Metabolic Engineering of Artemisinin Pathway into *L.sativa*

Andreas Pallidis

MSc Thesis

May 2014

Student Details

Name: Andreas Pallidis

MSc Program: Plant Biotechnology

Specialization: Plants for Human and Animal Health

Registration Number: 870710640020

Hoevestein 239 13A6

6708 AK Wageningen

Email: Andreas.pallidis@wur.nl

Thesis Details

Supervisors: Sander van der Krol, Thierry DeLatte, Bo Wang, Anna Undas

Course: MSc Thesis Plant Physiology (PPH 80439)

Department: Laboratory of Plant Physiology (PPH)

Droevendaalsesteeg 1

Building 107

6708 PB Wageningen

Preface

This report summarizes my work during my MSc Thesis in the Laboratory of Plant Physiology in Wageningen University and Research Centre (WUR).

I would like to thank all my supervisors for their help and guidance for all this time. In timeline order, I would like to thank Pr. Harro Bouwmeester who gave us the 1st lecture on medicinal plants and biotechnology and motivated me to arrange an internship with this group. Then, I started working with Anna Undas (Phd) who gave me the first instructions for sampling and data analysis/statistics. Most of my wet lab work and planning/guidance was supervised by Bo Wang (Phd). I would also like to thank Dr. Thierry DeLatte for his very useful comments and input since the beginning until the end of my work.

I would like to thank Dr. Sander van de Krol for all his support and guidance during my MSc Thesis and Internship. Major part of my data analysis was supervised by Sander, but also he was guiding and helping me plan every step in my experiments. Sander also helped me to arrange and successfully complete my internship.

Part of the analysis of my MSc Thesis samples took place in Max Planck Institute (MPI) for Molecular Plant Physiology in Golm. At this point I would like to thank my supervisors in the host institute. Firstly Dr. Kopka for all the nice communication we had for arranging the internship and giving me the chance to work in MPI and his group, analysing my samples. Secondly I would like to thank Alexander Erban (research assistant) for all his effort and patience to explain me all the details of the metabolomic analysis he performs in MPI. And finally I would like to thank Dr. Steffani Schmidt for her assistance in the wet lab experiments, and her guidance all around the workflow and strict rules/regulations of MPI.

Contents

Work Summary

Malaria is a parasitic disease infecting millions of people and causing thousands of deaths, each year, especially in the African continent. Artemisinin (the most potent antimalarial drug known to date) is not being sufficiently produced in order to be provided cheaply to those in need. We investigate the potential of lettuce (*L.sativa*) to be used as a production system for the synthesis of the antimalarial drug by transiently expressing the essential enzymes (ADS, FPS, HMGR and CYP71AV1) for the artemisinin biosynthesis. Firstly we optimized the lettuce agro-infiltration method, selecting the most efficient agro-strain (AGL-0) on the most suitable cultivar tissue (*L.sativa* cv Olof) harvesting leaves on the day the highest luciferase expression (9^{th} day after infiltration). We infiltrated fully expanded lettuce leaves using ATTA, sampled and extracted (MeOH:Formic acid) for less-polar metabolite untargeted metabolomic analysis and analyzed in LC-QTOF-MS. Samples were also analyzed in GC/MS facilities (MeOH extracted) in MPI, Golm. Artemisinin related compounds/conjugations were not between detectable levels in our samples, although our infiltration enhanced metabolic shifts mainly in the secondary metabolism, while primary metabolism was less affected with only some sugars' relative concentration being altered. Most probably the boosted FPP was utilized by endogenous P450s for costunolide accumulation, but in order to verify this, targeted analysis for costunolides should be performed, as well as evaluating the expression levels of the infiltrated genes.

1. Introduction

1.1 Malaria disease

Malaria is a deadly disease, affecting mainly children (<5 years old) and pregnant women (1). Although is one of the oldest known to humans disease (WHO estimates for 2012, 3.4 billion people were at risk of malaria, from which 627.000 cases were lethal) we are not able to fight malaria efficiently. Sub-Saharan Africa and SE Asia are the most vulnerable regions, where malaria was not eradicated (during global malaria eradication program of 1955) and citizens have limited reach to medication and preventive infrastructure (e.g. bed nets).

Infections are caused by the protozoans of *Plasmodium* genus which use female mosquitos of *Anopheles* genus, as vectors. The majority of infections and deaths are caused by the species of *P.falciparum*, *P.vivax*, *P.ovale* and *P.malariae* (by order of infection rate). Parasites have evolved an elegant way of infection, on which they are able to hide or become inapproachable from the immune cells (2).

Malaria prophylaxis includes a series of medications used for infection prevention. Most common medications used include quinine derived compounds and its synthetic analogues (Meflloquine) or antibiotics (doxycycline) that result in parasites abnormal cell division. Although the mentioned drugs were able to limit the parasites' spread in the bloodstream, few Plasmodium species developed resistance, resulting in the loss of efficacy (3).

A complete effective vaccine has not been developed yet for any type of malaria (cause by any of the *Plasmodium* species). The available malaria treatments (post-infection medications) are focused on quinine/choroquine and its derivatives or synthetic analogues. Since quinine resistance was reported mainly by *P.falciparum* parasite (1920s), combination therapies of antimalarial compounds were introduced (4). Even so, the parasite developed multidrug resistance and the need for an alternative medication became apparent (5).

1.2 Artemisinin as a solution

In 1970s, a promising solution was found by Chinese researchers: a naturally synthesized compound by the native Chinese species of *Artemisia annua* (figure 1), named artemisinin, was able to cure infected individuals (4) from both uncomplicated and severe malaria. Also known as sweet wormwood, *A.annua* is a member of Asteraceae family, with average 2 meters height with alternate branches and leaves with strong aromatic odour. These leaves are covered with characteristic glandular trichomes where the majority of the plant secondary metabolites are produced and stored. Artemisinin is a sesquiterpene lactone synthesized in the aerial parts of plants. Is a colourless crystalline substance, with molecular weight of 282 kDa and chemical formula of $C_{15}H_{22}O_5$ (figure 2). As for its molecular structure is defined as a sesquiterpene containing a lactone ring, commonly described as a sesquiterpenoid, deriving from the condensation of three 5-carbon isoprenoid molecules. A peroxide bridge (present in the molecule's structure), is responsible for the antiplasmodial potency, due to its capacity to release reactive oxygen species that damage the parasites lethally (5).

Figure 1. Drawing of *A.annua* L. plant found on the bank of the Rhine river near Arnhem in September 1697 (Reproduced from Kops J. et al. 1906). All aerial parts produce artemisinin, with the highest concentration in the flowers.

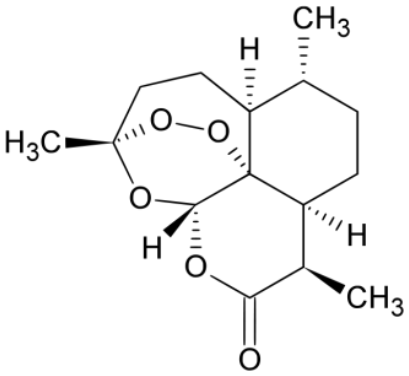


Figure 2. Structure of Artemisinin. Tricyclic structure, containing a lactone ring, and a peroxide bridge which is responsible for it antiplasmodiac activity (reproduced from free access website).

Today, artemisinin is used coupled with other antimalarial compounds (such as mefloquine and lumefantrine) in ACTs (Artemisinin Combination Therapies) in order to be more effective in the resistant parasites. This approach is highly suggested by WHO in order to avoid reduction of effectiveness and resistance development, as previously reported with other antimalarial medications. ACTs are the first line of treatment against all *P.falciparum* malaria infections (1) and its clinical use is proven to be safe (6).

1.3.1 Current production status of artemisinin

Growth of Demand: Current production systems of artemisinin cannot provide a low cost antimalarial alternative to the traditional antimalarial medication. The usage of artemisinin and derivatives in the malaria management is increasing which results in growth of demand (1). Currently, ACTs are suggested as a first line treatment for all *P.falciparum* treatments, while ACTs are also effective against *P.vivax*, *P.malariae* and *P.ovale* infections. But monotherapy is strongly discouraged in order to limit the possibilities of resistance development by the parasite as previously occurred with other antimalarial substances (1). Although *A.annua* is not the only species that accumulates artemisinin, is the one with highest registered concentration (7). As most of secondary metabolites, artemisinins can be found in trace amounts in *A.annua* plants, approximately 0.1-1% of the total dry leaf weight (8). More approaches are necessary in order to increase the global offer and therefore increase the accessibility to those in need.

1.3.2 Strategies for availability increase

In order to increase efficiency of current production systems, few approaches exist: selective breeding for richer artemisinin content hybrids, (semi-)synthetic synthesis, improve extraction/purification methods, and genetic engineering (GE) approaches. Breeding the high yield cultivars of *A.annua* plants is a relatively slow procedure, while total chemical synthesis is expensive (9). Several approaches contribute to elevated amounts of sesquiterpenes, but

we focus on GE organisms, since numerous works proved metabolic engineering can lead to accumulation of artemisinin related compounds.

1.4 Biosynthesis of Artemisinin

Metabolic engineering offers the big advantage to introduce novel metabolic pathways into organisms that naturally do not possess them. As a prerequisite, the metabolic pathway to be introduced has to be elucidated.

The biosynthetic pathway of artemisinin is almost "solidified" (10, 11), depicted in figure 3 (adapted from 12). Early precursor of artemisinin is farnesyl diphosphate or pyrophosphate (FPP), an important intermediate of the Mevalonate pathway (MEV) and terpene biosynthesis. Two key enzymes regulate the production of FPP: 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) and Farnesyl Pyrophosphate/Diphosphate Synthase (FPS/FDPS). HMGR is the rate-controlling enzyme of MEV since it is responsible for the conversion of 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) to mevalonate acid, which downstream converts to isopentenyl pyrophosphate (IPP). FPP is synthesized by FPS using IPP and DMAPP as substrates. IPP is generated mainly in cytosol (MEV) but also in plastids (MEP) (13).

The artemisinin biosynthetic pathway initiates in the glandular trichomes by the cyclization of FPP to the volatile amorpha-4,11-diene, a reaction catalysed by ADS (amorpha-4,11-synthase). Following, the CYP17AV1 enzyme (a P450 cytochrome) is oxidizing the volatile to artemisinic alcohol (AAOH) first, then to artemisinic aldehyde (AAA) and artemisinic acid (AA). Both ADS and CYP71AV1 are trichome specific enzymes (25, 26, 27). Dbr2 (Artemisinic aldehyde double-bond reductase) is responsible for the reduction of AAA to dihydroartemisinic aldehyde (DHAAA) (14) which is subsequently oxidized by aldh1 (aldehyde dehydrogenase 1) to dihydroartemisinic acid (DHAA)(15). The latter conversions of DHA to artemisinin and AA to arteannuin B, are non-enzymatic probably by spontaneous photo-oxidation (16). Although the terpene transport plays a very significant role in the accumulation of the precursors is outside of the scope of the present investigation. But is important to illustrate that AA and DHAA are accumulated in the secretory cells but transported in the apoplastic space where they are later imported in the subcuticular space. Investigating the underlying transport mechanisms of artemisinin precursors further is a topic that could give important information for pathway engineering.

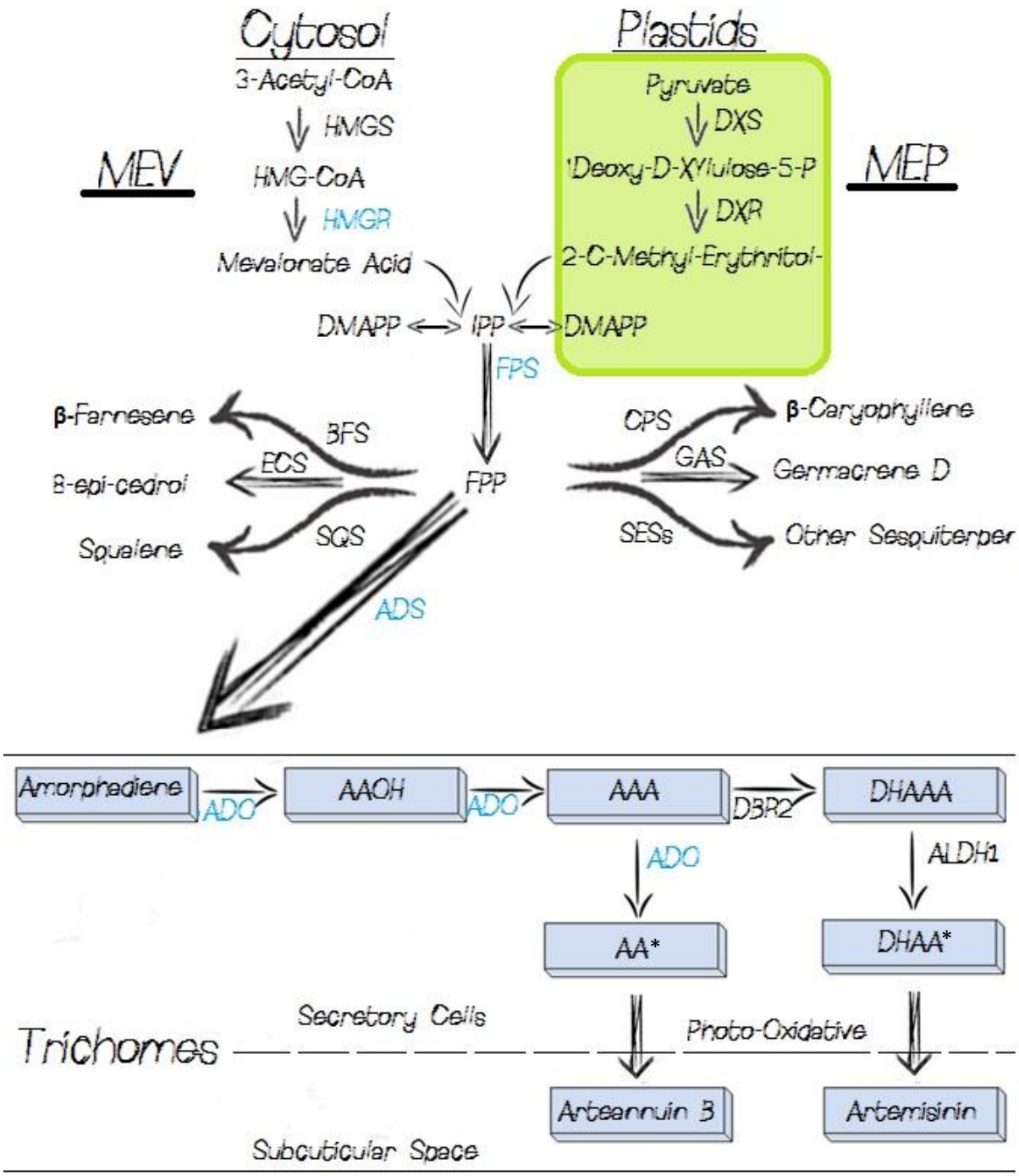


Figure 3. Biosynthetic pathway of Artemisinin in *A.annua* (adapted from Tang, K., et al.)(12). Up left: MEV pathway (cytosol) is supplying IPP which is used as substrate by FPS to synthesize FPP, the primary precursor of Artemisinins). Up right: MEP pathway (located in plastids) produces also IPP. In blue are the enzymes cloned in lettuce, in our apporach. [* symboled: The precursors are sequestred in the apoplast, before they enter the subcuticular space.]

1.5 Metabolic engineering of artemisinin pathway

1.5.1 Microbes

Approaches to use microbes as manufacturers for artemisinins exist. The initial step (13) was focused on introducing MEV and ADS to *E.coli* cells and it proved terpene precursors (amorphadiene) is possible to accumulate in microbes through metabolic engineering. A remarkable step to this direction is from Ro and co-workers (17) who used FPS, HMGR, ADS for amorphadiene accumulation and CYP71AV1/CPR for production of AA, in *S.cerevisiae* engineered strains. This strategy yielded AA and amorphadiene at "*a high biomass fraction comparable to the one produced by A.annua, but much faster*" (17). As discussed above, the latter step of the DHAA conversion to artemisinin is non enzymatic (17) in plants (photo-autoxidation) while in microbes needs exogenous chemical processing. This fact could realise plants as more suitable hosts for metabolic engineering of artemisinin.

1.5.2 *A.annua* engineering for artemisinin production boosting

Plants potential for homologous (*A.annua*)/heterologous (other genera) production of artemisinins, is mostly unrealised. As discussed above, in plants oxidation of DHAA is light and oxygen dependent, providing a 1-step-competitive advantage. The less post-harvest/extraction conversions needed, the more economically viable the system is. The advantages of plants as host systems include high scalability, high safety coupled with the low operating and capital costs.

For homologous systems (*A.annua*), the strategies are focusing on boosting the "pre-amorphadiene pools" like FPP, by overexpressing enzymes of MEV and/or MEP. Key enzymes for artemisinin precursors (FPP) production were overexpressed such as HMGR (18) where there was 22.5% increased artemisinin content, and FPS (19) and there was an up to 3fold increase in the artemisinin production (compared to control plants). Other approaches include the overexpression of ADS and CYP71AV1/CPR in *A.annua* leaves, which resulted in 2 fold increase of artemisinin (20) and in 1.5-2.4 fold increase (21) and overexpression of FPS and CYP71AV1/CPR resulted in 3fold increase of artemisinin content (22). Any attempt to boost any descendant molecules should include HMGR, due to its rate-limiting role. Boosting FPP production does not directly increase amorphadiene production, due to its utilization by endogenous competitive pathways. Squalene synthase (SQS) is utilizing FPP in favour of the production of sterols. In yeast experiments (17,23,24) down-regulation of SQS leaded to accumulation of AA, while in *A.annua* leaded to decrease of sterols and increase of artemisinin content (28,29). At this point should be illustrated that FPP is a key metabolite of primary metabolism, involved in feedback mechanisms. Accumulation of FPP may signify a negative feedback loop for its own production, minimizing the contribution of exogenous over-boosting mechanisms. Although the oxidizing capacity of CYP71AV1 is coupled with CPR, there is no experimental data highlighting the contribution of CPR to accumulation of oxidized intermediates (AOH, AAA, AA) (30). So far, the highest increase in Artemisia transgenic plants has been reported (31) was 7.65 fold higher compared to control plants, by fusing ADS/HMGR with the CaMV35S promoter. This research illustrates the rate-limiting activity of HMGR for artemisinin over-production. Other approaches include silencing of competitive pathways such as RNAi of squalene synthase gene (sterols biosynthesis), or RNAi for caryophyllene synthase. Competitive pathways are depicted in figure 3, and the

approaches to boost artemisinin production in A.annua are elaborated in detail in the review of Tang et al. (12).

Although *A.annua* plants contain all the necessary enzymatic tools to propagate artemisinin, its utilization as production platform may be of limited contribution. *Artemisia* species are not domesticated plants that can support high biomass, and the harvest/extraction procedures are time-consuming and costly. Genetic modification of more commercial crops, which cultivation is well established throughout the world, is necessary in order to propagate cheaply the antimalarial drug.

1.5.3 Heterologous plants engineering and subcellular targeting

Tobacco (*N.tabacum*) is a well-established cultivated crop, with rapid accumulation of biomass at low costs and has being extensively researched for the production of biopharmaceuticals (mainly proteins). Several remarkable efforts to use tobacco as host for artemisinin production have been made, which display diversity in the cloning genes and subcellular targeting.

As illustrated (figure 3), artemisinin accumulation takes place in the glandular trichomes of *A.annua*. Engineering the same set of genes in heterologous hosts (such as tobacco) is important to take into account the subcellular targeting, by fusing the cloning genes with targeting sequences. The potentially cloned enzymes derive from different cellular compartments: enzymes for boosting MEV pathway function in cytosol, while enzymes of MEP are in the plastids (figure 3). ADS probably is the most crucial enzyme for subcellular localization**:** when targeted in cytosol, ADS expression enhances the amorphadiene levels (34) but the accumulation of amorphadiene is higher when ADS and FPS are mitochondrial-targeted (34) and elevated amounts of AA in tobacco were displayed when ADS and CYP71AV1 was targeted in chloroplasts (33). Chloroplasts display some advantages such as high transgene expression, containment of transgenes due to maternal inheritance, and accurate protein folding (33). However, mitochondrial targeting of ADS and FPS in tobacco, displayed accumulation of artemisinin precursors (32). In these studies where CYP71AV1 was co-expressed in cytosol while ADS/FPS where fused with mitochondrial/plastid targetting proteins, still artemisinin precursors were detected. The significance of this relies on the fact that amorphadiene is accessible to CYP71AV1, and that subcellular targeting of ADS/FPS does not limit the access to amorphadiene. However, for optimizing the accumulation, a quantitative comparison of the artemisinin intermediates is necessary.

Tobacco is a well-studied research model, with available genomic/proteomic and metabolomic databases, but also its easy commercial with high biomass and well established inexpensive cultivation facilities in many parts of the world, while extraction techniques are easy to operate. The limitation of its usage as a host for artemisinin, relies on the form of artemisinin intermediates (sugar conjugative forms (32)) and obstacles of any metabolic engineering attempts (regulating mechanisms, feedback mechanisms etc.).

Other production systems can be evaluated, such as chicory (*Cichorium intybus*) or lettuce which both have quite high biomass and accumulate several sesquiterpene lactones in high quantities, and classified to the same family as *A.annua* (Asteraceae). Both plants grow quite fast (4 weeks). Another important capability of chicory/lettuce system is the containment of

endogenous cytochromes that successfully oxidize sesquiterpenes as amorphadiene (36). Cytochrome P450 hydroxylase was capable of converting amorphadiene to its corresponding alcohol, which proves that chicory P450s do not exhibit "narrow substrate specificity" (36). Moreover, the germacrene A synthase is using FPP as a substrate for the synthesis of (+)-geracryl A. One can see many similarities in the two pathways between costunolide and artemisinin, except the final step, where in the case of costunolide is an enzyme-dependent reaction and not photo-oxidation (as in artemisinin).

Our effort focuses on evaluating the efficiency of *L.sativa* leaves to accumulate artemisinin precursors. We firstly optimized the agro-infiltration technique by monitoring the luciferase expression in 4 time-points of 3 lettuce cultivars, infiltrated with 2 different agro-strains. Previous effort to detect artemisinin in ADS infiltrated lettuce has failed (37). Although lettuce leaves were expressing ADS transgene, GC/MS analysis of extracts did not reveal artemisinins, but the researchers focus was solely on artemisinin, while in our approach we attempt a wider look to any annotated artemisinin precursors and conjugations.

To do so, we infiltrate lettuce leaves with mitochondrial targeted ADS, FPS (from *ERG20*), and truncated HMGR in order to boost the FPS production (v5 cassette, see M&Ms). We co-infiltrated (in other vectors) CYP71AV1 and p19 anti-silencing vector (35). We also evaluated the ability of endogenous P450s to oxidize amorphadiene, by skipping its infiltration (substitute with p19 expressing vector).

Aim of Study

Evaluate the potency of lettuce to synthesize artemisinin related compounds, and the potential substrate drain by endogenous competitive pathway.

Research Questions

- Can we efficiently agro-infiltrate lettuce leaves with artemisinin related genes?
- Are artemisinin related compounds synthesized in lettuce leaves?
- Which artemisinin related compounds are synthesized and are they in free form or sugar conjugated?

- If artemisinins and conjugations not annotated, how the metabolic profile of lettuce leaves being altered when we make the infiltration treatments?

2. Materials and Methods

2.1 Lettuce agro-infiltration optimization

We evaluated the efficiency of lettuce agro-infiltration comparing 3 *Lactuca sativa* cultivars, 2 *Agrobacterium tumefaciens* strains, and 4 time-points. The agro-infiltration efficiency was tested using *Agrobacterium tumefaciens* strains that harbor the luciferase gene in pBin vector (AGL-0-pBin-Luc and C58C1-pBin-Luc) and the gene expression was quantified by measuring the light intensity of infiltrated leaves sprayed with luciferin (1nM). The luciferin solution was sprayed 10' before camera exposure.

2.2 Plant Material

Three different cultivars of *Lactuca sativa* (Olof /Cobham Green /Norden) were grown from seeds to soil, under greenhouse conditions with 16 h light at 28°C and 8 h at 25°C. The first 3 fully expanded leaves of 4 weeks old *L.sativa* plants were used for Agrobacterium transient transformation assay. The leaves were infiltrated on the abaxial side using a 1mL needleless syringe.

2.3 Cloning

pBin-Luc transfection. AGL-0 strains were already harbouring the luciferase gene in pBin construct (AGL-0/pBin-Luc), but cloning to C58C1 was necessary. The pBin-Luc construct was electroporated to electro-competent C58C1 cells. The luciferase intron is flanked by Pb35S promoter and T-Rbc terminator sequences. The plasmid was transferred from the AGL-0 strain to C58C1 by electroporation after plasmid isolation (using QIAprep Spin Miniprep Kit).

Luciferase expression (measurement of light intensity photos) was measured in 4 different time points, the 3rd/5th/7th/9th day after agro-infiltration.

We also evaluated luciferase activity concerning the number of genes co-infiltrated. We tested the light intensity when luciferase was infiltrated individually (AGL-0-pBin-LUC), with 1 more gene (AGL-0-pBin-LUC-EV), with 2 more genes (AGL-0-pBin-LUC-EV-P450) and with 4 more genes (AGL-0-pBin-LUC-EV-P450-Dbr2-Aldh1) depicted in figure 4.

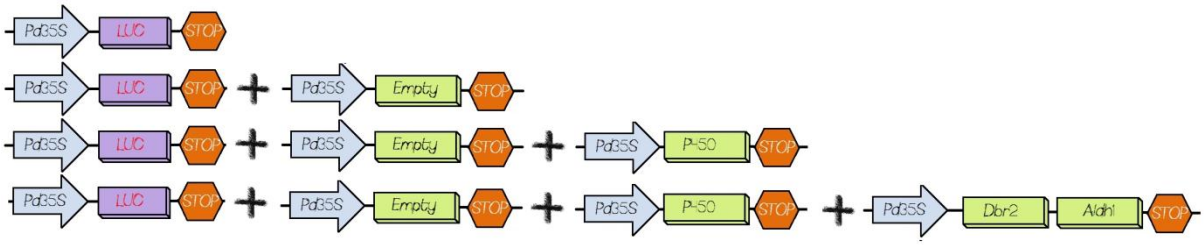


Figure 4. Constructs in pBin, used for lettuce agroinfiltration optimization. Pd35S: Cauliflower Mosaic Virus double 35S promoter. Luc: luciferase intron. P450: CYP71AV1 amorphadiene oxidase. Dbr2: aldehyde Delta11(13) double bond reductase. Aldh1: Aldehyde dehydrogonase.

Transfection to *A.tumefaciens* C58C1. The pbin-V5 construct (figure 5) along with p19 and pBin-P450 were electroporated in electro-competent C58C1 cells. Electroporation was

performed by adding 1 µl of 100 ng/µl plasmid solution to 50µl of thawed competent *A. tumefaciens* C58C1 cells in a 2 mm cuvette and using an electroporation machine set at 2.5V, 250µF and 200Ω. The electroporated cells were grown in plates containing LB non-selective LB medium (28°C), then single colonies picked and inoculated in liquid LB (shaking at 220 rpm at 28°C) containing kanamycin (50 mg/L) and rifampicillin (34 mg/L). The *A.tumefaciens* C58C1 cells are Rifampicillin resistant, while the pBin vector carries the Kanamycin resistance gene.

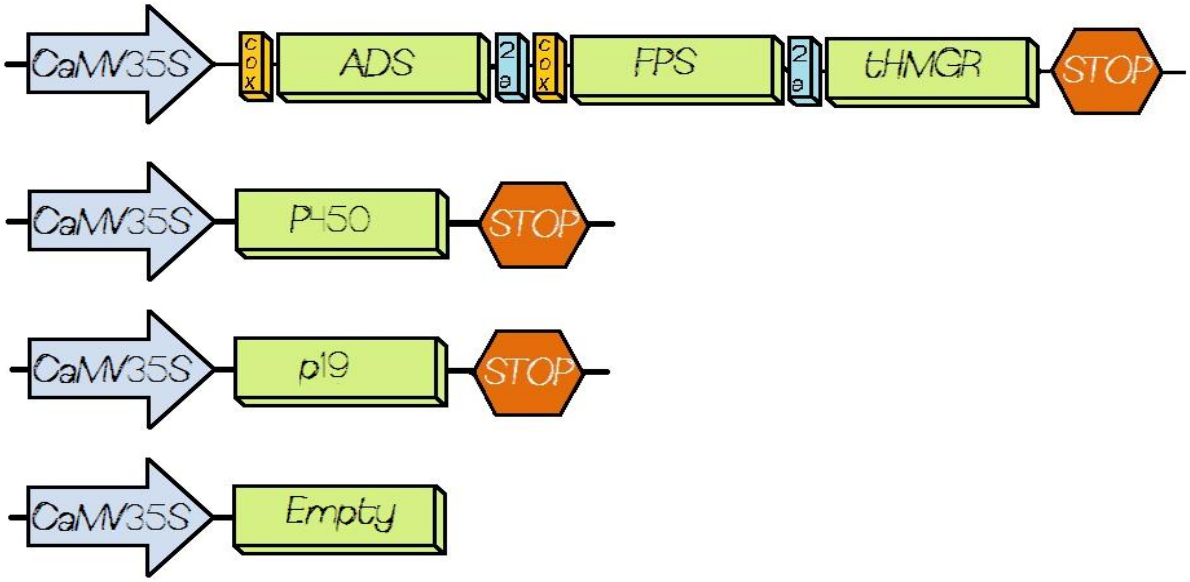


Figure 5. **Constructs of pBin, used for the artemisinin agro-infiltrations.** CaMV35S: Cauliflower Mosaic Virus 35S promoter. Cox: mitochondrial targetting signal. ADS: Amorphadiene synthase. Stop: Rubisco Terminator sequence. FPS: Farnesyl diphosphate synthase. tHMGR: truncated 3-hydroxy-3-methylglutaryl-CoA reductase. P450: CYP71AV1. P19: anti-silencing sequence

The cox sequence is a 44 amino acids long subunit which is recognized by the importing cellular machinery, allowing the transfer of the whole peptide to the mitochondrial matrix where is later cleaved. The 2a sequences allow the cassette to be expressed as three individual peptides (ribosomal skipping).

2.4 Luciferase activity assay

As a negative control we used a non-infiltrated leaf of each lettuce type (WT) and as a positive control we used an *Arabidopsis* leaf harbouring the luciferase gene. The evaluation for 1,2,3, 5 genes number was performed using 2 lettuce cultivars (Norden and Cobham Green) and the pictures were taken after 2 minutes of rest (to avoid chlorophyll emission detection) + 7 minutes of exposure. Leaves were 4 weeks old. Luciferase substrate (1 nM luciferin) was sprayed on top of the leaf 10' before camera observation. Each photograph was taken using sensitive camera, and firstly a photograph on light was taken (in order to adjust leaf surface). The background noise (level of negative control) was not subtracted, and it is shown in the figures. The photos were taken 6 days after infiltration. For the 5 genes construct we used also the Olof cultivar. All gene combinations were contained in AGL-0 strain.

2.4 Transient expression in *L.sativa* leaves

The agrobacterium transient transformation assay was performed based on Van Herpen et al. (32) method. After growth for 24h in LB liquid selective medium, C58C1 cells were harvested by centrifugation for 15 min at 5000 g and 28ºC and then resuspended in 0.1 mM MES buffer containing 10 mM MgCl2 and 0.1 mM acetosyringone. OD_{600} was set at 0.5 (adjusting using Agro-Infiltration buffer) followed by incubation on a roller-bank at room temperature for 3h. When co-infiltration was needed, we used equal volumes of each strain. In order to keep the gene dosage constant and comparability between the gene combinations, combinations with less unique candidate genes were supplemented with C58C1 cultures containing the empty pBin vector (EV). In all treatments (except control) a C58C1 strain harbouring a gene encoding the TBSV P19 protein was added to maximize protein production by suppression of gene silencing (35). *L.sativa* cv Olof plants were grown from seeds on soil in a greenhouse with 16 hours of light at 28ºC (16 h) / 25ºC (8 h). Strain mixtures were infiltrated into leaves of five-week-old plants using a 1 mL syringe. The bacteria were slowly injected into the abaxial side of the leaf. The plants were grown under greenhouse conditions and harvested 9 days after infiltration. The infiltration scheme of the gene combinations is depicted in the table below. For each treatment, 3 plants were used, and 3 leaves were infiltrated per plant, resulting in 9 replicates per treatment.

Table 1. Infiltration Scheme (per leaf)

	C58C1 Empty vector	**C58C1-V5**	**C58C1-p19**	**C58C1 CYP71AV1**
Control	3x	-	-	-
Treatment 1	-	1	1	1
Treatment 2	1	1	1	-

2.5 Sampling for LC-QTOF-MS/MS

The sampling was performed based on the protocols of Van Herpen et al. (32). Non-volatile compounds were analysed using a protocol for untargeted metabolomics of plant tissues. 100 mg of infiltrated leaf material from each plant was grounded in liquid N_2 and extracted with 300 µL methanol:formic acid (1000:1,v/v). The extracts were prepared by brief vortexing and sonication for 15 min. Then the extracts were centrifuged (5' at max speed) and the supernatant was filtered through 0.45 µm inorganic membrane filters (RC4, minisart). LC–PDA–MS analysis was performed using a Waters Alliance 2795 HPLC connected to a Waters 2996 PDA detector and subsequently a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS technologies, UK) operating in negative ionization mode. The column used was an analytical column (2.0 mm x 150 mm; Phenomenex, USA) attached with a C18 pre-column (2.0 mm x 4 mm; Phenomenex, USA). Degassed eluent A (ultra-pure water:formic acid [1000:1,v/v]) and eluent B (acetonitrile:formic acid [1000:1,v/v) were pumped at 0.19 mL min-1 into the HPLC system. The gradient started at 5% B and increased linearly to 35% B in 45 minutes. Then the column was washed and equilibrated for 15 minutes before the next injection. The injection volume was 5 µL. The MS–MS measurements were done with following collision energies of 10, 15, 25, 35 and 50 eV. Leucine enkaphalin ([M–H]-=554.2620) was used as a lock mass for on-line accurate mass correction.

2.6 Sampling for GC/MS analysis

After harvest the leaves were powdered using liquid N_2, and 150mg (+/- 5mg) of powdered material was stored in 2ml Eppendorf tubes (for non-volatile analysis) and 500mg for volatile analysis in the MPI including 1 back up sample (see appendix). The samples were stored in -80°C until the day of transfer (2 weeks) and dried ice was used for transport in insulated box.

2.6.1 Polar phase Non-Volatile extraction

For polar phase extraction of the samples, 720µl of Pre-Mix (see table 3) were added in each of the 2mL Eppendorf tubes and shook at 70°C for 15'. After 1' of incubation the caps were opened to relieve gas pressure and cooled down to room temperature. 200ul of CHCl3 were added, shake at 37°C for 5' and 400µL of H_2O were added. The samples were vortexed and centrifuged again for 5' at 14000 rpm. The upper (polar) phase was aliquoted in 1.5ml tapered Eppendorf tubes and the solvents were evaporated in centrifugal evaporator (SpeedVac).

Table 2. Chemical Composition of Premix for Polar phase extraction

Component	**Volume (µL)**
MeOH (100%)	600
Nonadecanoic acid methylesterr (2mg/ml in $CHCl_3$)	60
Premixture D-sorbitol-$^{13}C_6$ 0.2mg/ml in MeOH and d4-Alanine 1mg/ml in H_2O	60

Before measurement, the dried samples were derivatized by adding 40µL of Methoxyaminhydrochlorid (5mg/ml DMAP in pyridine added to 40 mg/ml MeOX), shake at 30°C for 90 minutes, spin down, added 80µL of DeMix (10µL alkan-mixture + 70µL of MSTFA/BSTFA), shook and spin down as before. The 80µL were transferred to GC sample vials including 1 empty vial with all previous solutions as wash, and an additional QC-K (Quality Control Mix) and MM (Multi-Mix) composed by all sample types of the experimental design for calibration. Along with them, 2 non-sample controls were prepared (sample-free vials) that contain all the rest of mixes and are used for removing any background noise and/or contaminations. The QC-mix and Master-mix are used for performance control/quality drifts of the machinery. Derivatization of the samples is necessary in order to convert non-volatile compounds to volatile compounds that are GC-MS detectable. Therefore, the detection limits of the analysis are higher and the specificity lower.

2.6.2 Non-polar extraction

Parallel to the classic polar extraction, another protocol was used, focusing primarily on artemisinin conjugations. 500µL of 2N KOH in MeOH were added to 150mg of sample and incubated for 1h in 70°C. After incubation the samples were transferred on ice and 80µL of HCl were added for neutralization. 300µL of Hexane were added, and the samples were mixed for 5 minutes in 14000 rpm. 200µL of the non-polar phase were transferred to new vials and dried out in SpeedVac (overnight).

2.6.3 Volatile analysis

The samples for volatile analysis were directly sent for analysis (after derivatization and addition of internal standard).

2.6 Data processing for LC/MS analysis

LC/MS data processing was based on van Herpen (32). LC–MS raw chromatograms were firstly processed using MassLynx 4.0 (Waters). This software allows the acquisition, visualisation and manual processing of raw chromatograms from both GC/LC-MS. In this step we manually checked for the presence of artemisinin related compound sugar conjugations using the results of previous experiments (32) on *N.benthamiana* plants (table C, appendix). After evaluation of raw chromatograms (check for empty graphs, removal of artefacts, peak overloads) the data is inserted in MetAlign (by RIKILT Wageningen UR, Plant Research International). This software allows pre-processing and comparison of full scan nominal or accurate mass LC-MS and GC-MS data. Using MetAlign we baseline corrected our chromatograms (up to the noise level) and align the mass peaks throughout all samples. The processing parameters data were set to analyse from scan numbers 60-2590 (corresponding to retention time 1.15-49.16min) and maximum amplitude 25.000. This analysis generated a mass peak alignment matrix of about 15000 metabolites per treatment, and also the possibility of a metabolite to be represented by more than one mass peak (isotopes, etc) is high. Therefore we used MS Clust (42) for data reduction and mass spectra retrieval. MS Clust, groups the metabolites based on retention time distance as well as intensity similarity distance. Multiple mass signals derived from the same compound were grouped in MSClust by Multivariate Mass Spectra Reconstruction (MMSR) (42).

2.7 Data analysis of LC/MS output

The generated file with the filtered metabolites from MSClust was used for data analysis. Firstly the mean, standard deviation and the ratio of standard deviation to mean were calculated. TWe used Multibase (Numerical Dynamics) for plotting the PCAs (Principal Component Analysis). The top 10 contributors of each treatment were used for visualization of data. The data were plotted pairwise.

2.7 Data Processing for GC/MS analysis

Preliminary (without pre-processing) evaluation of raw chromatograms is performed in Masslynx Mass Spectrometry Software (by Waters) or Tag Finder (for specific setting see 6.3 paragraph of Appendix). This software allows the acquisition, visualisation and manual processing of raw chromatograms from both GC/LC-MS. After evaluation of raw chromatograms (check for empty graphs, removal of artefacts, peak overloads) the data is inserted in MetAlign (by RIKILT Wageningen UR, Plant Research International). This software allows pre-processing and comparison of full scan nominal or accurate mass LC-MS and GC-MS data. Using MetAlign we baseline corrected our chromatograms (up to the noise level) and align the mass peaks throughout all samples, using default settings. For the main data processing we used Tag Finder (exact settings of processing shown in appendix). Most important settings should be mentioned is the range masses (outside range is considered as noise and excluded) 76-146; 150-600 (exclusion of nonspecific mass traces). The data were normalized according to the internal standard The RI was calculated and all compounds were shorted according to RI in order to increase speed of processing.

2.7.1 Targeted analysis and untargeted analysis

We manually search for known masses and retention indexes, using Golm Metabolome Database as reference (table 3):

Table 3. Analyte Masses and Retention times used in Targeted Analysis

Analyte Name (#Tetramethylsilane Groups)	Analyte Mass	Retention Time (seconds)
Amorpha-4,11-diene	204 (main mass)/189(secondary)	1564.08
Amorpha-4,11-diene (side product)	189(main mass)/204(secondary)	1564.08
AA (1 TMS*)	306.516	1853.33
AA (0 TMS)	234.335	1880.3
AOH (1TMS)	292.532	1785.76
AAOH (0 TMS)	220.35	1773.27
AAA (0 TMS)	218.34	959.54

*TMS stands for Trimethylsilyl, the derivatize agent for alcohols/phenols/carboxulic acids.

The mass spectra of these analytes are shown in figures A-E in appendix section.

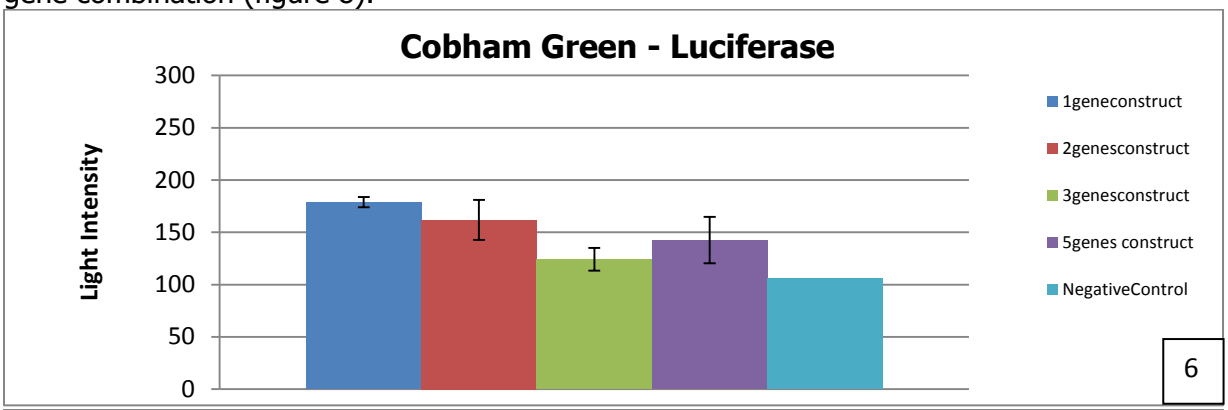
Since none of the above metabolites was detected, an untargeted analysis was performed. We focused on metabolites of primary metabolism that displayed variability in their relative abundance.

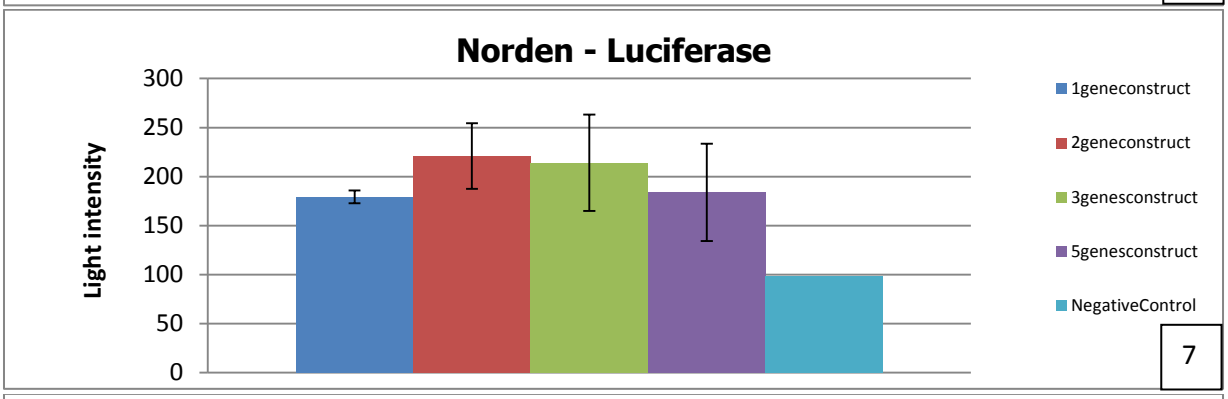
3. Results

3.1 Lettuce Agro-infiltration optimization

3.1.1 Infiltration using different numbers of genes

Here we show the results of lettuce agro infiltration optimization. Firstly we evaluated the gene expression of luciferase in lettuce infiltrated leaves, between 4 different numbers of genes (1, 2, 3, and 5) and the results are shown in figures 6-8. The leaves were sprayed with 1nM luciferin, 10' prior to camera exposure. All gene combinations were contained in AGL-0 strain. Results are shown per lettuce type (figures 6 and 7) and all lettuce types for 5 gene combination (figure 8).





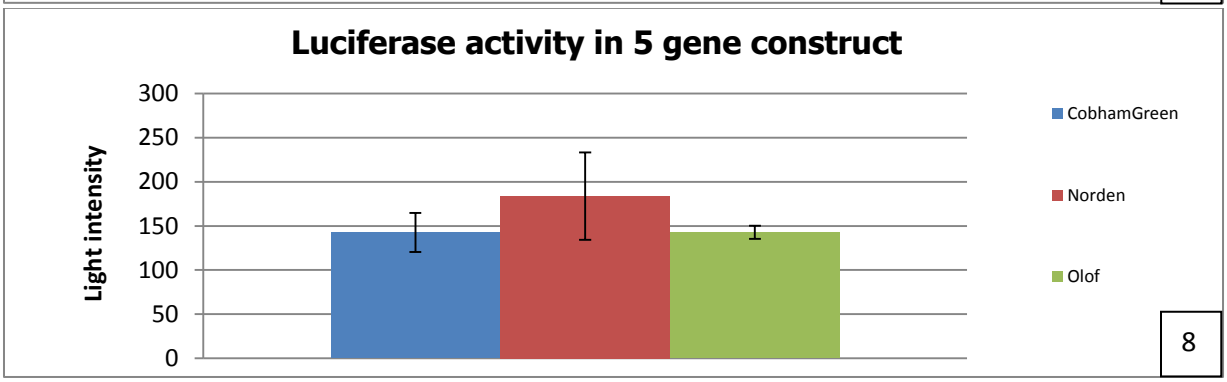


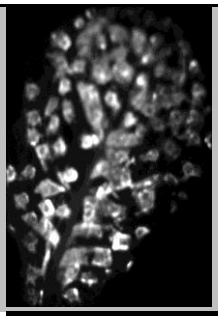
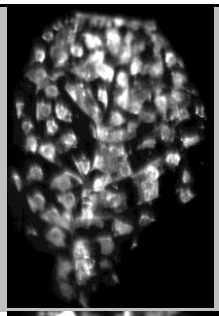
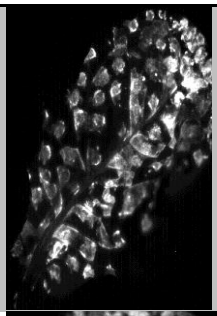
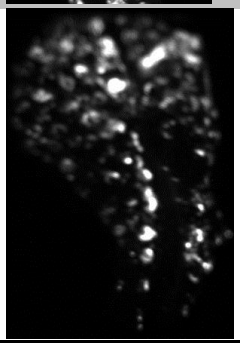
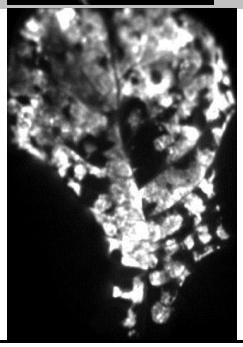
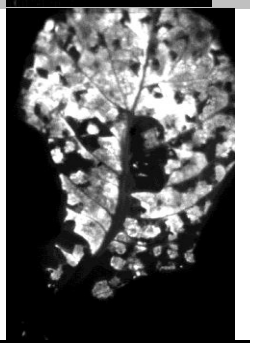
Figure 6. Cobham Green infiltration with luciferase, using different gene numbers (1,2,3 and 5). Negative control: non infiltrated leaf. Error bars represent the standard deviation.
Figure 7. Norden infiltration with luciferase, using different gene numbers (1,2,3 and 5). Negative control: non infiltrated leaf. Error bars represent the standard deviation.
Figure 8. Three lettuce cultivars, infiltrated with luciferase gene in combination with 4 more genes. Negative control: non infiltrated leaf. Error bars represent the standard deviation.

There is no significant difference between the gene numbers used. The constructs that are co-infiltrated do not influence the expression of luciferase in the three lettuce types, using the AGL-0 agro-strain.

3.1.2 Infiltration using 2 different agro-strains (C58C1 and AGL-0)

Here, we infiltrated the lettuce cultivars with two different agro-strains, C58C1 and AGL-0, both carrying the luciferase gene. The difference of luciferase expression in Olof cultivar is well illustrated the photos in table 4 (summarized photos of infiltrations in table B in appendix), and the measurement of light intensity is shown in figure 10. The data were normalized to the total leaf surface. 4 replicates per leaf were measured.

Table 4. Light Intensity of Olof infiltration with 2 agro strains harbouring the luciferase gene

	3rd	5th	7th	9th
AGL-0				
C58C1				

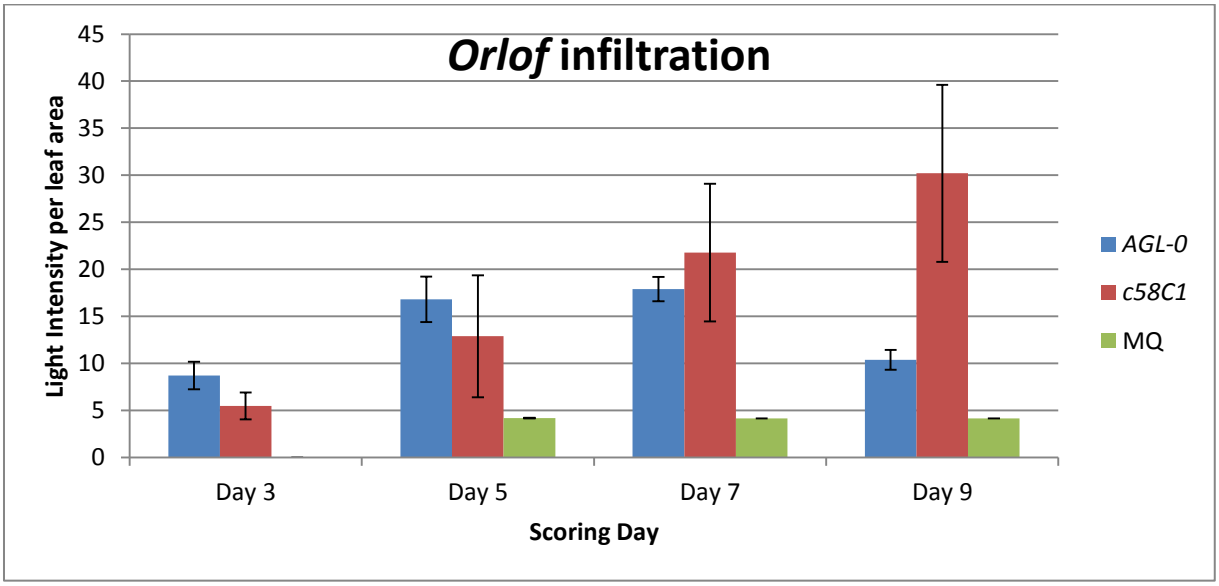


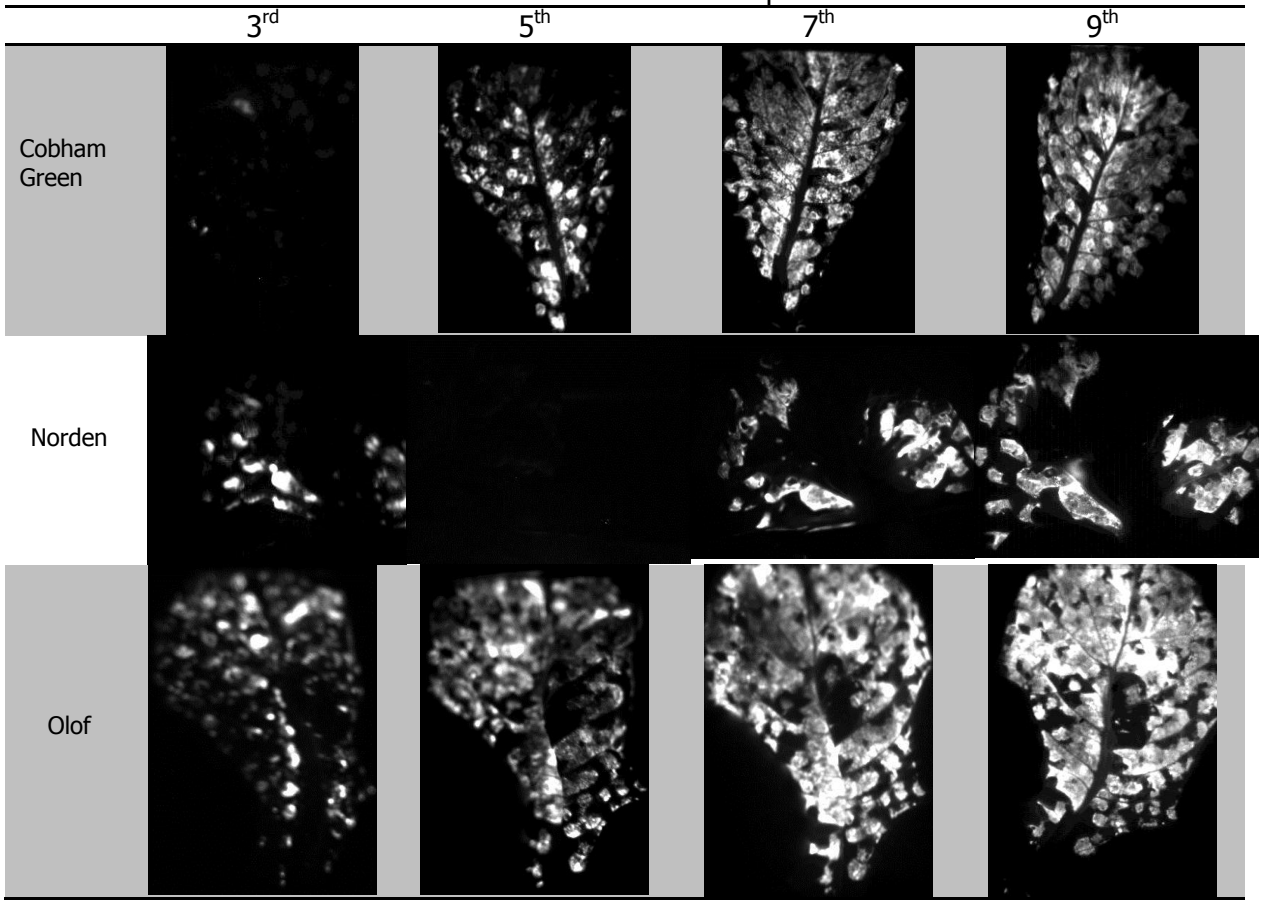
Figure 10. *Orlof* cultivar infiltration with two agrostrains, C58C1 and AGL-0, both harbouring the luciferase gene. Negative control: Orlof infiltrated with MQ. Error bars represent standard deviation of replicates.

The figures above reveal that C58C1 strain carrying the luciferase gene, yields higher luciferase activity (light intensity) than the AGL-0 strain, in Olof cultivar, on the 7th and 9th scoring days. Although the first two scoring days the C58C1 strain shows less expression, the 7th and 9th day the expression is considerably higher. The light intensity using the C58C1 agro strain was higher compared to any other lettuce cultivar (summarized data shown in appendix, figure F-G).

3.1.3 Infiltration using 3 different lettuce types (Olof, Cobham Green, Norden)

Here, we infiltrated the three lettuce cultivars (Olof, Cobham Green, Norden*)* with C58C1 agro-strain, carrying the luciferase gene. The difference of luciferase expression is well illustrated the photos in table 5, and the measurement of light intensity is shown in figure 11. The data were normalized to the total leaf surface. 4 replicates per leaf were measured.

Table 5. C58C1 Luciferase Expression



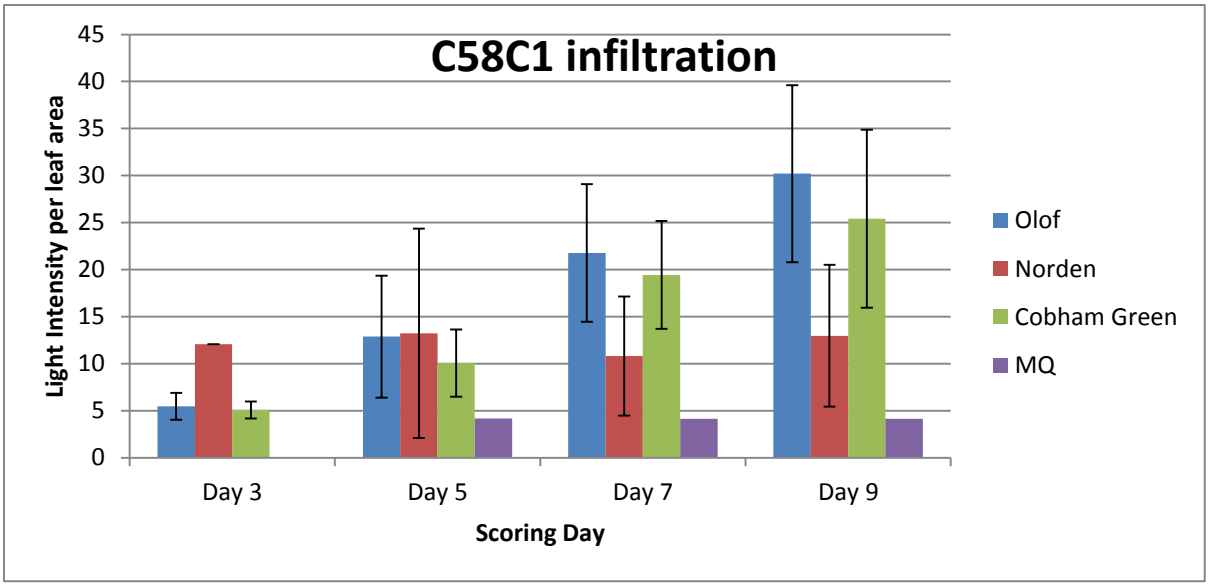


Figure 11. Three lettuce cultivars infiltration with two agrostrains, C58C1 harbouring the luciferase gene. Negative control: Orlof infiltrated with MQ. Error bars represent standard deviation of replicates.

3.1.4 Optimal day for scoring luciferase expression

Here we evaluated the optimal scoring date of lettuce infiltrations, using both agro-strains. The comparison of lettuce photos is well depicted in the table 4 (from previous section). In figure 12, the infiltration with C58C1 strain is shown.

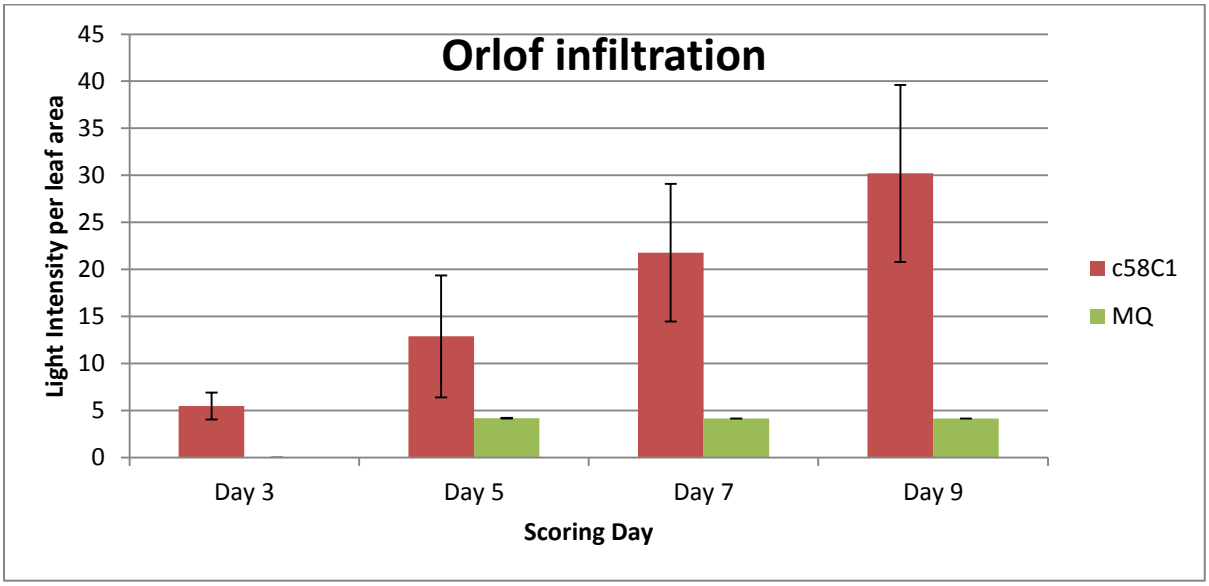


Figure 12. Olof infiltration using the C58C1 strain harbouring the luciferase gene.

From above figure is clearly shown that the highest scoring in light intensity was during the 9th day after infiltration.

Summary of optimization infiltrations. In the graphs above is illustrated the highest luciferin expression on the 9th day of infiltration in Olof cultivar, using the C58C1 strain. Therefore, this combination of parameters (Olof, 9th day, C58C1) was selected for the artemisinin related genes agro-infiltration. The complete comparison of all lettuce cultivars and both agro-strains in all scoring dates are shown in figures F-G and table B of appendix section.

3.2 LC/MS targeted analysis

We infiltrated lettuce leaves, using C58C1 *A.tumefaciens* strain, containing the artemisinin related genes (as shown in table 1). Firstly we searched for the free forms of artemisinin related compounds and conjugations (table C, in Appendix) based on data derived from *N.benthamiana* results (32). None of the listed conjugations was between detectable levels in our samples.

3.3 GC/MS targeted analysis

Here we used the same samples as previously described (table 1) for GC-MS analysis in the MPI facilities. These lettuce leaves were infiltrated with C58C1 *A.tumefaciens* strain containing the artemisinin related genes. Here we also searched for artemisinin precursors (AA (1TMS), AOH (1TMS), AAA, AA, AOH) the raw chromatograms, based on the retention times and masses listed in table 3. These metabolites were not in detection levels in our chromatograms, deriving from all treatments.

3.4 Untargeted analysis of GC/MS data

Here we used the same samples as previously described (table 1) for GC-MS analysis in the MPI facilities. Since targeted analysis for artemisinin related compounds and conjugations did not display results, we proceed with untargeted analysis of metabolites that are up/down regulated. Results are shown in figure 13 below. Rest of non-annotated metabolites that were detected in the GC/MS untargeted analysis are included in figure H of appendix.

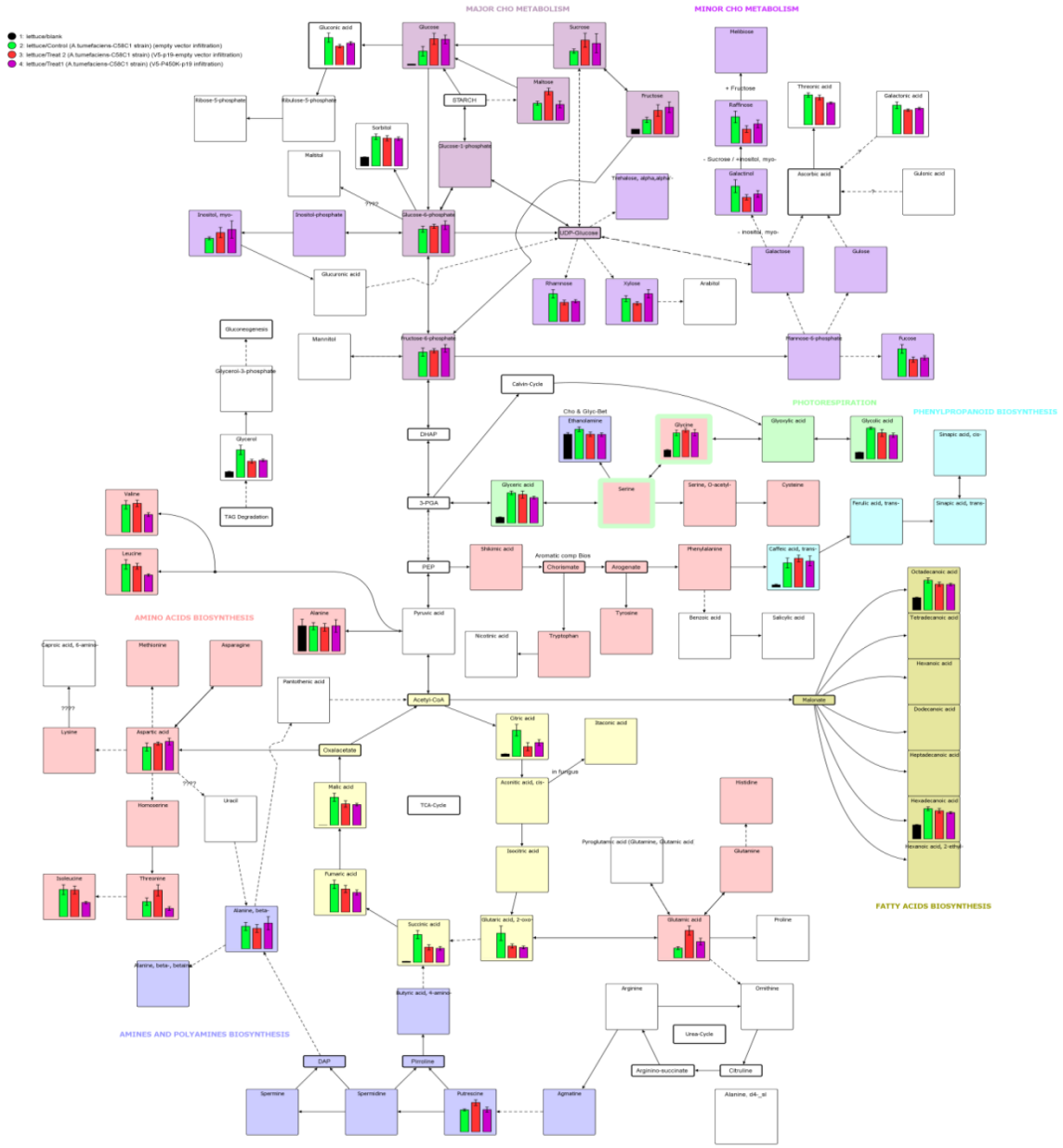


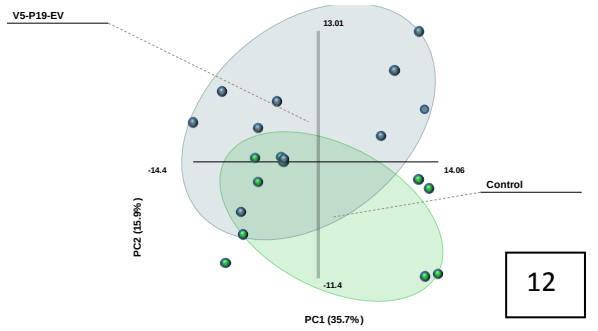
Figure 13. Scheme of untargetted metabolomic analysis in GC/MS. All the metabolites of primary metabolism that showed variation in relative abundance are depicted.

3.5 Untargeted analysis of LC/MS data

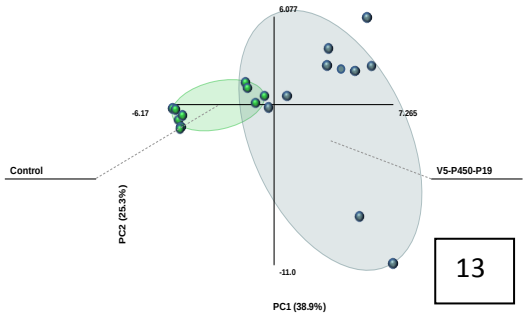
Here we describe the results of the lettuce agro-infiltrated leaves with artemisinin related genes, analysed on LC-QTOF-MS. The Olof cultivar leaves were infiltrated with the artemisinin related genes (V5-p19 and EV/P450) as shown below (pairwise comparison of treatments):

A. Control against V5-p19-Ev infiltration
B. Control against V5-P450-p19 infiltration
C. V5-P19-EV against V5-P19-EV infiltration

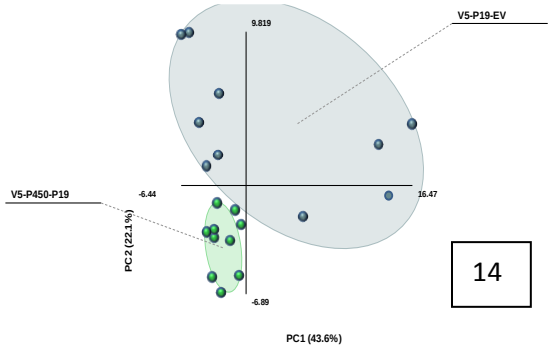
The selected figures represent only metabolites that have lower 0.5 (figures 12-14) or higher than 2 (figures 15-17) mean ratios between the two treatments. For the plot construction the 10 metabolites that contributed mostly in the graph are used.



12

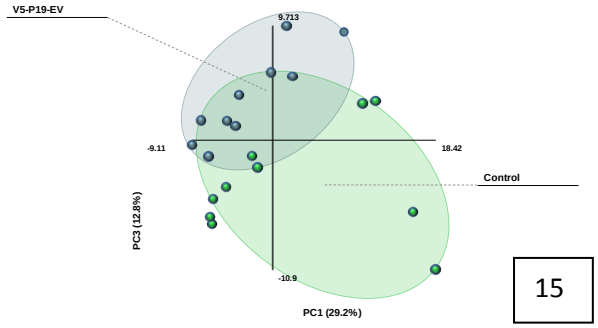


13

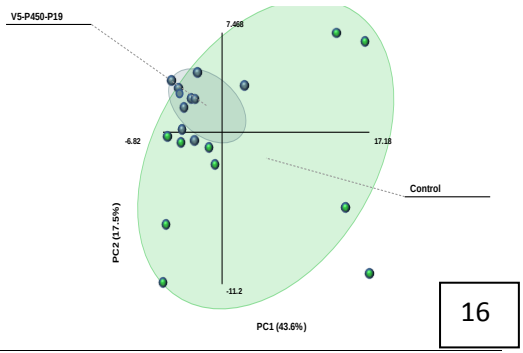


14

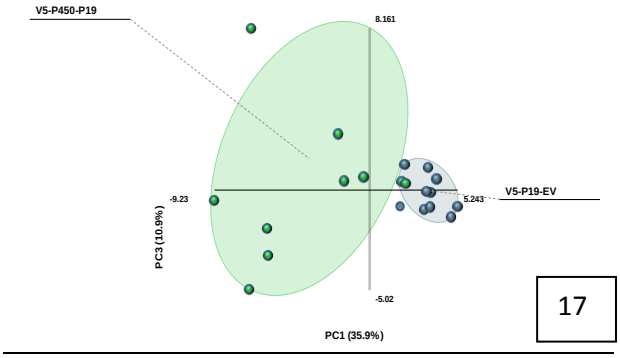
Figures 12-14. PCA plots of top 10 contributing metabolites, with mean ratios less than 0.5. **Figure 12**: Group analysis between control (infiltration with 3 EVs) and treatment 1 (V5-P19-EV). **Figure 13**: Group analysis between control (infiltration with 3 EVs) and treatment 2 (V5-P450-P19). **Figure 14**: Group analysis between treatment 1 (V5-P19-EV) and treatment 2 (V5-P450-P19).



15



16



Figures 15-17. PCA plots of top 10 contributing metabolites, with mean ratios higher than 2. **Figure 15**: Group analysis between control (infiltration with 3 EVs) and treatment 1 (V5-P19-EV). **Figure 16**: Group analysis between control (infiltration with 3 EVs) and treatment 2 (V5-P450-P19). **Figure 17**: Group analysis between treatment 1 (V5-P19-EV) and treatment 2 (V5-P450-P19).

4. Discussion

4.1 Lettuce agro-infiltration optimization.

In order to optimize the agro-infiltration of lettuce with artemisinin related genes, we firstly optimized the infiltration by measuring luciferase transgene expression in lettuce. We studied the variability of gene expression of luciferase by co-infiltrating with different numbers of genes, in three different lettuce cultivars, using 2 different *A.tumefaciens* strains (harbouring the luciferase gene), harvesting on 4 different time points (scoring days). The luciferase gene expression was evaluated by spraying luciferin (1nM) substrate on the infiltrated leaves and measuring the light intensity on different scoring days. This information provides more information about the transgene expression in lettuce, and we used that information in order to maximize the transgene expression of artemisinin related genes.

Concerning the permeability of leave's cuticle and homogenous agro-infiltration, the Cobham Green cultivar was the easiest to penetrate. The buffer containing the agro-strains was spread to the leaf surface, almost throughout the whole leaf. Following easiest was Olof and the lowest spread was observed in Norden which displays a quite waxy thick cuticle.

4.1.1 Lettuce infiltration with various gene number infiltrations.

We evaluated the infiltration with a different number of gene constructs (1, 2, 3 and 5). The light intensity of luciferase between these constructs did not differ significantly. Thus we conclude that between these number of constructs, the gene expression levels do not differ between the lettuce cultivars we infiltrated using the AGL-0 strain. Although repeating the same infiltration using C58C1 strain would be interesting, the gene constructs were not available in this strain. Cloning the same gene combinations in C58C1 and re-evaluating our results would be important.

4.1.2 Infiltration using 2 different agro-strains (C58C1 and AGL-0)

Here we used 2 different agro-strains (C58C1 and AGL-0) both harvesting the luciferase gene. We sprayed luciferin on top of each leaf and we measured the light intensity on sensitive camera. Concerning the agro-strains, C58C1 showed higher luciferin activity (light intensity) on the 7^{th} and 9^{th} scoring days. In the first two scoring days, the trend was different as the AGL-0 strain showed higher expression, but the expression on the latter scoring days was higher. Therefore, for infiltrating the artemisinin related genes, we selected C58C1 as strain.

4.1.3 Infiltration using 3 different lettuce cultivars (Olof, Cobham Green, Norden)

Here we tested the susceptibility of three lettuce cultivars (Olof, Cobham Green and Norden) to the infiltration with the agro-strains. As discussed earlier, the spread of the buffer during infiltration was not uniformal, with Cobham Green showing the highest leaf buffer spread followed by Olof and Norden. But this was not in accordance with the highest light intensity deriving from the leaves. Olof displayed the highest light intensity when infiltrated with the C58C1 strain, and each day the expression was higher (9^{th}>7^{th}>5^{th}>3^{rd}). The trend was not when the AGL-0 strain was used (data shown in appendix), where the intensity was higher on the 5^{th} and 7^{th} days of scoring. Nevertheless, comparing the lettuce cultivars light intensity we observed that Olof displays the highest expression when the C58C1 strain was used and was selected for the artemisinin related genes infiltrations.

3.1.4 Optimal day for scoring luciferase expression

Here we studied the optimal scoring date for highest gene expression. The selected days were the 3rd, 5th, 7th and 9th day after infiltration with the agro-strains containing the luciferase gene. The luciferase expression was quantified by measuring the light intensity of infiltrated leaves, sprayed with 1nM luciferin.

The highest light intensity was measured on the 9th scoring date in Olof cultivar, using the C58C1 strain. The trend between the two agro-strains used was different: when C58C1 strain was used, the highest expression was observed in the later scoring dates (7th and 9th) while using the AGL-0 strain, the gene expression was highest in the middle days (5th and 7th).

Luciferase expression between agro-strains, cultivars and time points. Concerning the highest luciferace expression (light intensity) between the two agro-strains, Olof cultivar displayed the highest luciferase expression levels when infiltrated with C58C1, on the 9th day of infiltration. It is important to illustrate that the leaves of the same cultivar have also the widest surface. Although higher leaf surface could correlate with higher spreading of the infiltration buffer, as is displayed from the pictures, the light intensity in Olof is displayed from the whole leaf, while the light intensity in the other two cultivars derived mostly from individual infiltration spots. The lettuce agro-infiltration was previously evaluated (38), and the C58C1 agro-strain displayed the highest gene expression with very low necrotic response displayed from the lettuce leaves.

Apparently, the buffer spread we observed during the infiltration (Cobham Green>Olof>Norden) does not match with the light intensity results (Olof>Cobham Green>Norden). As discussed above, this is due to the fact that luciferase expression in less easily infiltrated leaves (Cobham Green and Norden) was deriving from individual infiltration points and not the whole leaf surface. Which means that the gene was expressed mostly in the infiltration points and not from the whole buffer spread surface.

Based on these results, we designed the experiment for artemisinin genes infiltration using the Olof cultivar, the C58C1 agro-strain and harvesting on the 9th post-infiltration day.

4.2 Lettuce agro-infiltration with artemisinin genes

Here we show the results of lettuce infiltrated with artemisinin related genes. We evaluated the presence of artemisinin related compounds in free form (targeted analysis) as well as in conjugated form (untargeted analysis)

GC/MS analysis. In GC/MS no artemisinin related compounds were observed in the corresponding chromatograms of all our samples (control and treatments). In GC/MS analysis we searched for artemisinin precursors (AAA, AAOH, AOH, AA) as shown in table 3. The reference retention times and mass spectra for this analysis derived from the Golm Metabolome Database (GMD). None of the listed compounds was between detection levels in our samples. The GC/MS approach for untargeted analysis was focused in primary metabolism. As depicted in figure 10, the focus is given to major and minor carbohydrate metabolism, amino acid and fatty acid biosynthesis and citric acid cycle.

TCA metabolites display less relative concentration. As we explain in our hypothesis, our untargeted approach aimed on observing metabolic shifts in primary metabolism between our control and the two treatments. Concerning the more relevant to artemisinin biosynthesis pathway, focus should be given in TCA cycle, which is also the area that most metabolic shifts are observed. Trend is that main metabolites of the pathway (citric acid, glutaric acid and succinic acid) in our transgenic leaves (treatment #1 and #2) are in lower levels compared to the control ones. The same trend but in less extent is followed by the other two metabolites (malic and fumaric acid). Although our approach was not quantitative and targeted analysis of MEV and MEP was not performed, the reason for the repression of these metabolites could be a feedback regulation mechanism between MEV/MEP and TCA. Boosting the expression of HMGR and FPS enhances the accumulation of FPP which has previously discussed to be causing a negative feedback to MEV when accumulated (24). Moreover, a similar feedback mechanism was discussed (39) when AA was externally applied to *A.annua* leaves and ADS transcript levels dropped. A similar mechanism could exist between other MEV precursors and TCA cycle, but to fortify this argument a study focusing specifically on the MEV metabolites (while overexpressing HMGR and FPS) because MEV is a regulated pathway and conclusions should be made on experimental data.

Concerning the major carbohydrate metabolism, some sugars (glucose, fructose and sucrose) are following the opposite trend of the TCA cycle metabolites, but in less extent. The infiltrated leaves display higher abundance than control treatment. While in the minor carbohydrates metabolism, raffinose and galactinol show higher abundance in control treatment.

On the other hand, LC/MS analysis was focused on secondary metabolism. While establishing the experimental set up, we selected to analyse our samples in electrospray ionization (ESI) in negative mode. Although lettuce and Artemisia sesquiterpenes lactones are readily detected in positive ESI (41), we focused on the artemisinin conjugations that could be accumulated in our treatments. Moreover, targeted approach for sugar conjugations should include reference sugar conjugated precursors that we did not possess. Finally, study the metabolic alternations caused by the introduction of a novel biosynthetic pathway in lettuce most probably would generate metabolic shifts difficult to predict beforehand. As shown by the PCA plots, our treatments show constant variability.

Secondary metabolism altered wider than primary. Comparing the outputs of the two analyses (LC-QTOF-MS and LC/MS) we see that secondary metabolism is alternated in more points than central metabolism. Plants have regulating mechanisms in order to conserve/protect their primary metabolism against externally applied alternations, which this is one of the main bottlenecks of metabolic engineering of plants.

Analysis of significantly differ metabolites between GC/MS and LC/MS analysis. Here we attempted to link the metabolites that are up/down regulated in LC/MS with the metabolites that up/down regulated in GC/MS analysis. To do so, we randomly selected some centrotypes (clustered metabolites) from the PCAs top 10 metabolites, and trace back the metabolites they refer to. We then checked if the same retention time metabolite was present in the GC/MS datasheets but with no success. This is not evidence that the

metabolites are not present in GC/MS datasheets, but could not be manually found due to time shifts that may exist between LC and GC data.

Analysis of significantly differ metabolites between treatments in LC/MS data. In this analysis we attempt to link the metabolites that differ significantly between our lettuce treatments, and are depicted in LC/MS PCA analysis. Firstly, we compared the centrotypes that show significant difference between all pairwise treatments (Control vs Treatment#1, control vs Treatment#2 and Treatment#1 vs Treatment #2) with a ratio less than 0.5 (2 fold increase). The analysis showed that all the centrotypes that differ significantly were different between the treatments. There was no similar centrotype that differed significantly between the pre-mentioned comparisons. Although for the pairwise treatment comparison (treatment#1 vs treatment #2) this result was expected, one would expect that for the comparisons between the control treatment (3 empty vectors) against the treatment#1 and treatment#2 to show some similarities. Moreover we picked random metabolites from the ones that significantly differ in the treatment#1 vs treatment#2 comparisons and compare their mass (Da) with the mass of closely related centrotypes. We searched for mass changes close to 16Da (mass of hydroxylation group) in order to evaluate the activity of P450, but our effort did not yield any results.

Endogenous competitive pathway of lettuce sesquiterpenes (costunolides). Concerning the secondary metabolism, we observed alternations but not in favour of artemisinin conjugations/precursors. As discussed earlier, lettuce's capacity to accumulate a series of sesquiterpene lactones establishes it as an attractive platform for transgene expression of non-endogenous sesquiterpenes (as artemisinin) but at the same time there is a drawback. When MEV is boosted, the FPP accumulated can utilized for endogenous pathways, rather than the externally infiltrated, as in our case the artemisinin pathway. One of these competitive biopathways is the costunolide formation from FPP, as previously reported (36). The resemblance between the two pathways is illustrated in figure 6.

As depicted, the resemblance between the two pathways is high, and both pathways utilize FPP. Therefore, even if the accumulation of FPP in our transgenic lettuce leaves was successful, could be utilized by lettuce for the accumulation of costunolides.

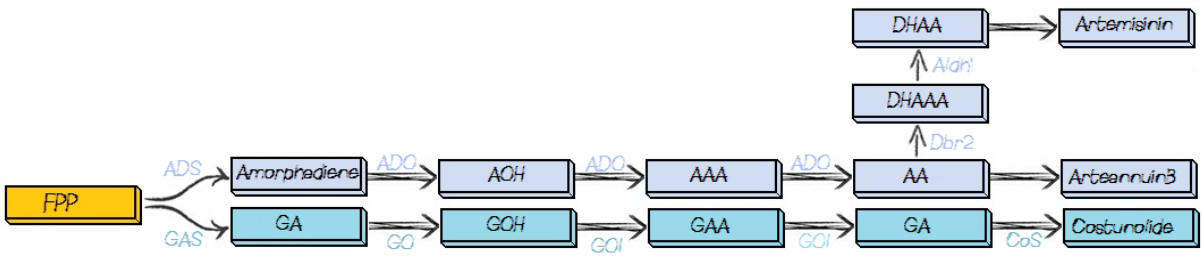


Figure 18. Comparison between artemisinin biosynthesis in A.annua and costunolide biosynthesis. Artemisinin pathway: ADS: amorphadiene synthase, ADO: CYP71AV1, AOH: Artemisinic alcohol, AAA: Artemisininc aldehyde, AA: Artemisinic Acid, DHAAA: Duhydroartemisinic aldehyde, DHAA: Dihydroartemisinic acid. Costunolide pathway: GAS: germacrene A synthase, GA: Germacrane A, GO: Germacrane A hydroxylase, GOH: Germacrane A alchocol, GOI: Germacrane A hydrogonase, GAA: Germacrane A aldehyde, GA: Germacrane A acid, CoS: Costunolide synthase.

Remarks for improvement of lettuce as host for production of artemisinin.

Cloning techniques. As discussed earlier, plastid targeting has its own advantages. Although not the traditional cellular compartment to accumulate FPP, it has been shown that introducing MEV pathway with plastid targeting can enhance the cytosolic MEV intermediates production. This should be achieved with the least vectors possible (ideally one) to increase cloning efficiency, ensure equivalent simultaneous transcription of enzymes in one cell, and circumvent potential genetic instability associated with large multi-gene constructs (11). A vacuum infiltration would be interesting to test, since it covers homogenously the whole leaf surface, providing a broad distribution of the transgene. After infiltration would be necessary to check the expression levels of the transgenes, before proceeding to metabolomic analysis (checking expression levels is an evaluation that lacked from our approach).

Light and oxygen*.* Is possible that the efficiency of P450s between the different conversion reactions they accommodate is not equal. As discussed in other research (40), the alcohol and aldehyde oxidases may show less activity compared to amorphadiene hydroxylases. Another interesting scenario discussed on the same paper is the activity of endogenous alcohol dehydrogonases that reduce aldehydes back to alcohols, in reduction-friendly environment. As proven that artemisinin biosynthesis in *A.annua* takes place mainly in glandular trichomes, probably this particular plant compartment provides a chemical environment that favours oxidations of aldehydes. In order to achieve the same conditions in plant hosts without trichomes, sufficient oxygenation and light (to favour the photo-oxidations) could be provided. Laticifer cells (abundant in lettuce) could also facilitate the production of sesquiterpene lactones, since accumulation of sesquiterpenes in the latex is high.

Test endogenous lettuce P450 activity. A simple and time-saving experiment could be to apply artemisinin or AA (in 70% EtOH) on top of mature WT lettuce leaves. Using already established approaches from *A.annua* (39), spraying AA on *L.sativa* leaves and measuring gene expression of endogenous P450s conclusions concerning the existence of regulation mechanisms could be made. Similar experiments could take place with AAOH/AAA and apart from monitoring gene expression; also targeted analysis of sugar conjugations would be interesting.

5. References

5.1 Reports

1. World-Health-Organization (2012) World Malaria Report 2012.

5.2 Journals

2. Tilley, L., M. W. A. Dixon and K. Kirk (2011). "The *Plasmodium falciparum*-infected red blood cell." International Journal of Biochemistry & Cell Biology 43(6): 839-842.
3. Le Bras, J. and R. Durand (2003). "The mechanisms of resistance to antimalarial drugs in *Plasmodium falciparum*." Fundam Clin Pharmacol 17(2): 147-153.
4. Faurant, C. (2011). "From bark to weed: the history of artemisinin." Parasite 18(3): 215-218.
5. Krishna, S., A. C. Uhlemann and R. K. Haynes (2004). "Artemisinins: mechanisms of action and potential for resistance." Drug Resist Updat 7(4-5): 233-244.
6. Efferth, T. and B. Kaina (2010). "Toxicity of the antimalarial artemisinin and its dervatives." Crit Rev Toxicol 40(5): 405-421.
7. Mannan, A., I. Ahmed, W. Arshad, M. F. Asim, R. A. Qureshi, I. Hussain and B. Mirza (2010). "Survey of artemisinin production by diverse *Artemisia* species in northern Pakistan." Malar J 9: 310.
8. Kumar, S., S. K. Gupta, P. Singh, P. Bajpai, M. M. Gupta, D. Singh, A. K. Gupta, G. Ram, A. K. Shasany and S. Sharma (2004). "High yields of artemisinin by multi-harvest of *Artemisia annua* crops." Industrial Crops and Products 19(1): 77-90.
9. White, N. J. (2008). "Qinghaosu (artemisinin): the price of success." Science 320(5874): 330-334.
10. Bertea, C. M., J. R. Freije, H. van der Woude, F. W. A. Verstappen, L. Perk, V. Marquez, J. W. De Kraker, M. A. Posthumus, B. J. M. Jansen, A. de Groot, M. C. R. Franssen and H. J. Bouwmeester (2005). "Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*." Planta Medica 71(1): 40-47.
11. Farhi, M., M. Kozin, S. Duchin and A. Vainstein (2013). "Metabolic engineering of plants for artemisinin synthesis." Biotechnol Genet Eng Rev 29(2): 135-148.
12. Tang, K., Q. Shen, T. Yan and X. Fu (2014). "Transgenic approach to increase artemisinin content in *Artemisia annua* L." Plant Cell Rep 33(4): 605-615.
13. Martin, V. J., D. J. Pitera, S. T. Withers, J. D. Newman and J. D. Keasling (2003). "Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids." Nat Biotechnol 21(7): 796-802.
14. Zhang, Y., K. H. Teoh, D. W. Reed, L. Maes, A. Goossens, D. J. Olson, A. R. Ross and P. S. Covello (2008). "The molecular cloning of artemisinic aldehyde Delta11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*." J Biol Chem 283(31): 21501-21508.
15. Teoh KH, Polichuk DR, Reed DW, Covello PS (2009) Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. Botany-Botanique 87: 635–642.
16. Sy, L. K. and G. D. Brown (2002). "The mechanism of the spontaneous autoxidation of dihydroartemisinic acid." Tetrahedron 58(5): 897-908.
17. Ro, D. K., E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong and J. D. Keasling (2006). "Production of the antimalarial drug precursor artemisinic acid in engineered yeast." Nature 440(7086): 940-943.
18. Aquil, S., A. M. Husaini, M. Z. Abdin and G. M. Rather (2009). "Overexpression of the HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic *Artemisia annua* plants." Planta Med 75(13): 1453-1458.

19. Chen, D., H. Ye and G. Li (2000). "Expression of a chimeric farnesyl diphosphate synthase gene in *Artemisia annua* L. transgenic plants via *Agrobacterium tumefaciens*-mediated transformation." Plant Sci 155(2): 179-185.
20. Ma, C. F., H. H. Wang, X. Lu, H. Wang, G. W. Xu and B. Y. Liu (2009). "Terpenoid metabolic profiling analysis of transgenic *Artemisia annua* L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry." Metabolomics 5(4): 497-506.
21. Lu, X., Q. Shen, L. Zhang, F. Y. Zhang, W. M. Jiang, Z. Y. Lv, T. X. Yan, X. Q. Fu, G. F. Wang and K. X. Tang (2013). "Promotion of artemisinin biosynthesis in transgenic *Artemisia annua* by overexpressing ADS, CYP71AV1 and CPR genes." Industrial Crops and Products 49: 380-385.
22. Chen, Y. F., Q. Shen, Y. Y. Wang, T. Wang, S. Y. Wu, L. Zhang, X. Lu, F. Y. Zhang, W. M. Jiang, B. Qiu, E. D. Gao, X. F. Sun and K. X. Tang (2013). "The stacked over-expression of FPS, CYP71AV1 and CPR genes leads to the increase of artemisinin level in *Artemisia annua* L." Plant Biotechnology Reports 7(3): 287-295.
23. Ro, D. K., M. Ouellet, E. M. Paradise, H. Burd, D. Eng, C. J. Paddon, J. D. Newman and J. D. Keasling (2008). "Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid." Bmc Biotechnology 8.
24. Paradise, E. M., J. Kirby, R. Chan and J. D. Keasling (2008). "Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulating squalene synthase." Biotechnology and Bioengineering 100(2): 371-378.
25. Wang, H., J. Han, S. Kanagarajan, A. Lundgren and P. E. Brodelius (2013). "Trichome-specific expression of the amorpha-4,11-diene 12-hydroxylase (cyp71av1) gene, encoding a key enzyme of artemisinin biosynthesis in *Artemisia annua*, as reported by a promoter-GUS fusion." Plant Mol Biol 81(1-2): 119-138.
26. Wang H., Olofsson L., Lundgren A., Brodelius P.E., (2011) "Trichome-Specific Expression of Amorpha-4,11-Diene Synthase, a Key Enzyme of Artemisinin Biosynthesis in *Artemisia* annua L., as Reported by a Promoter-GUS Fusion". Am.Journal of Pl.Sciences, 2, 619-628.
27. Kim, S. H., Y. J. Chang and S. U. Kim (2008). "Tissue specificity and developmental pattern of amorpha-4,11-diene synthase (ADS) proved by ADS promoter-driven GUS expression in the heterologous plant, *Arabidopsis thaliana*." Planta Med 74(2): 188-193.
28. Yang, R. Y., L. L. Feng, X. Q. Yang, L. L. Yin, X. L. Xu and Q. P. Zeng (2008). "Quantitative transcript profiling reveals down-regulation of A sterol pathway relevant gene and overexpression of artemisinin biogenetic genes in transgenic *Artemisia annua* plants." Planta Med 74(12): 1510-1516.
29. Zhang, L., F. Jing, F. Li, M. Li, Y. Wang, G. Wang, X. Sun and K. Tang (2009). "Development of transgenic *Artemisia annua* (Chinese wormwood) plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA-mediated gene silencing." Biotechnol Appl Biochem 52(Pt 3): 199-207.
30. Farhi, M., E. Marhevka, J. Ben-Ari, A. Algamas-Dimantov, Z. Liang, V. Zeevi, O. Edelbaum, B. Spitzer-Rimon, H. Abeliovich, B. Schwartz, T. Tzfira and A. Vainstein (2011). "Generation of the potent anti-malarial drug artemisinin in tobacco." Nat Biotechnol 29(12): 1072-1074.
31. Alam, P. and M. Z. Abdin (2011). "Over-expression of HMG-CoA reductase and amorpha-4,11-diene synthase genes in *Artemisia annua* L. and its influence on artemisinin content." Plant Cell Reports 30(10): 1919-1928.
32. van Herpen, T. W. J. M., K. Cankar, M. Nogueira, D. Bosch, H. J. Bouwmeester and J. Beekwilder (2010). "*Nicotiana benthamiana* as a Production Platform for Artemisinin Precursors." Plos One 5(12).

33. Saxena, B., M. Subramaniyan, K. Malhotra, N. S. Bhavesh, S. D. Potlakayala and S. Kumar (2014). "Metabolic engineering of chloroplasts for artemisinic acid biosynthesis and impact on plant growth." J Biosci 39(1): 33-41.
34. Wu, S., M. Schalk, A. Clark, R. B. Miles, R. Coates and J. Chappell (2006). "Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants." Nat Biotechnol 24(11): 1441-1447.
35. Voinnet, O., S. Rivas, P. Mestre and D. Baulcombe (2003). "An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus." Plant Journal 33(5): 949-956.
36. de Kraker, J. W., M. Schurink, M. C. R. Franssen, W. A. Konig, A. de Groot and H. J. Bouwmeester (2003). "Hydroxylation of sesquiterpenes by enzymes from chicory (*Cichorium intybus* L.) roots." Tetrahedron 59(3): 409-418.
37. Cho W.D., Park D.Y., Chunk H.K., (2005) "Agrobacterium-Mediated Transformation of Lettuce with a Terpene Synthase Gene". J. Kor. Soc. Hort. Sci. 46(3):169-175
38. Wroblewski, T., A. Tomczak and R. Michelmore (2005). "Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*." Plant Biotechnology Journal 3(2): 259-273.
39. Arsenault, P. R., D. Vail, K. K. Wobbe, K. Erickson and P. J. Weathers (2010). "Reproductive development modulates gene expression and metabolite levels with possible feedback inhibition of artemisinin in *Artemisia annua*." Plant Physiol 154(2): 958-968.
40. Zhang, Y., G. Nowak, D. W. Reed and P. S. Covello (2011). "The production of artemisinin precursors in tobacco." Plant Biotechnol J 9(4): 445-454.
41. Sessa, R. A., M. H. Bennett, M. J. Lewis, J. W. Mansfield and M. H. Beale (2000). "Metabolite profiling of sesquiterpene lactones from *Lactuca* species. Major latex components are novel oxalate and sulfate conjugates of lactucin and its derivatives." J Biol Chem 275(35): 26877-26884.
42. Tikunov, Y. M., S. Laptenok, R. D. Hall, A. Bovy and R. C. de Vos (2012). "MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data." Metabolomics 8(4): 714-718.

5.4 Photographic Material and Databases

- Cover Photo © Birgit Betzelt, Source: http://tinyurl.com/k8olstx (Last assessed 25/4/2014)
- Figure 1: Kops et al., J., Flora Batava, vol. 22: t. 1697 (1906), website: http://www.plantillustrations.org/illustration.php?id_illustration=137689 (*Last assessed: 22/04/2014)*
- Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/) (last assessed 22/04/2014)

5. Appendix

Primers for Luciferase intron

Luc+-F-NcoI	AGCAAACCATGGAAGACGCCAAAAAC
Luc+-R-NotI	AGCAAAGCGGCCGCTTACACGGCGATCTTTCCG

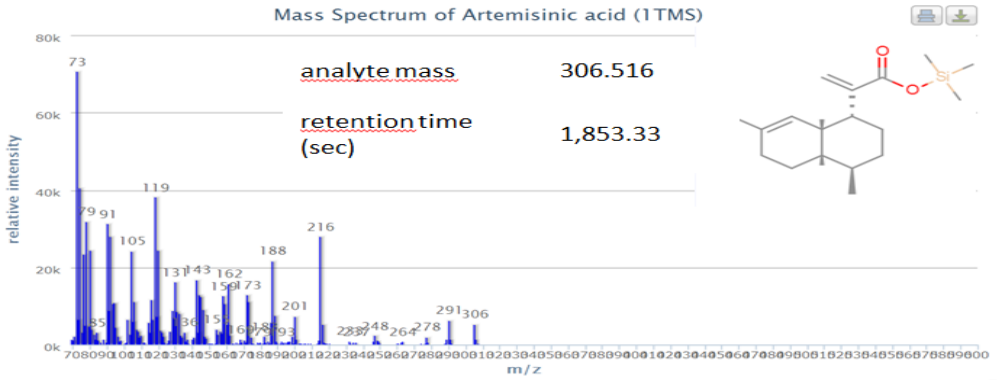


Figure A. Mass spectrum used for the detection of AA(1TMS) in targeted GC/MS analysis in MPI.

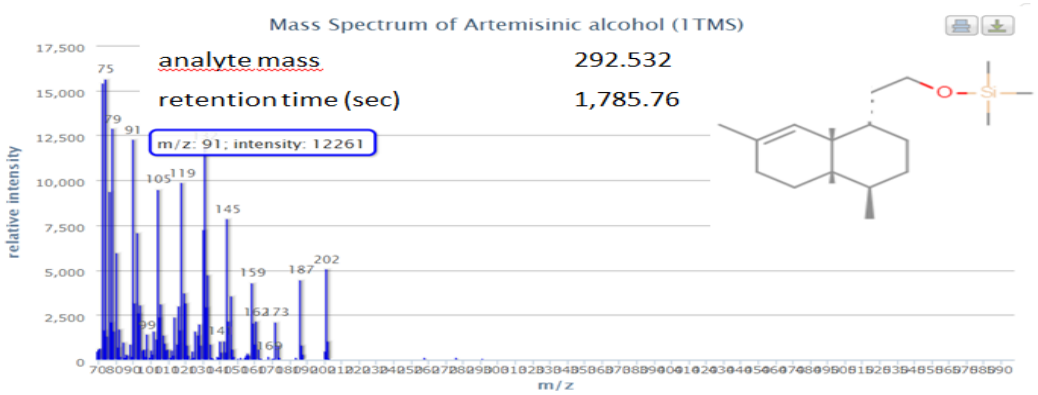


Figure B. Mass spectrum used for the detection of AOH(1TMS) in targeted GC/MS analysis in MPI.

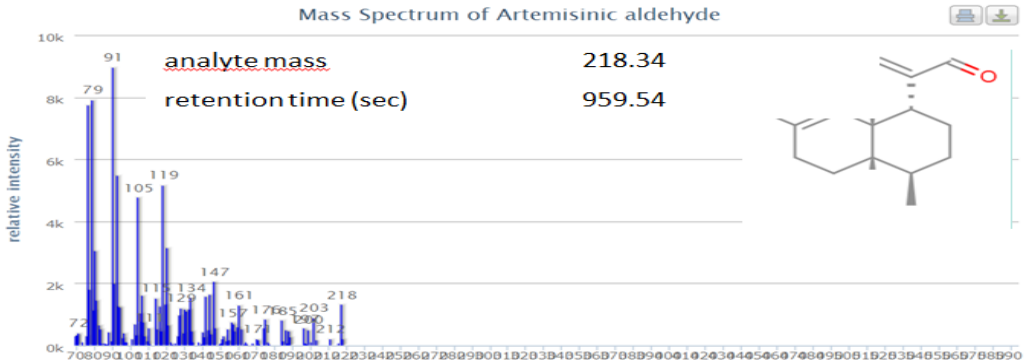


Figure C. Mass spectrum used for the detection of AAA in targeted GC/MS analysis in MPI.

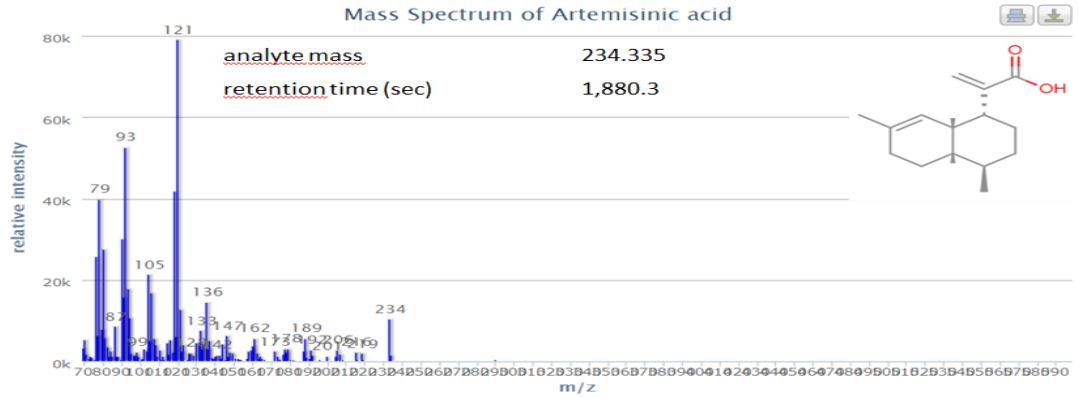


Figure D. Mass spectrum used for the detection of AA in targeted GC/MS analysis in MPI.

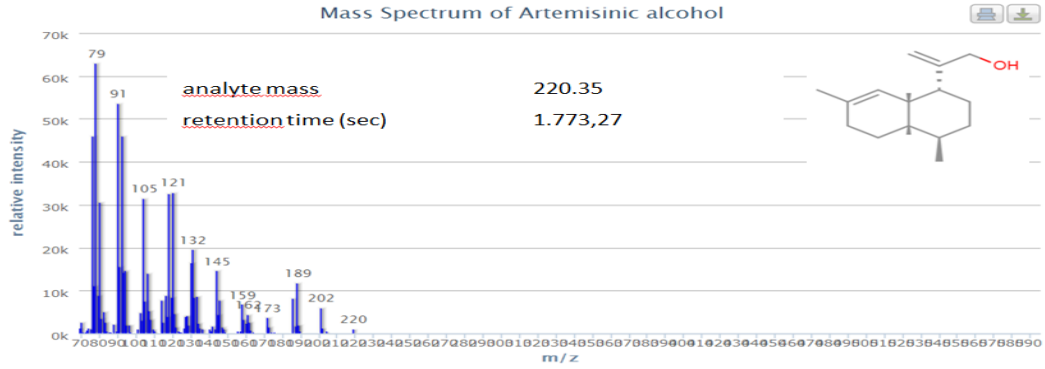
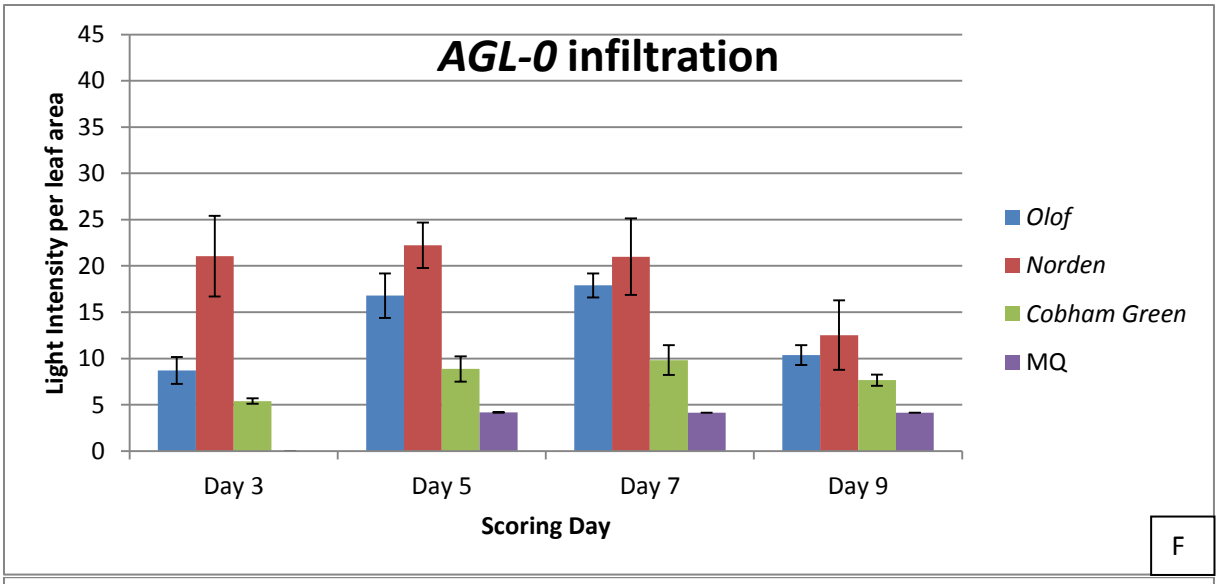
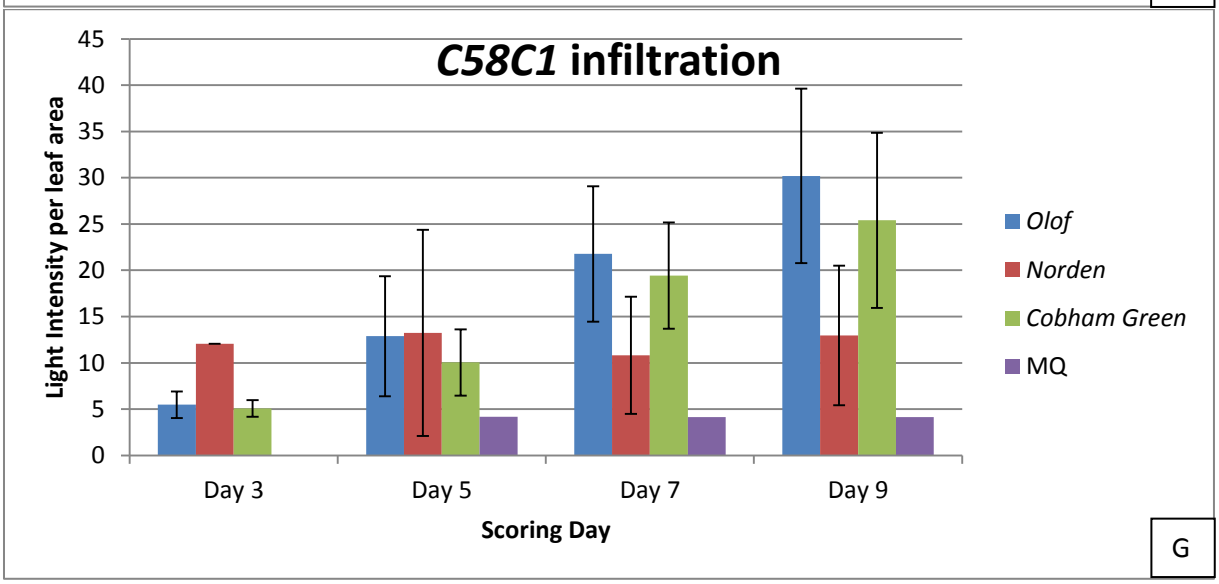
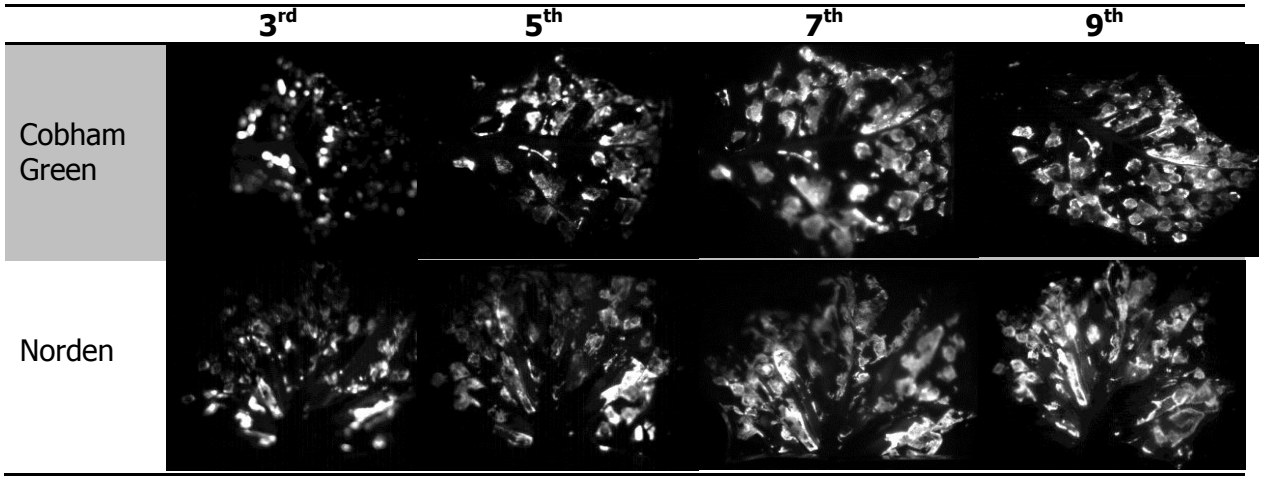


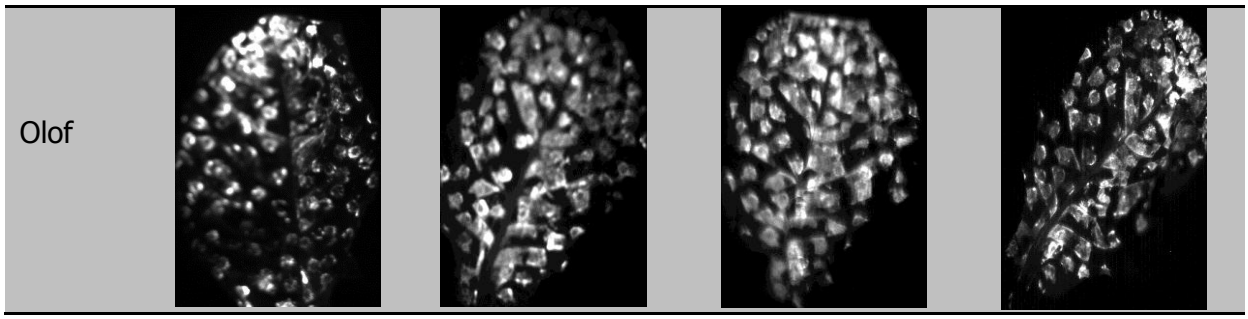
Figure E. Mass spectrum used for the detection of AOH in targeted GC/MS analysis in MPI.





Figures F-G. Summarizing figure of all light intensities deriving from luciferase expression, using all lettuce cultivars, both agro-strains and all scoring dates. **Figure F**: AGL-0 infiltration harbouring the luciferase gene. **Figure G**: C58C1 infiltration harbouring the luciferase gene.





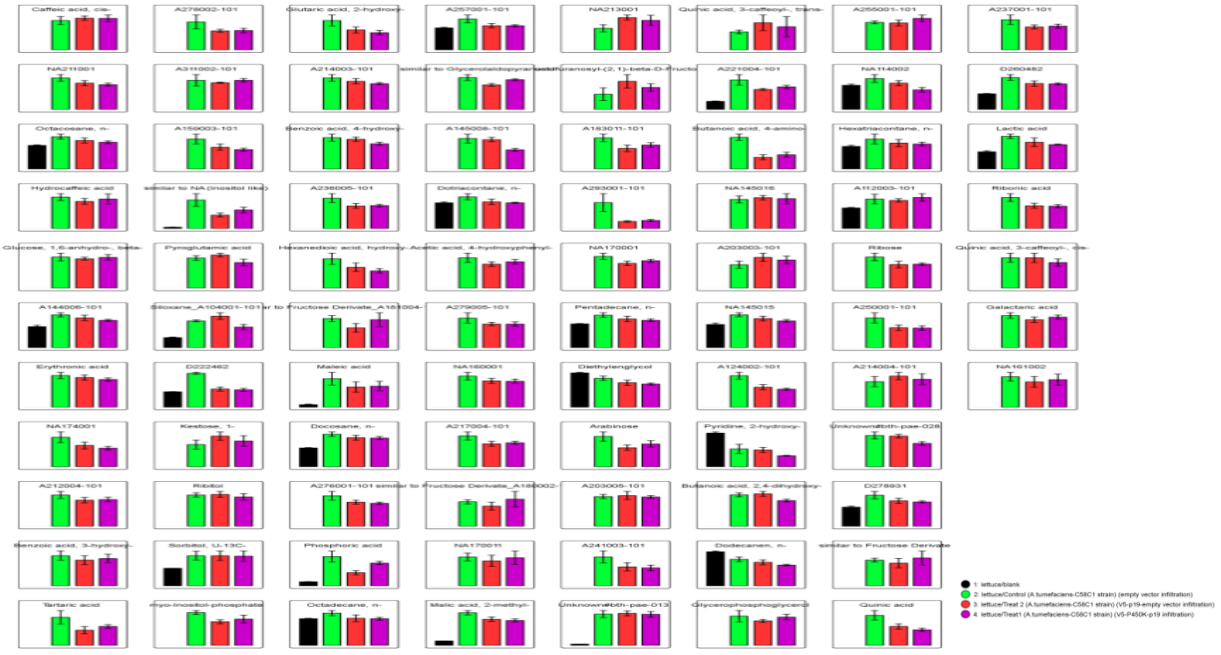


Figure H. Scheme of metabolites not annotaded and not included in the primary metabolism.

Table C. Conjugated artemisinin precursors produced in agro-infiltrated *N. benthamiana* leaf extracts. Non-volatile metabolites with mass intensity higher than 500 in LC-QTOF-MS, which were significantly increased in leaves agro-infiltrated with *AmFH+P450LAP*, *AmFH+P450HAP*, *AmFH+P450LAP+DBR2*, or *AmFH+P450HAP+DBR2* were targeted for analysis by LC-QTOF-MS/MS fragmentation.

Ret (min)	Detected Mass(D)*	MS-MS fragments	Mol form	ΔMass (ppm)	Putative ID	Intensity^{+}			
						AmFH +P450LAP	AmFH +P450HAP	AmFH +P450LAP +DBR2	AmFH +P450HAP +DBR2
27.85	543.2793	381[M-Hex-H]$^{-}$	$C_{27}H_{44}O_{11}$	1.3	AAOH-Hex2	319	3715	264	1753
28.87	629.2810	585[M-CO_2-H]$^{-}$,543[M-Mal-H]$^{-}$, 381[M-Mal-Hex-H]$^{-}$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (I) [a]	1096	2239	861	1378
29.35	629.2810	585[M-CO_2-H]$^{-}$,543[M-Mal-H]$^{-}$, 381[M-Mal-Hex-H]$^{-}$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (II)	606	24251	627	16215
29.69	629.2810	585[M-CO_2-H]$^{-}$,543[M-Mal-H]$^{-}$, 381[M-Mal-Hex-H]$^{-}$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (III)	439	1109	ND	515
23.80	542.2536	272,254,210,179,143,128 **	$C_{25}H_{41}N_3O_8S$	1.0	AAA-GSH-H_2O	2462	7891	130	1970
24.02	765.3181	719[M-H]$^{-}$,395[M-2Hex-H]$^{-}$, 233[M-3Hex-H]$^{-}$	$C_{34}H_{54}O_{19}$	0.1	(AA-Hex3) FA [b]	21190	612	100	ND

27.88	557.2598	395[M-Hex-H]$^-$,233[M-2Hex-H]$^-$	$C_{27}H_{42}O_{12}$	0	AA-Hex2 (I)[c]	5403	917	121	371
28.48	557.2598	395[M-Hex-H]$^-$,233[M-2Hex-H]$^-$	$C_{27}H_{42}O_{12}$	0	AA-Hex2 (II)	3773	110	28	ND
29.38	643.2602	599[M-CO_2-H]$^-$, 395[M-Mal-Hex-H]$^-$, 233[M-Mal-2Hex-H]$^-$	$C_{30}H_{44}O_{15}$	0.5	AA-Hex2-Mal (I)	16021	2972	1652	1783
29.78	1287.5282	643[M-H]$^-$,599[M-CO_2-H]$^-$, 395[M-Mal-Hex-H]$^-$,233[M-Mal-2Hex-H]$^-$	$C_{60}H_{88}O_{30}$	4.5	AA-Hex2-Mal ([2M-H]$^-$)	2611	ND	129	82
24.86	753.3545	707[M-H]$^-$,545[M-Hex-H]$^-$, 383[M-2Hex-H]$^-$,221[M-3Hex-H]$^-$	$C_{34}H_{58}O_{18}$	0.9	(DHAAOH-Hex3) FA	145	85	14248	11573
28.80	545.2962	383[M-Hex-H]$^-$,221[M-2Hex-H]$^-$	$C_{27}H_{46}O_{11}$	1.8	DHAAOH-Hex2 (I)	343	194	14321	14968
30.75	545.2962	383[M-Hex-H]$^-$,221[M-2Hex-H]$^-$	$C_{27}H_{46}O_{11}$	1.8	DHAAOH-Hex2 (II)	ND	18	855	919
30.27	631.2966	587[M-CO_2-H]$^-$,545[M-Mal-H]$^-$, 383[M-Mal-Hex-H]$^-$,221[M-Mal-2Hex-H]$^-$	$C_{30}H_{48}O_{14}$	0	DHAAOH-Hex2-Mal	479	266	24475	24394
23.87	767.3338	721[M-H]$^-$,397[M-2Hex-H]$^-$, 235[M-3Hex-H]$^-$	$C_{34}H_{56}O_{19}$	1.6	(DHAA-Hex3) FA	1237	ND	23356	861
27.28	559.2755	397[M-Hex-H]$^-$,235[M-2Hex-H]$^-$	$C_{27}H_{44}O_{12}$	0.3	DHAA-Hex2	ND	ND	5188	120
26.58	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$,	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (I)	1912	5539	236	1645

		235[M-Mal-2Hex-H]$^-$							
28.92	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$, 235[M-Mal-2Hex-H]$^-$	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (II)	ND	20	1303	76
29.83	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$, 235[M-Mal-2Hex-H]$^-$	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (III)	3424	64	ND	ND

+ Peak intensities are the mean of three agro-infiltrated leaves.

* Detected mass (D): The mass was detected in negative mode of LC-QTOF-MS.

** The ions of a number of representative GSH adducts in the negative ion mode.

Ret (min): retention time, in minutes; Mol form: molecular formula of the metabolite; ΔMass (ppm): deviation between the detected mass and real accurate mass, in ppm; Putative ID: putative identification of metabolite; ND: not detectable.