta °C	Expected Product size (bp)	Reference
SmTPS1	SmTps1_F_BamHI SmTps1_R_NotI	F: TTGGATCCGGATCATAATGATTGTAAACAAGGGGT R: TTGCGGCCGCTCATTTAATCGGATTTACGAGAAGTG	55 °	~1700	In this study
SmTPS2	SmTps2_F_BamHI SmTps2_R_NotI	F: TTGGATCCGATGGCTGCCGTAGAAGCTAA R: TTGCGGCCGCTTACATTGGTAAAGAGCCAACAAA	55 °	~1700	In this study
SmTPS3	SmTps3_F_BamHI SmTps3_R_NotI	F: TTGGATCCGATGGCAGCAGTTGAAGCTAC R: TTGCGGCCGCTTACATGGGAACAAAATAAAAAAACAAGA	55 °	~1700	In this study
SmTPS4	SmTps4_F_BamHI SmTps4_R_NotI	F: TTGGATCCGATGGCCACCGTTGAAGC R: TTGCGGCCGCCTATATAGGGACCCTATCAATCAACAAG	55 °	~1700	In this study
SmTPS5	SmTps5_F_BamHI SmTps5_R_NotI	F: TTGGATCCGATGGCAGCTGATCATGCAAC R: TTGCGGCCGCTCTAGGATTTAGGAGTAAAAAGTAAGGA	55 °	~1700	In this study
SmTPS6	SmTps6_F_BamHI SmTps6_R_NotI	F: TTGGATCCGATGGCCGTCACTGATCAAGA R: TTGCGGCCGCTTATATAACTTCTAGCTTCCTCTTCGG	55 °	~1700	In this study
SmTPS7	SmTps7_F_BamHI SmTps7_R_NotI	F: TTGGATCCGATGACAACTTCAAACTTAATACTTGATC R: TTGCGGCCGCTCATATACTTATATTATGAACGAGCAAAGA	55 °	~1700	In this study
SmTPS8	SmTps8_F_BamHI SmTps8_R_NotI	F: TTGGATCCGATGGCCACTATTGAAGCCAA R: TTGCGGCCGCTTATATAACAGGGACTGGAGTAATGA	55 °	~1700	In this study
SmTPS9	SmTps9_F_BamHI SmTps9_R_NotI	F: TTGGATCCGATGTTGACTGTTGAACAAGAGAGC R: TTGCGGCCGCTTAATGGTGGTGGTAGGGATGA	55 °	~1700	In this study
SmTPS10	SmTps10_F_BamHI SmTps10_R_NotI	F: TTGGATCCGATGTCTTTTAAACAAGAAGATGTTATCC R: TTGCGGCCGCCTAATTGATAGCATTAATGAGAATAGATTTGA	55 °	~1700	In this study
SmTPS11	SmTps11_F_SacI SmTps11_R_NotI	F: TTGAGCTCGATGTCCTCGGAACAGCTAATTTT R: TTGCGGCCGCTCAATGTAGCCCTTGGATTGG	55 °	~1700	In this study

Table 5. Cloning primer designing to isolate complete CDS.

Gene name	Primer name	Primer sequence (5'-3')	Ta °C	Expected Product size (bp)	Reference
SmTPS1	SmTps1_F GIB	F: CCACAGCCAGGATCATAATGATTGTAAACAAGGG	59 °C	~1700	In this study
	SmTps1_R GIB	R: GCATTATGCTCATTTAATCGGATTTACGAGAAGTG			
	pDuet T1_F GIB	F: CCGATTAAATGAGCATAATGCTTAAGTCGAACA	59 °C	4008	
	pDuet T1_R GIB	R: TTATGATCCTGGCTGTGGTGATG			
SmTPS2	SmTps2_F GIB	F: CCATCATCACCACAGCCAGATGGCTGCCGTAGAAGCTAATGATACC	68 °C	~1700	In this study
	SmTps2_R GIB	R: CAATACGATTACTTTCTGTTCGACTTAAGCATTATGCTTACATTGGTAAAGAGCCAACAAACAGGAGATTAATA			
	pDuet T2_F GIB	F: TCTCCTGTTTGTTGGCTCTTTACCAATGTAAGCATAATGCTTAAGTCGAACAGAAAGTAATCGTATTGTAC	69 °C	4008	
	pDuet T2_R GIB	R: CATTAGCTTCTACGGCAGCCATCTGGCTGTGGTGATGATGGTGATG			
SmTPS3	SmTps3_F GIB	F: GCCAGATGGCAGCAGTTGAAG	59 °C	~1700	In this study
	SmTps3_R GIB	R: GCATTATGCTTACATGGGAACAAAATAAAAAAACAA			
	pDuet T3_F GIB	F: GTTCCCATGTAAGCATAATGCTTAAGTCGAACA	59 °C	4008	
	pDuet T3_R GIB	R: CTGCTGCCATCTGGCTGTGGTGATG			
SmTPS4	SmTps4_F GIB	F: CACCATCATCACCACAGCCAGATGGCCACCGTTGAAGCCAATAC	69 °C	~1700	In this study
	SmTps4_R GIB	R: TACAATACGATTACTTTCTGTTCGACTTAAGCATTATGCCTATATAGGGACCCTATCAATCAACAAGAGTGTCAC			
	pDuet T4_F GIB	F: TGACACTCTTGTTGATTGATAGGGTCCCTATATAGGCATAATGCTTAAGTCGAACAGAAAGTAATCGTATTGTAC	69 °C	4008	
	pDuet T4_R GIB	R: GCTTCAACGGTGGCCATCTGGCTGTGGTGATGATGGTGATG			
SmTPS5	SmTps5_F GIB	F: CAGCCAGATGGCAGCTGATCATG	59 °C	~1700	In this study
	Sm Tps5_R GIB	R: GCATTATGCCTAGGATTTAGGAGTAAAAAGTAAGG			
	pDuet T5_F GIB	F: TCCTAAATCCTAGGCATAATGCTTAAGTCGAACA	59 °C	4008	
	pDuet T5_R GIB	R: GCTGCCATCTGGCTGTGGTGATG			

3.5.5. pACYCDuet-1 cloning by Gibson Assembly method

Gibson Assembly was developed by Dr. Daniel Gibson and his colleagues at the J. Craig Venter Institute and licensed to New England Biolabs by Synthetic Genomics, Inc. It allows for successful assembly of multiple DNA fragments, regardless of fragment length or end compatibility. This method has been successfully used by Gibson's group to assemble oligonucleotides, DNA with varied overlaps (15–80 bp) and fragments hundreds of kilobases long (1–2) and has been rapidly adopted by the synthetic biology community due to its ease-of-use, flexibility and suitability for large DNA constructs. Gibson Assembly Cloning Kit has been further optimized to increase the efficiencies for simultaneous assembly and cloning of one or two fragments into any vector. Gibson Assembly efficiently joins multiple overlapping DNA fragments in a single-tube isothermal reaction.

The Gibson Assembly Master Mix includes three different enzymatic activities that perform in the same buffer:

- the exonuclease creates single-stranded 3′ overhangs that facilitate the annealing of fragments that share complementarity at one end (overlap region);
- the proprietary DNA polymerase fills in gaps within each annealed fragment;
- the DNA ligase seals nicks in the assembled DNA.

The end result is a circular, double-stranded, fully sealed DNA molecule that can be used to transform *E. coli* Competent Cells (Fig. 22) (New England Biolabs Inc., 2018).

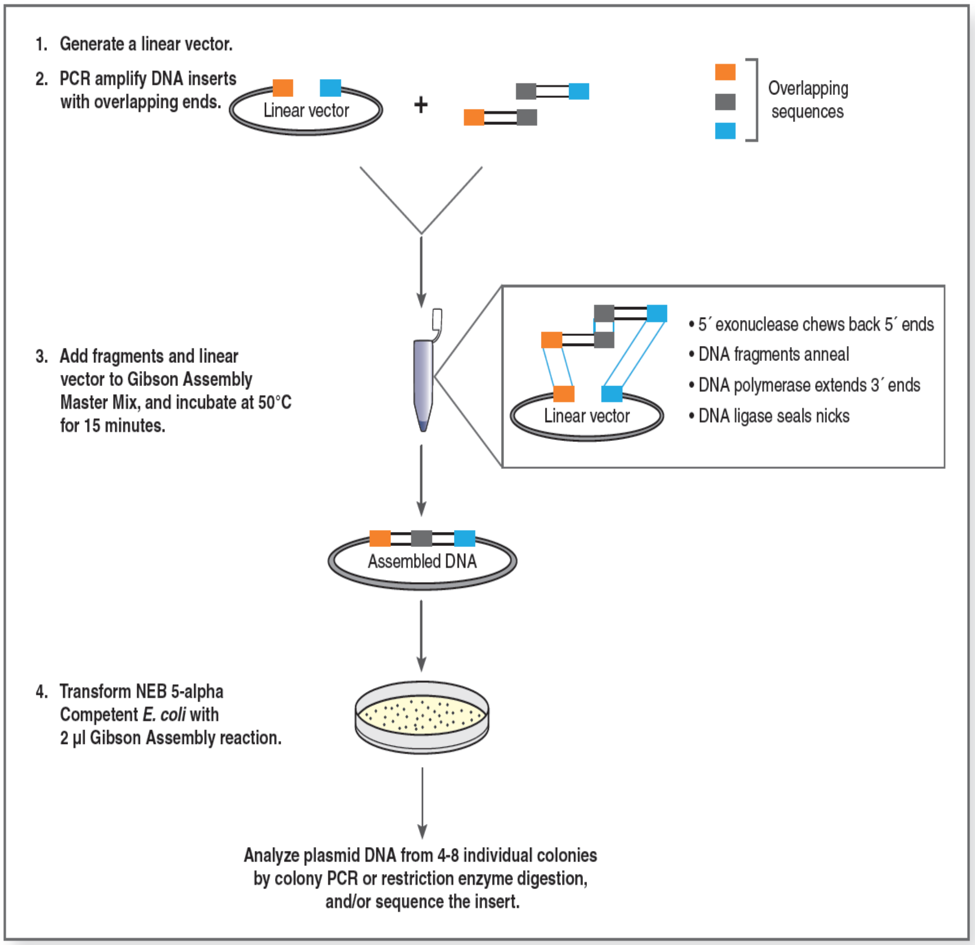

1. Generate a linear vector.
2. PCR amplify DNA inserts with overlapping ends.
Linear vector
+
Overlapping sequences
3. Add fragments and linear vector to Gibson Assembly Master Mix, and incubate at 50°C for 15 minutes.
Linear vector
• 5´ exonuclease chews back 5´ ends
• DNA fragments anneal
• DNA polymerase extends 3´ ends
• DNA ligase seals nicks
Assembled DNA
4. Transform NEB 5-alpha Competent E. coli with 2 µl Gibson Assembly reaction.
Analyze plasmid DNA from 4-8 individual colonies by colony PCR or restriction enzyme digestion, and/or sequence the insert.

Figure 22. Overview of the Gibson Assembly Cloning Method (New England Biolabs Inc., 2018).

The PCR products of SmTPS1, SmTPS2, SmTPS3, SmTPS4, SmTPS5 that showed a band in agarose gel, were cloned into the pACYCDuet-1 vector using the Gibson Assembly method according to the manufacturer's instruction with minor modifications, as follows:

1. design primers to amplify fragments (and/or vector) with appropriate overlaps (Tab. 5).

2. PCR amplify fragments using a high-fidelity DNA polymerase. Amplification of cDNA was carried out using Q5® High-Fidelity DNA Polymerase (M0491) (NEB) with the PCR program shown in paragraph 3.5.4.

3. Prepare linearized vector by PCR amplification using a high-fidelity DNA polymerase. Amplification of pACYCDuet-1 plasmid (Novagen) was carried out using Q5® High-Fidelity DNA Polymerase (M0491) (NEB) with the program shown in paragraph 3.5.4. This vector carries chloramphenicol (*Chl*) resistance genes (Fig. 23) that enables positive selection of recombinant plasmid. Only the cells with recombinant plasmids are able to propagate.
4. Confirm and determine concentration of fragments and linearized vector using agarose gel electrophoresis and a Nanodrop™ instrument. The amplified fragments were detected by agarose gel electrophoresis and PCR products of appropriate length were isolated from an agarose gel separation (Zymoclean™ Kit).
5. Add fragments and linearized vector to Gibson Assembly Master Mix and incubate at 50°C for 4 hours.
6. Transform into DH5α *E. coli* competent cell.

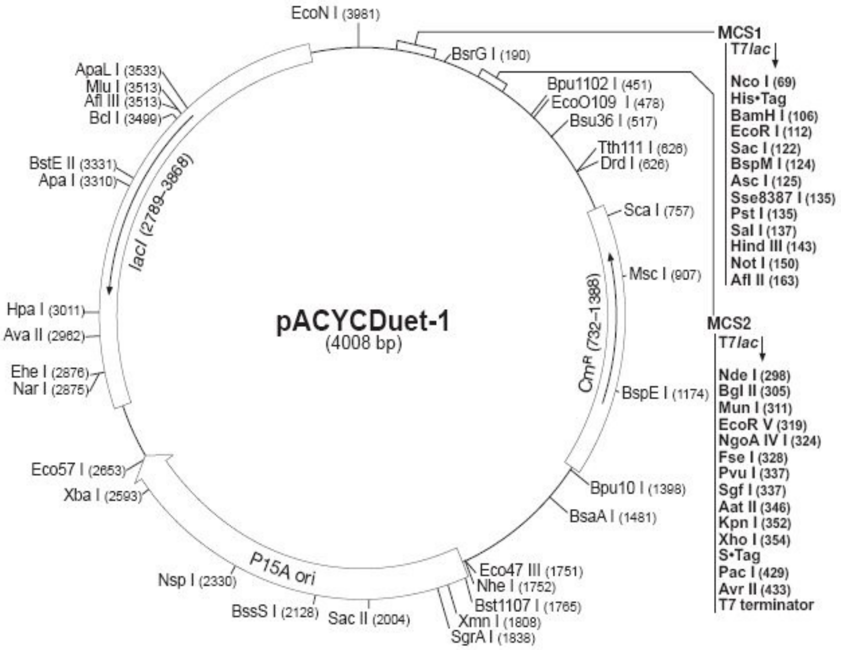


Figure 23. Map of pACYCDuet-1 Vector with the restriction sites evidenced. Cm^R is the gene that caused the chloramphenicol resistance (https://ecoliwiki.org/colipedia/index.php/File:PACYCDUET-1.jpg).

3.5.5.1. Chemical transformation in *Escherichia coli*

The resultant Gibson reaction products were transformed into *E. coli* DH5α competent cells according to the manufacturer's instruction with minor modifications, as follows:

- Thaw chemically competent cells on ice.
- Add 2 µL of the chilled assembly product (diluted 1:4) to the competent cells (200 µL). Mix gently by flicking the tube 4-5 times. Do not vortex.
- Place the mixture on ice for 30 minutes. Do not mix.
- Heat shock at 42 °C for 50 seconds. Do not mix.
- Transfer tubes to ice for 2 minutes.
- Add 900 µL of LB (Luria-Bertani: 1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.0) liquid medium.
- Incubate the tube at 37 °C for 60 minutes. Shake vigorously (250 rpm).
- Centrifuge at 3,000 rpm for 3 minutes.
- Spread 100 µL of the cells onto the prewarmed selection LB (Luria-Bertani: 1% tryptone, 0.5% yeast extract, 1% NaCl, 15 g/L agar, pH 7.0) plates with chloramphenicol 50 µg/mL.
- Incubate the plates at 37 ºC for 24 hours.

3.5.5.2. Colony PCR and miniprep

To verify if the transformation was successful, colony screening was performed using the Colony PCR technique.

After 24 hours at 37 °C the colonies were picked with a toothpick in sterile conditions, put in 7 µL sterile water (used as template for the screening of the recombinant clones during colony PCR) and a part spread onto LB plates with chloramphenicol (50 µg/mL), incubates at 37 °C for 24 hours and stored at 4 °C for the future analyses.

The amplification was carried out with Super Taq (Thermo Fisher Scientific) at Ta of 54 °C. The vector primers used were ACYCDuetS1 forward sequencing primer 5'-GGATCTCGACGCTCTCCCT- 3' and ACYCDuetAS1 reverse sequencing primer 5'-GATTATGCGGCCGTGTACAA- 3' (Novagen). These primers targeted the right and the left terminal of bacterial genome, respectively. The PCR products were verified on agarose gel at 1% in TAE buffer. Positive clones were incubated for 24 hours at 37 °C in LB liquid broth with chloramphenicol (50 μg/mL). Plasmid DNA was extracted by the means of QIAprep Miniprep Kits (Quiagen, Germany), that uses silica membrane technology (Fig. 24). Quantitative and qualitative analyses of extracted plasmid DNA were performed by using NanoDrop spectrophotometer (Isogen, Life Science).

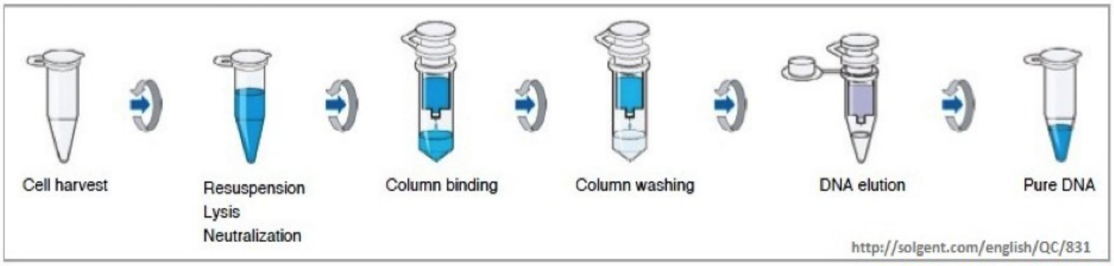


Figure 24. Scheme of miniprep protocol, with the resuspension, column binding, column washing, DNA elution phase showed (http://2014.igem.org/Team:Macquarie_Australia/WetLab/Protocols/PlasmidPreps).

The sequences were obtained using Sanger sequencing method by means of the following primers: ACYCDuetS1 forward sequencing primer 5’-GGATCTCGACGCTCTCCCT-3’ and ACYCDuetAS1 reverse sequencing primer 5’-GATTATGCGGCCGTGTACAA-3’ (Novagen). The contigs sequences were obtained with DNASTAR software (Lasergene, USA) (Supplementary Fig. S2) and then used for BLAST searches (Supplementary Tab. S1).

3.5.6. Heterologous expression of SmTPSs in *Escherichia coli*

To express *S. marianum* terpene synthase (SmTPS) proteins, an empty vector and vectors harboring different SmTPS genes were used for transformation of *E. coli* strain BL21 (DE3), as follows:

- Thaw chemically competent cells on ice.
- Add 1μL of plasmid (diluted 10X) to the competent cells (100 μL). Mix gently by flicking the tube 4-5 times. Do not vortex.
- Place the mixture on ice for 30 minutes. Do not mix.
- Heat shock at 42 °C for 50 seconds. Do not mix.

- Transfer tubes to ice for 2 minutes.
- Add 900 µL of autoclaved SOC (Super Optimal broth with Catabolite repression: 2% tryptone, 0.5% Yeast extract, 0.2% NaCl 5M, 0.25% KCl 1M, 1% $MgCl_2$ 1M, 1% $MgSO_4$ 1M,) liquid medium with 2% glucose 1M (filter sterilized).
- Incubate the tube at 37 °C for 60 minutes. Shake vigorously (250 rpm).
- Spread 100 µL of the cells onto the prewarmed selection LB (Luria-Bertani: 1% tryptone, 0.5% yeast extract, 1% NaCl, 15 g/L agar, pH 7.0) plates with chloramphenicol 50 µg/mL.
- Incubate the plates at 37 ºC for 24 hours.

After 24 hours at 37 °C the colonies were picked with a toothpick in sterile conditions and grown in LB liquid medium with chloramphenicol 50 µg/mL and 1% glucose at 37 °C at 250 rpm for 24 hours. A glycerol stock was stored at -80 °C. Afterward 24 hours, 500 µL of starter culture was diluted in 50 mL of 2xYT (Yeast Extract Tryptone: 1% NaCl, 1.6% tryptone, 1% yeast extract) liquid medium with chloramphenicol 50 µg/mL to OD_{600} (optical density measured at a wavelength of 600 nm) at 37 °C at 250 rpm. When the OD_{600} was in the range from 0.6 - 0.8 the culture was induced with 50 µL isopropyl-β-D-thiogalactopyranoside (IPTG) (1M) at 18 °C at 250 rpm for 24 hours. After induction, the cells were harvested by centrifugation and resuspended in 1 mL of Tris buffer (50 mM, pH 7.5) with 10 µL of β-mercaptoethanol and then the cells were disrupted using a FastPrep at 6.5 speed for 10 seconds and 0.2 g of zirconia beads. The crude protein extracts were collected into the supernatant and utilized in the following enzymatic activity assays.

3.5.6.1. *In vitro* enzyme assay

Assay of sesquiterpene synthase was performed in a volume of about 900 µL containing:

- 100 µL of crude enzyme solution (the precise protein amount was not determined);

- 800 μL MOPSO (3-Morpholino-2-hydroxypropanasulfonic acid) buffer (15 mM MOPSO, 12.5% glycerol, 1mM Ascorbic acid, 100 μL tween 20) adjusted at pH 7.0 before to add $MgCl_2$ (1 mM) and dithiothreitol (DTT) (2mM);
- 20 μL Na ortovanadate (250 mM);
- 5 μL FPP (10 mM) (Sigma-Aldrich) or GPP (Sigma-Aldrich) (10mM) as substrate.

The incubation mixture was overlaid with 1 mL of pentane to trap volatile products and incubated at 30 °C for 2 hours. After incubation, the pentane overlay (supernatant) was collected and centrifuge at 1200 rpm for 10 minutes.

The clear supernatant was extracted after centrifugation, dehydrated using anhydrous Na_2SO_4 and then subjected to analysis by gas chromatography-mass spectrometry (GC-MS) (Fig. 25).

Figure 25. Gas chromatography system coupled to mass-spectrometry (Agilent).

Analytes from 1 μL samples were separated using a gas chromatograph (5890 series II, Hewlett-Packard) equipped with a 30 m x 0.25 mm, 0.25 mm film thickness column (ZB-5, Phenomenex) using helium as carrier gas at flow rate of 1 mL/min. The injector was used in splitless mode with the inlet temperature set to 250 °C. The initial oven temperature of 45 °C was increased after 1 minute to 300 °C at a rate of 10 °C/min and held for 5 minutes at 300 °C. The GC was coupled to a mass-selective detector (model

5972A, Hewlett-Packard). Compounds were identified by comparison of mass spectra (Supplementary Fig. S9) and retention times (RT) with those of the following authentic standards: β-Elemene and ginger oil (Supplementary Fig. S8).

3.5.7. Comparison between RT-PCR and RNASeq data

The discovery of the reverse transcriptase enzyme led to the development of reverse transcription polymerase chain reaction (RT-PCR) technique which has since replaced Northern blot as the method of choice for RNA detection and quantification. This enzyme allows the transformation of the mRNA in cDNA which consents it to be quantified. In 2000, Bustin *et al.* describes a new technology based on this concept. RT-PCR is used to qualitatively detect gene expression through the creation of complementary DNA (cDNA) transcripts from mRNA, and quantitatively measure the amplification of cDNA by means of fluorescent probes. The process of RT-PCR involves three steps: reverse transcriptase-based conversion of RNA to cDNA, the amplification of cDNA by PCR, and the detection and quantification of amplified products-referred as amplicons (Jozefczuk and Adjaye, 2011). RT-PCR can be used to quantify mRNA in both relative (Kang *et al*., 2010) and absolute terms (Bustin *et al*., 2005). It has become one of the most extensively used methods of gene quantification as it has a large dynamic range, with high sensitivity and can be highly sequence-specific. The problems with this technique are the time consumed in the laboratory for the analysis of one gene and the variability of the results. Moreover, comparing expression levels across different experiments is often difficult and can request complicated normalization methods. One simple and effective way to measure transcriptome composition and to discover new exons or genes is by the direct ultra-high-throughput sequencing of cDNA. This method, called RNASeq, also termed "Whole Transcriptome Shotgun Sequencing" ("WTSS"), has clear advantages over existing approaches and allows the mapping and quantification of transcriptomes (Mortazavi *et al*., 2008). RNA-seq refers to the use of high-throughput sequencing technologies to sequence cDNA in order to get information about RNA content (Morin *et al*., 2008). A single experiment can provide information about gene expression, novel transcripts, novel isoforms, alternative splice sites, allele-specific expression, cSNPs, and rare transcripts. Ideally, it should be able to directly identify and quantify all RNAs, small or large (Pepke *et al*., 2009; Shah *et al*., 2009; Wiegand *et al*., 2010; Ngo *et al*., 2011). The sample analysis can be done with a variety of platforms, for example, the Illumina Genome Analyzer platform (Mortazavi *et al*., 2008), the ABI Solid Sequencing (Cloonan *et al*., 2008), or the Life Science's 454 Sequencing (Barbazuk *et al*., 2007). The principal steps of the technique are: isolate the RNAs from a sample, convert them to cDNA fragments using RT, then a high-throughput sequencer is used to generate millions of reads from the cDNA fragments, reads are mapped to a reference genome or transcript set with an alignment tool, and counts of reads mapped to each gene. For transcriptome sequencing, expression levels of different genes are determined by counting the number of reads mapped to the gene and then normalizing this read count by the length of the gene model and the total number of mapped reads in the sample. The primary advantages of this technique are high reproducibility, the large dynamic range, low background noise, requirement of less sample RNA, and the ability to detect novel transcripts, even in the absence of a sequenced genome (Wang *et al*., 2009). Nevertheless, this RNA-Seq technique will not probable replace current RT-PCR method but will be complementary. The results of the RNA-Seq will identify those genes that need to then be examined using RT-PCR methods, more so in those laboratories with resource constraints (Costa *et al*., 2013). For this reason, in our study we decide to exploit the RNASeq data already published on-line to do RT-PCR in specific milk thistle tissues and to compare our RT-PCR results with these RNASeq data.

RT-PCR was performed using primers listed in Tables 4 and 5 (for the PCR protocol see Paragraph 3.5.4.). PCR products were separated on a 1% agarose gel, and bands were visualized using high-performance chemiluminescence machine (G: Box, Syngene). At the end, PCR products were compared with the RNA-Seq data published on-line (https://www.ncbi.nlm.nih.gov/sra/?term=silybum+marianum).

3.6. Statistical analysis

For the metabolomic analysis, GC-MS raw data were processed using MetAlign software (Lommen, 2009) to extract and align the mass signals (s/n ≥ 3). The software tool MSClust was used to remove signal redundancy per metabolite and to reconstruct compound mass spectra as previously described (Tikunov *et al.*, 2012). Processed volatile profile data were log transformed, autoscaled (van den Berg *et al.*, 2006) and subjected to Principal Component Analysis (PCA) using the software package SIMCA-P (version 15.0.2 Sartorius Stedim Data Analytics, Umeå, Sweden). The number of significant PC's was determined using k-fold cross validation (Eriksson *et al.*, 2006).

4. Results

Ecophysiological and metabolomic studies were carried out to identify and quantify the different silymarin constituents (silybin, silychristin and silydianin) in milk thistle tissues. Metabolomic analyses were also integrated with bioinformatic and genetic methodologies to identify and quantify terpenes in milk thistle tissues and to characterize the genes involved in the biosynthesis of these bioactive compounds.

4.1. Germination of milk thistle seeds

The seeds of milk thistle genotype, used in this experiment, showed a high germination rate correlated with the afterripening. In fact, considering the seeds sown 7 months after the harvesting, already at the second day after sowing, they had a germination rate of 73.0% and on the fourth day they achieved a percentage of germination of 92.1%.

While, in the seeds sown 8 months after the harvesting, we observed a germination rate of 83.3% on the second day after sowing and the percentage of germination reached 96.5% on the eighth day after sowing (Fig. 26). In both germination tests, the rate of germination remained constant respectively after the fourth and eighth day from the beginning of the tests.

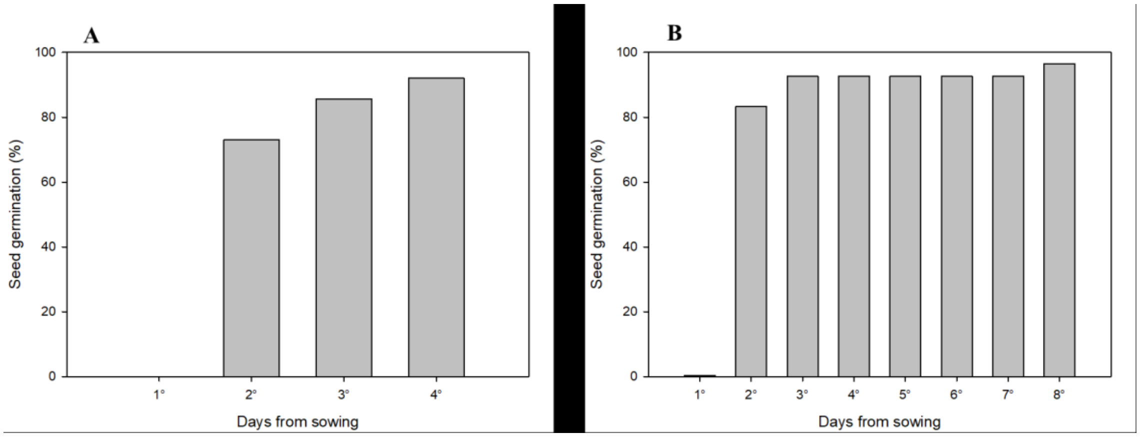


Figure 26. Germination of milk thistle seeds. Percentage of germination 7 (**A**) and 8 (**B**) months after the harvesting.

4.2. Metabolomic profiling of silymarin in milk thistle tissues

The liquid chromatography method associated with the Orbitrap allowed the identification and quantification of the different silymarin constituents in flowers from three different development stages, leaves, stems and roots of three diverse milk thistle chemotypes.

4.2.1. Detection and quantification of silymarin in milk thistle tissues

The different silymarin constituents were found in the various milk thistle tissues.
Typically, ion chromatography in negative mode of silymarin constituents was reported in Supplementary Figs. S6 and S7. Tab.6 shows the relative amounts of silymarin constituents in the six investigated milk thistle tissues. With regards to the flowers from the third development stage belonging to milk thistle chemotype A (FA3), the dominant silymarin constituent was silychristin (86.23 µg g^{-1} FW, fresh weight) followed by silybin (75.75 µg g^{-1} FW), while silydianin was not detected. In chemotype B (FB3), the silymarin constituent present in major amount was silydianin (117.43 µg g^{-1} FW) followed by silybin (48.62 µg g^{-1} FW), while silychristin was not detected. In chemotype C (FC3) the concentrations of the different constituents were consistent with both chemotype A and B with a predominance of silydianin (179.75 µg g^{-1} FW) followed by silychristin (90.43 µg g^{-1} FW) and silybin (59.18 µg g^{-1} FW).
The total content of silymarin (obtained from the sum of the three different silymarin constituents) was present in major amount in the flowers at the third development stage, with a high concentration in the chemotype C, followed by chemotype B and A.
In the other tissues the different silymarin constituents were present in trace or not detected (see Tab. 6).

Table 6. Quantification of different silymarin constituents in different milk thistle tissues.

	Silychristin	Silydianin	Silybin	Total Silymarin
Flowers Stage 1				
Chemotype A	0.07	0.36	0.05	0.48
Chemotype B	n.d.	n.d.	0.0001	0.0001
Chemotype C	0.002	n.d.	0.01	0.012
Flowers Stage 2				
Chemotype A	0.02	0.004	0.03	0.054
Chemotype B	0.01	n.d.	0.02	0.03
Chemotype C	0.02	n.d.	0.02	0.04
Flowers stage 3				
Chemotype A	**86.23**	**n.d.**	**75.75**	**161.98**
Chemotype B	**n.d.**	**117.43**	**48.63**	**166.06**
Chemotype C	**90.43**	**179.75**	**59.18**	**329.36**
Leaves				
Chemotype A	0.002	0.04	0.01	0.052
Chemotype B	0.005	0.02	0.002	0.027
Chemotype C	n.d.	0.002	0.0004	0.0024
Stems				
Chemotype A	n.d.	0.001	n.d.	0.001
Chemotype B	0.04	0.2	0.02	0.26
Chemotype C	n.d.	n.d.	n.d.	n.d.
Roots				
Chemotype A	0.003	0.008	n.d.	0.011
Chemotype B	0.002	0.01	0.001	0.013
Chemotype C	n.d.	n.d.	n.d.	n.d.

Total Silymarin is reported as the sum of silychristin, silydianin and silybin.

The unit was µg g $^{-1}$ FW.

n.d., not detected.

4.3. Metabolomic profiling of terpenes in different milk thistle tissues

Gas chromatography-quadrupole mass spectrometry (GC-qMS) of the headspace sampled by solid-phase micro extraction (SPME) allowed the identification and quantification of some terpenes in leaves, stems, roots and flowers from three different development stages of three diverse milk thistle chemotypes.

4.3.1. Detection and quantification of terpenes in milk thistle tissues

Generally speaking, the terpenes compounds identified in milk thistle tissues were mainly monoterpenes, sesquiterpenes, and terpenes derivates. A total of 17 terpenes were detected in milk thistle tissues (Supplementary Tab. S2). α-Thujene, β-Myrcene, α-Terpinene, β-Phellendrene, β-Ocimene, α-Terpineol, β-Cyclocitral, Caryophyllene, α-Selinene, β-Selinene were identified based on the spectral match with the NIST library and the retention indeces (RIs). Whereas the compounds of greatest interest, α-Pinene, α-Phellandrene, p-Cymene, D-Limonene, α-Copaene, β-Elemene and Valencene, were identified using commercial standards (see Paragraph 3.4.1. and Supplementary Fig. S3) as well as based on the spectral match with the NIST library and the retention indeces (RIs) (Supplementary Tab. S2). These volatile terpene compounds showed differing emission profiles in the different milk thistle tissues and sometimes even in the three biological replicates analyzed for each tissue (as shown in Tab. 7 and Supplementary Tab. S2 and Fig. S4). Typical MS spectra of some terpenes and chromatograms of some samples were reported in Supplementary Figs. S4 and S5. D-Limonene was the most abundant terpene compound detected in milk thistle tissues. D-Limonene was present in major amount in the stems (98.62 μg g^{-1} FW) and flowers at the second (95.31 μg g^{-1} FW) and third developmental stage (118.15 μg g^{-1} FW) (Tab. 7) than in other tissues. α-Pinene could be detected in all the tissues with a higher concentration in the flowers from different development stages (respectively 3.30 μg g^{-1}, 3.02 μg g^{-1}, 3.48 μg g^{-1} FW) and in the stems (2.33 μg g^{-1} FW). β-Elemene was found only in the roots (8.20 μg g^{-1} FW) and in trace in other tissues. Valencene was detected in the roots (0.67 μg g^{-1} FW) and was present in trace in leaves, stems and flowers at the second development stage. It was not detected in flowers at the first and second development stages. Indole, an aromatic heterocyclic organic compounds, was not detected or detected in trace in milk thistle tissues.

Table 7. Quantification of terpenes released from different milk thistle tissues.

Compounds	Flowers Stage 1	Flowers Stage 2	Flowers Stage 3	Leaves	Stems	Roots
α-Pinene	3.30 ±2.40	3.02 ±0.81	3.48 ±2.53	1.84 ±0.11	2.33 ±0.87	1.04 ±0.76
D-Limonene	77.28 ±40.21	95.31 ±46.99	118.15 ±91.38	60.03 ±9.61	98.62 ±37.76	83.77 ±71.19
β-Elemene	0.01 ±0.02	0.10 ±0.09	0.02 ±0.01	0.03 ±0.01	0.23 ±0.34	8.20 ±13.86
Valencene	n.d.	0.02 ±0.03	n.d.	0.03 ±0.03	0.02 ±0.04	0.67 ±0.74

Data were the means ±SD of three biologica replicates.
The unit was μg g^{-1} FW (fresh weight).
n.d., not detected.

4.3.2. Principal component analysis of compounds identified by GC-MS

In order to enable a simple visual interpretation of all untargeted metabolites (VOCs) found in different milk thistle tissues by Metalign-Galaxy workflow method, a principal component analysis (PCA) was performed and showed in Fig. 27.

The huge difference between the roots and other tissues didn't allow us to see the variability between flowers from three different stages, leaves and stems, for this reason we decide to process both the PCA with all the samples (Fig. 27A) and the PCA with all the samples except the roots (Fig. 27B). T1 and T2 explained the major percentage in the variability of milk thistle tissues (respectively 38.3% in Fig. 27A and 37.7% in Fig. 27B). Considering all the samples (Fig. 27A) flowers from three different development stages, leaves and stems were clearly separated based on metabolites (VOCs) from the roots (that grow underground), in fact two different clusters were identified (highlighted with a black circle). With regards to all the samples except the root (Fig. 27B), two cluster were also identified: the first cluster includes flowers at the third and second development stage and one flowers at the first development stage; the second cluster contains leaves, stems and two flowers at the first development stage (highlighted with a black circle).

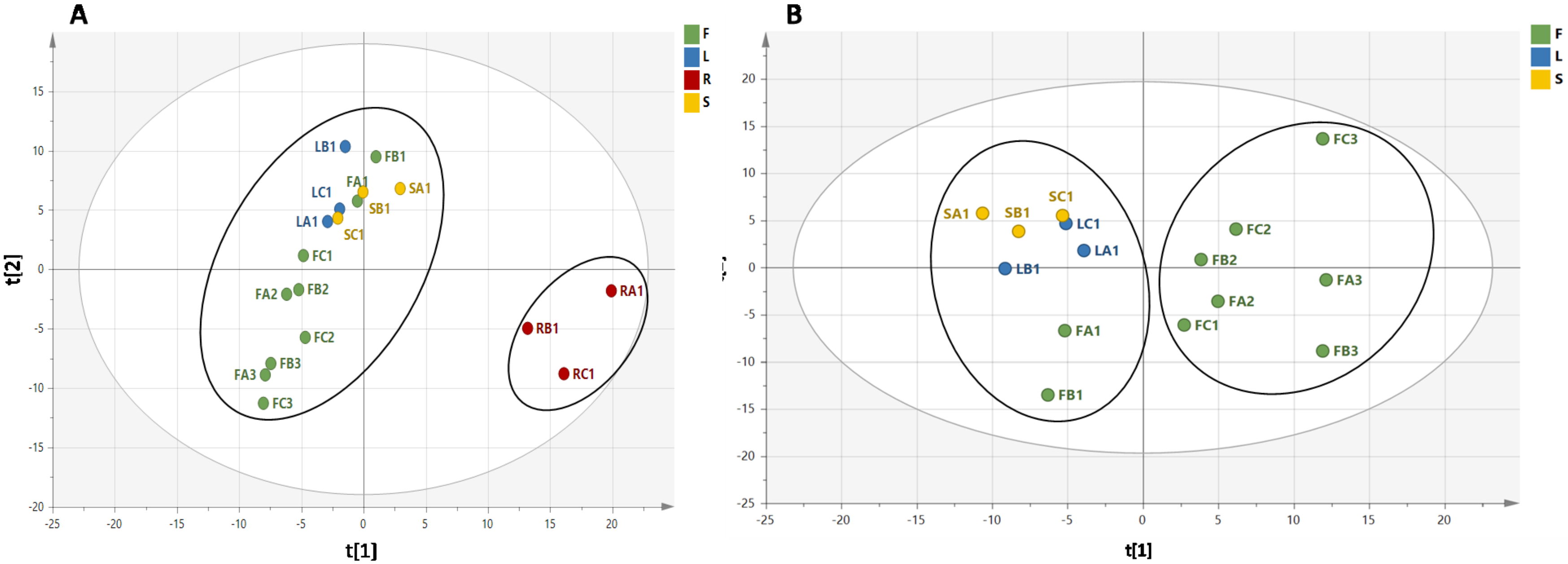


Figure 27. Principal component analysis (PCA) of six milk thistle tissues: flowers (Green); leaves (Blue); stems (Yellow); roots (Red) of 299 metabolites (VOCs). In the flowers the numbers 1, 2 and 3 indicate respectively the first, second and third flowers development stage. **A.** PCA of all the milk thistle tissues **B.** PCA of all the milk thistle tissue except the roots.

4.4. Functional genetics

Based on the terpenes profile obtained from the different tissues of milk thistle, we investigated the genes involved in the biosynthesis of these bioactive compounds.

4.4.1. RNA isolation

Total RNA was extracted from leaves, stems, roots and flowers from three different development stages of milk thistle (see Fig. 14). The extracts were analyzed by capillary electrophoresis (Fig. 28) to visually assess the quality of RNA.

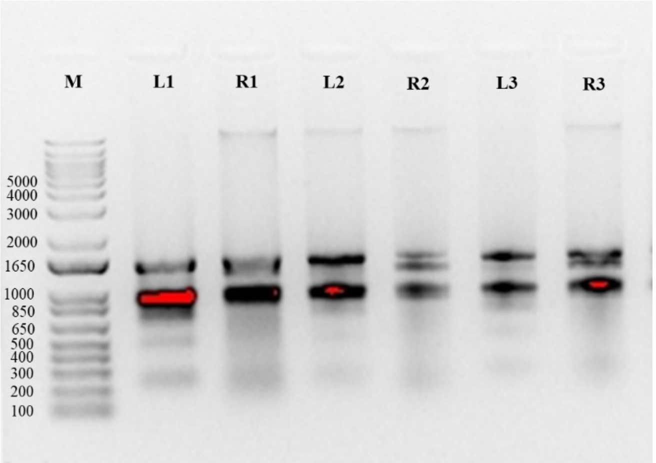


Figure 28. Gel view of RNA runs of leaves (L1, L2, L3), and roots (R1, R2, R3) from different milk thistle plants. Marker (M) used is 1 kb plus DNA Ladder (Invitrogen).

4.4.2. Identification and isolation of terpene synthase genes

The clones of pACYCDuet-1 plasmid were generated by the ligation of the PCR product of linearized pACYCDuet-1 and *SmTPSs* cDNA at 50 °C for 4 h (Fig. 22). Bacterial clones were obtained and screened using two pairs of primers which target both ends of pACYCDuet-1 plasmid. Several clones were identified as PCR positive (Fig. 29).

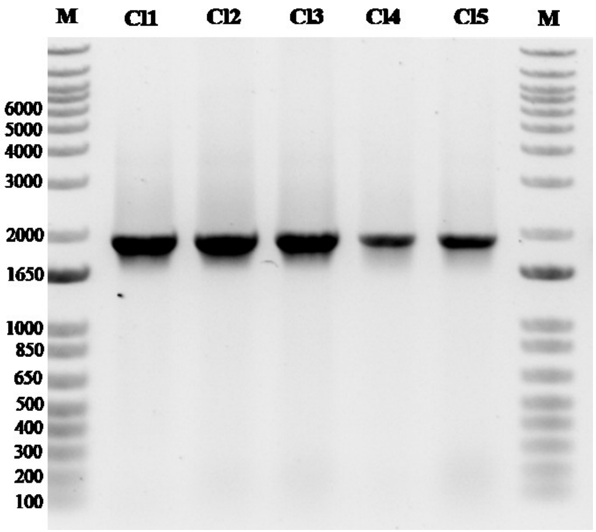


Figure 29. Identification of *SmTPS2* clones by PCR. PCR screening with detection primers (Paragraph 3.5.5.2.). Marker (M) used is 1 kb plus DNA Ladder (Invitrogen).

Three positive clones were then screened using primers targeting both ends of the pACYCDuet-1 plasmid resulted in the isolation of four full-length (*SmTPS1-SmTPS4*) and one not full-length (*SmTPS5*) *S. marianum* sesquiterpene synthase cDNAs, based on putative *SmTPSs* genes were amplified. As shown in Tabs. 4 and 5, *SmTPS1-TSmTPS5* ORF sequences had a length of about 1700 bp. As shown in Supplementary Fig. S1, *SmTPS1-SmTPS4* and *SmTPS5* ORF sequences encoded about 560 and 329 deduced amino acids, respectively, and showed high sequence identities with TPS proteins from other species of the Asteraceae (Supplementary Tab. S1). In fact, a Blast search in GenBank revealed that SmTPS1 has high homology with (E)-beta-farnesene synthase-like and terpene synthase, metal-binding domain-containing protein from *Cynara cardunculus* var. *scolymus* (highest identity, 92.55 %), (E)-beta-farnesene synthase-like from *Lactuca sativa* (identity 81.34 %), putative terpenoid cyclase/protein prenyltransferase alpha-alpha toroid from *Helianthus annus* (identity 77.05%), (E)-beta-farnesene synthase from *Chrysanthenum indicum* (identity 77.31%) and (E)-beta-farnesene synthase-like from *Helianthus annus* (identity 77.05 %); while SmTPS2 showed high homology with terpene synthase, metal-binding domain-containing protein from *Cynara cardunculus* var. *scolymus* and germacrene A synthase short form from *Lactuca sativa*, *Cichorium intybus* and germacrene A synthase from

Crepidiastrum sonchifolium and *Taraxacum officinale* (identity 100%) (Supplementary Tab. S1).

4.4.3. Biochemical characterization of the enzyme encoded by *SmTPS* genes

The main volatile terpenes in most flowers are monoterpenes and sesquiterpenes, synthesize by pathways catalyzed by TPS proteins using GPP or FPP as substrate, respectively. To further confirm the enzymatic properties of the SmTPS proteins and their dominant roles in the biosynthesis of terpenes in milk thistle, substrate specificity analyses were conducted using both GPP and FPP. To prepare recombinant proteins for the biochemical analysis, the four full-length *SmTPS* genes were expressed in *E. coli*. SmTPSs were induced into the supernatant and the crude protein extracts were utilized in the assays of enzymatic activity (Fig. 30), the crude protein extracts from the *E. coli* expression system containing empty vector were used as controls in the biochemical assays.

As shown in Fig. 30, upon incubation with FPP as a substrate, both SmTPS1 and SmTPS2 were confirmed to be versatile enzymes, respectively, with multiple and single products. SmTPS1 converted FPP in farnesane, amorpha-4,11-diene, sesquisabinene hydrate trans (Sesquiterpenoide), and β-Curcumene, zingiberene, β-sesquiphellandrene like the mainly catalyzed products; whereas SmTPS2 converted FPP in Germacrene A, the precursor of β-elemene. In contrast, no sesquiterpenes were detected in assays using SmTPS3 and SmTPS4 as enzymes. Catalytic activity analysis of the four SmTPS proteins was also performed using GPP as substrate. Results showed that all SmTPSs did not have the ability to synthesize monoterpenes.

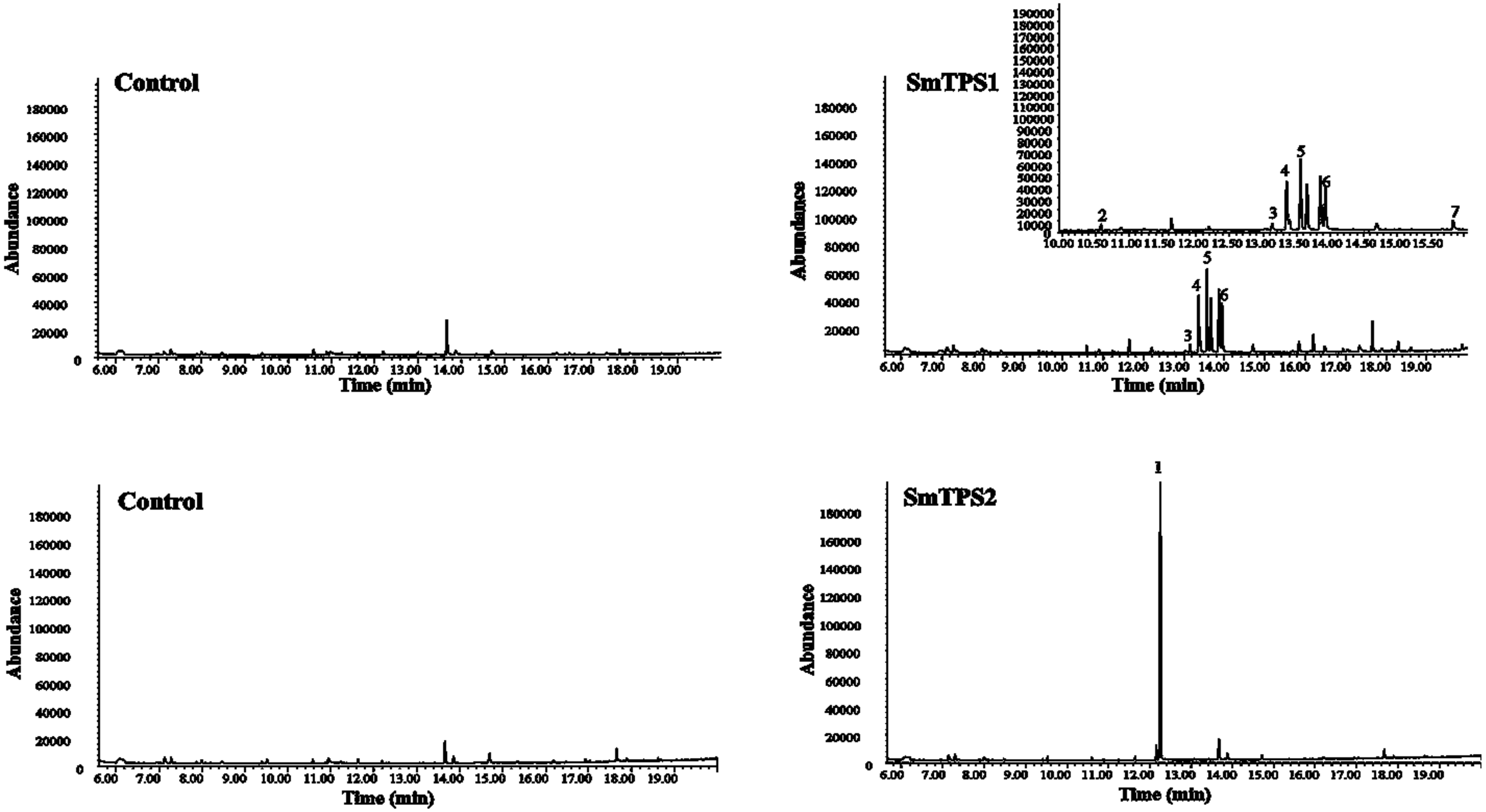


Figure 30. *In vitro* enzymatic activity analysis of SmTPS proteins using FPP as substrate. GC-MS chromatograms show the following terpenes peaks: 1. β-elemene; 2. farnesane; 3. amorpha-4,11 diene; 4. β-curcumene; 5. zingiberene; 6. β-sesquiphellandrene; 7. sesquisabinene hydrate trans.

4.4.4. Sequence and phylogenetic analysis of SmTPS proteins

Sequence alignment revealed that all the SmTPS proteins contained the conserved DDxxD and NSE/DTE region, that are essential for the binding of Mg^{2+} or Mn^{2+} cofactors to catalyze terpene biosynthesis (Supplementary Fig. S1). Moreover, SmTPS4 showed some amino acid residues changes in the NSE/DTE region and a deletion immediately after the DDxxD domain. Furthermore, the SmTPS proteins also shared another conserved motif, RRX_8W, which is usually found in TPSs catalyzing the cyclization of monoterpenes and RxR, which has been shown to become ordered upon FPP binding. SmTPS1 showed the largest number of residues changed in the RRX_8W domain. The biosynthesis of monoterpenes and sesquiterpenes is thought to be compartmentalized, with monoterpenes produced in the plastids, where GPP is synthesized, and sesquiterpenes formed in the cytosol, where FPP is generated (Chen *et al*., 2011; Dudareva *et al*., 2013). The three-dimensional structures of proteins obtained with Modeller (Eswar *et al*., 2007) software showed that all the SmTPS proteins had a transient peptide (cTP or chloroplast transit peptides) positioned upstream of the RRX8W motif and therefore had a high probability of localizing in the plastid (Fig. 31). This is in contrast with the subcellular localization analysis done with the online ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/) and Predotar (https://urgi.versailles.inra.fr/Tools/Predotar) software, which showed that SmTPS proteins have a cellular distribution (Tabs. 8 and 9) and with the residue changes of SmTPS1 in the RRX8W motif in SmTps1 (Supplementary Fig. S1).

Table 8. Results of ClhoroP subcellular localization analysis.

Sequence	Score	cTP	CS-score	cTP-lenght
SmTPS1	0.433	-	-0.250	11
SmTPS2	0.449	-	1.566	17
SmTPS3	0.473	-	1.267	2
SmTPS4	0.460	-	-0.825	5

Table 9. Results of Predotar subcellular localization analysis. ER, endoplasmatic reticulum.

Sequence	Mitochondrial	Plastid	ER	Elsewhere	Prediction
SmTPS1	0,01	0,00	0,00	0,99	none
SmTPS2	0,01	0,00	0,00	0,99	none
SmTPS3	0,01	0,04	0,00	0,96	none
SmTPS4	0,01	0,01	0,00	0,99	none

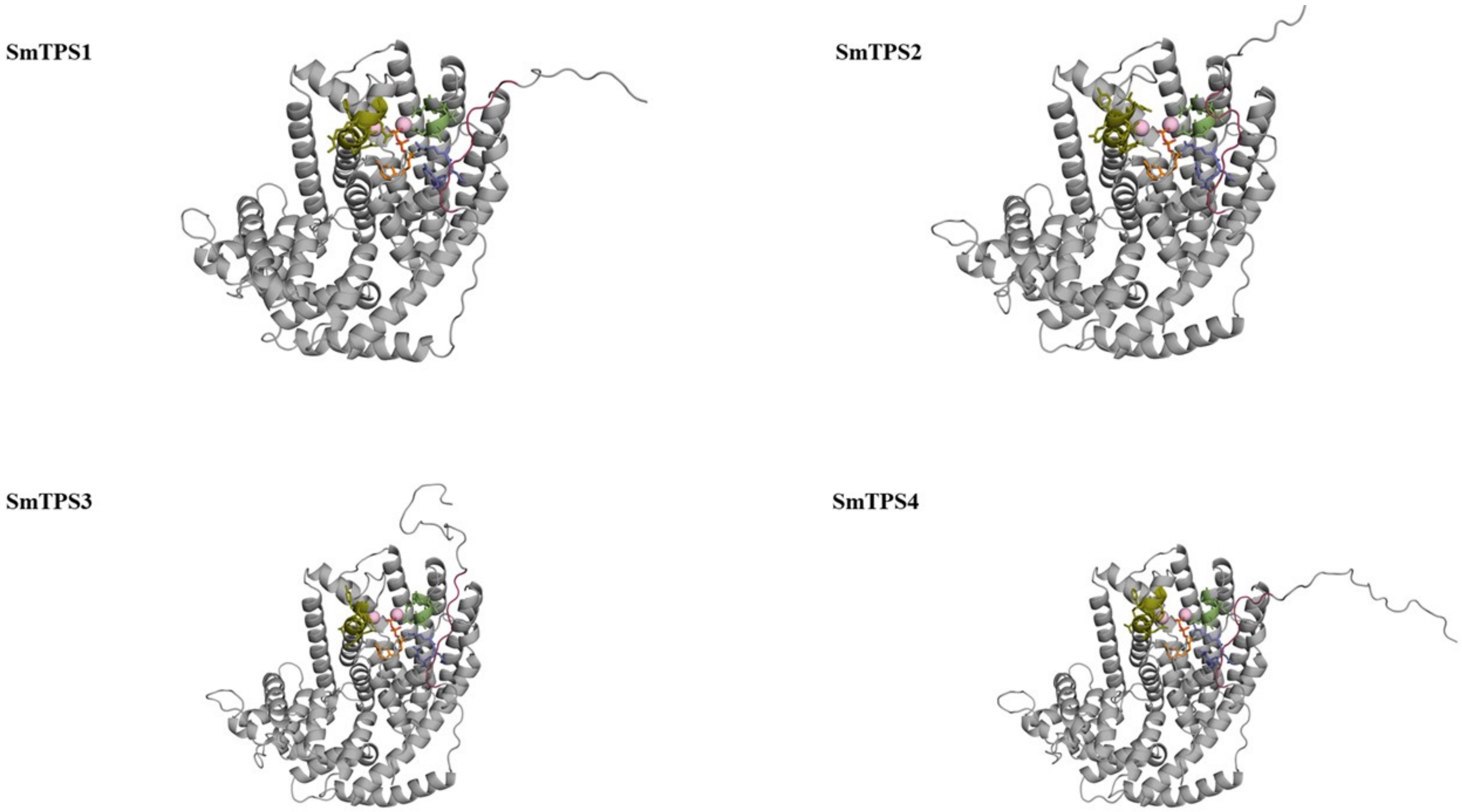


Figure 31. Three-dimensional structure predicted model of SmTPSs. Known TPS motifs RxR (blue), DDxxD (light green) and NSE/DTE (dark green) shown on the structure of SmTPS. The N- and C-terminal domains are in grey while the *hypothetical* target/signal peptide that's used for localisation is in red. Pink spheres represent Mg^{2+} ions. A substrate analog, farnesylhydroxyphosphonate (FHP) is in orange.

To further clarify the potential roles of the five SmTPS proteins, a phylogenetic tree was generated (Fig. 32).

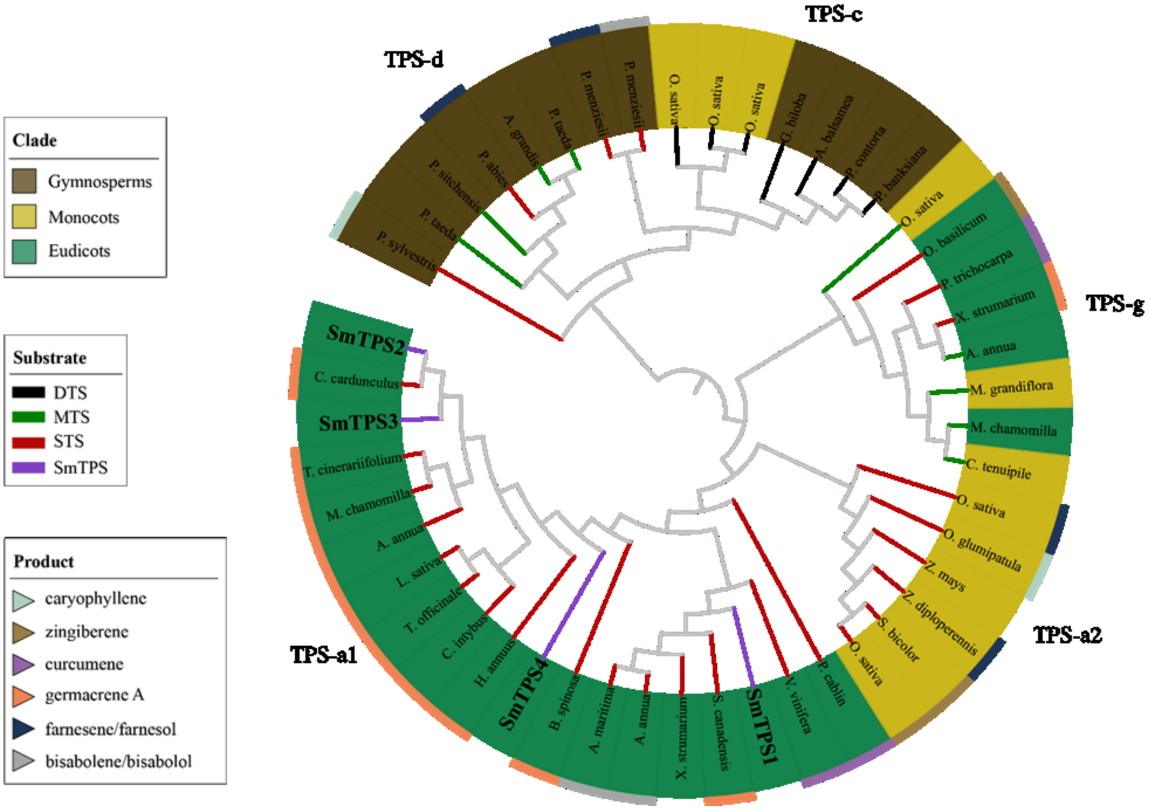


Figure 32. Results of phylogenetic analysis of TPS proteins from *S. marianum* (SmTPS1–SmTPS4) and other plants (Supplementary Tab. S3). The TPS-a1, TPS-a2, TPS-c, TPS-d, and TPS-g clades and plant species are reported. *S. marianum* TPS proteins identified in this study are highlighted in bold.

The results showed that TPS proteins from various species (Supplementary Tab. S3) were clearly classified into five different clades, including clades TPS-c (most conserved among land plants), TPS-d (gymnosperm specific), and three angiosperm-specific clades, TPS-g, and TPS-a (Fig. 32). The TPS-a clade was further divided into a dicot-specific subclade (eudicots) (TPS-a1) and a monocot specific subclade (TPS-a2). All the SmTPS proteins identified in the present study clustered into angiosperm-specific clades. In particular, SmTPS1-SmTPS4 clustered into the TPS-a eudicot subclade together with other TPS proteins from eudicot plant species.

4.4.5. Comparison between RT-PCR and RNASeq data

The RT-PCR (Reverse Transcription-Polymerase Chain Reaction) results showed a difference in expression of *SmTPS* candidate genes between the various milk thistle tissues and also based on the grown conditions used (field-grown or laboratory-grown conditions) (Paragraph 3.1.). To confirm our data obtained from RT-PCR, we compared them with the RNASeq data currently available on line (https://www.ncbi.nlm.nih.gov/sra/?term=silybum+marianum).

Through our comparison, we were able to validate 7 out of the 11 terpene synthase candidate genes amplified by RT-PCR (Tab. 10 and Fig. 33). As shown in Tab. 10, *SmTPS1* was amplified only from flowers at the third stage, while the RNASeq data showed an expression in all the milk thistle tissues. *SmTPS2* was amplified from the flowers at the first and third stage, in leaves, stems, roots and young plant (leaves and roots of plant grown in MS medium), a similar result was shown in the RNASeq data. *SmTPS3* was amplified only in roots and in young plant, while the RNASeq data also showed an expression in other tissues (Tab. 10 and Fig. 33). *SmTPS4* were amplified in all the milk thistle tissues, a similar result was obtained from RNASeq data. *SmTPS5* was amplified only in the flowers, RNASeq showed an expression in shoot+flower. *SmTPS9* was expressed only in tissues of plants grown in laboratory conditions (light band on agarose gel), RNASeq data showed an expression in all the tissues excepted shoot+flower. *SmTPS10* was expressed in all the milk thistle tissues except the flowers at the third development stage, regarding the RNASeq data the expression was very low and not detectable.

Table 10. RT-PCR results.

Candidate genes	Flowers sage 1	Flowers stage 2	Flowers stage 3	Leaves	Stems	Roots	Young plant
SmTPS1			*				
SmTPS2	*		*	*	*	*	**
SmTPS3						**	**
SmTPS4	*	*	*	*	*	*	**
SmTPS5	*	*	*				
SmTPS9				**		**	**
SmTPS10	*	*		*	*	**	**

Youg plant, plant one month from sowing.

*Gene amplified.

**Gene amplified only in plant grown in laboratory conditions.

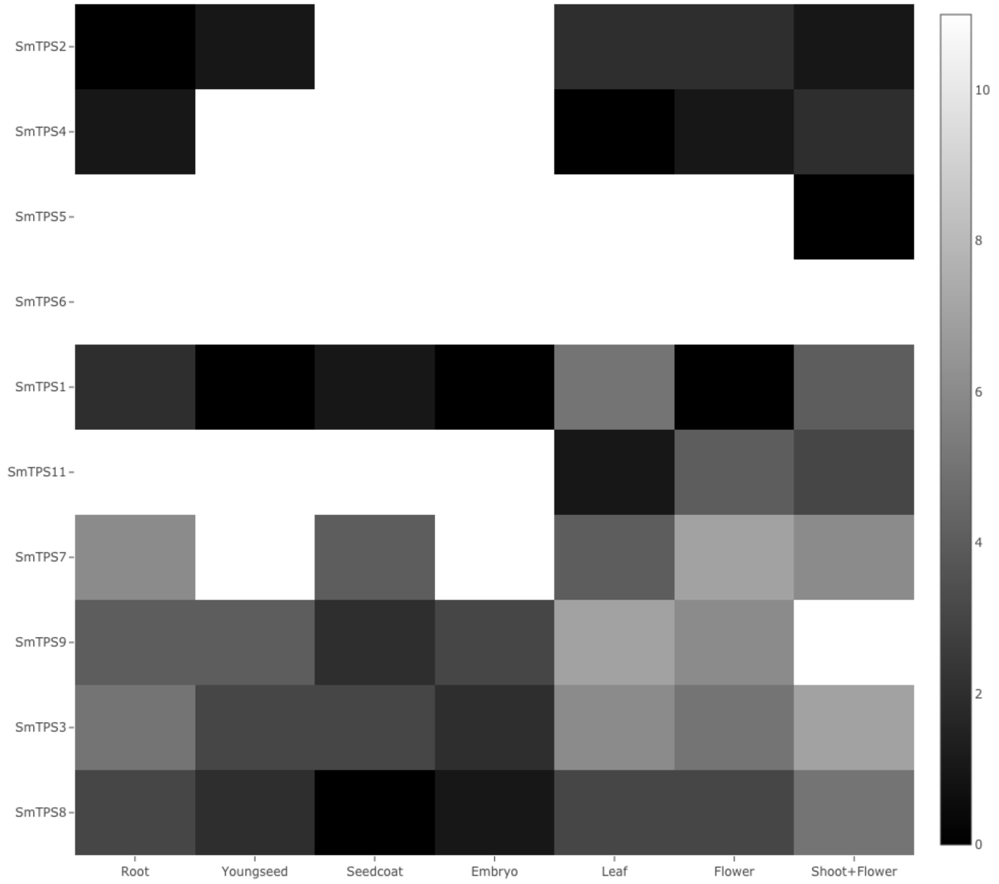


Figure 33. Heatmap of a selection of 10 of the candidate genes, used in this study, expressed by tissue, which shows the expression level of some milk thistle tissues (from on-line RNA-seq data). The number on the scale bar represents the rank of that gene in that tissue. Saturated black: most expressed, White: least expressed (Paragraph 3.5.3.).

5. Discussion

Ecophysiological and metabolomic studies allowed to identify and quantify the different silymarin constituents (silybin, silychristin and silydianin) in milk thistle tissues. Metabolomic analyses, integrated with bioinformatic and genetic methodologies, permitted to identify and quantify terpenes in milk thistle tissues and to characterize the genes involved in the biosynthesis of these bioactive compounds.

5.1. Germination of milk thistle seeds

Ecophysiological studies on germination milk thistle seeds allowed to confirm the correlation between the germination temperature and the afterripening requirements. Generally, the higher the incubation temperature during germination, the longer the afterripening requirement (up to a maximum of five months) (Young *et al*., 1978).
Our results were consistent with the data of Young *et al.* (1978), which showed that at 25°C, five months were required for maximum germination. In fact, in our study, seven months after harvesting seeds had a high percentage of germination and we also observed an increase in the percentage of germination in seeds sown eight months after harvesting compared to those sown seven months after harvesting.

5.2. Metabolomic profiling of silymarin in milk thistle tissues

The LC-MS analysis allowed to identify three different silymarin chemotypes and to detect for the first time silymarin not only in the seed extracts (Yongkun *et al*., 2017; Giuliani *et al*., 2018) and *in vitro* cultures (Poppe and Peterson, 2016) but also in flowers and sometimes in other milk thistle tissues, even if in traces (Tab. 6). These results were not consistent with the data reported by Yongkun *et al.* (2017), that detected silybin only in seeds. As a whole, the difference in the amounts of silymarin constituents in the flowers at the third stage of chemotypes A and B was consistent to the results reported by Giuliani *et al*. (2018), that showed a higher concentration of silychristin in the chemotype A and of silydianin in the chemotype B. In our results the chemotype C showed high silydianin content exhibiting a silymarin profile that matched with chemotype B, while regarding the content of silychristin and silybin it showed a silymarin profile that matched with chemotype A identified in this study (Tab. 6). This difference of the silymarin composition in the different chemotypes can be explain by the presence of regulating factors that during the biosynthesis or the transport into the apoplastic space lead to a variation in the amounts of the different isomeric flavonolignans (Poppe and Petersen, 2016). Moreover, the difference in the amounts of silymarin in the flowers from three different development stages was consistent to the results reported by Carrier *et al.* (2002), which showed a high concentration of silymarin constituents in the flowers at the last stage of development compared to the other flower development stages. In fact, the ripe seeds (Fig. 14 and Tab. 3) are present in this flower development stage, where silymarin has been principally detected by previous studies (Cappelletti and Caniato, 1984; Giuliani *et al.*, 2018). The total content of silymarin was present in major amount in the flowers at the third development stage, with a high concentration in the chemotype C, followed by chemotype B and A (Tab. 6). This result showed a correlation between the synthesis of silymarin and the three different silymarin constituents.

In our study, the amount of silybin in leaves, stems and roots was low or not detected, these results were in accordance with prior investigations reported by Yongkun *et al.* (2017), where the silybin was not detected in the milk thistle tissues (Tab. 6).

5.3. Metabolomic profiling of terpenes in milk thistle tissues

The GC-MS analysis permitted to identify a wide range of terpenes, including 9 monoterpenes ($C_{10}H_{16}$ and $C_{10}H_{14}$), 6 sesquiterpenes ($C_{15}H_{24}$), and 2 terpenes derivatives ($C_{10}H_{18}O$ and $C_{10}H_{16}O$), for the first time not only in seeds extract, but also in other milk thistle tissues (Supplementary Tab. S2).

D-Limonene was the predominant compound detected in milk thistle tissues (Tab. 7). It was mainly released from stems and flowers at the second and third developmental stage (Tab. 7), as well as smaller amounts from the roots, leaves and flowers at the first stage (Tab. 7). Moreover, it showed an increase in concentration from the first to the third flower development stage. Limonene has been found in essential oil of milk thistle seeds (Mhamdi *et al*., 2016), indicating that it has a role in this plant. It has also been identified as volatile terpene in other plants, such as *Cynara scolymus* L. (MacLeod *et al*., 1982; Nassar *et al*., 2013), *Matricaria chamomilla* (Heuskin *et al*., 2009) etc., using scent to attract pollinators, and it contributes to the sweet fragrance noted by humans.

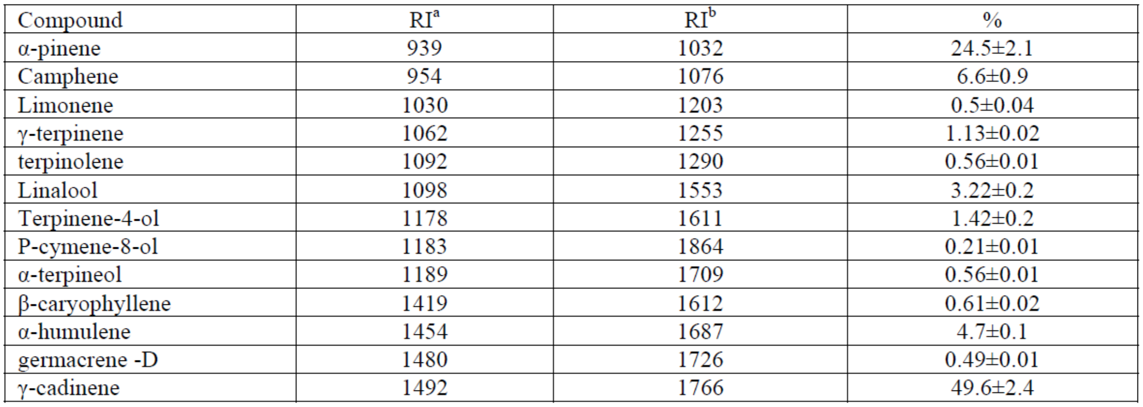
Many plant species have evolved highly specific biosynthetic pathways to produce a vast number of terpenoids. These terpenoids seemingly have evolved for many functions, such as to increase the resistance against microorganisms or insects, to improve the attraction of pollinators or seed-dispersing organisms (flower fragrance, fruit flavor), or to improve the fragrance or flavor of commercial products (under breeding selective pressure). Several plant species have also evolved biosynthetic pathways geared for the synthesis of limonene and a collection of different limonene-derived compounds (Seigler 1998; Chapman and Hall, 2002; Duez *et al*, 2003). Probably for this reason, we found a different distribution of limonene in the same tissues of different milk thistle chemotype.

α-Pinene was another major monoterpene detected in milk thistle flowers, while was released in smaller amounts from stems, leaves and roots.

Unlike the situation with the generally present D-limonene and α-Pinene, other volatile components were found to be differentially released in the milk thistle tissues; whereas the sesquiterpenes β-Elemene and Valencene, as major components, were emitted from roots and were only in trace or not detected in other tissues (Tab. 7). Valencene was found in *Cynara scolymus* tissues (MacLeod *et al*., 1982). We carried out a selection of all the terpenes identified in milk thistle tissues, for this reason some volatile compounds have been detected, but not quantified (Supplementary Tab. S2).

α-Copaene was found to be differentially released in the milk thistle tissues, while the other terpenes compounds were detected in all the tissues. A previous study (Mhamdi *et al*., 2016) found some terpenes in the essential oil of milk thistle seeds (Tab. 11); in according with this study, we found some of these compounds (α-Pinene, α-Terpinene, p-Cymene, α-Terpineol, β-Cariophillene and β-Elemene) in the flowers at the third development stage, which contains the ripe seeds. Mhamdi *et al*. (2016) also found in the essential oil of milk thistle germacrene, a precursor of β-elemene (Kraker *et al*., 1998), the sesquiterpenes identified, in this study, in milk thistle roots. This shows a clear correlation between genetic and metabolomic results obtained in this study (see Paragraph 5.4.1).

Table 11. Essential oil composition of milk thistle seeds. RI[a], RI[b]: retention indices calculated using respectively an apolar column (HP-5) and polar column (HPInnowax) (Mhamdi *et al*., 2016).



Compound	RI[a]	RI[b]	%
α-pinene	939	1032	24.5±2.1
Camphene	954	1076	6.6±0.9
Limonene	1030	1203	0.5±0.04
γ-terpinene	1062	1255	1.13±0.02
terpinolene	1092	1290	0.56±0.01
Linalool	1098	1553	3.22±0.2
Terpinene-4-ol	1178	1611	1.42±0.2
P-cymene-8-ol	1183	1864	0.21±0.01
α-terpineol	1189	1709	0.56±0.01
β-caryophyllene	1419	1612	0.61±0.02
α-humulene	1454	1687	4.7±0.1
germacrene -D	1480	1726	0.49±0.01
γ-cadinene	1492	1766	49.6±2.4

PCA data showed an obvious separation based on volatile metabolites between the roots and flowers from the three different development stages, leaves and stems (Fig. 27A) and between the flowers in an advanced development stage and the other tissues (Fig. 27B). Probably these differences were correlated at the different functions of terpenes in the plant. In fact, as mentioned above, sesquiterpenes are synthesized by sesquiterpene synthases and play a variety of ecological roles in higher plants. Many sesquiterpenes are volatile compounds that are commonly emitted from flowers serving as attractants to pollinators (Morse *et al*., 2012), but also as repellents against nectar thieves (Junker and Bluethgen, 2008). In addition, sesquiterpene emissions from leaves of several plant species play important roles in direct and indirect chemical defense against pathogens and herbivores (Schnee *et al*., 2002; Cheng *et al*., 2007; Chappell and Coates, 2010). They can serve both as repellents (Huang *et al*., 2012; Scala *et al*., 2013) or as attractants of herbivore predators and parasitoids (Schnee *et al*., 2002). Sesquiterpenes are also synthesized and accumulated in underground organs like rhizomes and roots (De Kraker *et al*., 1998; Kovacevic *et al*., 2002; Rasmann *et al*., 2005) where they participate in attracting nematode predators (Rasmann *et al*., 2005;

Pazouki *et al*., 2015). In according with these previous studies, β-elemene found in the milk thistle roots could play a defense role against nematodes and other harmful organisms.

5.4. Functional genetics

5.4.1. Biochemical characterization of the enzyme encoded by *SmTPS* genes

Biochemical characterization of the enzyme encoded by *SmTPS* genes allowed, for the first time, to identify terpene synthase proteins in milk thistle.

One SmTPS protein, SmTPS1 was verified to have the ability to catalyze the formation of multiple volatile sesquiterpenes, typically present in ginger oil (Fig. 30). In fact, TPSs are responsible for producing the huge diversity of terpenes formed by plants (McGarvey and Croteau, 1995). Many TPSs have been found to be capable of producing multiple terpenes from a single prenyl diphosphate substrate *in vitro* (Martin *et al*., 2010; Green *et al*., 2012; Nieuwenhuizen *et al*., 2013) or *in vivo* (Davidovich-Rikanati *et al*., 2008; Green *et al*., 2012), and the profile of terpenes generated by a given species usually comprises one or two compounds that dominate as major products with others as minor components (Tholl *et al*., 2005; Garms *et al*., 2010).

The oil of ginger (*Zingiber officinale* Roscoe) rhizome is an important raw material for the food and pharmaceutical industries. It contains several active ingredients such as zingiberene and curcumene, which have also been synthesized by SmTPS1. The sesquiterpene hydrocarbon zingiberene, [5-(1,5-dimethyl-4-hexenyl)-2-methyl-1,3-cyclohexadiene] is associated with resistance to the Colorado potato beetle (*Leptinotarsa decemlineata* Say) (Carter *et al*., 1989) and beet armyworm (*Spodoptera exigua*) (Eigenbrode *et al*., 1996; 1993); Curcumene, a C15 sesquiterpene hydrocarbon in ginger rhizomes, also has insecticidal effects (Agarwal *et al*., 2001; Antonious *et al*., 2003). In according with these studies, the multiple volatile sesquiterpenes (typically present in ginger oil) synthesized by SmTPS1 could have an important defense role as insecticides.

Instead, SmTPS2 was shown to be a single-product enzyme that could covert FPP to Germacrene A. Germacrene A is formed by cyclization of farnesyl diphosphate (FPP). The cyclization reaction is catalyzed by the terpene synthase (+)-germacrene A synthase (GAS). Germacrene A synthases from a number of Asteraceae species have been described, including chicory (Bouwmeester *et al*., 2002), lettuce (Bennett *et al*., 2002), goldenrod (Prosser *et al*., 2002), *Ixeris dentata* (Kim *et al*., 2005), *Artemisia annua* (Bertea *et al*., 2006), sunflower (Gopfert *et al*., 2009) and *Cynara cardunculus* (Menin *et al*., 2012). The product of SmTPS2 was Germacrene A, the precursor of β-Elemene. In fact, β-Elemene is not the direct product of a sesquiterpene synthase, but it is transformed from the precursor germacrene A by a one-step molecular rearrangement under heating and/or acidic conditions (Fig. 34). Even at room temperature *in vitro*, the conversion yield can reach 98% or more (Weinheimer *et al*., 1970; Kraker *et al*., 1998; Yating *et al*., 2016).

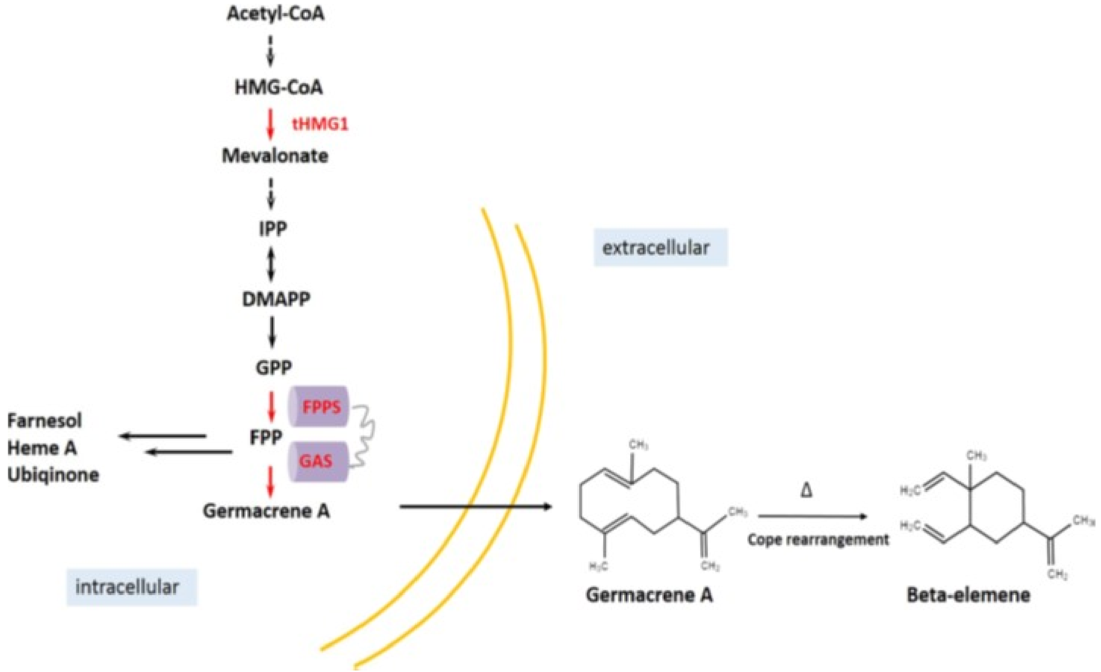


Figure 34. Production of Germacrene A and β-Elemene. Germacrene A is produced through the intracellular MVA pathway, and then transformed to β-Elemene through Cope rearrangement under heat condition *in vitro* (Yating *et al*., 2016).

Sesquiterpenes, including germacrenes, are particularly abundant in the Asteraceae family. In numerous species belonging to Asteraceae, germacrenes perform a central role in the formation of different sesquiterpene derivatives, in particular, sesquiterpene lactones (Adio, 2009; Pazouki *et al*., 2015). Therefore, these results are in agreement with these previous studies.

By means of gas chromatography-quadrupole mass spectrometry (SPME-GC-qMS), β-Elemene was also found mainly in roots, and β-Phellandrene (a typical terpene found in ginger essential oil) (Kamaliroosta *et al*., 2013) was identified in the tissues of milk thistle chemotype studied in this research (respectively in Tab. 7 and Supplementary Table S2). This shows, again, a clear correlation between genetic and metabolomic results obtained in this study.

Therefore, it can be concluded that milk thistle has the capability to synthesize a wide variety of distinct floral scents, derived from the terpenoid biosynthetic pathway.

5.4.2. Sequence and phylogenetic analysis of SmTPS proteins

In the past two decades, terpene synthases have been widely examined in terrestrial plants, and are usually divided into seven clades, designated as TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f, TPS-g, and TPS-h. TPS-a, TPS-b, and TPS-g are recognized as angiosperm-specific clades (Chen *et al*., 2011). The TPS-a clade was further divided into a dicot-specific (eudicots in Fig. 32) subclade and a monocot-specific subclade, this clade is composed of angiosperm-specific sesquiterpene synthases (Hyatt *et al*., 2007; Chen *et al*., 2011).

In this study, SmTPS1-SmTPS4 proteins were clustered into the eudicots-specific TPS clades, TPS-a1 subclade (Fig. 32). As expected, SmTPS1 and SmTPS2 proteins were found to be capable of generating sesquiterpenes using FPP as substrate (Dudareva *et al*., 2003; Gao *et al*., 2018). While SmTPS3, SmTPS4 and SmTPS5 proteins were not functional. In fact, SmTPS3 showed the lack of two amino acid residues in the NSE/DTE region, essential for the binding of Mg^{2+} or Mn^{2+} cofactors to catalyze terpene biosynthesis. While, SmTPS4 showed the lack of one amino acid immediately after the DDXXD motif, also essential for the binding of Mg^{2+} or Mn^{2+} cofactors to catalyze terpene biosynthesis (Gao *et al*., 2018) and a small number of residues changes in NSE/DTE motif. Probably for this reason, they were not functional. Regarding SmTPS5, it was a short protein (Supplementary Fig. S1).

5.4.3. Comparison between RT-PCR and RNASeq data

The comparison of the data obtained in this study from the RT-PCR and the complementary RNASeq data currently available on line (https://www.ncbi.nlm.nih.gov /sra/?term=silybum+marianum) has been useful to find in specific milk thistle tissues the genes involved in the biosynthesis of terpenes.

Our results (Tab. 10 and Fig. 33) showed a similarity between RNASeq data and RT-PCR products for some *SmTPS* genes (*SmTPS2*, *SmTPS4*, *SmTPS5* and *SmTPS10*) and a difference for other *SmTPS* genes (*SmTPS1*, *SmTPS3* and *SmTPS9*). This confirm that the RNA-Seq technique cannot likely replace current RT-PCR methods, but will be complementary (Costa *et al*., 2013). In fact, the results of the RNA-Seq allows the identification of those genes that need to then be examined using RT-PCR, which allows to qualitatively detect gene expression through the creation of complementary DNA (cDNA) transcripts from mRNA, and quantitatively measure the amplification of cDNA by means of fluorescent probes (Bustin, 2000).

The present study carrying out a comparison between RNASeq on-line data and RT-PCR products of TPS genes provided a more comprehensive analysis of milk thistle transcriptome. These data may represent a starting point for further gene expression studies of *SmTPS* genes.

6. Conclusion

The researches carried out in this study were essential to know better: the ecophysiology of milk thistle seeds germination; the silymarin and terpenes metabolomic profiling in different tissues of Sicilian milk thistle; the genes involved in the biosynthesis of terpenes and their respective encoded enzyme and a semiquantitative expression of terpenes genes in Sicilian milk thistle tissues.

The seeds germination results show a correlation between afterripening and percentage of germination as shown in other studies. This confirm the importance of afterripening in the ecophysiology of milk thistle seeds.

Concerning the localization and composition of silymarin constituents in milk thistle tissues, our results show that the composition of the silymarin mixture depends on the milk thistle chemotypes and on the presence of ripe seeds in the sample analyzed. In fact, the major amount of silymarin was found in the flowers at the third development stage, which contain the ripe seeds.

Regarding the terpenes localization in milk thistle tissues, our results show a different distribution of sesquiterpenes, monoterpenes and terpene derivates, a major concentration of D-Limonene in milk thistle tissues and the presence of β-elemene mainly in the roots. These results suggest that the different localization of various terpenes in diverse milk thistle tissues is related to the different functions of this bioactive compounds in plants.

In our study, for the first time, five *SmTPS* genes were isolated and two were found to be functional: SmTps1, a multi-product enzyme, catalyzing the formation of sesquiterpenes, typically present in ginger oil and SmTPS2, a single-product enzyme catalyzing the formation of Germacrene A.

As mentioned above, volatile compounds have important roles in direct and indirect plant defense against herbivores and pathogens, in reproduction by attraction of pollinators and seed disseminators, and in plant thermotolerance. Apart from their importance in plant physiology and ecology, volatile terpenoids are also used as natural flavor and aroma compounds and have beneficial effect on humans as health promoting compounds. In fact, terpenes are also known for their anti-inflammation, anti-carcinogenesis and neuroprotection effects.

Zingiberene (found in oil of ginger) has been shown to have a considerable spectrum of biological activity such as antiviral, antiulcer, and antifertility effects; Curcumene, a C15 sesquiterpene hydrocarbon in ginger rhizomes, also has antiulcer and insecticidal effects and (2)-b-sesquiphellandrene has been recognized as antiulcer active principles in ginger too. Furthermore, germacrene A itself, and in particular its rearrangement product β-elemene, has been shown to possess anticancer activity.

This study not only provides a molecular basis for the production of volatile terpenes in milk thistle tissues, but it also supplies a new promising plant model system to investigate the regulation and evolution of sesquiterpene synthesis. Therefore, these results could lead to further uses of milk thistle not only for new therapeutic purposes, but also as insecticide.

Further studies are needed to understand better: the expression of silymarin in the different chemotypes, the functions of terpenes in the different milk thistle tissues and the expression of genes involved in the biosynthesis of terpenes in this plant.

We hope that this study will also pave the way to understand the evolutionary and environmental significance of the volatile terpenes in *Silybum* species and also in other Asteraceae.

7. Apppendix

Table S1. BLASTp of candidate *S. marianum* terpene synthases (SmTPSs) against all NCBI protein database.

Protein	Predicted Protein	Species name	Max score	Total score	Query cover	E value	Per Ident	Accession number
SmTPS1	(E)-beta-farnesene synthase-like	*Cynara cardunculus* var. scolymus	1068	1068	99%	0.0	92.55%	XP_024994027.1
	Terpene synthase, metal-binding domain-containing protein	*Cynara cardunculus* var. scolymus	1068	1068	99%	0.0	92.55%	KVH96380.1
	E)-beta-farnesene synthase-like	*Cynara cardunculus* var. scolymus	1060	1060	99%	0.0	91.82%	XP_024994078.1
	Terpene synthase, metal-binding domain-containing protein	*Cynara cardunculus* var. scolymus	1042	1042	99%	0.0	89.98%	KVH96394.1
	(E)-beta-farnesene synthase-like	*Lactuca sativa*	944	944	99%	0.0	81.34%	XP_023749309.1
	putative terpenoid cyclases/protein prenyltransferase alpha-alpha toroid	*Helianthus annuus*	893	893	99%	0.0	77.05%	OTF85122.1
	E-B-farnesene synthase	*Chrysanthemum indicum*	892	892	99%	0.0	77.31%	AUJ87600.1
	(E)-beta-farnesene synthase-like	*Helianthus annuus*	892	892	99%	0.0	77.05%	XP_022025245.1
SmTPS2	Terpene synthase, metal-binding domain-containing protein	*Cynara carduncul*u s var. scolymus	958	958	100%	0.0	84.90%	KVH92099.1
	germacrene A synthase short form	*Cynara cardunculus* var. scolymus	957	957	100%	0.0	84.90%	XP_024977692.1
	germacrene A synthase short form	***Lactuca sativa***	910	910	100%	0.0	79.96%	XP_023734561.1
	RecName: Full=Germacrene A synthase short form; Short=CiGASsh	*Cichorium intybus*	909	909	100%	0.0	80.14%	Q8LSC2.1
	germacrene A synthase	*Crepidiastrum sonchifolium*	907	907	100%	0.0	78.73%	ABB00361.1
	germacrene A synthase LTC1	*Lactuca sativa*	907	907	100%	0.0	79.79%	AAM11626.1
	germacrene A synthase	***Taraxacum officinale***	904	904	100%	0.0	79.75%	ALY05868.1
	germacrene A synthase short form-like	*Lactuca sativa*	903	903	100%	0.0	79.96%	XP_023734566.1

Table S2. Identification of terpenes in milk thistle tissues. [a]Compounds identified based on RI, NIST match and references standards. [b]Compounds identified based on RI and NIST match only. Compounds identified respectively in one (*), two (**) or three biological replicates (***). n.d., not detected.

Compounds	Molecular formula	Retention time (minutes)	Flowers stage 1	Flowers stage 2	Flowers stage 3	Leaves	Stems	Roots
α-Thujene[b]	$C_{10}H_{16}$	13.00	***	***	***	***	***	**
α-Pinene[a]	$C_{10}H_{16}$	13.33	***	***	***	***	***	***
β-Myrcene[b]	$C_{10}H_{16}$	15.07	***	***	***	***	***	***
α-Phellandrene[a]	$C_{10}H_{16}$	15.83	***	***	***	***	***	***
α-Terpinene[b]	$C_{10}H_{16}$	16.18	***	***	***	***	***	***
p-Cymene[a]	$C_{10}H_{14}$	16.44	***	***	***	***	***	***
D-Limonene[a]	$C_{10}H_{16}$	16.62	***	***	***	***	***	***
β-Phellandrene[b]	$C_{10}H_{16}$	16.72	***	***	***	***	***	**
β-Ocimene[b]	$C_{10}H_{16}$	17.00	***	***	***	***	***	**
α-Terpineol[b]	$C_{10}H_{18}O$	21.98	***	***	***	***	***	***
β-Cyclocitral[b]	$C_{10}H_{16}O$	22.76	***	***	***	***	***	***
α-Copaene[a]	$C_{15}H_{24}$	27.36	*	***	n.d.	*	*	*
β-Elemene[a]	$C_{15}H_{24}$	27.64	***	***	***	***	***	***
Caryophyllene[b]	$C_{15}H_{24}$	28.67	***	***	*	***	***	***
Valencene[a]	$C_{15}H_{24}$	30.37	n.d.	*	n.d.	**	*	***
β-Selinene[b]	$C_{15}H_{24}$	30.49	***	***	**	***	**	***
α-Selinene[b]	$C_{15}H_{24}$	30.63	**	**	***	***	***	***

Table S3. TPS proteins from other plant species used in phylogenetic analysis.

Protein name	Species name	Protein ID in Uniprot
Caryophyllene/humulene synthase	*Pinus sylvestris*	B4XAK4
(-)-alpha-terpineol synthase, chloroplastic	*Pinus taeda*	Q84KL4
Carene synthase 2, chloroplastic	*Picea sitchensis*	F1CKI8
E,E-alpha-farnesene synthase	*Picea abies*	Q675K8
Terpinolene synthase, chloroplastic	*Abies grandis*	Q9M7D0
(+)-alpha-pinene synthase, chloroplastic	*Pinus taeda*	Q84KL3
(E)-beta-farnesene synthase	*Pseudotsuga menziesii*	Q4QSN5
(E)-gamma-bisabolene synthase	*Pseudotsuga menziesii*	Q4QSN4
Ent-copalyl diphosphate synthase 1	*Oryza sativa*	Q6ET36
Ent-copalyl diphosphate synthase 2	*Oryza sativa*	Q5MQ85
Syn-copalyl diphosphate synthase	*Oryza sativa*	Q6E7D7
Bifunctional levopimaradiene synthase, chloroplastic	*Ginkgo biloba*	Q947C4
Bifunctional abietadiene synthase, chloroplastic	*Abies balsamea*	H8ZM70
Bifunctional levopimaradiene synthase, chloroplastic	*Pinus contorta*	M4HYC6
Bifunctional levopimaradiene synthase, chloroplastic	*Pinus banksiana*	M4HXU6
S-(+)-linalool synthase, chloroplastic	*Oryza sativa*	Q6ZH94
Alpha-zingiberene synthase	*Ocimum basilicum*	Q5SBP4
Terpene synthase	*Populus trichocarpa*	A0A076GAU9
Putative monoterpene synthase	*Xanthium strumarium*	A0A142BX72
R-linalool synthase QH5, chloroplastic	*Artemisia annua*	Q9SPN1
Alpha-terpineol synthase, chloroplastic	*Magnolia grandiflora*	B3TPQ7
(E)-beta-ocimene synthase, chloroplastic	*Matricaria chamomilla*	I6RE61
Geraniol synthase, chloroplastic	*Cinnamomum tenuipile*	Q8GUE4
Farnesol synthase	*Oryza sativa*	A0F0B7
Terpene synthase	*Oryza glumipatula*	A0A076L503
(S)-beta-macrocarpene synthase	*Zea mays*	A0A1Q0XLC1
(E)-beta-farnesene synthase	*Zea diploperennis*	C7E5V9
Zingiberene synthase	*Sorghum bicolor*	C5YHH7
Zingiberene synthase	*Oryza sativa*	C5H3I7
Gamma-curcumene synthase	*Pogostemon cablin*	Q49SP7
Beta-curcumene synthase	*Vitis vinifera*	E5GAG1
Germacrene A synthase	*Solidago canadensis*	Q9AR67
Sesquiterpene synthase	*Xanthium strumarium*	A0A142BX71
Alpha-bisabolol synthase	*Artemisia annua*	M4HZ33
(+)-alpha-bisabolol synthase	*Artemisia maritima*	A0A1L7NYF8
Germacrene A synthase 1	*Barnadesia spinosa*	A0A0U1XNG2
Germacrene A synthase 2	*Helianthus annuus*	Q4U3F7
Germacrene A synthase short form	*Cichorium intybus*	Q8LSC2
Germacrene A synthase	*Taraxacum officinale*	A0A0U4IMN4
Germacrene A synthase LTC1	*Lactuca sativa*	Q8S3A6
Germacrene A synthase	*Artemisia annua*	Q1PDD2
Germacrene A synthase	*Matricaria chamomilla*	I6QSN0
Germacrene A synthase	*Tanacetum cinerariifolium*	R9WSX5
Germacrene A synthase	*Cynara cardunculus*	I1TE91

Figure S1. Conserved residues analysis and subcellular localization of *S. marianum* TPS proteins. Full-length protein sequence alignment of SmTPS proteins (SmTPS1-SmTPS4). Functionally important conserved residues (in order: RRx$_8$W, RxR, DDxxD, and NSE/DTE) are highlighted with a colored red background.

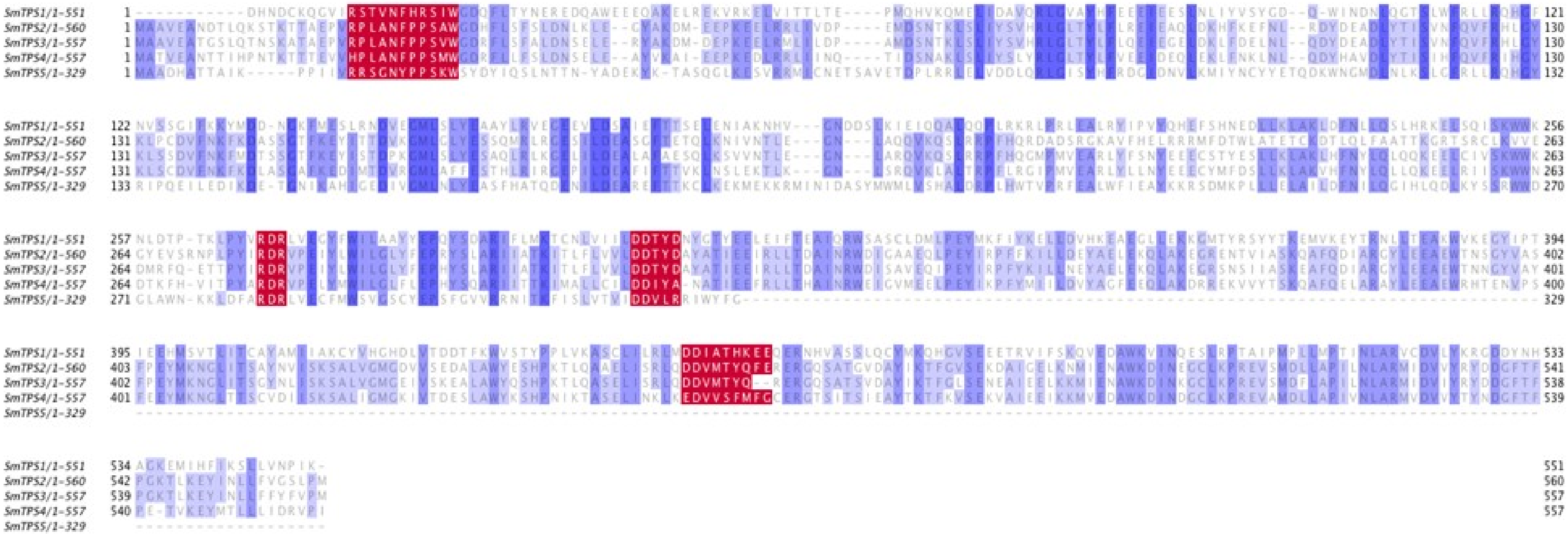


Figure S2. Sequence contigs obtained from pACYCDuet-1 cloning.

SmTPS1_contig

```
GATCATAATGATTGTAAACAAGGGGTAATCCGCAGCACAGTGAACTTCCATCGTAGCATTTGGGGAGATCAGTTTCTC
ACCTACAACGAGAGAGAGGATCAAGCTTGGGAGGAGGAACAAGCAAAAGAACTAAGAGAGAAAGTGAGAAAAGAGCTA
GTGATCACAACTTTAACCGAACCAATGCAACATGTGAAGCAAATGGAACTCATTGATGCGGTTCAACGTCTCGGTGTG
GCTTATCACTTTGAGGAGGAAATTGAAGAATCCTTAAACCTTATCTATGTTTCTTATGGGGATCAATGGATCAACGAT
AACCTCCAAGGCACTTCACTTTGGTTTCGACTCCTACGACAACATGGCTTCAATGTTTCAAGTGGAATATTCAAGAAG
TACATGGACGACAATGGAAAATTTATGGAATCCTTAAGAAACGATGTCGAAGGCATGCTTTCTTTGTACGAAGCAGCA
TACTTGAGGGTGGAAGGAGAAGAAGTTTTAGATTCCGCCATCGAATTTACAACTTCGGAACTTGAAAACATAGCCAAA
AATCATGTAGGTAACGACGATTCTCTTAAGATTGAAATACAACAAGCACTACAACAGCCTCTCCGAAAAAGGTTGCCA
AGGCTTGAGGCGTTGCGCTACATACCGGTCTACCAACATGAATTTTCTCATAATGAGGACTTACTAAAGCTTGCCAAG
TTAGATTTTAACTTGCTCCAATCATTGCACAGAAAAGAGCTTAGCCAAATCAGCAAATGGTGGAAGAATTTGGATACG
CCAACCAAGCTACCATATGTTCGAGATAGATTGGTTGAAGGTTACTTTTGGATATTGGCAGCCTATTACGAACCTCAA
TATTCTGATGCGAGGATCTTTCTCATGAAAACATGCAACCTCGTAATCATATTGGATGACACGTATGATAATTATGGT
ACTTATGAGGAGCTCGAGATCTTTACTGAAGCTATTCAAAGATGGTCTGCAAGCTGCTTGGATATGCTTCCGGAATAC
ATGAAATTCATATATAAAGAACTTTTGGATGTTCACAAAGAAGCTGAGGGGTTACTAGAAAAGAAAGGAATGACATAT
CGTAGCTACTATACTAAAGAGATGGTGAAAGAGTATACTAGAAATCTTCTAACAGAAGCCAAATGGGTAAAAGAGGGT
TATATTCCAACCATAGAGGAACACATGTCTGTAACCCTGATAACTTGTGCCTATGCCATGATCATCGCGAAATGTTAT
GTTCATGGACATGATTTGGTCACAGATGATACCTTTAAGTGGGTGTCCACCTATCCTCCTCTTGTAAAGGCTTCATGT
TTAATTTTAAGGCTCATGGATGATATTGCTACCCATAAGGAGGAACAAGAAAGGAACCATGTAGCTTCTAGCTTACAA
TGCTACATGAAGCAACATGGTGTCTCGGAGGAGGAAACGCGTGTAATATTTTCAAAACAAGTCGAAGATGCGTGGAAA
GTTATAAATCAAGAATCTCTTAGGCCTACTGCTATCCCGATGCCTTTGCTTATGCCTACAATCAACCTTGCCCGTGTC
TGTGACGTACTTTATAAACGTGGCGACGATTACAATCATGCAGGGAAAGAAATGATCCATTTCATCAAGTCACTTCTC
GTAAATCCGATTAAATGA
```

SmTPS2_contig

```
ATGGCTGCCGTAGAAGCTAATGATACCCTCCAAAAAAGCACAAAAACCACTGCTGAACCGGTGCGTCCTTTGGCCAAC
TTTCCTCCTTCGGCATGGGGTGATCACTTCCTATCGTTCTCTCTTGACAATTTGAAATTGGAAGGATATGCCAAAGAC
ATGGAGGAGCCAAAAGAAGAACTGAGAAGATTGATCGTTGATCCAGAAATGGATTCAAATACGAAACTAAGTTTGATT
TATTCTGTACACCGTCTTGGTTTGACGTATCTATTTTTGCGTGAGATTGAGGCACAACTTGACAAACATTTCAAAGAG
TTCAACTTGCGAGATTATGATGAAGCTGATTTGTACACAATCTCCGTTAACTTTCAAGTTTTCCGACACCTTGGTTAC
AAATTGCCTTGTGATGTGTTTAACAAATTCAAGGATGCAAGCTCAGGTACTTTCAAAGAATACATCACCACTGATGTG
AAGGGTATGTTAGGCTTATATGAATCTTCACAAATGAGACTAAGGGGAGAATCTATTTTGGATGAAGCCTCGGGTTTC
ACCGAAACTCAACTTAAAAACATTGTAAACACTCTCGAAGGTAACCTCGCACAGCAAGTGAAACAATCGTTGAGGAGG
CCCTTCCATCaAAGGGATGCCGATAGTCGAGGCAAGGCTGTATTTCACGAACTACGAAGAAGAATGTTCGACACATGG
CTCGCTACTGAAACTTGCAAAGATACACTTCAACTATTTGCAGCTACAACAAAAGGAAGAACTTCGCGTTGTCTCAAA
GTGGTGGAAGGATATGAGGTTTCAAGAAACCcACTcCCATATATAAGAGATAGAGTACCAGAGATTTATTTATGGATA
TTGGGGTTATACTTTGAGCCTCGTTACTCTTTGGCACGAATCATTGCCACCAAAATTACATTGTTTCTTGTGGTGCTA
GATGACACATATGATGCATATGCTACTATTGAAGAGATTCGACTTCTAaCAGACgCCATAAATAGGTGGGATATTGGT
GCTGCAGAGCAACTTCCGGAATATATTAGACCATTCTTCAAAATTCTCCTGGATGAATACGCTGAACTTGAGAAACAA
CTAGCAAAAGAAGGAAGGGAAAATACTGTTATTGCATCAAAACAAGCGTTTCAAGACATAGCTAGAGGTTACCTTGAA
GAGGCTGAATGGACAAACAGTGGATATGTCGCTTCGTTTCCTGAATATATGAAGAATGGGTTGATCACTTCTGCCTAC
AATGTTATTTCAAAATCTGCTTTAGTCGGTATGGGCGATGTAGTTAGTGAAGATGCTTTGGCTTGGTATGAAAGTCAT
CCGAAGACTCTTCAAGCTGCAGAGTTAATTTCAAGACTCCAAGATGATGTTATGACTTATCAGTTCGAGCGTGAAAGA
GGACAATCAGCCACAGGAGTTGATGCATATATCAAGACGTTTGGTGTGTCGGAAAAAGATGCTATTGGTGAGCTCAAG
AATATGATTGAAAATGCATGGAAAGATATAAACGAGGGGTGTCTCAAGCCAAGAGAAGTTTCCATGGATTTGCTTGCC
CCGATTCTTAATCTTGCACGAATGATAGATGTTGTATACAGGTATGACGATGGGTTCACTTTTCCGGGAAAGACCCTT
AAAGAGTATATTAATCTCCTGTTTGTTGGCTCTTTACCAATGTAA
```

SmTPS3_contig

```
ATGGCAGCAGTTGAAGCTACTGGTTCCCTCCAAACGAACTCCAAAGCCACTGCTGAGCCGGTGCGTCCTTTGGCCAAC
TTCCCTCCTTCTGTATGGGGTGATCGCTTTCTATCGTTCGCCCTTGACAATTCTGAATTGGAAAGATATGCTAAAGAT
ATGGATGAGCCAAAAGAAGAATTGAGAATGCTGATTCTTGATCCAGCAATGGATTCAAATACAAAGCTAAGTTTGATT
TATTCTGTACACCGTCTTGGTTTGACATATCTGTTCTTGCAAGAGATTGAAGGCGAGCTTGACAAGCTTTTCGACGAG
CTAAACTTGCAAGATTACGACGAGGCTGATCTTTACACGATTTCGGTTAACTTTCAAGTTTTCCGACACCTTGGTTAC
AAACTGTCTTCTGACGTGTTTAACAAATTCATGGACACTAGCTCGGGTACATTCAAGGAATACATTTCCACCGATCCG
AAGGGTATGTTAAGTTTATACGAATCTGCACAGTTGAGATTAAAAGGAGAATTAATTTTAGATGAAGCATTGGCATTC
GCTGAAAGTCAACTCAAGAGTGTCGTAAACACTCTAGAAGGCAATCTCGCACGGCAGGTGAAACAATCGTTGAGGAGA
CCATTCCATCAAGGGATGCCAATGGTgGAAGCAAGGCTATATTTCTCCAACTATGAAGAAGAATGTTCAACATATGAG
TCACTACTAAAGCTTGCAAAGTTGCATTTTAACTATTTGCAGCTACAACAAAAGGAAGAACTTTGCATTGTTTCAAAG
TGGTGGAAGGATATGAGGTTTCAAGAAACTACTCCTTACATAAGGGATAGAGTTCCAGAGATTTACTTATGGATACTG
GGATTATATTTCGAGCCTCATTACTCATTGGCACGAATCATCGCCACCAAAaTTACATTGTTTCTTGTGGTGCTAGAT
GACACGTATGATGCGTATGCTACCATTGAAGAGATTCGGCTTCTAACAGATGCCATAAATAGGTGGGACATTAGTGCT
GTGGAGCAAATTCCAGAATATATTAGACCATTCTACAAAATTCTGCTTAATGAGTATGCTGAACTTGAGAAACAACTA
GCTAAAGAAGGAAGATCAAATAGCATTATTGCTTCAAAAGAAGCGTTTCAAGACATAGCTAGAGGCTATCTTGAAGAG
GCTGAATGGACAAACAATGGATATGTAGCATATTTTCCTGAGTATATGAAGAACGGGTTAATCACTTCTGGCTACAAT
CTTATTTCAAAATCTGCTTTAGTAGGTATGGGGGAGATAGTTAGTAAAGAAGCTTTGGCTTGGTACCAAAGTCATCCA
AAGACTTTGCAAGCTTCAGAGTTAATTTCAAGACTCCAAGATGATGTCATGACATATCAGCGTGAAAGAGGGCAATCA
GCCACCAGTGTTGATGCATATATCAAGACCTTTGGTCTGTCAGAAAATGAAGCAATTGAAGAGCTCAAGAAAATGATT
GAAAATGCATGGAAGGATATAAACAAGGGGTGTCTCAAGCCACGAGAAGTCTCCATGGATTTCCTTGCACCAATTCTC
AATCTTGCACGAATGATAGATGTGATATATAGGTATGACGATGGGTTCACTTTTCCCGGAAAGACCCTCAAAGAGTAT
ATTAATCTCTTGtttttttattttgttcccatgtaa
```

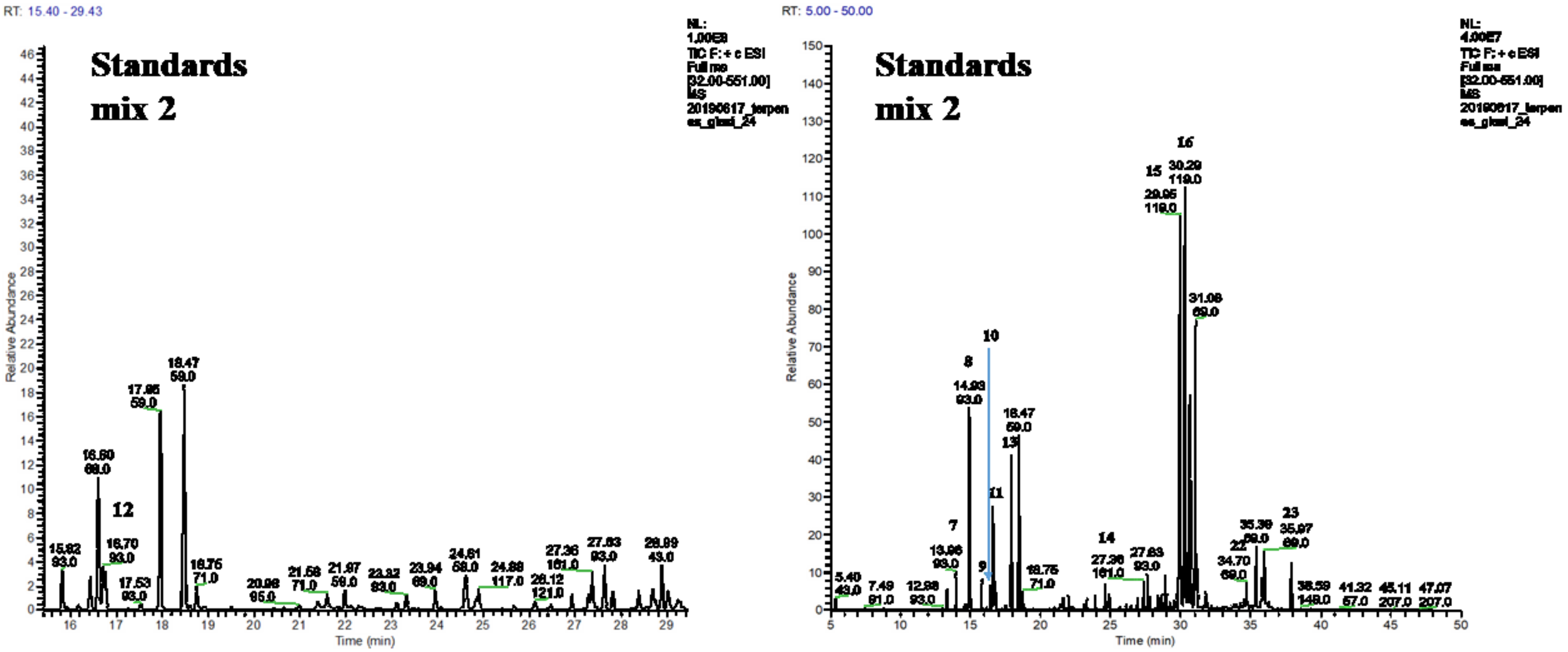
SmTPS4_contig

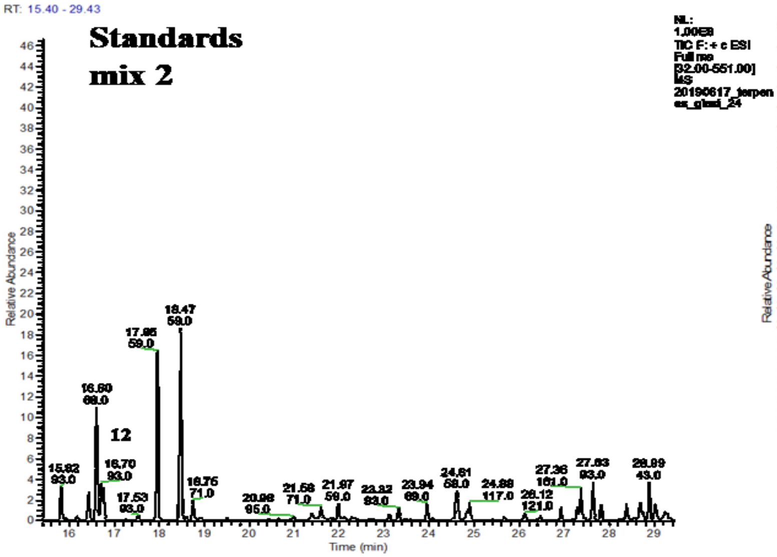
```
ATGGGCAGCAGCCATCACCATCATCACCACAGCCAGATGGCCACCGTTGAAGCCAATACTACCATCCATCCGAACACC
AAAACAACCACAGAGGTGGTGCATCCTTTGGCCAACTTCCCTCCTTCGATGTGGGGTGATCGCTTCTTATTATTCTCT
CTTGACAACTCGGAATTGGAAGCATATGTTAAAGCCATTGAGGAGCCAAAGGAAGATCTAAGAAGATTGATAATCAAC
CAAACTATTGATTCAAATGCAAAATTGAGTTTGATTTACTCTTTATATCGTCTTGGTTTGACCTATCTTTTCGTGGAA
GAGATTGATGAACAGCTTGAAAAACTCTTCAATAAGCTTAACTTGCAAGATTATCATGCTGTTGATCTATATACAATC
TCTATTCACTTTCAAGTTTTCAGAATCCATGGTTACAAATTGTCTTGTGATGTGTTTAACAAGTTCAAGGATTTGGCA
TCCGGTGCATTCAAGGAAGATATTATGACGGATGTGAGGGGTATGCTAGCTTTCTTTGAGTCTACGCATTTGAGGATA
AGAGGAGAGCCTATTTTAGATGAAGCGTTCATATTTACAACTGTTAAACTGAATAGTTTAGAAAAAACTCTCAAAGGA
AATCTTTCAAGGCAAGTGAAACTTGCTTTGACTAGACCATTTCTTCGAGGCATTCCAATGGTAGAGGCAAGGTTATAT
TTACTCAACTATGAAGAAGAATGCTACATGTTCGACTCATTATTAAAGCTTGCAAAAGTGCACTTCAACTACTTGCAA
CTACTACAAAAGGAAGAACTTCGTATTATTTCAAAGTGGTGGAATGACACAAAGTTCCATGTAATAACTCCTTATGCA
AGAGATAGAGTACCTGAGCTTTACATGTGGATATTGGGATTATTTTTGGAGCCACATTACTCTCAAGCCCGAATTATA
ACAACCAAAATTATGGCATTACTATGTATATTAGATGACATATATGCCAATGCTACAATTGAAGAGTTtCGTCTTCTA
ACaCATGCGATCaATAGGTGGGAAATtGGTGTCATGGAGGAACTTCCAGAATATATTAAACCATTCTATATGATTATA
TTGGACGTGTATGCCGGATTTGAAGAACAACTAGCTAAAGACAGAAGAGAAAAAGTGGTTTACACTTCAAAACAAGCG
TTTCAAGAACTAGCTAGAGCCTATCTTGAGGAGGCAGAATGGAGACATACTGAAAATGTGCCATCATTTGAAGAATAT
ATGAAAAACGGTTTGACTACATCTTGTGTTGACATTATATCAAAATCGGCTTTGATCGGTATGGGCAAGATTGTCACT
GACGAGAGTTTGGCATGGTACAAAAGTCATCCAAATATCAAGACGGCATCAGAGTTAATTAATAAACTCAAAGAAGAC
GTAGTGTCTTTTATGTTTGGGTGCGAAAGAGGAACCTCAATCACAAGCATAGAGGCATATACCAAGACTTTTAAAGTG
TCAGAAAAGGTAGCTATTGAGGAGATCAAGAAAATGGTTGAAGATGCATGGAAAGATATAAATGATGGATGTCTAAAG
CCAAGGGAAGTTGCAATGGATTTACTTGCTCCAATTGTTAATCTTGCACGAATGGTAGATGTGGTATATACGTACAAT
GATGGGTTCACTTTTCCCGAGACTGTCAAAGAGTATATGACACTCTTGTTGATTGATAGGGTCCCTATATAG
```

SmTPS5_contig

```
ATGGCAGCTGATCATGCAACCACTGCTATTAAGCCACCTATTATTGTTAGACGATCAGGAAACTATCCCCCTTCAAAA
TGGTCCTATGATTATATCCAGTCACTCAATACCACCAACTATGCCGATGAAAAATACAAGACAGCATCACAAGGTTTG
AAAGAAAGTGTGAGGAGGATGATTTGTAATGAAACAAGTGCAGTGGAGACGGACCCGTTGAGAAGACTTGAATTAGTG
GATGATTTGCAGAGGCTTGGAATATCATATCACTTCCGAGATGGAATAGATAATGTGTTGAAGATGATATACAATTGC
TACTATGAAACTCAAGATAAATGGAACGGAATGGATTTGAACCTTAAATCCCTTGGCTTTAGACTCTTACGACAACAT
GGTTATCGTATCCCTCAAGAAATACTTGAGGACATCAAGGATGAGACGGGAAACATCAAGGCTCATATAGGTGAGGAC
ATTGTAGGAATGCTTAACTTGTATGAGGCTTCATTTCATGCTACCCAGGATGAAAATaTACTAGACGAAGCCAGGGAA
TTCACAACCAAATGTCTCAAaGAAAAGATGGAGAAGAAGAGAATGATTaATATTGATGCAAGTTATATGTGGATGTTA
GTAAGTCATGCCTTGGACCGTCCATTGCATTGGACGGTTCCAAGGTTTGAGGCACTATGGTTTATAGAAGCATACAAG
AAGAGAAGCGACATGAAACCGTTGTTGTTAGAGCTTGCTATATTAGATTTCAACATTCTGCAAGGAATACACCTACAA
GATCTTAAGTACTCGTCAAGGTGGTGGGATGGTCTAGCTTGGAACAaGAAGCTGGATTTTGCTCGAGATAGGCTGGTT
GAGTGTTTTATGTGGTCTGTTGGTTCATGTTACGAACCATCATTTgGTGTTGTAAGGAGAAACATCACTAAGTTTATT
TCCCTAGTGACTGTCATAGACGATGTCCTCCGACGtATATGGTACTTTGGATGA
```

Figure S3. GC-MS chromatograms of standards mix show the following volatile compound peaks: 1. α-pinene; 2. sabinene; 3. limonene; 4. indole; 5. β-elemene; 6. valencene; 7. Camphene; 8. β-pinene; 9. α-phellandrene;10. p-cymene; 11. limonene; 12 β-phellandrene; 13. Linalool oxide B; 14. α-copaene; 15. α-curcumene; 16. α-zingiberene; 17. α-farnesene; 18. β-bisabolene; 19. γ-cadinene; 20. Myristicin; 21. β-sesquiphellandrene; 22. farnesal; 23. Nootkatone.




RT: 15.40 - 29.43
Standards
mix 2
NL:
1.00E8
TIC F: + c ESI
Full ms
[32.00-551.00]
MS
20190617_terpen
es_giusi_24
Relative Abundance
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
15.82
93.0
16.60
69.0
12
16.70
93.0
17.53
93.0
17.95
59.0
18.47
59.0
18.75
71.0
20.96
95.0
21.58
71.0
21.97
59.0
23.32
93.0
23.94
69.0
24.61
59.0
24.88
117.0
26.12
121.0
27.36
161.0
27.63
93.0
28.89
43.0
16
17
18
19
20
21
22
23
24
25
26
27
28
29
Time (min)

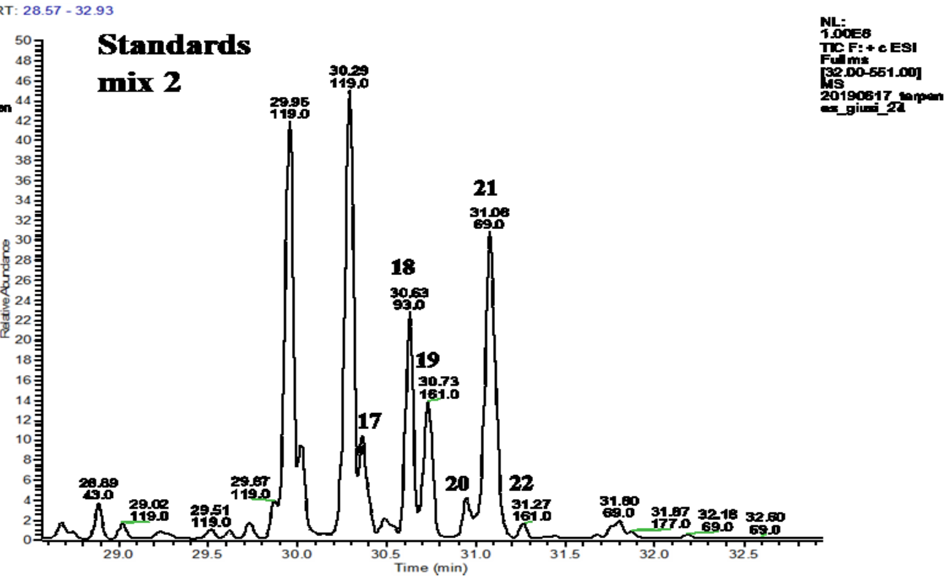

RT: 28.57 - 32.93
Standards
mix 2
NL:
1.00E8
TIC F: + c ESI
Full ms
[32.00-551.00]
MS
20190617_terpen
es_giusi_24
Relative Abundance
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
28.89
43.0
29.02
119.0
29.51
119.0
29.87
119.0
29.95
119.0
30.29
119.0
17
18
30.63
93.0
19
30.73
161.0
20
21
31.08
69.0
22
31.27
161.0
31.80
69.0
31.87
177.0
32.18
69.0
32.60
69.0
29.0
29.5
30.0
30.5
31.0
31.5
32.0
32.5
Time (min)

Figure S4. GC-MS chromatograms showing examples of terpenes peaks in roots (RA1, RB1, RC1) and flowers from different development stages (FA1, FA2, FA3), highlighted as follows: α-copaene in green; β-elemene in red; α-phellandrene in blue; p-cymene in yellow; D-limonene in black.

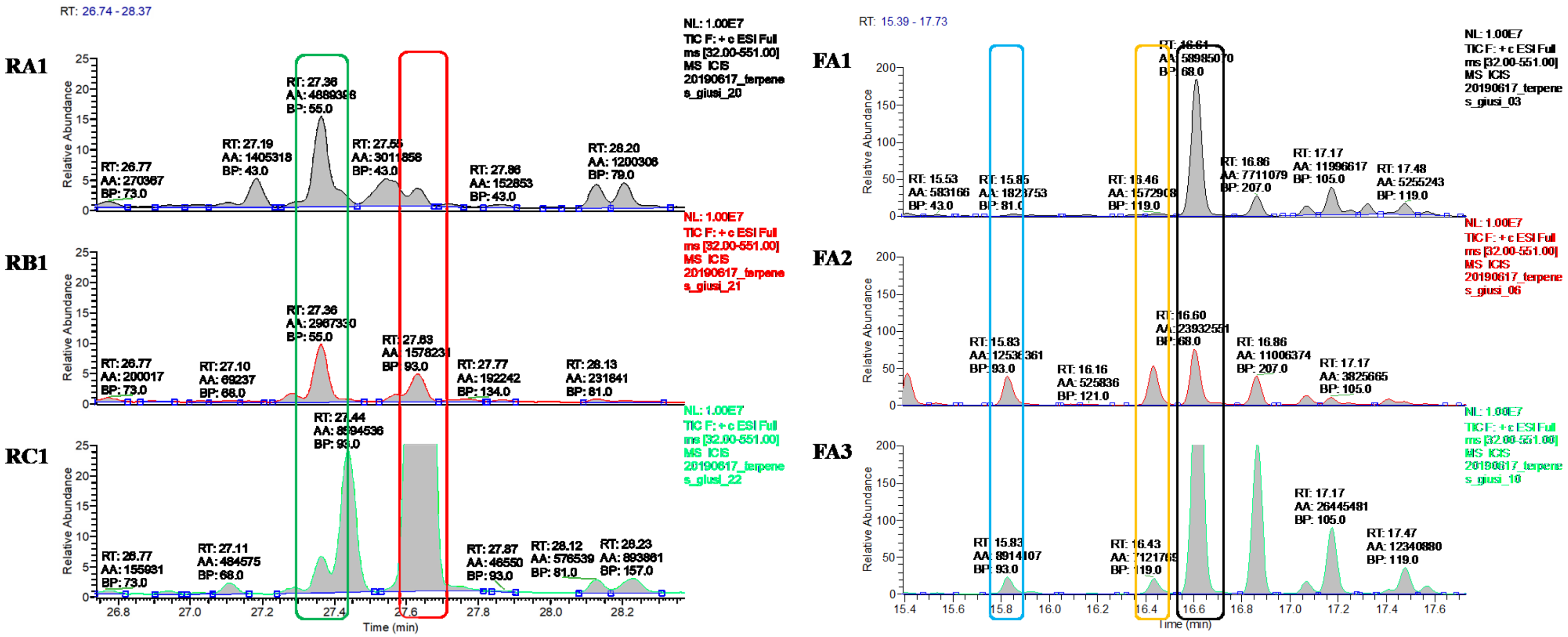
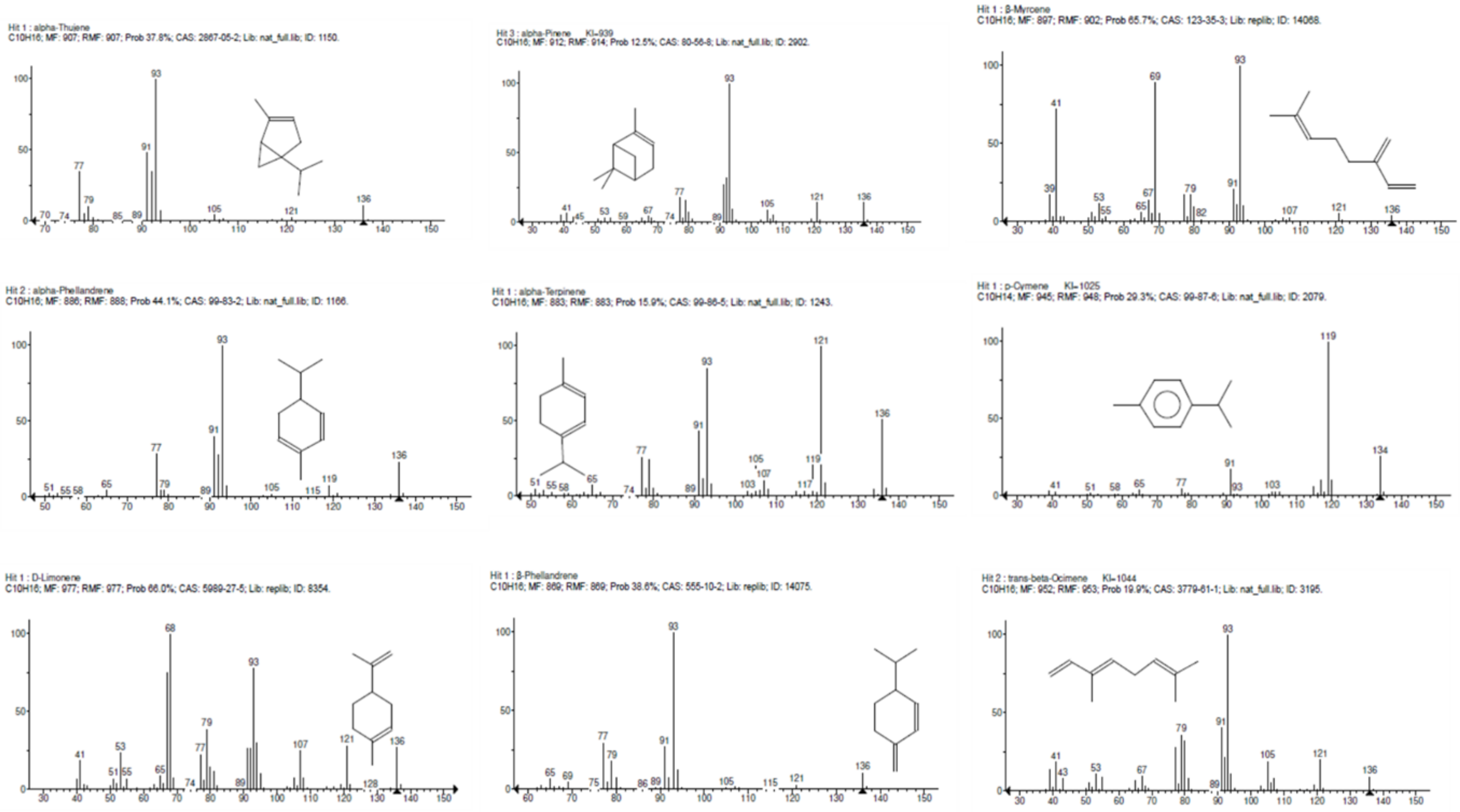
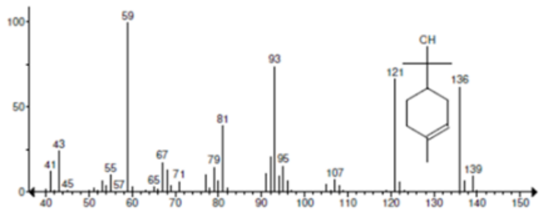


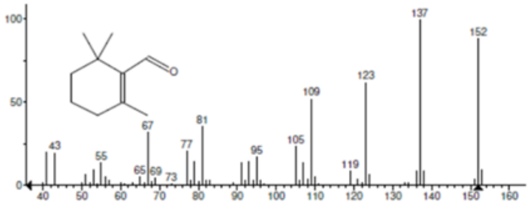
Figure S5. Mass spectra and structures of some terpenes released from milk thistle tissues.



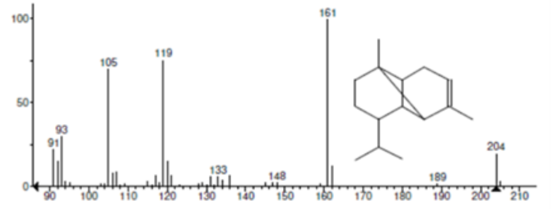
Hit 2 : alpha-Terpineol   KI=1189
C10H18O; MF: 800; RMF: 800; Prob 46.9%; CAS: 98-55-5; Lib: nat_full.lib; ID: 2251.



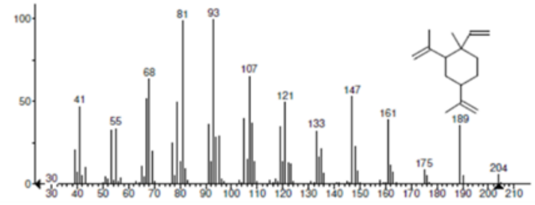
Hit 2 : beta-Cyclocitral
C10H16O; MF: 891; RMF: 896; Prob 83.4%; CAS: 432-25-7; Lib: nat_full.lib; ID: 1318.



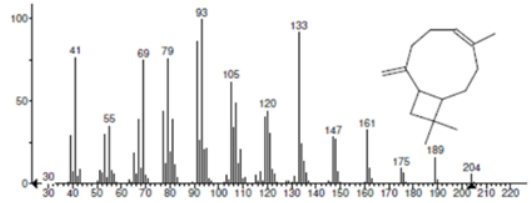
Hit 1 : alpha-Copaene
C15H24; MF: 835; RMF: 838; Prob 35.2%; CAS: 3856-25-5; Lib: nat_full.lib; ID: 1164.



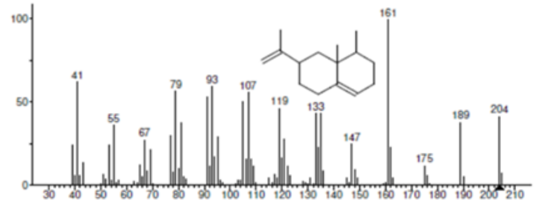
Hit 1 : beta-Elemene   KI=1391
C15H24; MF: 925; RMF: 926; Prob 42.9%; CAS: 515-13-9; Lib: nat_full.lib; ID: 2448.



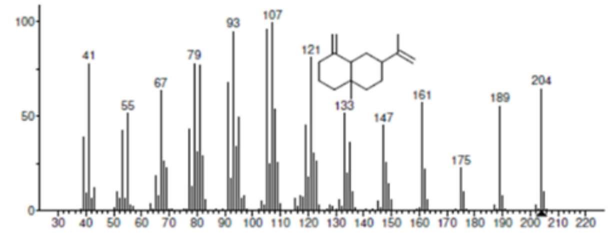
Hit 1 : Caryophyllene
C15H24; MF: 860; RMF: 860; Prob 35.6%; CAS: 87-44-5; Lib: mainlib; ID: 66572.



Hit 1 : Valencene   RI=1494
C15H24; MF: 786; RMF: 826; Prob 7.91%; CAS: 4630-07-3; Lib: nat_full.lib; ID: 293.



Hit 2 : beta-Selinene   RI=1486
C15H24; MF: 837; RMF: 838; Prob 22.8%; CAS: 17066-67-0; Lib: nat_full.lib; ID: 268.



Hit 1 : alpha-Selinene   RI=1494
C15H24; MF: 958; RMF: 959; Prob 26.4%; CAS: 473-13-2; Lib: nat_full.lib; ID: 266.

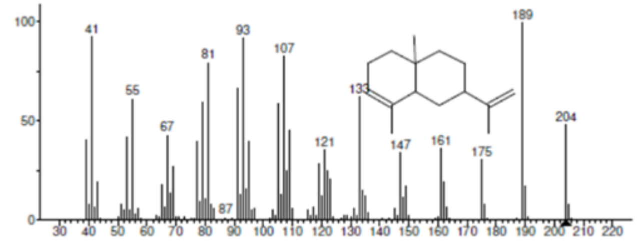


Figure S6. LC-MS chromatograms of silybin, silychristin and silydianin standards.

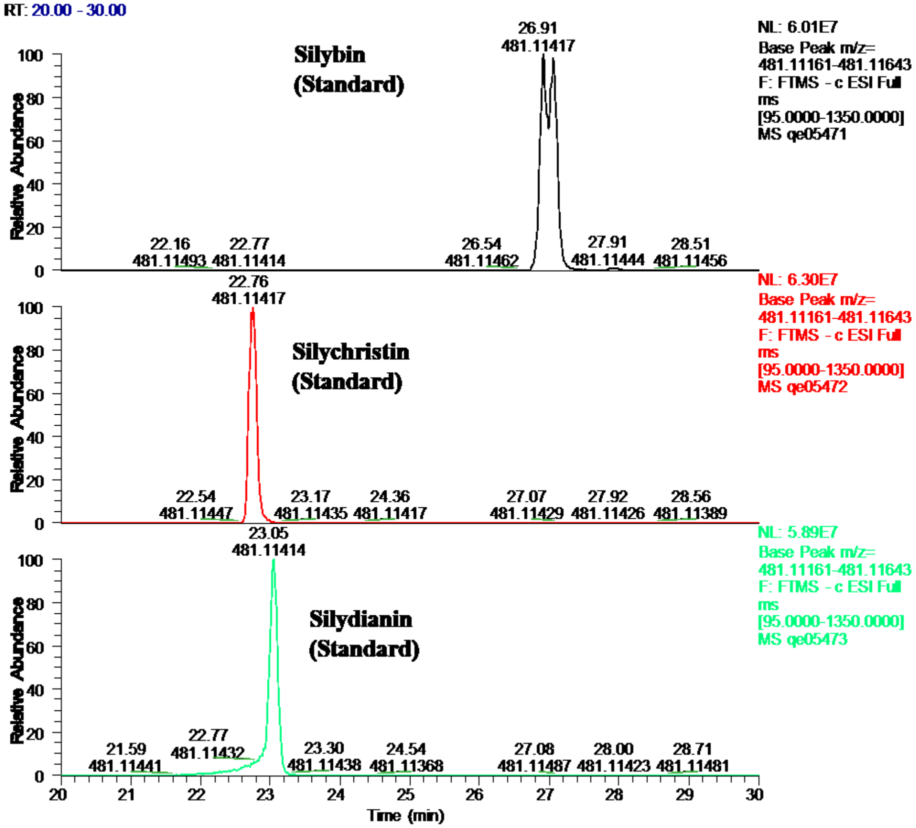


Figure S7. LC-MS chromatograms showing silymarin constituents peaks of milk thistle flowers at the third stage, highlighted as follows: silydianin (RT 23.05) and silychristin (RT 22.76) in green, silybin (RT 26.91) in black.

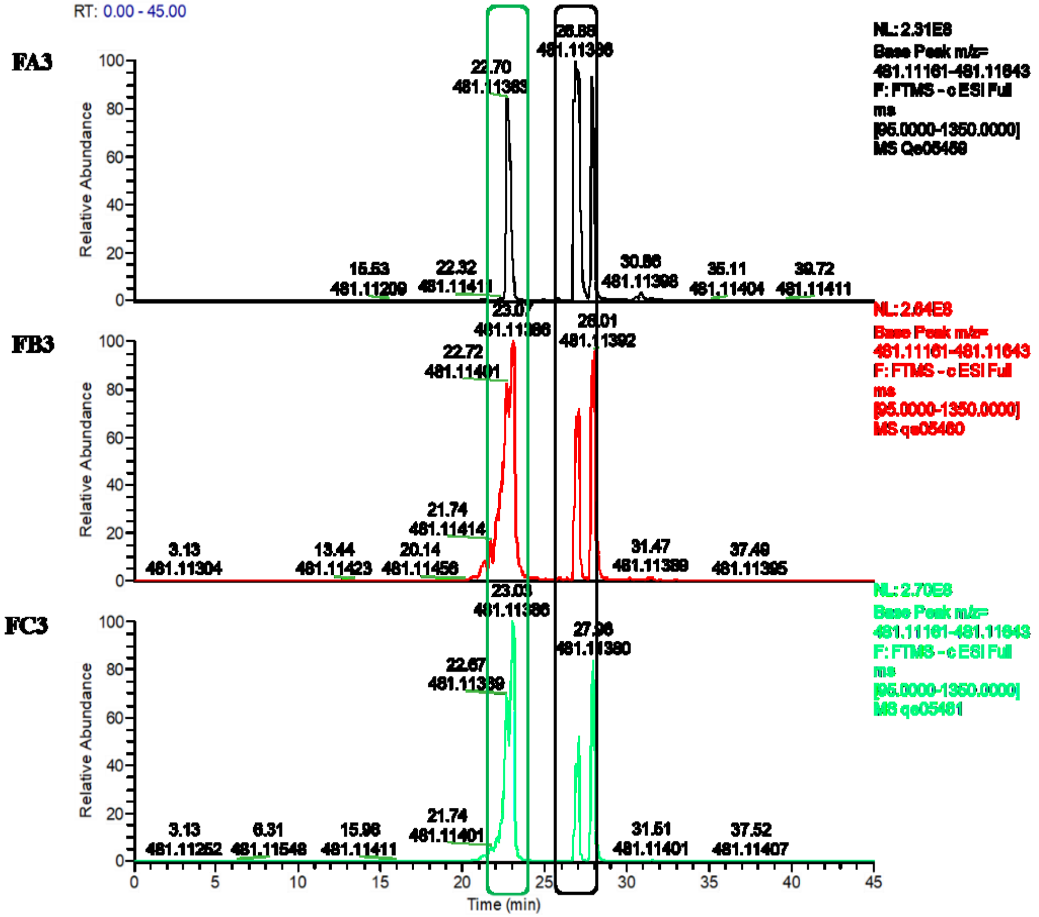


Figure S8. *In vitro* assay of authentic standards (β-elemene and ginger oil) by GC-MS. GC-MS chromatograms show the following terpenes peaks: 1. β-elemene; 2. farnesane; 4. α-curcumene; 5. zingiberene; 6. β-sesquiphellandrene.

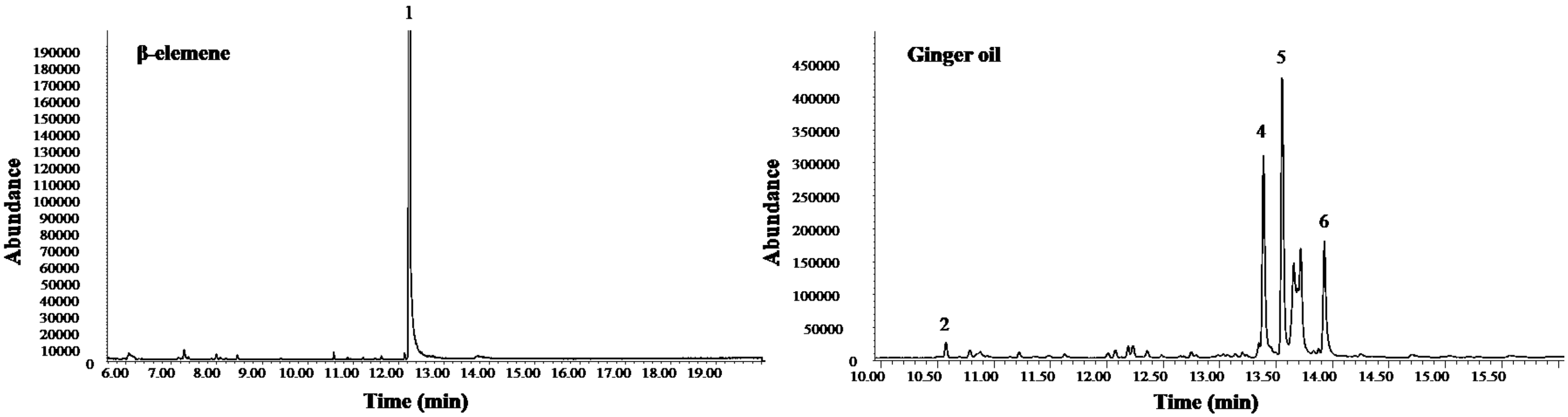
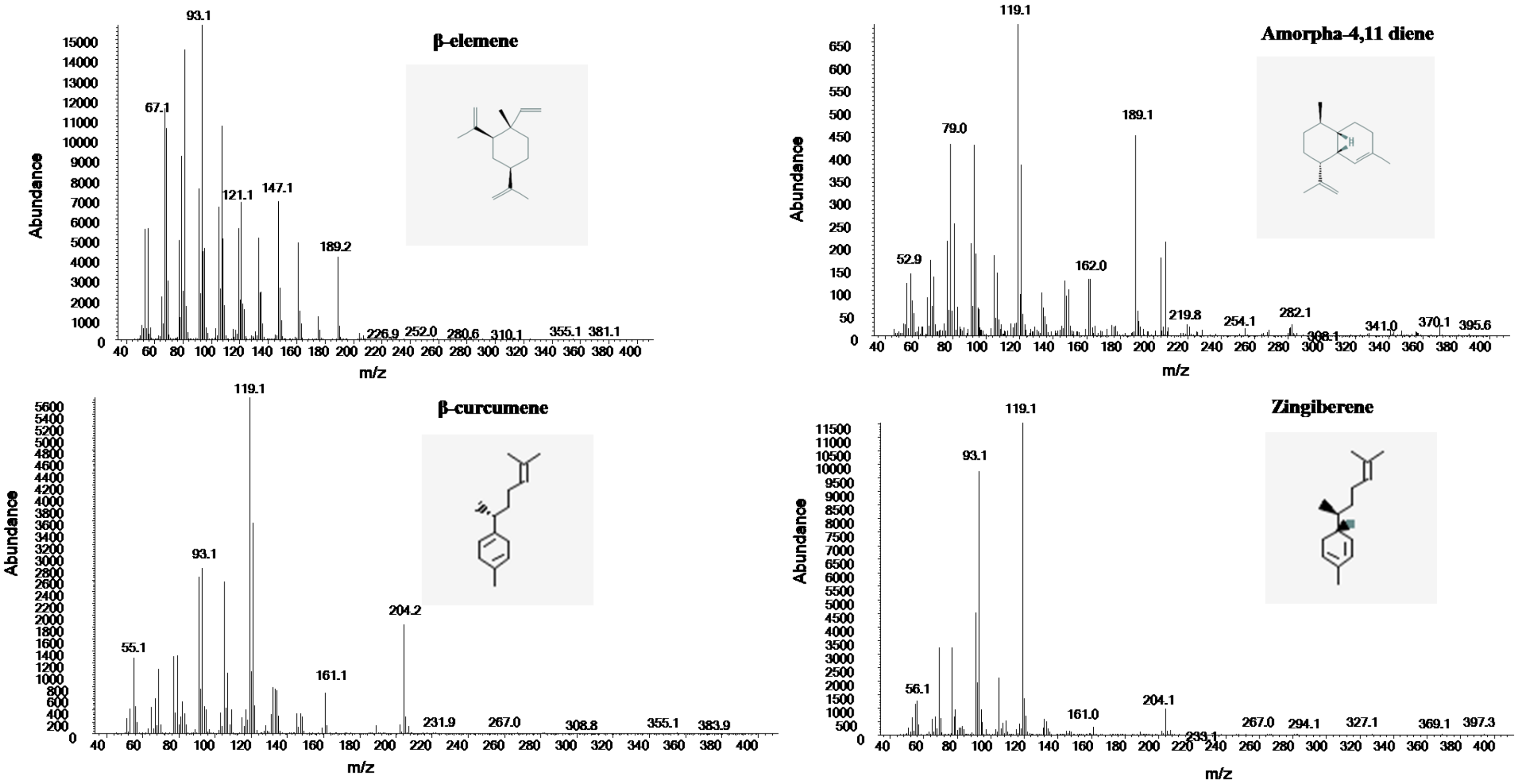


Figure S9. *In vitro* enzymatic activity analysis of SmTPS proteins using FPP as substrate. Mass spectra of some terpenes identified.



8. References

Abenavoli L., A. Izzo A.A., Milić N., Cicala C., Santini A., Capasso R., 2018. Milk thistle (*Silybum marianum*): A concise overview on its chemistry, pharmacological, and nutraceutical uses in liver diseases. Phytotherapy Research 1-12.

Abenavoli L., Capasso R., Milic N., and Capasso F., 2010. Milk Thistle in Liver Diseases: Past, Present, Future. Phytother. Res. 24: 1423-1432.

Adam K-P, Thiel R., Zapp J. 1999. Incorporation of 1-[1-13C]deoxy-dxylulose in chamomile sesquiterpenes. Archives of Biochemistry and Biophysics 369: 127-132.

Adams and TeBeest., 2017. Geographic variation in volatile leaf oils (terpenes) in natural populations of *Helianthus annuus* (Asteraceae, Sunflowers). Phytologia.

Adeyemi M.M.H., 2010. The potential of secondary metabolites in plant material as deterents against insect pests: A review. African Journal of Pure and Applied Chemistry Vol. 4(11): 243-246.

Adio A. M., 2009. Germacrenes A–E and related compounds: thermal, photochemical and acid induced transannular cyclizations. Tetrahedron 65: 1533-1552.

Agarwal M., Walia S., Dhingra S., Khambay B., 2001. Insect growth inhibition, antifeedant and antifungal activity of compounds isolated/derived from *Zingiber officinale* Roscoe rhizomes. Pest Manag. Sci. 57 (3): 289-300.

Aharoni A., Giri A.P., Verstappen F.W.A., Bertea C.M., Sevenier R., Sun Z.K., Jongsma M.A., Schwab W., Bouwmeester H.J., 2004. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell 16: 3110-3131.

Ahmed E., Malik A., Ferheen S., Afza N., Azhar-ul-haq, Lodhi M.A. and Choudhary I. M., 2006. Chymotrypsin Inhibitory Triterpenoids from *Silybum marianum*. Chem. Pharm. Bull. 54(1): 103-106.

Ahmed E., tun Noor A., Malik A., Ferheen S. and Afza N., 2007. Spectral Assignments and Reference Data. Structural determination of silymins A and B, new pentacyclic triterpenes from *Silybum marianum*, by 1D and 2D NMR spectroscopy. Magn. Reson. Chem. 45: 79-81

Andrew R. and Izzo A.A., 2017. Principles of pharmacological research of nutraceuticals. British Journal of Pharmacology 174: 1177-1194.

Andriotis V.M., Kruger N.J., Pike M.J., Smith A.M., 2010. Plastidial glycolysis in developing Arabidopsis embryos. New Phytologist 185: 649-662.

Andrzejewskaa J., Sadowska K., Mielcarek S., 2011. Effect of sowing date and rate on the yield and flavonolignan content of the fruits of milk thistle (*Silybum marianum* L. Gaertn.) grown on light soil in a moderate climate. Industrial Crops and Products 33: 462-468.

Antonious G.F. and Kochhar T.S., 2003. Zingiberene and Curcumene in Wild Tomato. Journal of Environmental Science and Health, Part B: Pesticides, Food Contaminants, and Agricultural Wastes 38(4): 489-500.

Astley S.B., Lindsay D.G., 2002. European Research on the Functional Effects of Dietary Antioxidants - EUROFEDA. Mol Asp Med 23: 1-38.

Aubourg S., Lecharny A., Bohlmann J., 2002. Genomic analysis of the terpenoid synthase (AtTPS) gene family of Arabidopsis thaliana. Molecular Genetics and Genomics 267: 730-745.

Azmir J., Zaidul I.S.M., Rahman M.M., Sharif K.M., Mohamed A., Sahena F., Jahurul M.H.A., Ghafoor K., Norulaini N.A.N., Omar A.K.M., 2013. Techniques for extraction of bioactive compounds from plant materials: A review. Journal of Food Engineering.

Balasubramanian S. and Panigrahi S., 2011. Solid-Phase Microextraction (SPME) Techniques for Quality Characterization of Food Products: A Review. Food Bioprocess Technol 4: 1-26.

Barbazuk W.B., Emrich S.J., Chen H.D. *et al*., 2007. SNP discovery via 454 transcriptome sequencing. Plant J 51: 910-8.

Bayer R.G., Stael S., Csaszar E., Teige M., 2011. Mining the soluble chloroplast proteome by affinity chromatography. Proteomics 11: 1287-1299.

Bennett M., Mansfield J., Lewis M., Beale M., 2002. Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). Phytochemistry 59: 255-261.

Bernhoft, 2010. A brief review on bioactive compounds in plants. In: Bioactive compounds in plants - benefits and risks for man and animals. Aksel Bernhoft editor. The Norwegian Academy of Science and Letters, Oslo.

Bertea C., Voster A., Verstappen F., Maffei M., Beekwilder J., Bouwmeester H., 2006. Isoprenoid biosynthesis in *Artemisia annua*: cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. Archives of Biochemistry and Biophysics 448: 3-12.

Bezman Y., Mayer F., Takeoka G.R., Buttery R.G., Ben-Oliel G., Rabinowitch H.D., Naim M., 2003. Differential effects of tomato (Lycopersicon esculentum Mill) matrix on the volatility of important aroma compounds. J Agric Food Chem 51: 722-726

Bick J.A., Lange B.M., 2003. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Archives of Biochemistry and Biophysics 415: 146-154.

Biedermann D., Vavříková E., Cvak L. and Křen V., 2014. Chemistry of silybin. Natural Product Reports 31, 1138-1157.

Bijak M., 2017. Silybin, a major bioactive component of milk thistle (*Silybum marianum* L. Gaernt.) chemistry, bioavailability, and metabolism. Molecules 22: 1942.

Bleeker P.M., Spyropoulou E.A., Diergaarde P.J., Volpin H., De Both M.T., Zerbe P., Bohlmann J., Falara V., Matsuba Y., Pichersky E. *et al*., 2011. RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from *Solanum lycopersicum* and *Solanum habrochaites* trichomes. Plant Molecular Biology 77: 323-336.

Blomhoff, 2010. Role of dietary phytochemicals in oxidative stress. In: Bioactive compounds in plants – benefits and risks for man and animals. Aksel Bernhoft editor. The Norwegian Academy of Science and Letters, Oslo.

Bohlmann J., Meyer-Gauen G., Croteau R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proceedings of the National Academy of Sciences, USA 95: 4126-4133.

Boller T., Wiemken A., 1986. Dynamics of vacuolar compartmentation. Annual Review of Plant Physiology 37: 137-164.

Bouwmeester H., Kodde J., VerstappenF., Altug I., de KrakeJ.r, Wallaart T., 2002. Isolation and characterization of two germacrene A synthase cDNA clones from chicory. Plant Physiology 129: 134-144.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology 34(5): 525.

Breitmaier E., 2006. Terpenes. Wiley-VCH, Weinheim, Germany.

Bustin S.A., 2000. Absolute quantification of mRNA using real time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25: 169-93.

Bustin S.A., Benes V., Nolan T. *et al*., 2005. Quantitative real-time RT-PCR--a perspective. J Mol Endocrinol 34: 597-601.

Cacciapuoti F., Scognamiglio A., Palumbo R., Forte R. and Cacciapuoti F., 2013. Silymarin in non alcoholic fatty liver disease. World J Hepatol. 5(3): 109-113.

Camarasu C.C., 2000. Headspace SPME method development for the analysis of volatile polar residual solvents by GC–MS. Journal of Pharmaceutical and Biomedical Analysis 23: 197-210.

Cane D.E. 1999. Sesquiterpene biosynthesis: cyclization mechanisms. In: Cane D.D., ed. Comprehensive natural products chemistry: isoprenoids including carotenoids and steroids. Amsterdam, the Netherlands: Elsevier 155-200.

Cankar K., van Houwelingen A., Bosch D., Sonke T., Bouwmeester H., Beekwilder J., 2011. A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene. FEBS Letters 585: 178-182.

Cappelletti E.M., Caniato R., 1984. Silymarin localization in the fruit and seed of *Silybum marianum* (L.) Gaertn. Herba Hungar 23: 53-62.

Carrier D.J., Crowe T., Sokhansanj S., Wahab J., Barl B., 2002. Milk Thistle, Silybum marianum (L.) Gaertn., Flower Head Development and Associated Marker Compound Profile. Journal of Herbs, Spices & Medicinal Plants 10: 1.

Carter C. D., Gianfagna T. J., Sacalis J. N., 1989. Sesquiterpenes in glandular trichomes of wild tomato species and toxicity to the Colordao potato beetle. J. Agric. Food Chem. 37(5): 1425-1428.

Chambers C.S., Holečková V., Petrásková L., Biedermann D., Valentová K., Buchta M. and Křen, V., 2017. The silymarin composition and why does it matter??? Food Research International, 100: 339-353.

Chang S., Puryear J. and Cairney J., 1993. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter 11(2): 113-116.

Chapman and Hall, 2002. Dictionary of natural products on CDROM, version 11:1.

Chappell J. and Coates R.M., 2010. “Sesquiterpenes,” in Comprehensive Natural Products II, eds M. Lew and L. Hung-Wen (Oxford: Elsevier), 609-641.

Chen F., Tholl D., Bohlmann J., Pichersky E., 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. The Plant Journal 66: 212-229.

Cheng A.X., Lou Y.G., Mao Y.B., Lu S., Wang L.J. and Chen, X.Y., 2007. Plant terpenoids: biosynthesis and ecological functions. J. Integr. Plant Biol. 49: 179-186.

Cho K.S., Lim Y.R., Lee K., Lee J., Lee J.H., Lee I.S., 2017. Terpenes from forests and human health. Toxicol. Res. 33(2): 97-106.

Cloonan N., Forrest A.R., Kolle G. *et al*. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 5: 613-9.

Cordovez V., Carrion V.J.J., Etalo D. W., Mumm R., Zhu H., Van Wezel G.P., Raaijmakers J. M., 2015. Diversity and functions of volatile organic compounds produced by Streptomyces from a disease-suppressive soil. Frontiers in Microbiology 6, 111.

Costa A.G.C., Karachaliou N., Rosell R., 2013. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl Lung Cancer Res* 2(2): 87-91.

Croteau R., Kutchan T.M., Lewis N.G. Natural Products (Secondary Metabolites). Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, R. Jones, 2000. Eds. American Society of Plant Physiologists.

Davidovich-Rikanati R., Lewinsohn E., Bar E., Iijima Y., Pichersky E., Sitrit Y., 2008. Overexpression of the lemon basil α-zingiberene synthase gene increases both mono- and sesquiterpene contents in tomato fruit. The Plant Journal 56, 228-238.

Davoli E., Gangai M. L., Morselli L. and Tonelli D., 2003. Characterization of odorants emissions from landfills by SPME and GC/MS. Chemosphere, 51, 357-368.

De Kraker J., Franssen M., De Groot A., König W. and Bouwmeester H., 1998. (+)-Germacrene A biosynthesis. The committed step in the biosynthesis of bitter sesquiterpene lactones in chicory. Plant Physiol. 117: 1381-1392.

De Vos R.C.H., Moco S., Lommen A., Keurentjes J.J.B., Bino R.J. and Hall R.D., 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2: 778-791.

Degenhardt J., Köllner T.G., Gershenzon J., 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70: 1621-1637.

Dudareva N., Andersson S., Orlova I., Gatto N., Reichelt M., Rhodes D., Boland W., Gershenzon J., 2005. The non-mevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proceedings of the National Academy of Sciences, USA 102: 933-938.

Dudareva N., Klempien A., Muhlemann J.K., Kaplan I., 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytologist 198: 16-32

Dudareva N., Martin D., Kish C.M., Kolosova N., Gorenstein N., Faldt J., Miller B., Bohlmann J., 2003. (E)-β-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. The Plant Cell 15: 1227-1241.

Dudareva N., Negre F., Nagegowda D.A. and Orlova, I., 2006. Plant volatiles: recent advances and future perspectives. Crit. Rev. Plant Sci. 25: 417-440.

Dudareva N., Pichersky E., 2000. Biochemical and molecular genetic aspects of floral scent. Plant Physiology 122 (3): 627-633.

Dudareva N., Pichersky E., Gershenzon J., 2004. Biochemistry of plant volatiles. Plant Physiology 135: 1893-1902.

Duetz W.A., Bouwmeester H., van Beilen J.B., Witholt B., 2003. Biotransformation of limonene by bacteria, fungi, yeasts, and plants. Appl Microbiol Biotechnol 61:269-277.

Ebrahimnezhad Z., Zarghami N., Keyhani M., Amirsaadat S., Akbarzadeh A., Rahmati M., Mohammad Taheri Z. and Nejati-Koshki K., 2013. Inhibition of hTERT gene expression by silibinin-loaded PLGA-PEG-Fe3O4 in T47D breast cancer cell line. Bioimpacts 3(2): 67-74.

Eigenbrode S.D., Trumble J.T., 1993. Antibiosis to beet armyworm (*Spodoptera exigua*) in Lycopersicon accessions. HortScience 28(9): 932-934.

Eigenbrode S.D., Trumble J.T., White K.K. 1996. Trichome exudates and resistance to beet armyworm (Lepidoptera: Noctuidae) in *Lycoperiscon hirsutum* f. typicum accessions. Environ. Entomol. 25: 90-95.

El Sherif F., Khattab S., Ibrahim A.K. and Ahmed S.A., 2013. Improved silymarin content in elicited multiple shoot cultures of *Silybum marianum* L. Physiol Mol Biol Plants, 19(1): 127-136.

Eljounaidi K., Cankar K., Comino C., Moglia A., Hehn A., Bourgaud F, Bouwmeester H., Menin B., Lanteri S., Beekwilder J., 2014. Cytochrome P450s from *Cynara cardunculus* L. CYP71AV9 andCYP71BL5, catalyze distinct hydroxylations in the sesquiterpenelactone biosynthetic pathway. Plant Science 223: 59-68.

Emanuelsson O., Nielsen H., von Heijne G., 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8: 978-984.

Eriksson L., Johansson E., Kettaneh-Wold N., Wikström C. and Trygg J., 2006. Multi- and megavariate data analysis part I: Basic principles and applications (Vol. 1). Umea: Umetrics AB.

Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., EramianD., Min-yi Shen, Pieper U., and Sali A., 2007. Comparative Protein Structure Modeling *UNIT 2.9* Using MODELLER. *Current Protocols in Protein Science* 2.9.1-2.9.31.

Fan H., Chen J., Lv H., Ao X., Wu Y., Ren B., Li W., 2017. Isolation and identification of terpenoids from chicory roots and their inhibitory activities against yeast α-glucosidase. European Food Research and Technology 243(6): 1009-1017.

Finn R.D. *et al*. 2013. "Pfam: the protein families database." *Nucleic acids research* 42(1): 222-230.

Flora K., Hahn M., Rosen H. and Benner K., 1998. Milk thistle (*Silybum marianum*) for the therapy of liver disease. The American Journal of Gastroenterology 93: 139-143.

Flϋgge U.I. and Gao W., 2005. Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biology 7: 91-97.

Furumoto T., Yamaguchi T., Ohshima-Ichie Y., Nakamura M., Tsuchida-Iwata Y., Shimamura M., Ohnishi J., Hata S., Gowik U., Westhoff P. *et al*., 2011. A plastidial sodium-dependent pyruvate transporter. Nature 476: 472-475.

Gao F., Liu B., Li M., Gao X., Fang Q., Liu C., Ding H., Wang L., and Gao X., 2018. Identification and characterization of terpene synthase genes accounting for volatile terpene emissions in flowers of *Freesia x hybrida*. Journal of Experimental Botany.

Garms S., Kollner T.G., Boland W., 2010. A multiproduct terpene synthase from *Medicago truncatula* generates cadalane sesquiterpenes via two different mechanisms. Journal of Organic Chemistry 75: 5590-5600.

Gazak R., Walterova D., Kren V., 2007. Silybin and silymarin-new and emerging applications in medicine. Curr Med Chem 14: 315-338.

Giuliani C., Tani C., Maleci Bini L., Fico G., Colombo R., Martinelli T., 2018. Localization of phenolic compounds in the fruits of *Silybum marianum* characterized by different silymarin chemotype and altered colour. Fitoterapia 130: 210-218.

Gopfert J., MacNevin G., Ro D., Spring O., 2009. Identification, functional characterization and developmental regulation of sesquiterpene synthases from sunflower capitate glandular trichomes. BMC Plant Biology 9: 86.

Green S.A., Chen X., Nieuwenhuizen N.J., Matich A.J., Wang M.Y., Bunn B.J., Yauk Y.K., Atkinson R.G., 2012. Identification, functional characterization, and regulation of the enzyme responsible for floral (E)-nerolidol biosynthesis in kiwifruit (*Actinidia chinensis*). Journal of Experimental Botany 63: 1951-1967.

Gresta F., Avola G., Guarnaccia P., 2006. Agronomic Characterization of Some Spontaneous Genotypes of Milk Thistle (*Silybum marianum* L. Gaertn.) in Mediterranean Environment. Journal of Herbs, Spices and Medicinal Plants 12 (4).

Groves R.H., and Kaye P.E., 1989. Germination and phenology of seven introduced thistle species in southern Australia. Aust. J. Bot. 37: 351-359.

Gu W.X., Chen X.C., Pan X.F., Chan A.S.C. and Yang T.K., 2000. First enantioselective syntheses of (2R,3R)- and (2S,3S)-3-(4-hydroxy-3-methoxyphenyl)-2-hydroxymethyl-1,4-benzodioxan-6-carbaldehyde. Tetrahedron: Asymmetry 11: 2801-2807.

Gutensohn M., Nagegowda D.A., Dudareva N., 2013. Involvement of compartmentalization in monoterpene and sesquiterpene biosynthesis in plants. In: Bach T.J., Rohmer M., eds. Isoprenoid synthesis in plants and microorganisms. New York, NY, USA: Springer 155-169.

Hamid S., Sabir A., Khan S. and P. Aziz P., 1983. Experimental cultivation of *Silybum marianum* and chemical composition of its oil. Pakistan Journal of Scientific and Industrial Research 26: 244-246.

Hampel D., Mosandl A., Wüst M., 2005. Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. Phytochemistry 66: 305-311.

Harborne J.R., 1993. Introduction to Ecological Biochemistry, forth ed. Academic Press, Elsevier, London 1-32.

Hartmann T., 2007. From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry 68: 2831-2846.

Hemmerlin A., Hoeffler J-F, Meyer O., Tritsch D., Kagan I.A., Grosdemange-Billiard C., Rohmer M., Bach T.J., 2003. Cross-talk between the cytosolic evalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. Journal of Biological Chemistry 278: 26666-26676.

Heuskina S., Godin B., Leroy P., Capella Q., Wathelet J.P., Verheggen F., Haubruge E., Lognay G., 2009. Fast gas chromatography characterisation of purified semiochemicals from essential oils of *Matricaria chamomilla* L. (Asteraceae) and *Nepeta cataria* L. (Lamiaceae). Journal of Chromatography A 1216(14): 2768-2775.

Hsieh M.H., Chang C.Y., Hsu S.J., Chen J.J., 2008. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in IspD and IspE albino mutants in Arabidopsis. Plant Molecular Biology 66: 663-673.

Huang M., Sanchez-Moreiras A.M., Abel C., Sohrabi R., Lee S., Gershenzon J., Tholl D., 2012. The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)-b-caryophyllene, is a defense against a bacterial pathogen. New Phytologist 193: 997-1008.

Huerta-Cepas, Jaime, François Serra, and Peer Bork, 2016. "ETE 3: reconstruction, analysis, and visualization of phylogenomic data." *Molecular biology and evolution* 33(6): 1635-1638.

Hyatt D.C., Youn B., Zhao Y., *et al*. 2007. Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. Proceedings of the National Academy of Sciences, USA 104:5360-5365.

Isman M., 2002. Insect Antifeedants: In Pesticide Outlook. Royal Soc. Chem. 152-157.

Ivica L. and Bork P., 2006. "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation." *Bioinformatics* 23(1): 127-128.

Jacobson M., 1975. Insecticides from Plants: A Review of the Literature, 1954-1971. Agricultural Handbook, U.S Department of Agriculture, Washington, D.C. 461: 138.

Jacobson M.,1989. Botanical pesticides, past present and future. In: Insecticides of plant origin. Ed. Arnason, J.T. Proceeding of the American Chemical Society, Washington, D.C. 1-10.

Joyard J., Ferro M., Masselon C., Seigneurin-Berny D., Salvi D., Garin J., Rolland N., 2010. Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Progress in Lipid Research 49: 128-158.

Jozefczuk J. and Adjaye J., 2011. Quantitative real-time PCRbased analysis of gene expression. Methods Enzymol 500: 99-109.

Junker R.R. and Bluethgen N., 2008. Floral scents repel potentially nectar thieving ants. Evol. Ecol. Res. 10: 295-308.

Kadereit J.W., Körner C., Kost B., Sonnewald U., 2014. Strasburger-Lehrbuch der Pflanzenwissenschaften, thirty seventh ed. Heidelberg: Springer-Spektrum.

Kamaliroosta Z., Kamaliroosta L., Elhamirad A.H., 2013. Isolation and Identification of Ginger Essential Oil. Journal of Food Biosciences and Technology 3: 73-80.

Kang X.P., Jiang T., Li Y.Q. *et al*., 2010. A duplex real-time RTPCR assay for detecting H5N1 avian influenza virus and pandemic H1N1 influenza virus. Virol. J. 7: 113.

Kim M., Chang Y., Bang M., Baek N., Jin J., Lee C., Kim S., 2005. cDNA isolation and characterization of (+)-germacrene A synthase from *Ixeris dentata* form. Albiflora Hara. Journal of Plant Biology 48: 178-186.

Kiyama R., 2017. Estrogenic terpenes and terpenoids: Pathways, functions and applications. European Journal of Pharmacology.

Knudsen J.T., Eriksson R., Gershenzon J., Stahl B., 2006. Diversity and distribution of floral scent. Botanical Review 72: 1-120.

Kovacevic N., Pavlovic M., Menkovic N., Tzakou O. and Couladis M., 2002. Composition of the essential oil from roots and rhizomes of *Valeriana pancicii*. Halacsy & Bald. Flavour Fragr. J. 17: 355-357.

Kraker J.W., Franssen M.C.R., de Groot A., König W.A., and Bouwmeester H.J., 1998. (1)-Germacrene A Biosynthesis. Plant Physiol 117: 1381-1392.

Krauss G.J., Nies D.H., 2014. Ecological Biochemistry – Environmental and Interspecific Interactions. Weinheim: Wiley-VCH.

Kumar T., Kumar Larokar Y., Kumar Iyer S., Kumar A., Tripathi D.K., 2011. Phytochemistry and Pharmacological Activities of *Silybum marianum*: A Review. International Journal of Pharmaceutical and Phytopharmacological Research 1(3): 124-133.

Kutchan T.M., 2005. A role for intra-and intercellular translocation in natural product biosynthesis. Current Opinion in Plant Biology 8: 292-300.

Kvasnicka F., Bíba B., Sevcík R., Voldrich M. and Krátká J., 2003. Analysis of the active components of silymarin. Journal of Chromatography. 990: 239-245.

Lange B.M., Rujan T., Martin W., Croteau R., 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences, USA 97: 13172-13177.

Laule O., Furholz A., Chang H.S., Zhu T., Wang X., Heifetz P.B., Gruissem W., Lange M., 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, USA 100: 6866-6871.

Levin D.A., 1976. The chemical defences of plants to pathogens and herbivores. Annual Review of Ecology and Systematics 7: 121-159.

Li B., Ruotti V., Stewart R.M. *et a*l., 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26: 493-500.

Lommen A., 2009. MetAlign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Analytical Chemistry 81: 3079-3086.

Lucini L., Kane D., Pellizzoni M., Ferrari A., Trevisi E., Ruzickova G., Arslan D., 2016. Phenolic profile and in vitro antioxidant power of different milk thistle [*Silybum marianum* (L.) Gaertn.] cultivars. Industrial Crops and Products 83: 11-16.

Lv Y, Gao S., Xu S., Du G., Zhou J. and CheJ., 2017. Spatial organization of silybin biosynthesis in milk thistle [*Silybum marianum* (L.) Gaertn]. The Plant Journal. John Wiley and Sons Ltd.

Macleod A.J., Nirmala M. Pieris N.M. and De Troconis N.G., 1982. Aroma volatiles of *Cynara scolymus* and *helianthus tuberosus.* Pkytochemistr 21(7): 1647-1651.

Manach C., Scalbert A., Morand C., Remesy C., Jimenez L., 2004. Polyphenols: food sources and bioavailability. Am J Clin Nutr. 79: 727-747.

Manach C., Williamson G., Morand C., Scalbert A., Remesy C., 2005. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. Am J Clin Nutr. 81: 230-242.

Mann J., 1992. Murder, Magic and Medicine. London: Oxford University Press.

Martin D.M., Aubourg S., Schouwey M.B., Daviet L., Schalk M., Toub O., Lund S.T., Bohlmann J., 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biology 10: 226.

Martin J.F., Demain A.L., 1978. The filamentous fungi. In: Smith, J.E., Berry, D.R. (Eds.), Developmental Mycology, vol. 3. Edward Arnold, London.

Martin R.J., Lauren D.R., Smith W.A., Jensen D.J., Douglas B., Deo and J.A., 2010. Factors influencing silymarin content and composition in variegated thistle (*Silybum marianum*). New Zealand Journal of Crop and Horticultural Science.

Martin R.J., Deo B. and Douglas J.A., 2000. Effect of time of sowing on reproductive development of variegated thistle. Agron. New Zealand 30: 1-5.

Martinelli T., Potenza E., Moschella A., Zaccheria F., Benedettelli S., and Andrzejewska J., 2016. Phenotypic Evaluation of a Milk Thistle Germplasm Collection: Fruit Morphology and Chemical Composition. Crop science 56.

McGarvey D.J., Croteau R., 1995. Terpenoid metabolism. The Plant Cell 7: 1015-1026.

Menin B., Comino C., Ezio Portis E., Andrea Moglia A, Cankar K., Bouwmeester H.J., Lanteri S., Beekwilder J., 2012. Genetic mapping and characterization of the globe artichoke (+)-germacrene A synthase gene, encoding the first dedicated enzyme for biosynthesis of the bitter sesquiterpene lactone cynaropicrin. Plant Science 190: 1-8.

Mhamdi B., Abbassi F., Smaoui A., Abdelly C. and Marzouk B., 2016. Fatty acids, essential oil and phenolics composition of *Silybum marianum* seeds and their antioxidant activities. Pak. J. Pharm. Sci. 29 (3): 951-959.

Morazzoni P., Bombardelli E., 1995. *Silybum marianum* (*Carduus marianus*). Fitoterapia 66: 3-42.

Morin R., Bainbridge M., Fejes A. *et al*., 2008. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45: 81-94.

Morse A., Kevan P., Shipp L., Khosla S. and Mcgarvey, B., 2012. The impact of greenhouse tomato (Solanales: Solanaceae) floral volatiles on bumble bee (Hymenoptera: Apidae) pollination. Environ. Entomol. 41: 855-864.

Mortazavi A., Williams B.A., McCue K. *et al*., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-8.

Nabeta K., Kawae T., Saitoh T., Kikuchi T., 1997. Synthesis of chlorophyll a and bcarotene from 2H- and 13C-labelled mevalonates and 13C-labelled glycine in cultured cells of liverworts, Heteroscyphus planus and Lophocolea heterophylla. Journal of the Chemical Society-Perkin Transactions 1: 261-267.

Nagegowda D.A., 2010. Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. Federation of European Biochemical Societies.

Nagegowda D.A., Gutensohn M., Wilkerson C.G., Dudareva N., 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant Journal 55: 224-239.

Naim M., *et al*., 2014. "Data access for the 1,000 Plants (1KP) project." Gigascience 3.1: 17.

Nakamura A., Shimada H., Masuda T., Ohta H., Takamiya K., 2001. Two distinct isopentenyl diphosphate isomerases in cytosol and plastid are differentially induced by environmental stresses in tobacco. FEBS Letters 506: 61-64.

Nassar M.I., Mohamed T.K., Elshamy A.I., El-Toumy S.A., Lateef A.M.A and Farrag A.R.H., 2013. Chemical constituents and anti-ulcerogenic potential of the scales of *Cynara scolymus* (artichoke) heads. Wiley Online Library.

New England Biolabs Inc., 2018.

Ngo V.N., Young R.M., Schmitz R. *et al*., 2011. Oncogenically active MYD88 mutations in human lymphoma. Nature 470: 115-9.

Nieuwenhuizen N.J., Green S.A., Chen X., Bailleul E.J., Matich A.J., Wang M.Y., Atkinson R.G., 2013. Functional genomics reveals that a compact terpene synthase gene family can account for terpene volatile production in apple. Plant Physiology 161: 787-804.

Ning-Bo Q., Sheng-Ge Li, Xue-Yuan Yang, Chi Gong, Xiang-Yu Zhang, Jian Wanga, Da-Hong Li, Yuan-Qiang Guo, Zhan-Lin Li, Hui-Ming Hua, 2017. Bioactive terpenoids from *Silybum marianum* and their suppression on NO release in LPS-induced BV-2 cells and interaction with iNOS. Bioorganic & Medicinal Chemistry 27: 2161-2165.

Oliver D.J., Nikolau B.J., Wurtele E.S., 2009. Acetyl-CoA-Life at the metabolic nexus. Plant Science 176: 597-601.

Orlova I., Nagegowda D.A., Kish C.M., Gutensohn M., Maeda H., Varbanova M., Fridman E., Yamaguchi S., Hanada A., Kamiya Y. *et al*., 2009. The small subunit of snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. Plant Cell 21: 4002-4017.

Owen T., 2004. Geoponika: Agricultural Pursuits. In: Moore, S.J. and Langlet, A., 2004. An Overview of Plants as Insect Repellents. In: Wilcox, M. and Bodeker, G. Eds. Traditional Medicine, Medicinal Plants and Malaria. Taylor and Francis, London. http://www.ancientlibrary.com/geoponica/index.html 1805.

Paulsen B.S., 2010. Highlights through the history of plant medicine. In: Proceedings from a Symposium Held at The Norwegian Academy of Science and Letters, Oslo, Norway.

Pawliszyn J., 2002. Solid phasemicroextraction. In: Issaq (Ed.), A century of separation science. New York: Marcel Dekker Inc. 399-419.

Pazouki L., Memari H. R., Kännaste A., Bichele R. and Niinemets Ü. Germacrene A synthase in yarrow (*Achillea millefolium*) is an enzyme with mixed substrate specificity: gene cloning, functional characterization and expression analysis. Frontiers in Plant Science, 2015.

Pepke S., Wold B., Mortazavi A., 2009. Computation for ChIPseq and RNA-seq studies. Nat Methods 6: S22-32.

Pichersky E., Noel J.P., Dudareva N., 2006. Biosynthesis of plant volatiles: nature's diversity and ingenuity. Science 311: 808-811.

Poppe L. and Petersen M., 2016. Variation in the flavonolignan composition of fruits from different *Silybum marianum* chemotypes and suspension cultures derived there from. Phytochemistry 131: 68-75.

Proksch P., Ebel R., 1998. Ecological significance of alkaloids from marine invertebrates. In: Roberts M.F., Wink M. Eds., Alkaloids: Biochemistry, Ecological Functions and Medical Applications. New York: Plenum 379-394.

Prosser I., PhillipsA., Gittings S., Lewis M., Hooper A., PickettJ., BealeM., 2002. (+)-(10R)-germacrene A synthase from goldenrod, *Solidago canadensis*: cDNA isolation, bacterial expression and functional analysis. Phytochemistry 59: 691-702.

Pulido P., Perello C., Rodriguez-Concepcion M., 2012. New insights into plant isoprenoid metabolism. Molecular Plant 5: 964-967.

Raccuia S.A. and De Mastro G. Cardo mariano. In Mosca G. *et al*., 2019. Oli e grassi. Fonti oleaginose per gli utilizzi food e no food. Edagricole 68-69.

Rasmann S., Köllner T. G., Degenhardt J., Hiltpold I., Toepfer S., Kuhlmann U., *et al*., 2005. Recruitment of entomopathogenic nematodes by insect-damaged maize roots. Nature 434: 732-737.

Rattan R.S., 2010. Mechanism of Action of Insecticidal Secondary Metabolites of Plant Origin. Crop Protect. 29(9): 913-920.

Rea P.A., 2007. Plant ATP-binding cassette transporters. Annual Review of Plant Biology. 58: 347-375.

Roberts M.F., Wink M., 1998. Alkaloids: Biochemistry, Ecology and Medicinal Applications. New York: Plenum.

Rohdich F., Zepeck F., Adam P., Hecht S., Kaiser J., Laupitz R., Gräwert T., Amslinger S., Eisenreich W., Bacher A. *et al*., 2003. The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. Proceedings of the National Academy of Sciences, USA 100: 1586-1591.

Saito K., Kaneko S., Furuya Y., Asada Y., Ito R. Sugie K., Akutsu M. and Yanagawa Y., 2019. Confirmation of synthetic cannabinoids in herb and blood by HS-SPME-GC/MS. Forensic Chemistry.

Scala A., Allmann S., Mirabella R., Haring M. A. and Schuurink R. C., 2013. Green leaf volatiles: a plant's multifunctional weapon against herbivores and pathogens. Int. J. Mol. Sci. 14: 17781-17811.

Schadewaldt H., 1969. The history of silymarin. Contribution to the history of liver therapy. Die Medizinische Welt 20: 902-914.

Schmeller T., Wink M., 1998. Utilization of alkaloids in modern medicine. In: Roberts, M.F., Wink M. Eds. Alkaloids: Biochemistry, Ecological Functions and Medical Applications. New York: Plenum 435-458.

Schnee C., Köllner T. G., Gershenzon J. and Degenhardt J., 2002. The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (E)-β-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. Plant Physiol. 130: 2049-2060.

Seigler D., 1998. Plant secondary metabolism. Kluwer, Dordrecht, The Netherlands 759.

Shah S.P., Köbel M., Senz J. *et al*., 2009. Mutation of FOXL2 in granulosa-cell tumors of the ovary. N Engl J Med 360: 2719-29.

Shakeri A. and Mahsa Ahmadian M., 2014. Phytochemical studies of some terpene compounds in roots of *Cynara scolymus.* International Journal of Farming and Allied Sciences 3(10): 1065-1068.

Sievers F. *et al*., 2011. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." *Molecular systems biology* 7.1.

Simkin A.J., Schwartz S.H., Auldridge M., Taylor M.G., Klee H.J., 2004. The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles β-ionone, pseudoionone, and geranylacetone. Plant Journal 40: 882-892.

Singh B., Ram Sharma R.A., 2015. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. Biotech 5: 129-151.

Slater G.S. and Birney E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

Small I., Peeters N., Legeai F., Lurin C., 2004. Predotar: A tool for rapidly screening proteomes for *N*-terminal targeting sequences. Proteomics.

Sonnenbichler J., Scalera F., Sonnenbichler I., and Weyhenmeyer R., 1999. Stimulatory effects of silibinin and silicristin from the milk thistle (*Silybum marianum*) on kidney cells. Journ. of Pharmacology and Experimental Therapeutics 290: 1375-1383.

Strehmel N., Hummel J., Erban A., Strassburg K. and Kopka, J., 2008. Retention index thresholds for compound matching in GC-MS metabolite profiling. Journal of Chromatography 871: 182-190.

Tamayo M.D. and Diamond S., 2007. Review of clinical trials evaluating safety and efficacy of milk thistle (*Silybum marianum* [L.] Gaertn.). Integrative cancer therapies 6 (2): 146-157.

Tholl D., Chen F., Petri J., Gershenzon J., Pichersky E., 2005. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. The Plant Journal 42: 757-771.

Tholl D., 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Current Opinion In Plant Biology 9: 297-304.

Tholl D., Chen F., Petri J., Gershenzon J., Pichersky E., 2005. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant Journal 42: 757-771.

Tiaz L., Zeiger E., 2006. Secondary metabolites and plant defense. In: Plant Physiology, 4th ed. Sinauer Associates, Inc. Sunderland, Massachusetts 283-308 (Chapter 13).

Tikunov Y.M., Laptenok S., Hall R. D., Bovy A., de Vos R. C. H., 2012. MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8: 714-718.

Tikunov Y., Lommen A., de Vos C.H. Ric, Verhoeven H.A., Bino R.J., Robert D. Hall, and Bovy A.G., 2005. A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. Plant Physiology 139: 1125-1137.

Tuduri L., Desauziers V. and Fanlo J. L., 2003. A simple calibration procedure for volatile organic compounds sampling in air with adsorptive solid-phase microextraction fibers. Analyst 128: 1028-1032.

van den Berg R., Hoefsloot H., Westerhuis J., Smilde A. and van der Werf M., 2006. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. BMC Genomics 7: 142.

van Treuren R., van Eekelen H.D.L.M., Wehrens R. and de Vos R.C.H., 2018. Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. Metabolomics 14: 146.

Van Wyk B.E., Wink M., 2004. Medicinal Plants of the World. Pretoria: Briza.

Vender C., 2001. Indagine sulla consistenza e le caratteristiche della produzione di piante officinali in Italia. ISAFA 3: 72.

Verdonk J.C., de Vos C.H.R., Verhoeven H.A., Haring M.A., van Tunen A.J., Schuurink R.C., 2003. Regulation of floral scent production in petunia revealed by targeted metabolomics. Phytochemistry 62: 997-1008.

Verhoeven H.A., Jonker H.H., de Vos R.C.H. and Hall R.D., 2012. Solid-phase micro-extraction (SPME) GC-MS analysis of natural volatile components in melon and rice. In N. G. Hardy and R. D. Hall (Eds.), Plant metabolomics methods 85-100. Ithaca, NY: Humana Press.

Vinatoru M., 2001. An overview of the ultrasonically assisted extraction of bioactive principles from herbs. Ultrasonics Sonochemistry 8 (3): 303-313.

Wagner K.H. and Elmadfa I., 2003. Biological relevance of terpenoids. Overview focusing on mono-, di- and tetraterpenes. Ann. Nutr. Metab. 47: 95-106.

Wang Z., Gerstein M., Snyder M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.

Ward J.L., Baker J.M., Llewellyn A.M., Hawkins N.D., Beale M.H., 2011. Metabolomic analysis of Arabidopsis reveals hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow. Proceedings of the National Academy of Sciences, USA 108: 10762-10767.

Waterhouse A.M. *et al*., 2009. "Jalview Version 2-a multiple sequence alignment editor and analysis workbench." *Bioinformatics* 25(9): 1189-1191.

Weinheimer A.J., Youngblood W.W., Washecheck P.H., Karns T.K.B., Ciereszko L.S., 1970. Isolation of the elusive (-)-germacrene-A from the gorgonian, *Eunicea mammosa*: chemistry of coelenterates XVIII. Tetrahedron Lett 7: 497-500.

Wiegand K.C., Shah S.P., Al-Agha O.M. *et al*., 2010. ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med 363: 1532-43.

Williamson G, Manach C., 2005. Bioavailability and bioefficacy of polyphenols in humans. II. Review of 93 intervention studies. Am J Clin Nutr. 81: 243S-255S.

Wink M., 1988. Plant breeding: Importance of plant secondary metabolites for protection against pathogens and herbivores. Theoretical and Applied Genetics 75: 225-233.

Wink M., 1992. The role of quinolizidine alkaloids in plant insect interactions. In: Bernays, E.A. Ed. Insect−Plant Interactions, vol. IV. Boca Raton: CRC Press 133-169.

Wink M., 1993. Allelochemical properties and the raison d'être of alkaloids. In: Cordell, G. Ed. The Alkaloids, vol.43. San Diego: Academic press 1-118.

Wink M., 1997. Compartmentation of secondary metabolites and xenobiotics in plant vacuoles. Advances of Botanical Research 25: 141-169.

Wink M., 2000. Interference of alkaloids with neuroreceptors and ion channels. In: Atta-Ur-Rahman Ed. Bioactive Natural Products. Amsterdam: Elsevier 1-127.

Wink M., 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64 (1): 3-19.

Wink M., 2007a. Bioprospection: The search for bioactive lead structures from nature. In: Kayser, O., Quax, W. Eds. Medical Plant Biotechnology. From Basic Research to Industrial Applications, vol.1. Weinheim: Wiley-VCH 97-116.

Wink M., 2007b. Molecular modes of action of cytotoxic alkaloids − From DNA intercalation, spindle poisoning, topoisomerase inhibition to apoptosis and multiple drug resistance. In: Cordell, G. Ed. The Alkaloids, vol.64. San Diego: Academic press 1-48.

Wink M., 2008a. Plant secondary metabolism: Diversity, function and its evolution. Natural Products Communications 3: 1205-1216.

Wink M., 2008b. Evolutionary advantage and molecular modes of action of multi-component mixtures used in phytomedicine. Current Drug Metabolism 9: 996-1009.

Wink M., 2010a. Functions and Biotechnology of Plant Secondary Metabolites. Annual Plant Reviews. Chichester: Wiley-Blackwell.

Wink M., 2010b. Biochemistry of Plant Secondary Metabolism. Annual Plant Reviews. Chichester: Wiley-Blackwell.

Wink M., 2013. Evolution of secondary metabolites in legumes (Fabaceae). South African Journal of Botany 89: 164-175.

Wink M., 2015. Modes of action of herbal medicines and plant secondary metabolites. Medicines 2: 251-286.

Wink M., 2016. Secondary metabolites, the role in plant diversification of. Elselvier Inc.

Wink M., Ashour M., El-Readi M.Z., 2012. Secondary metabolites inhibiting ABC transporters and reversing resistance of cancer cells and fungi to cytotoxic and antimicrobial agents. Frontiers in Microbiology 3: 1-15.

Wink M., Schimmer O., 2010. Molecular modes of action of defensive secondary metabolites. In: Wink, M. Ed. Functions and Biotechnology of Plant Secondary Metabolites, Annual Plant Reviews 39. Chichester: Wiley-Blackwell 21-161.

Wink M., Van Wyk B.E., 2008. Mind-Altering and Poisonous Plants of the World. Pretoria: Briza

Winterhalter P., Rouseff R., 2001. Carotenoid-derived aroma compounds: an introduction. In: Winterhalter P, Rouseff R, eds. Carotenoid-derived aroma compounds. Washington, DC, USA: American Chemical Society 1-17.

Wise M.L., Croteau R., 1999. Monoterpene biosynthesis. In: Cane DD, ed. Comprehensive natural products chemistry: isoprenoids including carotenoids and steroids. Amsterdam, the Netherlands: Elsevier 97-153.

Wu S., Schalk M., Clark A., Miles R.B., Coates R., Chappell J., 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. Nature biotechnology 24: 1441-1447.

Yating H.,·Yongjin J.Z., Bao J., Huang L., Nielsen J., Krivoruchko A., 2017. Metabolic engineering of *Saccharomyces cerevisiae* for production of germacrene A, a precursor of beta-elemene. J Ind Microbiol Biotechnol 44:1065-1072.

Yazaki K., 2006. ABC transporters involved in the transport of plant secondary metabolites. FEBS Letters 580: 1183-1191.

Yongkun L., Song G., Sha X., Guocheng D., Jingwen Z. and Jian C., 2017. Spatial organization of silybin biosynthesis in milk thistle (*Silybum marianum* (L.) Gaertn.). The plant journal.

Young J.A., Evans R.A., and Hawkes R.B., 1978. Milk thistle (*Silybum marianum*) seed germination. Weed Sci. 26: 395-398.

Zenk M.H. and Jünger M., 2007. Evolution and current status of the phytochemistry of nitrogenous compounds. Phytochemistry 68: 2757-2772.

Zhang Z. and Pawliszyn J., 1993. Headspace solid-phase microextraction. Analytical Chemistry, 65, 1843-1852.

Zi X. and R. Agarwal., 1999. Silibinin decreases prostate-specific antigen with cell growth inhibition via G 1 arrest, leading to differentiation of prostate carcinoma cells: Implications of prostate cancer intervention. Proc. National Academy of Sciences of USA 96:7490-7495.

Zulak K.G. and Bohlmann J., 2010. Terpenoid biosynthesis and specialized vascular cells of conifer defense. J. Integr. Plant Biol. 52: 86-97.

9. Websites

http://2014.igem.org/Team:Macquarie_Australia/WetLab/Protocols/PlasmidPreps

http://www.cbs.dtu.dk/services/ChloroP/

http://www.ncbi.nlm.nih.gov

https://ecoliwiki.org/colipedia/index.php/File:PACYCDUET-1.jpg

https://urgi.versailles.inra.fr/Tools/Predotar

https://www.discoverlife.org/mp/20m?kind=Silybum+marianum&guide=Wildflowers&cl=US/CA/Monterey/Hastings_Reserve

https://www.ncbi.nlm.nih.gov/sra/?term=silybum+marianum

https://www.wur.nl/en/product/Q-ExactivePlus-Orbitrap-LC-MSMS.htm