



FACULTY OF PHARMACEUTICAL SCIENCES

Trichomes, the key to an increased production of artemisinin in *Artemisia annua*.

Sandra Soetaert

Ghent University

Faculty of Pharmaceutical Sciences

Laboratory of Pharmaceutical Biotechnology

Promotor: Prof. Dr. Apr. Dieter Deforce

Co-Promotor: Prof. Dr. Alain Goossens

Co-Promotor: Dr. Apr. Filip Van Nieuwerburgh

Thesis submitted in the fulfilment of the requirements for the degree of
Doctor in the Pharmaceutical Sciences

Thesis supported by an FWO Vlaanderen doctoral grant

Cover photo:

Scanning electron microscopic image of *Artemisia annua* flower heads covered with filamentous and glandular trichomes.

COMPOSITION OF THE JURY

Chairman:

Prof. Dr. Apr. Jan Van Bocxlaer

Laboratory of Medical Biochemistry and Clinical Analysis, Faculty of Pharmaceutical Sciences, Ghent University

Members of the examination committee:

Prof. Dr. Apr. Dieter Deforce

Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Dr. Alain Goossens

Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University

Dr. Apr. Filip Van Nieuwerburgh

Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Dr. Gert Laekeman

Research Centre for Pharmaceutical Care and Pharmaco-Economics, Faculty of Pharmaceutical Sciences, KU Leuven

Prof. Dr. Jo Vandesompele

Centre of Medical Genetics, Faculty of Medicine and Health Sciences, Ghent University

Prof. Dr. Paul Coucke

Centre of Medical Genetics, Faculty of Medicine and Health Sciences, Ghent University

Wim Van Criekinge

Department of Mathematical Modelling, Statistics and Bio-informatics, Faculty of Bioscience Engineering, Ghent University

Prof. Dr. Dirk Inzé

Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University

Part I.	Background	1
Chapter I	Malaria general introduction	3
1	Discovery of artemisinin as anti-malaria compound.....	5
2	Malaria	6
Chapter II	Artemisinin and terpene production.....	15
1	Artemisinin supply	17
2	Terpenoids	17
3	Artemisinin production	21
4	Increased artemisinin production	28
References	33
Part II.	Technical optimizations for sequencing.....	39
Chapter I	Introduction to 2 nd generation sequencing	41
1	Comparison of sequencing platforms.....	46
2	Illumina sequencing	51
Chapter II	Optimizations for RNA preparation.....	55
1	Introduction and methods.....	59
2	Optimization results and discussion	66
3	Conclusions	71
Chapter III	Quantitative bias in Illumina TruSeq and a Novel Post Amplification Barcoding Strategy for Multiplexed DNA and Small RNA Deep Sequencing [12].....	73
1	Introduction	77
2	Methods	78
3	Results	81
4	Discussion	85
5	Supporting information	88
References	99
Part III.	Trichome analysis.....	103
Chapter I	Transcriptome analysis of glandular and filamentous trichomes.....	105
1	Introduction	109
2	Methods	112
3	Results and discussion.....	116
4	Conclusions	127
5	Supporting information	128

Chapter II Metabolite and transcriptome analysis of apical and sub-apical cells of glandular trichomes	135
1 Introduction	139
2 Materials and methods	140
3 Results and discussion.....	145
4 Conclusions	153
5 Supporting information	154
References	155
Part IV. Overall conclusions	161
References	169
Part V. Summary	171
Part VI. Samenvatting	177
Part VII. Abbreviations	183
List of abbreviations.....	185
Part VIII. Curriculum vitae.....	189
Part IX. Dankwoord	195

Part I. Background



Chapter I Malaria general introduction

Sandra Soetaert

1 Discovery of artemisinin as anti-malaria compound

Artemisia annua L. also named Sweet wormwood or Sweet Annie (Figure 1), belongs to the family of Asteraceae and is native to Asia. As the name indicates, this plant has an annual cycle. Greenish or yellowish flowers appear in August or September and are induced by a short-day photoperiod. It is a single-stemmed shrub with an aromatic smell that can grow over 2 m high and has fern-like leaves.

As early as 168 B.C., *Artemisia annua* was used in Chinese medicine for the treatment of hemorrhoids [1]. The first recommendation to use *A. annua* to cure fevers with sweating and jaundice, symptoms resembling malaria, dates from 150-219 A.D. in a text entitled *On Cold Damage (Shang Han Lun)* and written by Zhang Ji [1]. In 1596, the Chinese *materia medica* described the use of *A. annua* to treat malaria [1].

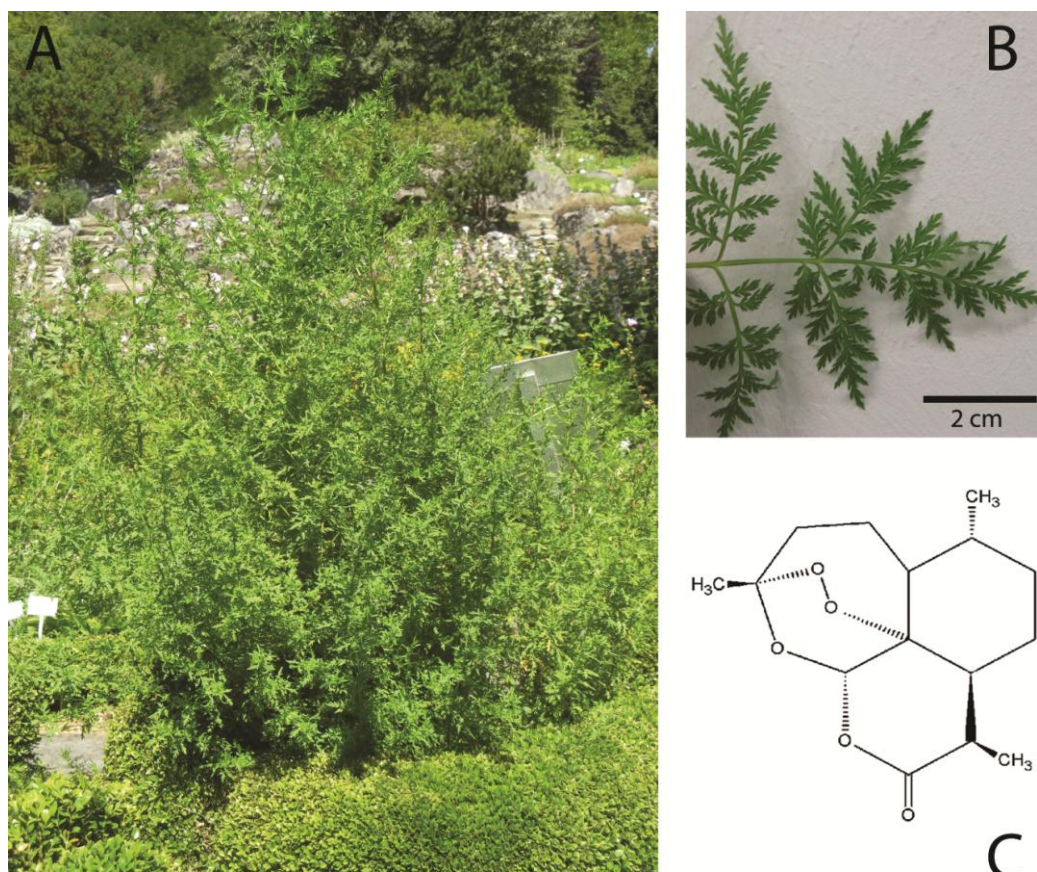


Figure 1: Morphology of *Artemisia annua* L. with its most active anti-malarial compound.

A: Picture of *A. annua* L. in the botanical garden of Ghent University; **B:** Detail of a leaf from *A. annua*; **C:** Molecular structure of artemisinin.

Malaria was more deadly for the soldiers during the Vietnam war than armed conflicts [3], therefore Vietnam asked the Chinese government for help to find anti-malaria treatments. A systematic screening of the Chinese *materia medica* was set up [2] and extracts from *A. annua* were one of the first substances tested but initially no anti-malarial activity was found [3]. This was due to the use of traditional methods of boiling and high-temperature extraction which damage the active ingredient. An extraction at lower temperature with ether revealed 100% inhibition of mouse malaria [4]. The active compound artemisinin was isolated and characterized by the Qinghaosu Antimalarial Coordinating Research Group (1979) as a sesquiterpene with endoperoxide bridge (Figure 1) [4]. Artemisinin production was reported in *Artemisia annua*, *Artemisia apiacea* and *Artemisia lancea* [6, 7].

2 Malaria

2.1 General facts

Roughly half of the world's population is at risk of malaria and the World Health Organization (WHO) estimated 216 million episodes in 2010 [5]. Despite the fact that malaria is preventable and curable, the WHO assesses that malaria was responsible for approximately 655,000 deaths in 2010 [5]. Areas at risk for malaria transmission are shown in Figure 2. Malaria is caused by an infection with parasites from the genus *Plasmodium*.

Five species of the genus *Plasmodium* infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. The most deadly species is *P. falciparum* which is predominant in Africa [5]. *P. vivax* is the most widespread species that counts for 80% of the infections in Southeast Asia and 70% of the infections in America [7]. This species can form dormant hypnozoite forms in the liver [7]. *P. malariae* infections are observed in all major malaria-endemic regions [6] and *P. ovale* is mainly present in Sub-Saharan Africa but occurs also in the western Pacific and the Asian mainland [7]. *P. malariae* and *P. ovale* are relatively mild infections [6]. *P. knowlesi* is a malaria parasite in monkeys which is increasingly associated with human infections in Southeast Asia [8]. This species can cause severe malaria comparable with falciparum malaria [8].

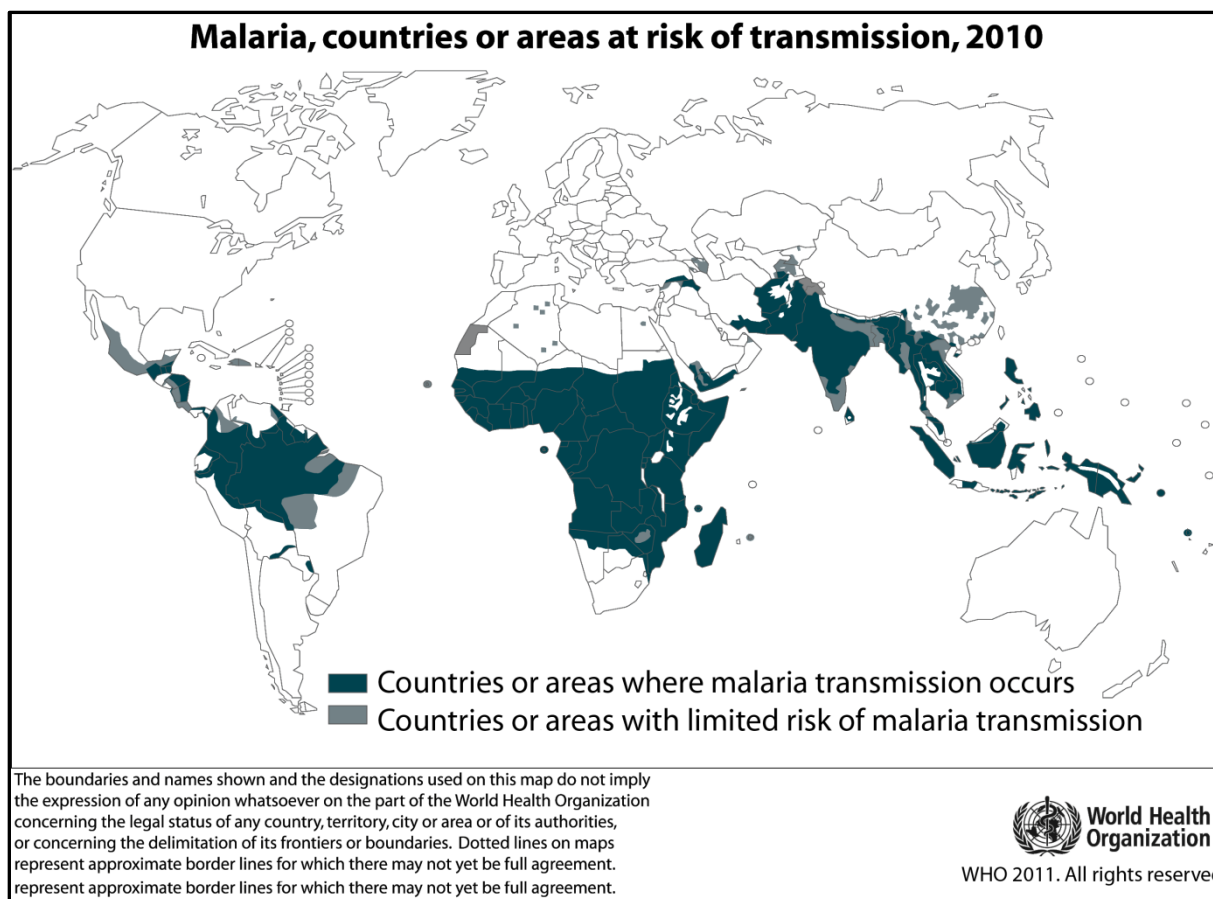


Figure 2: Areas at risk of malaria.

(http://gamapserver.who.int/mapLibrary/Files/Maps/Global_Malaria_2010.png)

2.2 Infection cycle and diagnosis

Plasmodium parasites are transmitted by mosquitoes from the genus *Anopheles*. Before the mosquito starts consuming her blood meal, saliva is injected as anticoagulant and *Plasmodium* sporozoites -motile and infective stages of the parasite- can be introduced (Figure 3) [9]. Malaria further develops in an exo-erythrocytic and erythrocytic (red blood cell) phase. The exo-erythrocytic phase starts with sporozoite migration to the liver to invade hepatocytes, a differentiation and asexual multiplication to make exo-erythrocytic schizonts containing merozoites. This phase is asymptomatic and takes 8-30 days [9]. Some malaria species can remain dormant in the liver for extended periods and cause relapses weeks or months later [10].

When merozoites are released in the bloodstream, they invade the erythrocytes (red blood cells) and develop through different stages. The first stage is the immature trophozoite with a characteristic ring structure (Figure 4) and mature trophozoite stage in which the parasite is highly metabolic active [9]. In the next stage, asexual division forms erythrocytic schizonts

containing merozoites [9]. These merozoites break out synchronously every 1-3 days and invade new erythrocytes. This event is linked to a wave of fever. After asexual amplification, a small proportion of the merozoites will differentiate into ♂ micro- and ♀ macro-gametocytes [9].

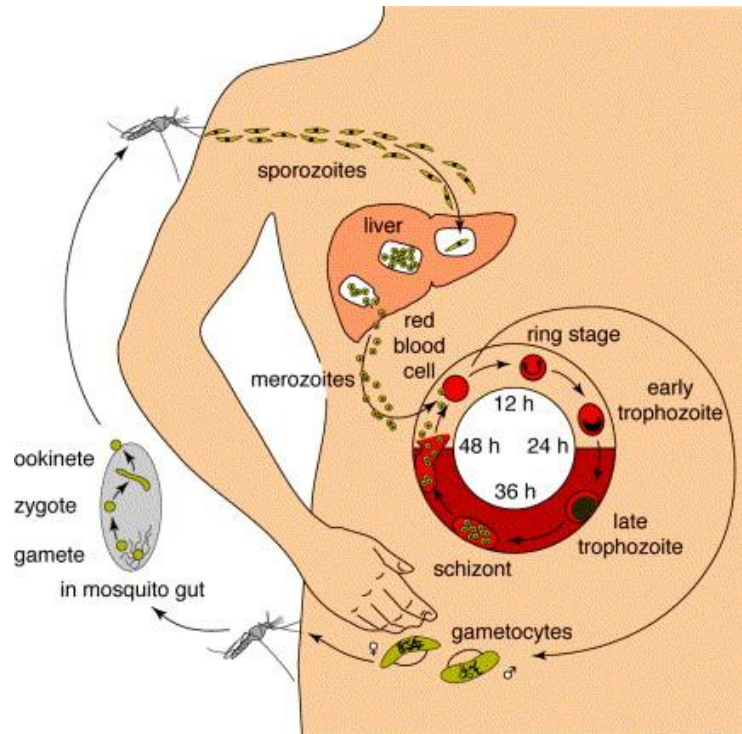


Figure 3: Infection cycle of *Plasmodium falciparum* [11].

Three developmental stages of *P. falciparum* can be distinguished: sporozoite-formation in the mosquito, an exo-erythrocytic phase in the liver and an erythrocytic phase in humans.

If these gametocytes are ingested from the blood stream by a female mosquito, micro- and macro-gametocytes will form a zygote in the mosquito gut. The zygote develops into an ookinete which can penetrate the gut and an oocyst is formed in the gut-wall of the mosquito. After 8-15 days, sporozoites are formed and they migrate to the salivary glands of the mosquito [10].

Accurate clinical diagnosis of malaria is challenging as the first symptoms of malaria are nonspecific and common to many diseases e.g. fatigue, headache, vomiting, chills and muscle and joint pains [12]. In a further stage: fever, chills, perspiration, anorexia, vomiting and worsening malaise are observed [12]. If not treated, particularly in *P. falciparum* or *P. knowlesi* infections, this can lead to severe malaria. In *P. falciparum* infections, parasite-infected erythrocytes accumulate and sequester in various organs such as brain, heart, lung

and kidney [9]. Severe malaria is complex and characterized by one or more of the following symptoms: coma in the case of cerebral malaria, metabolic acidosis, severe anaemia, hypoglycaemia, acute renal failure or acute pulmonary oedema [12]. Without appropriate treatment, severe malaria is fatal.

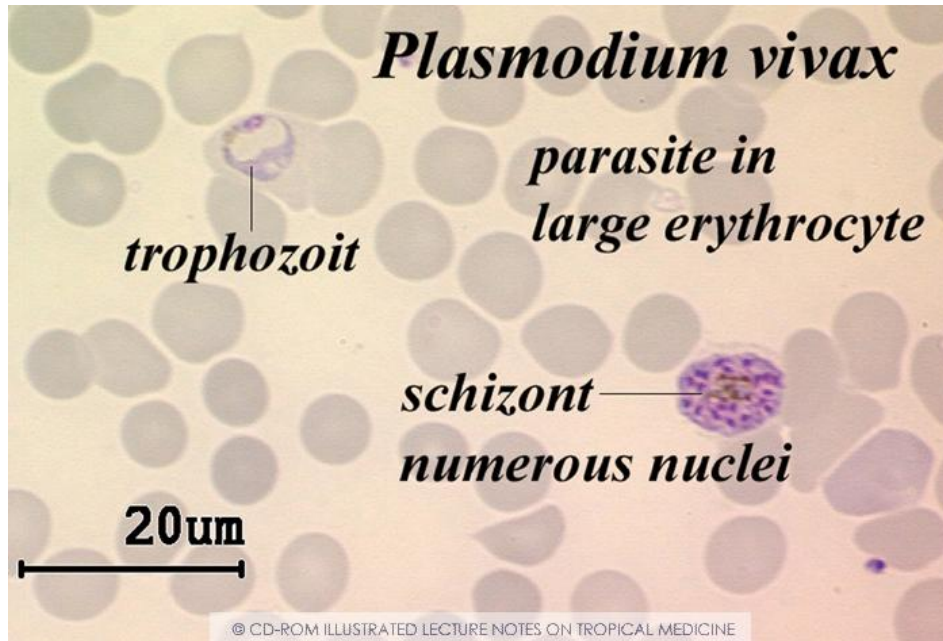


Figure 4: Light microscopy view of malaria parasites in blood.

A trophozoit- and schizont-stage of *P. vivax* are detected in red blood cells

(http://www.itg.be/itg/distancelearning/lecturenotesvandenendene/imagehtml/ppages/kabisa_1522.htm).

Since it is difficult to base the diagnosis of malaria on the symptoms alone, clinical suspicion of malaria should be confirmed by the detection of parasites. Two methods routinely used are light microscopy and rapid diagnostic tests (RDTs). With light microscopy, blood samples are screened on parasite presence (Figure 4) while with RDT parasite-specific antigens or enzymes are detected. Which method is recommended depends on local circumstances.

For light microscopy, a well-trained and skilled staff is required as well as a power source. An advantage is that with light microscopy the amount of parasites in the blood can be quantified and that the response to treatments can be assessed [12]. Additionally, it can lead to the identification of other causes of disease.

RDTs are more expensive but easier to use outside health facilities because no special equipment is needed and the tests are relatively simple to perform and interpret [12]. A disadvantage is that the specificity and sensitivity of RDTs can vary due to storage conditions

such as temperature and humidity. It is also difficult to distinguish new infections from effectively treated infections as some antigens can persist in the blood for a few weeks after treatment [13].

Nucleic acid-based tests (NATs) e.g. PCR are also used and can detect lower infections-levels than light microscopy and RDTs [13].

2.3 Anti-malarial compounds

Medicines useful for prophylaxis (prevention) are chloroquine, doxycycline, mefloquine and proguanil with atovaquone [14]. Chloroquine (e.g. Nivaquine) is used in regions where only *P. vivax*, *P. ovale* and *P. malariae* are causing infections since chloroquine resistance in *P. falciparum* is widespread [12]. Doxycycline is effective against malaria but must be administered daily [12] and during 4 weeks after returning home. Mefloquine (e.g. Lariam) is effective against all forms of malaria but can cause serious neuropsychiatric disturbances in approximately 1 out of 10 000 travellers [12]. Proguanil in combination with atovaquone (Malarone) is as efficient as mefloquine but with less side effects [14]. This medicine inhibits also pre-erythrocytic development in the liver [12]. It has to be noted that prophylactic medicines are not active against dormant liver stages of *P. vivax* and *P. ovale* [12]. Artemisinin has a half-life in the order of one hour [12] and is therefore not suited for prophylaxis of malaria.

To treat malaria infections, artemisinin-derivatives are not recommended as monotherapy since their half-life is too short. Therefore, artemisinin-derivatives are combined with slowly eliminated anti-malaria compounds. Another advantage of using a combination therapy with different modes of action is to delay the emergence of resistance against artemisinin. Five combinations with artemisinin-derivatives are recommended by the WHO: lumefantrine, amodiaquine, mefloquine, sulfadoxine-pyrimethamine and piperaquine [5].

Pure artemisinin is poorly soluble in water and oil [15] but can be administered orally. Oral administration is not suited for patients with severe malaria since extreme vomiting occurs. Therefore, artemisinin-derivatives are synthesized which are more soluble in water e.g. artesunate or in oil e.g. arteether or artemether (Figure 5) [16]. These compounds are *in vitro* more potent against *Plasmodium* parasites than artemisinin itself [17]. *In vivo*: artesunate,

arteether and artemether are converted into a number of metabolites with antimalarial activity such as dihydroartemisinin [18].

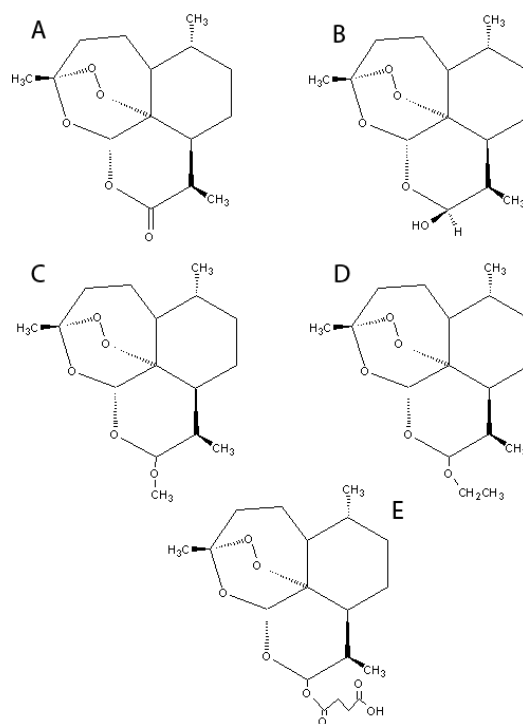


Figure 5: Artemisinin derivatives.

A: artemisinin, **B:** dihydroartemisinin, **C:** artemether, **D:** arteether, **E:** artesunate.

The WHO recommends that uncomplicated *P. falciparum* infections are treated with artemisinin-based-combination therapy (ACT) [12]. For severe malaria, the treatment should be started with parenteral artesunate and followed by an oral ACT. To cure patients from *P. vivax*, the preferred treatment is chloroquine in areas where there is no resistance for chloroquine; otherwise ACT should be used [12]. Because *P. vivax* has developed resistance to sulfadoxine-pyrimethamine, other ACTs are recommended [12]. *P. ovale* and *P. malariae* infections are preferably treated with chloroquine [12].

2.4 Targets for artemisinin

For the anti-malarial activity of artemisinin(-derivatives), several modes of action have been proposed but these are still under debate [19]. A schematic overview is given in Figure 6.

A first hypothesis is related to hemoglobin metabolism. *Plasmodium* parasites degrade hemoglobin present in the erythrocytes and in this reaction, amino acids are released which are used for parasite protein synthesis. In addition to amino acids, free heme is released which

is toxic for the parasite. The cleavage of the artemisinin endoperoxide bond produces free radicals which alkylate heme molecules. This interferes with the crystallization of heme into the non toxic haemozoin [19, 20]. Another proposed reaction model is that artemisinin can alkylate proteins [20].

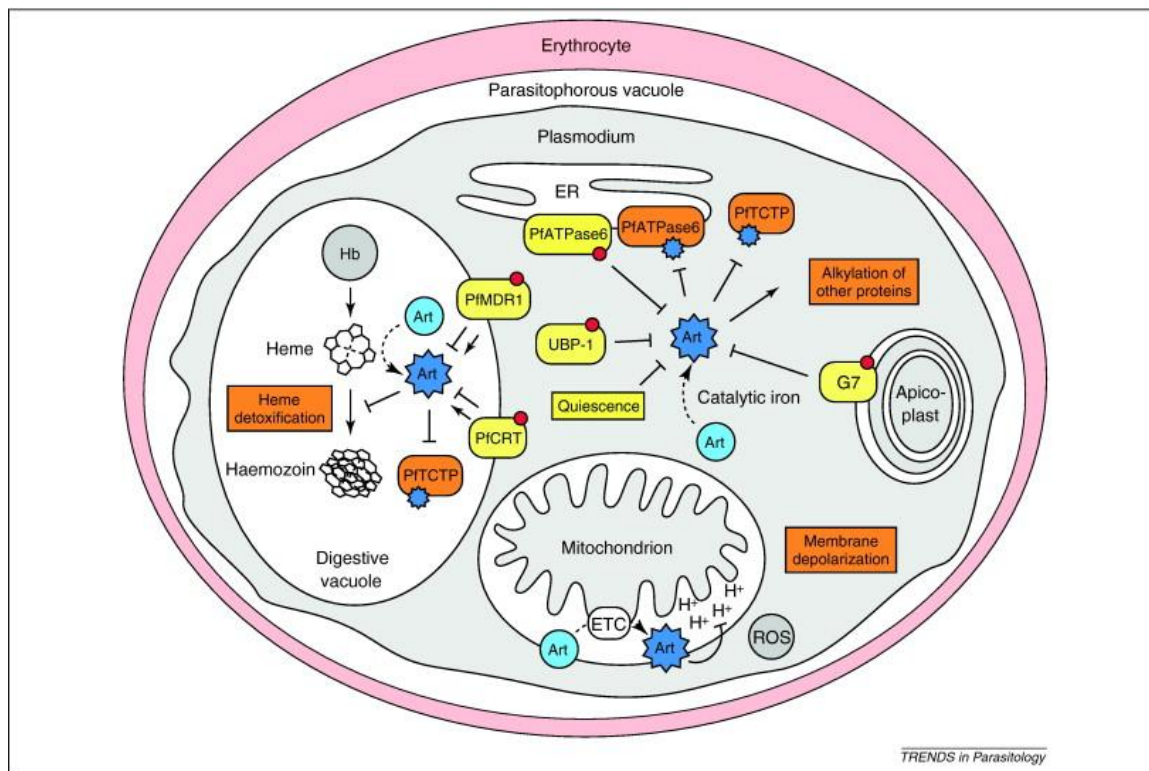


Figure 6: Overview of proposed modes of action and resistance against artemisinins [20]. Putative modes of action are coloured in orange, resistance mechanisms in yellow.

Artemisinin (Art) is activated into radicals. Activated artemisinins might block heme detoxification, block the Ca^{2+} storage in the ER by interfering with PfATPase6, alkylate proteins or cause membrane depolarization. Proteins with mutations (red dots) can increase (arrow) or decrease (inhibition line) susceptibility to artemisinins.

Another reaction model proposes the inhibition of sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCAs) [20]. The function of SERCA is to reduce free calcium in the cytosol and to concentrate it into membrane bound stores. It is known that SERCAs in mammals are inhibited by the sesquiterpene lactone thapsigargin which lacks an endoperoxide bridge [19]. Since artemisinin has a sesquiterpene lactone structure too, it was reasoned that artemisinin might act on SERCAs. Only one enzyme in *P. falciparum* is orthologous to SERCA: PfATPase6 [19]. This target is present in the endoplasmic reticulum of the parasite and not in

the food vacuole [19]. It is reported that PfATPase6 is indeed inhibited by artemisinin and that this supposedly can mediate parasite death [20].

Interference with mitochondrial functions is also suggested. The electron transport chain of malaria parasites is proposed to directly activate artemisinin which leads to an accumulation of reactive oxygen species [21]. In *Plasmodium berghei* but not in mammal cells, reactive oxygen species induce mitochondrial membrane depolarization and ultimately parasite death [20]. This membrane depolarization is endoperoxide bridge dependent since this effect was not observed with deoxyartemisinin [21]. In addition to this, artemisinin is suspected to cause damage of the parasite membrane [19].

2.5 Artemisinin resistance

Despite diverse action mechanisms proposed for artemisinin and its derivatives, resistance is emerging. Resistance is mainly reported in specific regions in Asia. In Laos, resistance is not yet observed as the half life of parasite decline after a treatment with artemisinin is only 2 hours [22]. In contrast with this, the half life of parasite decline is 6 hours at the Thai-Cambodian border [22]. This confirms previous reports [23] of artemisinin resistance at the Thai-Cambodian border. In western Thailand, artemisinin resistance has recently emerged (half life of 3 hours) [22]. At the Thai-Cambodian border, artemisinins have been used as monotherapy for more than 30 years [24] whether in western Thailand (Thai-Burumese border), artemisinin derivatives have been used almost exclusively in combination with mefloquine [24].

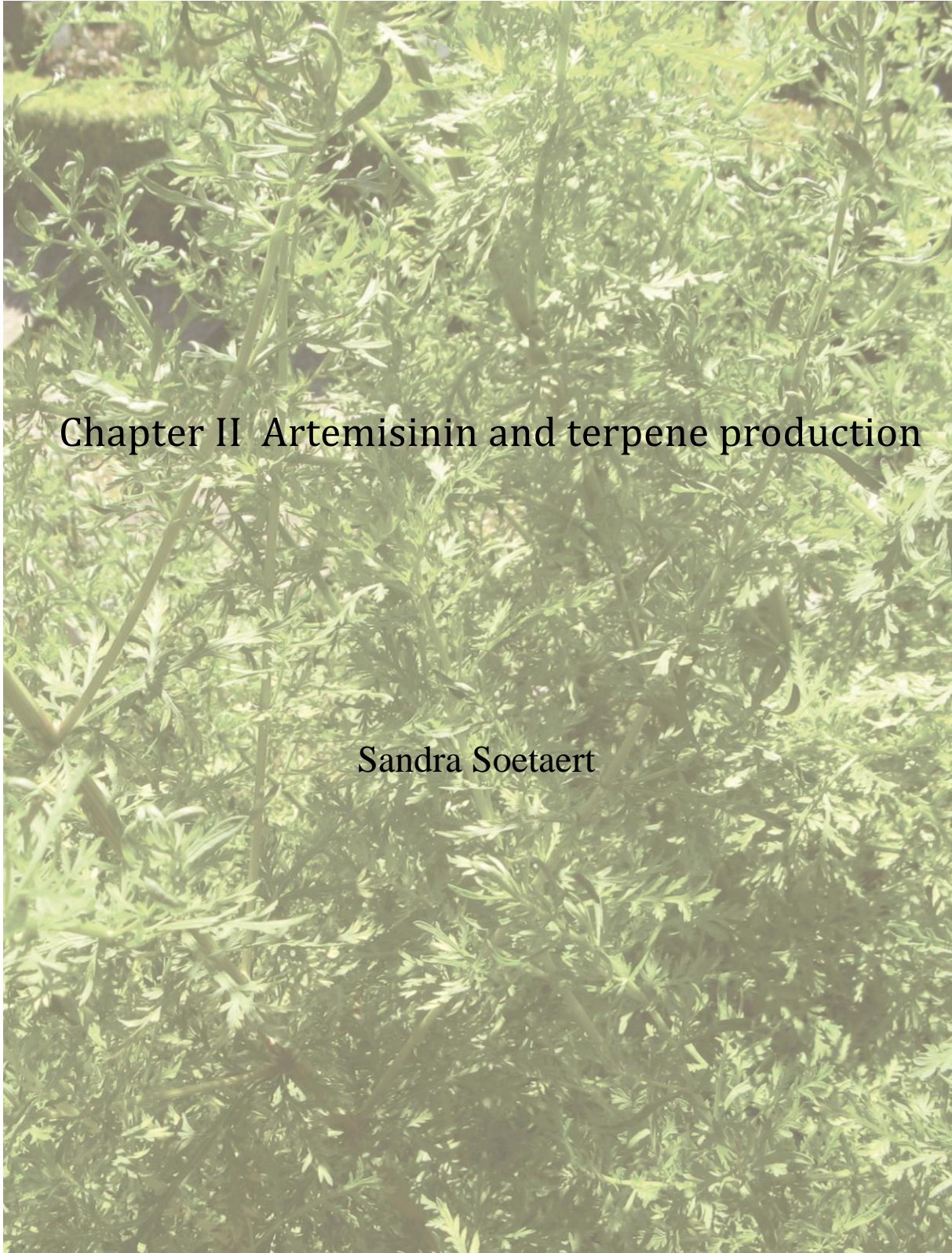
A gene associated with reduced artemisinin efficacy is *pfmdr1* which encodes an ATP-binding cassette transporter. This transporter is present on the membrane of the digestive vacuole of the parasite. As illustrated in Figure 6, there are indications that this transporter pumps drugs from the cytoplasm of the parasite to the food vacuole [25]. In vitro, it was shown that *pfmdr1* gene amplification is correlated with reduced susceptibility to artemisinin-derivatives and higher recrudescence after treatment with artesunate-mefloquine [23]. Removal of drug pressure can result in deamplification of *pfmdr1* and reduced resistance [23].

As illustrated in Figure 6, artemisinin can inhibit PfaTPase6. Therefore, it is not surprising that mutations observed in PfaTPase6 *in vitro* correlate with an increased mean IC₅₀ for artemether [26]. Cheeseman *et al.* identified 33 genome regions associated with artemisinin

resistance. From these regions, 10 were known anti-malarial resistance genes but *pfmdr1* and *pfatpase6* were not amongst them. This shows that the mechanism of artemisinin resistance is still under debate [22].

A mechanism by which *P. falciparum* might survive artemisinin treatment is by arresting its development at the ring stage (Figure 6). If drug pressure is removed, the parasite can resume its normal cell cycle [27]. There are indications that the parasite might enter the quiescent stage by reducing the rate of hemoglobin degradation [23, 28] which can explain why ring-stage parasites can survive short-lasting drug exposure and why high recrudescence is reported after artemisinin monotherapy [23].

Despite the emerging artemisinin resistance, the WHO underlines that the clinical efficacy of ACTs is not yet substantially reduced. Along the Thai-Cambodian border, the cure rate for patients treated with a combination of mefloquine and artesunate was above 90% [29]. However, it is important to closely monitor the spread of resistance.



Chapter II Artemisinin and terpene production

Sandra Soetaert

1 Artemisinin supply

A high supply of artemisinin at a low price is needed to cure patients from malaria. Chemical synthesis of artemisinin is very complicated and expensive [30] and therefore not cost-effective. In *A. annua*, the concentration of artemisinin is only 0.01-0.8% of dry weight [11]. Because of this, artemisinin production is enhanced by crossing high-producing plants [31]. Another strategy to increase artemisinin production is synthesis of artemisinic acid in engineered yeast and subsequent (photo)chemical conversion to artemisinin [32, 33]. Direct synthesis of artemisinin in yeast is only possible if the artemisinin biosynthesis pathway is completely unravelled. This chapter will discuss the general biosynthesis of terpenes and focus on the biosynthesis of artemisinin itself.

2 Terpenoids

2.1 Structure and function

Artemisinin belongs to the group of terpenoid metabolites. Terpenoids (isoprenoids) were named after turpentine [34], the first compound isolated from this group. Another name frequently used is terpene. Originally, the name terpene was used to refer to molecules assembled from five-carbon units based on the isopentane skeleton (Figure 7) and terpenoid or isoprenoid to refer to the entire class of terpenes as well as terpenes that were chemically modified but nowadays, the words terpenes and terpenoids are used interchangeably [35]. Terpenoids are the largest class of all known plant metabolites and includes more than 40 000 structures [36]. Based on their function, terpenoids are classified as primary- or secondary-metabolites.

Primary metabolites are essential for growth, development and general metabolism of plants. An example of such terpenoids are carotenoids [35, 37] which are important in photosynthesis. The phytol side chain of chlorophyll is another terpenoid [37] that enables the chlorophyll to be anchored in the membrane. Gibberellins are terpenoid phytohormones [38] involved in plant growth and development. [35, 37, 38].

In contrast, secondary metabolites, also called natural products, are not essential for plant growth but they influence interactions between plants and its environment. Bad-tasting or toxic metabolites can protect plants against herbivores and good-smelling volatiles attract

pollinators. Secondary plant terpenoids are industrially relevant in pharmaceuticals, flavour, fragrances and pesticides [36]: artemisinin and paclitaxel (Taxol) [35] are well known pharmaceutical terpenoids and menthol [35] is widely used in flavour and fragrance industry. Besides terpenoids, secondary metabolites comprise molecules from the alkaloids, phenylpropanoids and phenols [34].

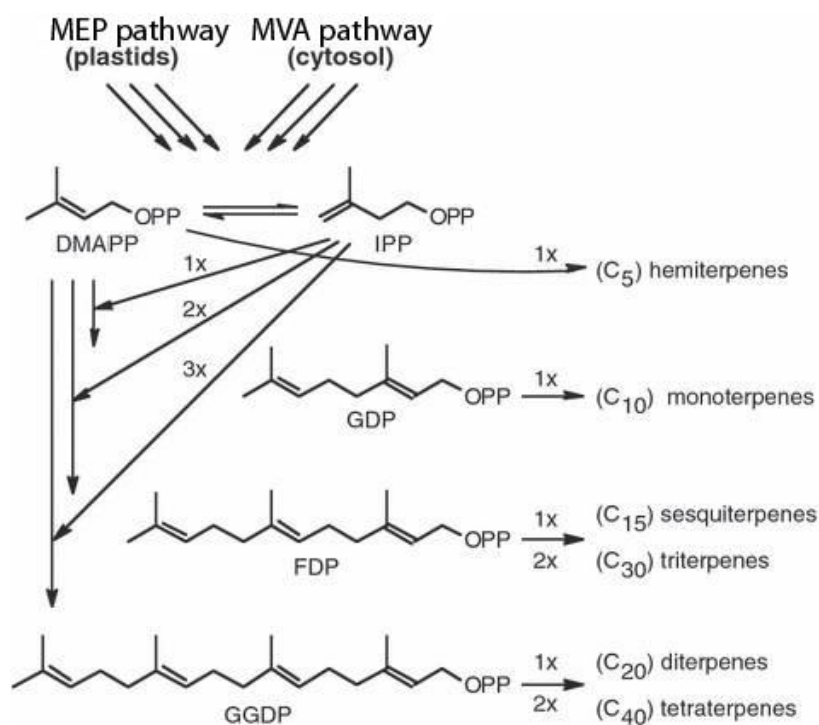


Figure 7: General scheme of plant terpenoid biosynthesis (adapted from [36]).

Isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) are the building blocks for terpenes. DMAPP, geranyl diphosphate (GDP), farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGDP) are precursors for hemi-, mono-, sesqui- and diterpenes. Two FDP molecules linked will form triterpenes and the association of two GGDP molecules leads to tetraterpenes.

Despite the structural diversity of terpenoids, as illustrated in Figure 7, isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl pyrophosphate (DMAPP) are the C₅-building blocks of all terpenoids. Terpenoid classification is based on the number of C₅-units in their structure. Hemiterpenoids contain only one C₅-unit whereas molecules assembled from two C₅-units are called monoterpenoids. In sesquiterpenes such as artemisinin (C₁₅H₂₂O₅), three building blocks are linked and di- and triterpenes contain respectively 4- and 6- C₅-units. Two pathways are involved in the production of IPP: the mevalonate (MVA) pathway and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway.

2.2 Biosynthesis of isoprene units

The mevalonate (MVA) pathway is present in all eukaryotes and some Gram-positive prokaryotes [35]. A schematic overview of the MVA pathway is shown in (Figure 8). This pathway is expressed in the cytosol and starts with the condensation of 3 molecules acetyl-CoA by acetyl-CoA acetyltransferase (AACT) and 3-hydroxy-3-methyl-glutaryl-CoA synthase (HMGS) to HMG-CoA. Subsequently, HMG-CoA is reduced by HMG-CoA reductase (HMGR) to mevalonic acid. HMGR is the key regulatory step from the mevalonate pathway [39, 40]. It has been shown that this step also limits artemisinin biosynthesis [40, 41]. Mevalonate is converted by mevalonate kinase (MVK) to mevalonic acid-5-phosphate and further by phosphomevalonate kinase (PMK) to mevalonic acid-5-diphosphate. Finally, mevalonic acid-5-diphosphate is converted by diphosphomevalonate decarboxylase (PMD) to IPP which can be isomerized to DMAPP by isopentenyl diphosphate isomerase (IDI).

The MVA pathway was the only known source of terpenoids before 1993 [42]. Rohmer *et al.* showed with isotope-labelling studies that there is an alternative pathway which does not originate from acetyl-CoA [42]. An alternative 2-C-methyl-D-erythritol 4-phosphate (MEP) or deoxyxylulose-5-phosphate pathway was discovered that is generally found in many prokaryotes, apicomplexa parasites and photosynthetic eukaryotes [43]. Plants have thus both MVA and MEP pathways [39]. Despite the prokaryotic origin of the MEP pathway, all genes coding for enzymes of this pathway are encoded in the nucleus [43] but the biosynthetic pathway itself is located in plastids [37].

The enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXS) combines pyruvate and glyceraldehyde-3-phosphate to 1-deoxy-D-xylulose-5-phosphate (DXP) (Figure 8). The next step is the conversion of 1-deoxy-D-xylulose-5-phosphate into 2-C-methyl-D-erythritol 4-phosphate by DXP-reductoisomerase (DXR) and this is subsequently converted by 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase (MCT) to 4-diphosphocytidyl-2-C-methyl-D-erythritol. This metabolite is phosphorylated by 4-cytidine 5'-diphospho-2-C-methyl-D-erythritol kinase (CMK) and 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate is formed which is cyclised by 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS) to form 2C-methyl-D-erythritol 2,4-cyclodiphosphate. This is converted by hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (HDS) and hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR) to IPP and DMAPP in a 5:1 ratio [35].

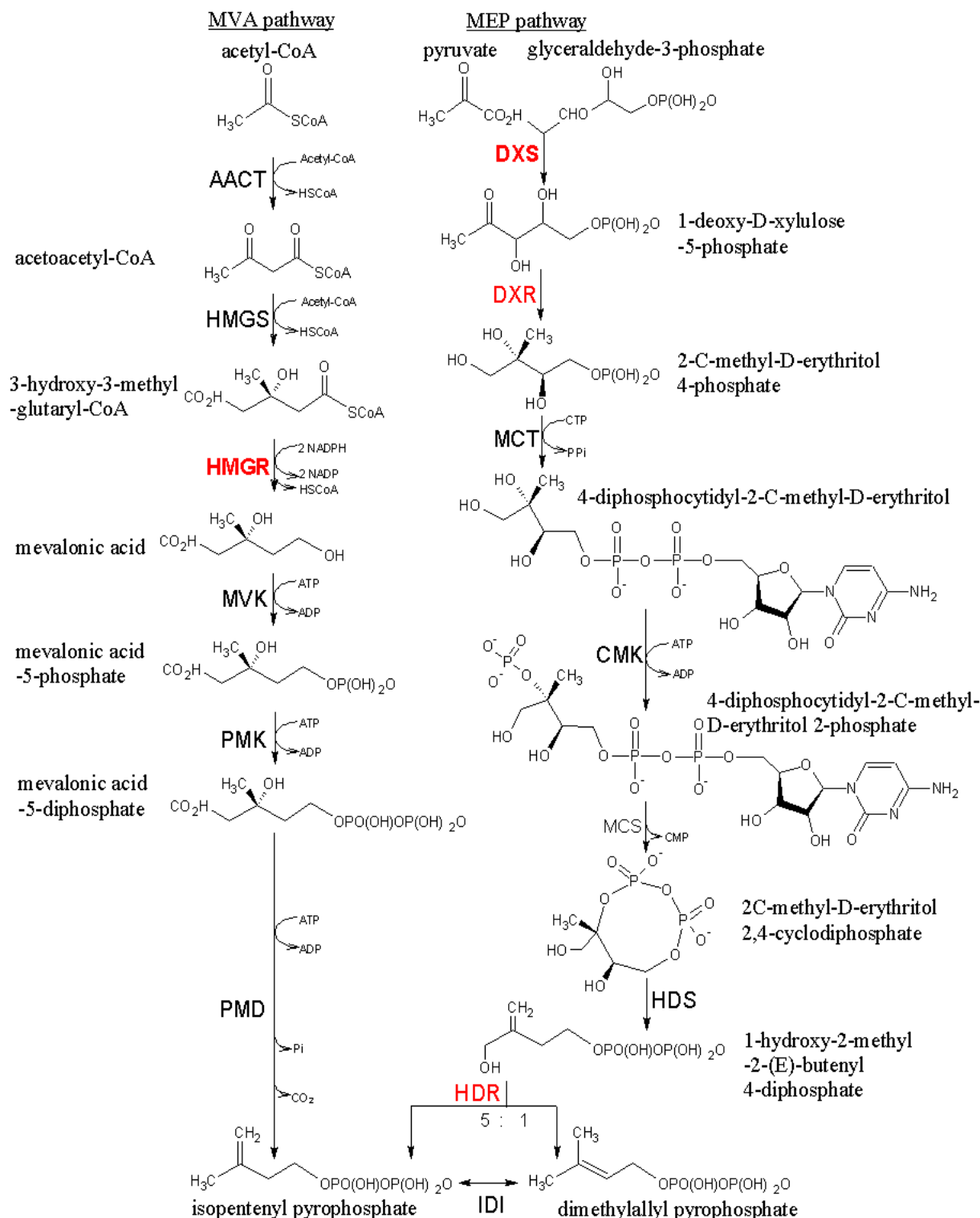


Figure 8: Schematic overview of MVA and MEP pathways.

The MVA pathway starts with acetyl-CoA to synthesize IPP. The MEP pathway uses as input pyruvate and glyceraldehyde-3-phosphate to form IPP and DMAPP. Rate limiting enzymes are labelled in red.

It is well established that DXS plays a role in regulation of the MEP pathway [43]. In addition to this, experimental evidence suggests that DXR and HDR are also rate-limiting in this pathway but their regulatory role appears to vary among plants and in different conditions [43]. Both MVA and MEP pathway are involved in artemisinin production [44] but the MVA pathway is the main contributor [41].

Farnesyl diphosphate synthase (FDS) catalyzes the condensation of two units of IPP and one of DMAPP to form farnesyl diphosphate, the precursor for all sesquiterpenoid molecules such as artemisinin [45]. The artemisinin biosynthesis pathway will be reviewed in the next paragraphs.

3 Artemisinin production

3.1 Trichomes and localization of artemisinin biosynthesis

The first important breakthrough to unravel the biosynthetic pathway of artemisinin was the discovery that the biosynthesis is localized in the glandular trichomes [46]. Trichomes, named after the Greek word for hair, are small outgrowths of epidermal origin on the surfaces of leaves and other organs of many plants. As illustrated in Figure 9, they range in size, shape and number of cells. Well known are the long trichomes that cover the seeds of cotton plants (Figure 9 B) and that are used in the textile industry. Other trichomes on leaf surfaces release the characteristic smell from basil, mint or thyme.

Based on secretory capacity, trichomes are grouped in two categories: non-glandular and glandular [47]. Both types as shown in Figure 9 C are present on leaves, stems and inflorescences of *A. annua*. Non-glandular trichomes from *A. annua* count five cells in a T-shape. The T-top is formed by an elongated cell [48]. Filamentous or non-glandular trichomes are assumed to form a physical barrier by steric hindrance of herbivores [49]. In numerous plant species, a negative correlation is observed between trichome density, insect feeding and oviposition [49, 50]. Furthermore, leaf hairs can serve as a reflector to decrease the light absorbed by leaves and reduce the heat load [51, 52], or impact water retention [53]. Non-glandular trichomes are mainly described for taxonomic and phylogenetic purposes [54-58] but little is known about their production of secondary metabolites [59].

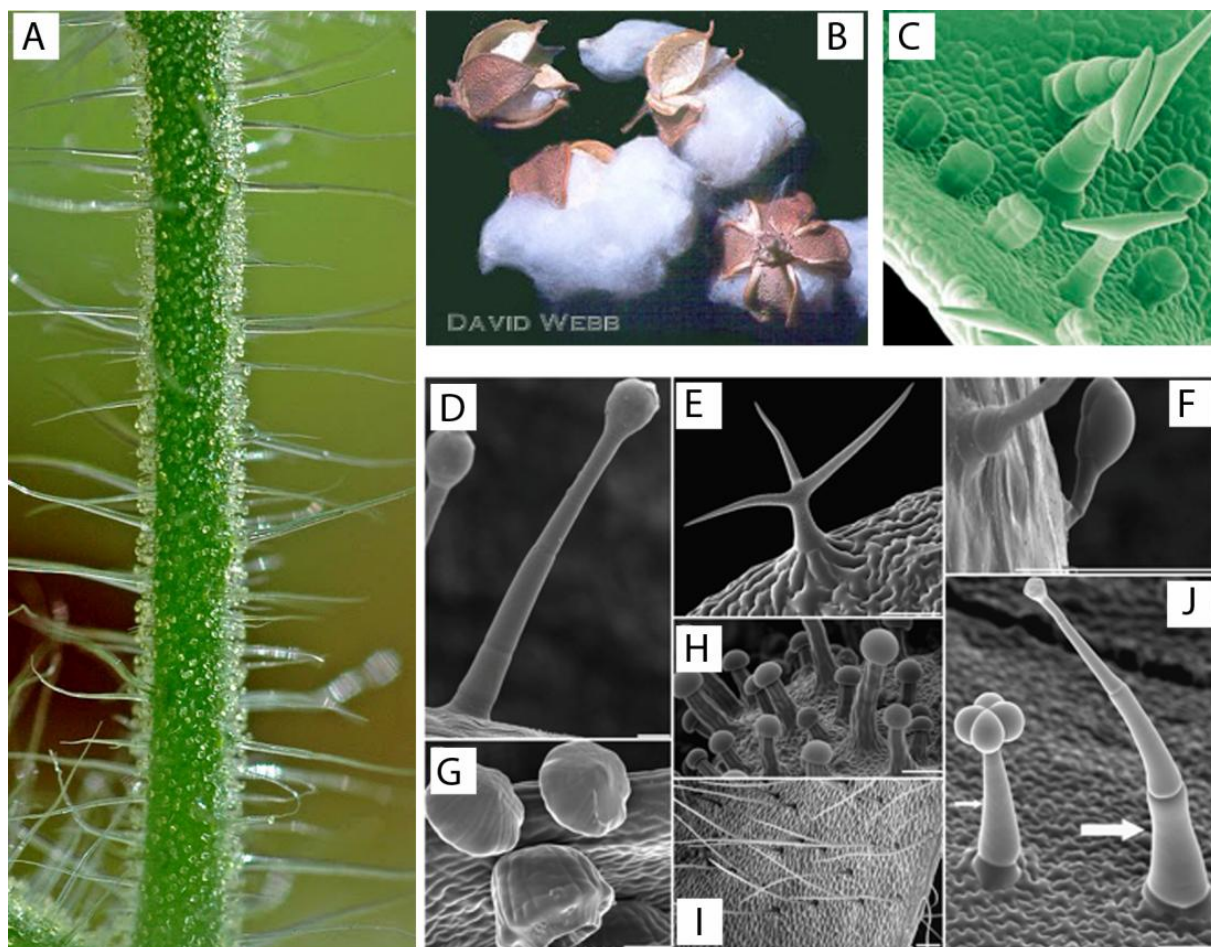


Figure 9: Morphological diversity of plant trichomes.

Pictures illustrate the diversity of plant trichomes in and amongst species. **A:** short gleaming glandular trichomes and long filamentous trichomes on flower pedicel of *Solanum lycopersicum* (Tomato), **B:** long white trichomes on the seeds of cotton, **C:** peltate glandular and T-shaped filamentous trichomes on the surface of *A. annua* (scale bar not available but glandular trichomes are approximately 37 μ m in size) **D:** long glandular trichome from *Medicago sativa* (Alfalfa), **E:** non-glandular trichome from *Arabidopsis* (Mouse-ear cress), **F:** procumbent trichome in *Medicago truncatula* (barrelclover), **G:** glandular trichome from *Humulus lupulus* (Hop), **H:** glandular trichomes on a female bract of *Cannabis sativa* (Marijuana), **I:** non-glandular trichomes on *Medicago truncatula*, **J:** glandular trichomes on *Solanum lycopersicum* (C-J: SEM pictures; D-J: scale bar = 100 μ m).

(A: http://digitalbotanicgarden.blogspot.be/2011_12_01_archive.html, B: http://www.biologie.uni-hamburg.de/b-online/library/webb/BOT311/BOT311-00/anthophyta_significant_life_cycl.htm, C: <http://www.biorenewables.org/about-us/feedstock-developmentunit/microsoft-powerpoint-photos-for-phil-ian-compatibility-mode/>, D-J: [60])

This is in contrast with glandular trichomes which are well characterized as production sites for a variety of secondary metabolites. These metabolites form a first-line defence at the surface of the plant through their capacity to entrap, deter or poison pathogens and herbivores [61]. Besides this, they can attract pollinators or avoid water loss [60]. Several compounds produced by glandular trichomes are commercially interesting for example menthol [62], cannabinoids [63, 64] and essential hop oils used in beer brewing [65].

Glandular trichomes from *A. annua* have a peltate morphology (shield shape) with 5 cell pairs arranged in two rows (biseriate): stalk cells, basal cells and 3 pairs of secretory cells (Figure 10). Secretory cells are bordered by a cuticula which separates from the cell walls and forms a sub-cuticular sac. In mature glandular trichomes, the cuticula splits and releases its content [66]. The sub-cuticular space is probably a storage compartment for phytotoxic secondary metabolites to avoid cell exposure to high levels of autoallelopathic compounds.

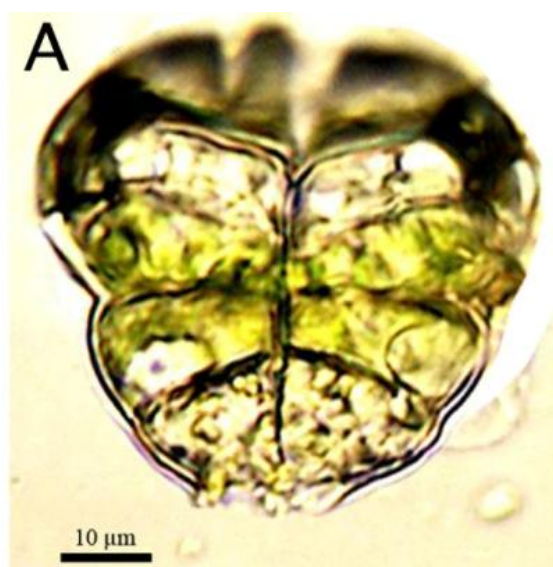


Figure 10: Morphology of a glandular trichome from *A. annua*.

Light microscopy picture of a glandular trichome from *A. annua* with on top a pair of white apical cells and two pairs of green sub-apical cells surrounded by a sub-cuticular cavity (cells at the basis of the trichome are partially removed).

As artemisinin is toxic to *A. annua* itself [67], this compound might be produced in the secretory cells and stored in the sub-cuticular space of glandular trichomes. This hypothesis was tested with a biotype of *A. annua* with only filamentous trichomes that arose spontaneously among field-cultivated plants. Duke *et al.*, compared this biotype to a normal biotype with both filamentous and glandular trichomes. Only in the presence of glandular trichomes, artemisinin was detected [46]. In addition to this, a 5-s dip in chloroform causes

collapse of the sub-cuticular sac from glandular trichomes and extracted 97% of the artemisinin from leaf tissue [46]. These results indicate that artemisinin is produced in glandular trichomes. In the next sections, the biosynthesis of artemisinin is discussed.

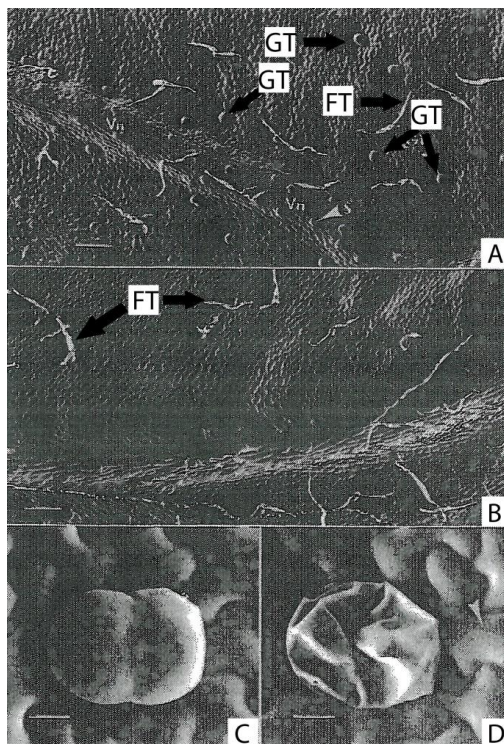


Figure 11: Leaf surface of *A. annua*.

Cryo-SEM of *Artemisia annua* leaf surface adapted from Duke *et al.* [46], glandular trichomes are marked as GT, filamentous trichomes as FT. A: Leaf surface of glanded biotype (scale bar = 100 μm); B: Leaf surface of glandless biotype (scale bar = 100 μm); C: Glandular trichome (scale bar = 10 μm); D: Glandular trichome after 5-s dip in chloroform (scale bar = 10 μm).

3.2 Known artemisinin biosynthesis enzymes

Step by step, the biosynthesis pathway of artemisinin is elucidated (Figure 12). The first step is the cyclization of farnesyl diphosphate to amorpha-4,11-diene. Amorpha-4,11-diene synthase (ADS), the enzyme responsible for this conversion was characterized by Bouwmeester *et al.* [68].

Other breakthroughs in unravelling the artemisinin biosynthesis, were the result of an expressed sequence tag (EST) approach [69]. Three EST libraries were constructed: glandular trichome, flower bud and glandular trichome-minus-flower-bud subtracted library. Genes were identified that are preferentially expressed in glandular trichomes. As was suggested that in the following step(s) of the artemisinin biosynthesis an enzyme of the cytochrome P450

family is involved [68, 70], ESTs corresponding to cytochromes from this family were investigated. This led to the characterization of CYP71AV1 as an enzyme that converts amorpha-4,11-diene via alcohol and aldehyde intermediates to artemisinic acid [32, 69]. Additionally, CYP71AV1 is able to oxidize dihydroartemisinic alcohol to dihydroartemisinic aldehyde but the rate is 50% slower than for artemisinic alcohol [71]. This enzyme was also investigated for its ability to oxidize dihydroartemisinic aldehyde to dihydroartemisinic acid but no net activity was detected [71].

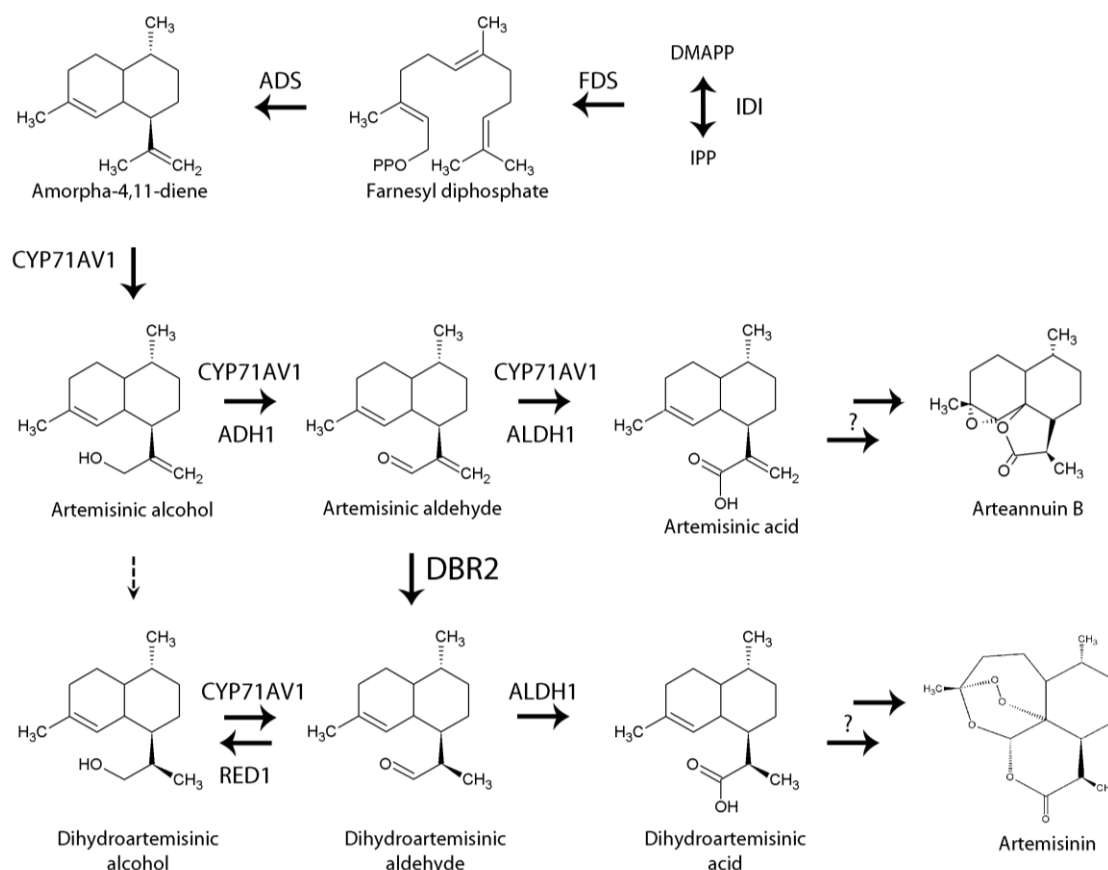


Figure 12: Biosynthetic pathway of artemisinin.

The biosynthesis of artemisinin is initialized by the formation of amorpha-4,11-diene which is subsequently converted to alcohol, aldehyde and acid forms. Double bond reduction leads to dihydroartemisinic forms. Dihydroartemisinic acid is proposed as precursor for artemisinin biosynthesis.

In the EST libraries, a transcript highly expressed in glandular trichomes popped out as a candidate aldehyde dehydrogenase. This enzyme was named aldehyde dehydrogenase 1 (ALDH1) and *in vitro* it was shown to oxidize the substrates artemisinic aldehyde and dihydroartemisinic aldehyde to their acid forms [71].

To convert artemisinic- to dihydroartemisinic-compounds, the double bond is reduced. In an effort to find the enzyme catalyzing this reaction, extracts of *A. annua* were investigated. With partial protein purification, mass spectrometry and EST analysis, Zhang *et al.* found an enzyme that has an artemisinic aldehyde Δ 11(13) double bond reductase activity [72]. The corresponding gene *Dbr2* (double bond reductase 2) was found to be more expressed in glandular trichomes. Purified recombinant *Dbr2* from *Escherichia coli* interacted with artemisinic aldehyde and formed dihydroartemisinic aldehyde. No double bond reductase activity was detected with arteannuin B, artemisinic acid or artemisinic alcohol [72].

An enzyme was characterized that can reduce dihydroartemisinic aldehyde to dihydroartemisinic alcohol: RED1 [73]. This enzyme may have a negative impact on artemisinin production. Olofsson *et al.* reported however that RED1 does not appear to have a significant influence on artemisinin biosynthesis since it is only partially localized in the trichomes and has a relative low turnover potential [74].

An enzyme proposed to oxidize artemisinic alcohol to artemisinic aldehyde is ADH1. A patent is filed for this gene (United States Patent Application 20110162097) but no article is published yet.

In yeast, an experiment was performed with co-expression of farnesyl diphosphate synthase, amorpho-4,11-diene synthase, CYP71AV1, cytochrome P450 reductase and DBR2. This strain was capable of synthesizing dihydroartemisinic acid [72]. It is noteworthy that dihydroartemisinic acid was accumulated even without co-expression of ALDH1.

3.3 Artemisinin biosynthesis from dihydroartemisinic acid

Dihydroartemisinic acid is proposed as precursor of artemisinin [75]. *In vitro*, dihydroartemisinic acid can autooxidize slowly and spontaneously to artemisinin and other products reported in *A. annua* [76]. This occurs in an organic solution (CHCl_3). As illustrated in Figure 13, the reaction mechanism involves 4 steps: initial reaction with molecular oxygen to yield a tertiary allylic hydroperoxide, Hock cleavage, oxygenation of this product and cyclization to the 1,2,4-trioxane system of artemisinin. Conversion was facilitated in open vessels and no transformations were observed when solutions were kept in the dark [76].

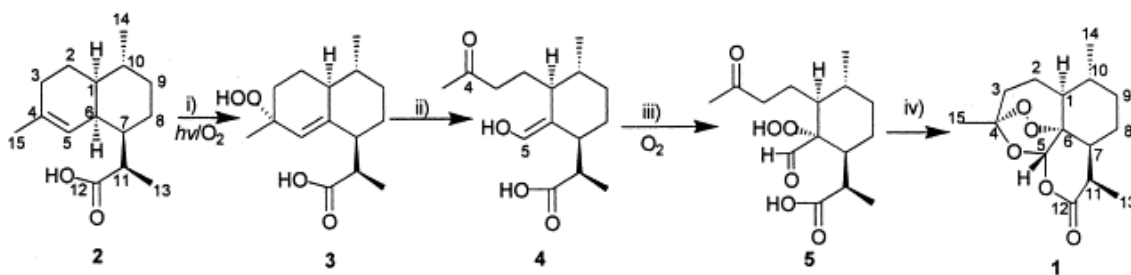


Figure 13: *In vitro* autooxidation of dihydroartemisinic acid to artemisinin [76].

Reaction mechanism for the autooxidation of dihydroartemisinic acid (2) to artemisinin (1) in an organic CHCl_3 solution. Intermediates are tertiary allylic hydroperoxide (3), enolic intermediate (4) and vicinal hydroperoxy-aldehyde (5).

To investigate whether dihydroartemisinic acid is *in vivo* also converted to artemisinin, labelled dihydroartemisinic acid has been fed via the root to intact *A. annua* plants [75]. This approach was chosen since preparation of cell-free extracts causes exposure to the atmosphere and can lead to artificial spontaneous oxygenations. As plants incised just above the root showed a similar distribution of dihydroartemisinic acid, the root is not a physiological barrier for dihydroartemisinic acid uptake [75].

In *A. annua*, close parallels were observed with the spontaneous *in vitro* autooxidation of dihydroartemisinic acid to artemisinin as e.g. labelled allylic hydroperoxide intermediates were detected. In spite of this, spontaneous autooxidations of dihydroartemisinic acid in organic solutions were completed only after several weeks whereas this takes just a few days in *A. annua* [75]. A possible explanation for this is that due to the presence of photosensitizers (under the form of pigments), the relative abundance of singlet oxygen ($^1\text{O}_2$) is high [75]. Singlet oxygen is generally considered to be the reactive species in autooxidation [75].

For the 16 metabolites with an incorporated label, the proportion was estimated in which labels were included. This was 70% with exception for the metabolites artemisinin and arteannuin H. From these compounds, only 5-15% incorporated a label [75].

In vitro, no spontaneous oxidation of dihydroartemisinic acid to artemisinin was observed in aqueous medium, therefore Brown and Sy proposed that this reaction occurs only in a predominant lipophilic environment such as glandular trichomes [75]. In lipophilic solutions, hydroperoxide intermediates were more stable [75]. These lipophilic compartments were probably the last to be accessed after root feeding and this might explain the relatively low label incorporation. This raised the question whether the 14 other metabolites with 70% label

incorporation were synthesized in the more accessible aqueous compartments. In aqueous solutions, dihydroartemisinic acid was mainly converted to *seco*-cadinane or dihydro-*epi*-deoxyarteannuin B [75]. *In vivo*, approximately 70% of these products incorporated a label during the feeding experiment. For the other metabolites, more detailed studies are needed to see if they are products of spontaneous autooxidations in an aqueous environment [75].

The main argument of the authors against enzyme involvement was a feeding experiment with live plants and plants that were dried. They assume that if only 30% of the initial plant weight was present, the intracellular environment was sufficiently dehydrated to inhibit enzymatic processes. Labelled artemisinin levels were too low to compare the live and dead groups but in both groups, a similar isotope concentration of tertiary allylic hydroperoxide was detected [75]. As this is an intermediate of the *in vitro* chemical conversion (see compound 3 in Figure 13) to artemisinin, this indicated also a chemical conversion from dihydroartemisinic acid to artemisinin *in vivo*. However, this did not prove that chemical conversion *in vivo* is the only or major pathway that leads to artemisinin. Since only 5-15% of labelled artemisinin was detected, it is possible that the conversions described in this article represent a minor alternative pathway [75] but that the major conversion to artemisinin was catalyzed by enzymes which were difficultly reached by root feeding of labelled compounds.

4 Increased artemisinin production

4.1 Increased artemisinin production in *A. annua*

The cultivation of *A. annua* is labour-intensive, plants are grown six months and harvested just before flowering [77] as plants in reproductive phase contain more artemisinin in comparison with vegetative plants which contain more dihydroartemisinic acid [78]. Dried leaves and inflorescences are extracted with organic solvents such as hexane and petroleum ether. Each hectare yields around 5 kg artemisinin from 1000 kg dry leaves [77]. The conversion to artemisinin-derivatives such as artesunate and artemether results in approximately 3 kg [77]. To have enough for 216 million adult artemisinin-based-combination (ACT) treatments; 37,000 hectares of *A. annua* plantations are needed. In 2006, around 25,000 hectares of *A. annua* were grown or under construction [77].

To increase the production of artemisinin in *Artemisia annua*, high-producing cultivars were obtained by selection and breeding of high-yielding plants. Mediplant developed a cultivar

called “Artemis” containing up to 1.4% artemisinin (on dry weight of leaves) by crossing two clones with high artemisinin content [79]. In a field trial with this cultivar, approximately 3000 kg of dried leaves were produced per hectare which yields 38 kg of artemisinin [79]. In the CNAP Artemisia Research Project, a huge breeding program was set up to produce new hybrids. Their starting material was the Artemis cultivar as well as several natural populations of *A. annua*. Additionally artificial variation was induced with ethylmethane sulfonate to find beneficial mutations. Genes and markers were identified for fast-track breeding and a quantitative trait loci map was composed [31]. Artemisinin production in *A. annua* plants can be augmented by increasing the total leaf biomass per hectare, by increasing the density of glandular trichomes and by producing more artemisinin in each glandular trichome.

Another possibility to increase the production in *A. annua* is to use genetic modification to overexpress enzymes that limit artemisinin biosynthesis. Since HMGR limits artemisinin biosynthesis, Aquil *et al.* transformed *A. annua* plants with HMGR from *Catharanthus roseus*. These transgenic lines showed up to 22% increase in artemisinin content [40]. MVA and MEP pathways synthesize IPP and DMAPP units and these are converted by farnesyl diphosphate synthase (FDS) to farnesyl diphosphate. Chen *et al.* transformed *A. annua* plants with FDS of *Gossypium arboreum* and in these overexpressed plants, 2-3 times higher artemisinin levels were detected [80].

Farnesyl diphosphate is not only the precursor for artemisinin synthesis but also for the biosynthesis of sterols and other sesquiterpenes such as germacrene A and β -caryophyllene [81]. Therefore, overexpression of ADS which guides farnesyl diphosphate to the artemisinin biosynthesis pathway resulted in approximately 2-fold elevated artemisinin levels [82].

Yang *et al.* followed an alternative strategy by downregulating the competing sterol pathway [83]. Antisense squalene synthase was expressed in *A. annua* to suppress the expression of endogenous squalene synthase and this induced overexpression of ADS, CYP71AV1 and cytochrome P450 reductase [83]. Although downregulation of squalene synthase did not proportionally increase artemisinin content. Suppression of squalene synthase may narrow the route to squalene but does not certainly guide to extra artemisinin [83]. However, treatment of *A. annua* seedlings with the squalene synthase inhibitor miconazole did significantly increase artemisinin concentrations [84]. An RNA silencing method was used by Chen *et al.* to

downregulate expression of β -caryophyllene in *A. annua* and this resulted in 55% increase of artemisinin [81].

4.2 Increase artemisinin production in heterologous organisms

Total chemical synthesis of artemisinin is very complex and not cost-effective [30]. Therefore, artemisinin supply in *A. annua* plants is complemented by semi-synthetic production. Semi-synthetic is used as designation to indicate biologically produced precursors which are further chemically converted to the molecule of interest. Total biosynthesis of artemisinin in bacteria or yeast is still impracticable as the artemisinin biosynthesis pathway is not completely elucidated. With this semi-synthetic system, a more stable artemisinin supply is assured and more farmland is available for food production.

First attempts were made in engineering *Escherichia coli*. *E. coli* contains the MEP pathway but lacks the MVA pathway to produce IPP. Strains were engineered to express the MVA pathway from yeast and additionally ADS [85]. This worked well but, it was very difficult to express membrane-bound cytochrome P450s such as CYP71AV1 due to the absence of an endoplasmic reticulum in *E. coli* [84, 86, 87].

Saccharomyces cerevisiae has an endoplasmic reticulum and produced higher yields of artemisinic acid, therefore it became the preferred system for semi-synthetic production of artemisinin [32, 87]. In yeast, overexpression of the MVA rate-limiting enzyme (a truncated form) HMGR increased the production of amorpha-4,11-diene five-fold [32]. Additionally, amorpha-4,11-diene production was elevated two-fold by downregulating the competing squalene synthase. ADS, CYP71AV1 and cytochrome P450 oxidoreductase were expressed in yeast and the resulting yeast strain produced 4.5% compared to 1.9% artemisinic acid per dry weight in *A. annua* plants and this in only 4-5 days instead of 6 months [32]. Thereafter, artemisinic acid was chemically converted to artemisinin. A semi-synthetic production system for artemisinin was developed by Amyris and commercialized by Sanofi-Aventis. At their website, Sanofi announces their goal is to produce 40 tons of artemisinin in 2013.

Lévesque and Seeberger developed a continuous flow system to convert dihydroartemisinic acid, derived from artemisinic acid to artemisinin [33]. Intermediates must not be isolated and purified prior to flow chemistry in thin tubes wrapped around a light source. Reaction steps were similar as in Figure 13. After initial reduction of artemisinic acid to dihydroartemisinic

acid, this compound was mixed with oxygen and tetraphenylporphyrin as photosensitizer into a tube and photooxidized to form a tertiary allylic hydroperoxide. Trifluoroacetic acid is added which cleaves a carbon ring, this ring reacts with molecular oxygen and the molecule is further condensed to form artemisinin. Dihydroartemisinic acid was converted to artemisinin in 4.5 minutes with a yield of 39% [33].

Recently, Westfall *et al.* overexpressed in addition to HMGR, all other enzymes of the MVA pathway in yeast. This increased the production of artemisinic acid 2-fold. Despite this increase, artemisinic acid was 10-fold lower than amorpha-4,11-diene production [88]. Therefore, the authors developed a method for chemical conversion of amorpha-4,11-diene to dihydroartemisinic acid. However it would be advantageous to produce artemisinic acid in yeast in levels similar to amorpha-4,11-diene as the chemical conversion to artemisinin is easier [88]. For ACTs, artemisinin is converted to its derivatives artesunate, arteether or artemether.

4.3 Aims and objectives

Straightaway production of artemisinin in yeast would be enabled if enzymes are discovered which convert dihydroartemisinic acid to artemisinin. Therefore, the main aim of this PhD project was to search for candidate genes involved in the last step(s) of the artemisinin biosynthesis. The workflow of this PhD project is summarized in Figure 14.

The quest was focussed on the transcriptome level of two types of trichomes on *A. annua*. With 2nd generation sequencing, a comparison was made between glandular trichomes which produce artemisinin and filamentous trichomes which do not produce artemisinin. For the latter, little was known about their production of other secondary metabolites. Therefore, the comparison of glandular and filamentous trichomes provided both biologically relevant information on specific secondary metabolite production in filamentous trichomes (additional aim of this project) and genes upregulated in glandular trichomes and less expressed in filamentous trichomes which might have a function in artemisinin biosynthesis. Another approach to find candidate genes for artemisinin biosynthesis was to elicitate *A. annua* plants with jasmonic acid since jasmonic acid treatment was shown to increase artemisinin production [89]. Glandular trichomes from plants treated with and without jasmonic acid were compared and candidate genes should have been upregulated after jasmonic acid treatment. Results of these experiments are reported in Part III Chapter I.

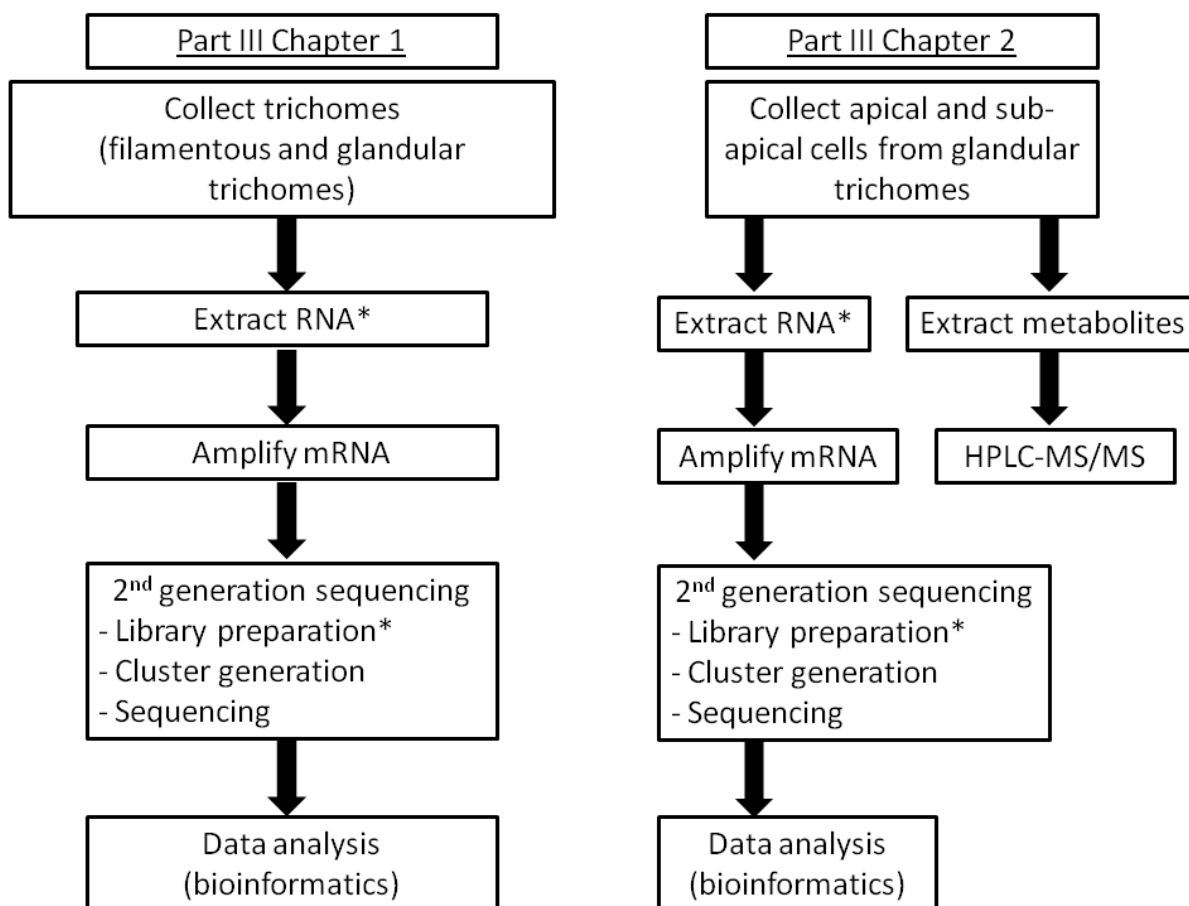


Figure 14: Workflow of this PhD project.

In Part III, Chapter 1: glandular and filamentous trichomes were collected and analyzed at the transcript level. In Part III, Chapter 2: apical and sub-apical cells from glandular trichomes were collected and analyzed at the transcriptome and metabolite level. Steps marked with * were optimized in Part II Chapters 2 and 3.

The last experiment designed to find candidate genes, was based on an article from Olsson *et al.* [90] which claimed that artemisinin production is specifically localized in the apical secretory cells from glandular trichomes. Therefore, two types of secretory cells: apical and sub-apical cells, from the glandular trichomes were compared to discover new candidates. These results are discussed in Part III Chapter II.

In Part II Chapter II and III, technical optimizations were made to find the best experimental setup to collect and prepare transcriptome material for 2nd generation sequencing.

References

Articles and books:

1. Willcox ML, Bodeker G, Bourdy G, Dhingra V, Falquet J, Ferreira JFS, Graz B, Hirt H-M, Hsu E, Magalhães PMD *et al*: **Artemisia annua as a traditional herbal antimalarial**. In: *Traditional Medicinal Plants for Malaria*. Edited by Willcox ML, Bodeker G, Rasoanaivo P, Raton B: CRC Press; 2004: 43-59.
2. Willcox M: **Artemisia species: From traditional medicines to modern antimalarials--and back again**. *J Altern Complement Med* 2009, **15**(2):101-109.
3. Hsu E: **Reflections on the 'discovery' of the antimalarial qinghao**. *British Journal of Clinical Pharmacology* 2006, **61**(6):666-670.
4. Miller LH, Su XZ: **Artemisinin: Discovery from the Chinese Herbal Garden**. *Cell* 2011, **146**(6):855-858.
5. WHO: *World malaria report* Geneva; 2011.
6. Mueller I, Zimmerman PA, Reeder JC: **Plasmodium malariae and Plasmodium ovale - the 'bashful' malaria parasites**. *Trends in Parasitology* 2007, **23**(6):278-283.
7. Collins WE, Jeffery GM: **Plasmodium ovale: Parasite and disease**. *Clin Microbiol Rev* 2005, **18**(3):570-581.
8. Cox-Singh J: **Zoonotic malaria: Plasmodium knowlesi, an emerging pathogen**. *Curr Opin Infect Dis* 2012, **25**(5):530-536.
9. Tuteja R: **Malaria - an overview**. *Febs J* 2007, **274**(18):4670-4679.
10. Su XZ, Hayton K, Wellems TE: **Genetic linkage and association analyses for trait mapping in Plasmodium falciparum**. *Nature Reviews Genetics* 2007, **8**(7):497-506.
11. van Aghtmael MA, Eggelte TA, van Boxtel CJ: **Artemisinin drugs in the treatment of malaria: from medicinal herb to registered medication**. *Trends in Pharmacological Sciences* 1999, **20**(5):199-205.
12. WHO: *Guidelines for the treatment of malaria* Geneva; 2010.
13. McMorro ML, Aidoo M, Kachur SP: **Malaria rapid diagnostic tests in elimination settings-can they find the last parasite?** *Clin Microbiol Infect* 2011, **17**(11):1624-1631.
14. Fernando SD, Rodrigo C, Rajapakse S: **Chemoprophylaxis in malaria: drugs, evidence of efficacy and costs**. *Asian Pac J Trop Med* 2011, **4**(4):330-336.
15. Balint GA: **Artemisinin and its derivatives - An important new class of antimalarial agents**. *Pharmacology & Therapeutics* 2001, **90**(2-3):261-265.
16. de Ridder S, van der Kooy F, Verpoorte R: **Artemisia annua as a self-reliant treatment for malaria in developing countries**. *Journal of Ethnopharmacology* 2008, **120**(3):302-314.
17. deVries PJ, Dien TK: **Clinical pharmacology and therapeutic potential of artemisinin and its derivatives in the treatment of malaria**. *Drugs* 1996, **52**(6):818-836.
18. Meshnick SR, Taylor TE, Kamchonwongpaisan S: **Artemisinin and the antimalarial endoperoxides: From herbal remedy to targeted chemotherapy**. *Microbiol Rev* 1996, **60**(2):301-315.

19. O'Neill PM, Barton VE, Ward SA: **The Molecular Mechanism of Action of Artemisinin-The Debate Continues.** *Molecules* 2010, **15**(3):1705-1721.
20. Ding XC, Beck HP, Raso G: **Plasmodium sensitivity to artemisinins: magic bullets hit elusive targets.** *Trends in Parasitology* 2011, **27**(2):73-81.
21. Wang J, Huang LY, Li J, Fan QW, Long YC, Li Y, Zhou B: **Artemisinin Directly Targets Malarial Mitochondria through Its Specific Mitochondrial Activation.** *PLoS One* 2010, **5**(3):A158-A169.
22. Cheeseman IH, Miller BA, Nair S, Nkhoma S, Tan A, Tan JC, Al Saai S, Phyto AP, Moo CL, Lwin KM *et al*: **A major genome region underlying artemisinin resistance in malaria.** *Science* 2012, **336**(6077):79-82.
23. O'Brien C, Henrich PP, Passi N, Fidock DA: **Recent clinical and molecular insights into emerging artemisinin resistance in *Plasmodium falciparum*.** *Curr Opin Infect Dis* 2011, **24**(6):570-577.
24. Dondorp AM, Nosten F, Yi P, Das D, Phyto AP, Tarning J, Lwin KM, Arie F, Hanpithakpong W, Lee SJ *et al*: **Artemisinin Resistance in *Plasmodium falciparum* Malaria.** *New England Journal of Medicine* 2009, **361**(5):455-467.
25. Rohrbach P, Sanchez CP, Hayton K, Friedrich O, Patel J, Sidhu ABS, Ferdig MT, Fidock DA, Lanzer M: **Genetic linkage of *pfmdr1* with food vacuolar solute import in *Plasmodium falciparum*.** *Embo J* 2006, **25**(13):3000-3011.
26. Jambou R, Legrand E, Niang M, Khim N, Lim P, Volney B, Ekala MT, Bouchier C, Esterre P, Fandeur T *et al*: **Resistance of *Plasmodium falciparum* field isolates to in-vitro artemether and point mutations of the SERCA-type PfATPase6.** *Lancet* 2005, **366**(9501):1960-1963.
27. Witkowski B, Lelievre J, Barragan MJL, Laurent V, Su XZ, Berry A, Benoit-Vical F: **Increased tolerance to artemisinin in *Plasmodium falciparum* is mediated by a quiescence mechanism.** *Antimicrob Agents Chemother* 2010, **54**(5):1872-1877.
28. Klonis N, Crespo-Ortiz MP, Bottova I, Abu-Bakar N, Kenny S, Rosenthal PJ, Tilley L: **Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(28):11405-11410.
29. Satimai W, Sudathip P, Vijaykadga S, Khamsiriwatchara A, Sawang S, Potithavoranan T, Sangvichean A, Delacollette C, Singhasivanon P, Kaewkungwal J *et al*: **Artemisinin resistance containment project in Thailand. II: responses to mefloquine-artesunate combination therapy among *falciparum* malaria patients in provinces bordering Cambodia.** *Malar J* 2012, **11**(300):1-13.
30. Covello PS: **Making artemisinin.** *Phytochemistry* 2008, **69**(17):2881-2885.
31. Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM *et al*: **The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin.** *Science* 2010, **327**(5963):328-331.
32. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J *et al*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440**(7086):940-943.
33. Levesque F, Seeberger PH: **Continuous-flow synthesis of the anti-malaria drug artemisinin.** *Angew Chem-Int Edit* 2012, **51**(7):1706-1709.

34. Buchanan BB, Gruissem W, Jones RL: **Biochemistry and molecular biology of plants**. In.: Rockville (MD) : American society of plant physiologists; 2000: 1250-1318.
35. Withers ST, Keasling JD: **Biosynthesis and engineering of isoprenoid small molecules**. *Applied Microbiology and Biotechnology* 2007, **73**(5):980-990.
36. Bohlmann J, Keeling CI: **Terpenoid biomaterials**. *Plant Journal* 2008, **54**(4):656-669.
37. Dubey VS, Bhalla R, Luthra R: **An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants**. *J Biosci* 2003, **28**(5):637-646.
38. Tholl D: **Terpene synthases and the regulation, diversity and biological roles of terpene metabolism**. *Current Opinion in Plant Biology* 2006, **9**(3):297-304.
39. Weathers PJ, Elkholy S, Wobbe KK: **Artemisinin: The biosynthetic pathway and its regulation in *Artemisia annua*, a terpenoid-rich species**. *In Vitro Cellular & Developmental Biology-Plant* 2006, **42**(4):309-317.
40. Aquil S, Husaini AM, Abdin MZ, Rather GM: **Overexpression of the HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic *Artemisia annua* plants**. *Planta Medica* 2009, **75**(13):1453-1458.
41. Ram M, Khan MA, Jha P, Khan S, Kiran U, Ahmad MM, Javed S, Abdin MZ: **HMG-CoA reductase limits artemisinin biosynthesis and accumulation in *Artemisia annua* L. plants**. *Acta Physiologiae Plantarum* 2010, **32**(5):859-866.
42. Rohmer M, Knani M, Simonin P, Sutter B, Sahn H: **Isoprenoid biosynthesis in bacteria - a novel pathway for the early steps leading to isopentenyl diphosphate**. *Biochemical Journal* 1993, **295**:517-524.
43. Cordoba E, Salmi M, Leon P: **Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants**. *Journal of Experimental Botany* 2009, **60**(10):2933-2943.
44. Towler MJ, Weathers PJ: **Evidence of artemisinin production from IPP stemming from both the mevalonate and the nonmevalonate pathways**. *Plant Cell Reports* 2007, **26**(12):2129-2136.
45. Brown GD: **The Biosynthesis of Artemisinin (Qinghaosu) and the Phytochemistry of *Artemisia annua* L. (Qinghao)**. *Molecules* 2010, **15**(11):7603-7698.
46. Duke MV, Paul RN, Elsohly HN, Sturtz G, Duke SO: **Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua* L.** *International Journal of Plant Sciences* 1994, **155**(3):365-372.
47. Schilmiller AL, Last RL, Pichersky E: **Harnessing plant trichome biochemistry for the production of useful compounds**. *Plant Journal* 2008, **54**(4):702-711.
48. Ferreira JFS, Janick J: **Floral morphology of *Artemisia annua* with special reference to trichomes**. *International Journal of Plant Sciences* 1995, **156**(6):807-815.
49. Levin DA: **The role of trichomes in plant defence**. *Quarterly Review of Plant Biology* 1973, **48**(1):3-15.
50. Mauricio R, Rausher MD: **Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense**. *Evolution* 1997, **51**(5):1435-1444.
51. Ehleringer J, Bjorkman O, Mooney HA: **Leaf pubescence: Effects on absorptance and photosynthesis in a desert shrub**. *Science* 1976, **192**(4237):376-377.

52. Manetas Y: **The importance of being hairy: the adverse effects of hair removal on stem photosynthesis of *Verbascum speciosum* are due to solar UV-B radiation.** *New Phytologist* 2003, **158**(3):503-508.
53. Meinzer F, Goldstein G: **Some consequences of leaf pubescence in the andean giant rosette plant *Espeletia timotensis*.** *Ecology* 1985, **66**(2):512-520.
54. Victorio CP, Moreira CB, Souza MD, Sato A, Arruda RDD: **Secretory Cavities and Volatiles of *Myrrhinium atropurpureum* Schott var. *atropurpureum* (Myrtaceae): An Endemic Species Collected in the Restingas of Rio de Janeiro, Brazil.** *Natural Product Communications* 2011, **6**(7):1045-1050.
55. Porto NM, De Figueiredo R, Oliveira AFM, Agra MD: **Leaf Epidermal Characteristics of *Cissampelos* L. (Menispermaceae) Species from Northeastern Brazil.** *Microscopy Research and Technique* 2011, **74**(4):370-376.
56. Bhatt A, Naidoo Y, Nicholas A: **The foliar trichomes of *Hypoestes aristata* (Vahl) Sol. ex Roem. & Schult var *aristata* (Acanthaceae) a widespread medicinal plant species in tropical sub-Saharan Africa: with comments on its possible phylogenetic significance.** *Biological Research* 2010, **43**(4):403-409.
57. Duarte MR, Lopes JF: **Leaf and stem morphoanatomy of *Petiveria alliacea*.** *Fitoterapia* 2005, **76**(7-8):599-607.
58. Bonzani NE, Barboza GE, Bugatti MA, Espinar LA: **Morpho-histological studies in the aromatic species of *Chenopodium* from Argentina.** *Fitoterapia* 2003, **74**(3):207-225.
59. Wagner GJ, Wang E, Shepherd RW: **New approaches for studying and exploiting an old protuberance, the plant trichome.** *Annals of Botany* 2004, **93**(1):3-11.
60. Dai XB, Wang GD, Yang DS, Tang YH, Broun P, Marks MD, Sumner LW, Dixon RA, Zhao PX: **TrichOME: A Comparative Omics Database for Plant Trichomes.** *Plant Physiology* 2010, **152**(1):44-54.
61. Wagner GJ: **Secreting glandular trichomes: more than just hairs.** *Plant Physiology* 1991, **96**(3):675-679.
62. Gershenzon J, Maffei M, Croteau R: **Biochemical and histochemical localization of monoterpenes biosynthesis in the glandular trichomes of Spearmint (*Mentha spicata*).** *Plant Physiology* 1989, **89**(4):1351-1357.
63. Dayanandan P, Kaufman PB: **Trichomes of *Cannabis sativa* L. (Cannabaceae).** *American Journal of Botany* 1976, **63**(5):578-591.
64. Turner JC, Hemphill JK, Mahlberg PG: **Quantitative determination of cannabinoids in individual glandular trichomes of *Cannabis sativa* L. (Cannabaceae).** *American Journal of Botany* 1978, **65**(10):1103-1106.
65. Wang GD, Tian L, Aziz N, Broun P, Dai XB, He J, King A, Zhao PX, Dixon RA: **Terpene Biosynthesis in Glandular Trichomes of Hop.** *Plant Physiology* 2008, **148**(3):1254-1266.
66. Duke SO, Paul RN: **Development and fine-structure of the glandular trichomes of *Artemisia annua* L.** *International Journal of Plant Sciences* 1993, **154**(1):107-118.
67. Duke SO, Vaughn KC, Croom EM, Elsohly HN: **Artemisinin, a constituent of annual wormwood (*Artemisia-annua*), is a selective phytotoxin.** *Weed Science* 1987, **35**(4):499-505.
68. Bouwmeester HJ, Wallaart TE, Janssen MHA, van Loo B, Jansen BJM, Posthumus MA, Schmidt CO, De Kraker JW, Konig WA, Franssen MCR: **Amorpha-4,11-diene**

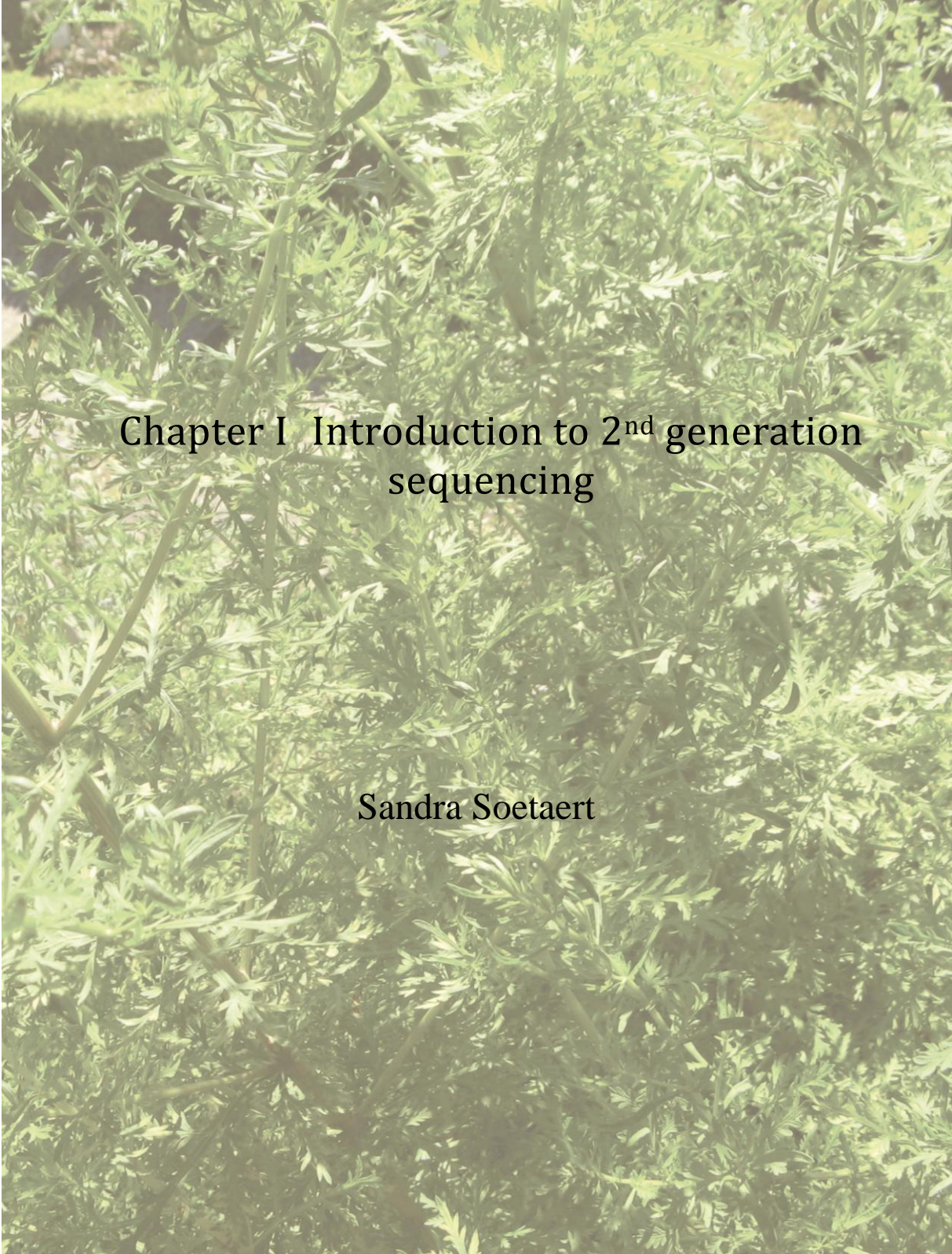
- synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochemistry* 1999, **52**(5):843-854.
69. Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS: **Artemisia annua L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin.** *Febs Letters* 2006, **580**(5):1411-1416.
70. Berteau CM, Freije JR, van der Woude H, Verstappen FWA, Perk L, Marquez V, De Kraker JW, Posthumus MA, Jansen BJM, de Groot A *et al*: **Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in Artemisia annua.** *Planta Medica* 2005, **71**(1):40-47.
71. Teoh KH, Polichuk DR, Reed DW, Covello PS: **Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in Artemisia annua.** *Botany* 2009, **87**(6):635-642.
72. Zhang Y, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJH, Ross ARS, Covello PS: **The molecular cloning of artemisinic aldehyde Delta 11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in Artemisia annua.** *Journal of Biological Chemistry* 2008, **283**(31):21501-21508.
73. Ryden AM, Ruyter-Spira C, Quax WJ, Osada H, Muranaka T, Kayser O, Bouwmeester H: **The Molecular Cloning of Dihydroartemisinic Aldehyde Reductase and its Implication in Artemisinin Biosynthesis in Artemisia annua.** *Planta Medica* 2010, **76**(15):1778-1783.
74. Olofsson L, Engstrom A, Lundgren A, Brodelius PE: **Relative expression of genes of terpene metabolism in different tissues of Artemisia annua L.** *BMC Plant Biol* 2011, **11**(45):1-12.
75. Brown GD, Sy LK: **In vivo transformations of dihydroartemisinic acid in Artemisia annua plants.** *Tetrahedron* 2004, **60**(5):1139-1159.
76. Sy LK, Brown GD: **The mechanism of the spontaneous autoxidation of dihydroartemisinic acid.** *Tetrahedron* 2002, **58**(5):897-908.
77. WHO: *Meeting on the production of artemisinin and artemisinin-based combination therapies* Tanzania, 2005-2006.
78. Arsenault PR, Vail D, Wobbe KK, Erickson K, Weathers PJ: **Reproductive Development Modulates Gene Expression and Metabolite Levels with Possible Feedback Inhibition of Artemisinin in Artemisia annua.** *Plant Physiology* 2010, **154**(2):958-968.
79. Delabays N, Simonnet X, Gaudin M: **The genetics of artemisinin content in Artemisia annua L. and the breeding of high yielding cultivars.** *Current Medicinal Chemistry* 2001, **8**(15):1795-1801.
80. Chen DH, Ye HC, Li GF: **Expression of a chimeric farnesyl diphosphate synthase gene in Artemisia annua L. transgenic plants via Agrobacterium tumefaciens-mediated transformation.** *Plant Science* 2000, **155**(2):179-185.
81. Chen JL, Fang HM, Ji YP, Pu GB, Guo YW, Huang LL, Du ZG, Liu BY, Ye HC, Li GF *et al*: **Artemisinin Biosynthesis Enhancement in Transgenic Artemisia annua Plants by Downregulation of the beta-Caryophyllene Synthase Gene.** *Planta Med* 2011, **77**(15):1759-1765.
82. Ma CF, Wang HH, Lu X, Wang H, Xu GW, Liu BY: **Terpenoid metabolic profiling analysis of transgenic Artemisia annua L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry.** *Metabolomics* 2009, **5**(4):497-506.

83. Yang RY, Feng LL, Yang XQ, Yin LL, Xu XL, Zeng QP: **Quantitative Transcript Profiling Reveals Down-Regulation of A Sterol Pathway Relevant Gene and Overexpression of Artemisinin Biogenetic Genes in Transgenic *Artemisia annua* Plants.** *Planta Medica* 2008, **74**(12):1510-1516.
84. Arsenault PR, Wobbe KK, Weathers PJ: **Recent Advances in Artemisinin Production Through Heterologous Expression.** *Current Medicinal Chemistry* 2008, **15**(27):2886-2896.
85. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids.** *Nature Biotechnology* 2003, **21**(7):796-802.
86. Leonard E, Koffas MAG: **Engineering of artificial plant cytochrome p450 enzymes for synthesis of isoflavones by *Escherichia coli*.** *Applied and Environmental Microbiology* 2007, **73**(22):7246-7251.
87. Chang MCY, Eachus RA, Trieu W, Ro DK, Keasling JD: **Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s.** *Nature Chemical Biology* 2007, **3**(5):274-277.
88. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A *et al*: **Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(3):111-118.
89. Maes L, Van Nieuwerburgh FCW, Zhang YS, Reed DW, Pollier J, Castele S, Inze D, Covello PS, Deforce DLD, Goossens A: **Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants.** *New Phytologist* 2011, **189**(1):176-189.
90. Olsson ME, Olofsson LM, Lindahl AL, Lundgren A, Brodelius M, Brodelius PE: **Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L.** *Phytochemistry* 2009, **70**(9):1123-1128.

Websites:

1. http://gamapserver.who.int/mapLibrary/Files/Maps/Global_Malaria_2010.png (25-10-2012).
2. http://www.itg.be/itg/distancelearning/lecturenotesvandenendene/imagehtml/ppages/kabisa_1522.htm (25-10-2012).
3. http://digitalbotanicgarden.blogspot.be/2011_12_01_archive.html (05-11-2012)
4. http://www.biologie.uni-hamburg.de/b-online/library/webb/BOT311/BOT311-00/anthophyta_significant_life_cycl.htm (05-11-2012)
5. <http://www.biorenewables.org/about-us/feedstock-developmentunit/microsoft-powerpoint-photos-for-phil-ian-compatibility-mode/> (05-11-2012)

Part II. Technical optimizations for sequencing



Chapter I Introduction to 2nd generation sequencing

Sandra Soetaert

To enhance the artemisinin supply, it would be beneficial if direct production of artemisinin in yeast is enabled. Due to gaps in our knowledge of the biosynthesis pathway, this was not yet possible. Therefore, the aim of this project was to discover new candidate genes which can close this gap. This research project was mainly focussed on trichomes from *A. annua*. As it was known that plants with both glandular and filamentous trichomes contain artemisinin and plants with only filamentous trichomes do not, it is clear that artemisinin biosynthesis takes place in glandular trichomes. Therefore, a comparison of filamentous and glandular trichomes can generate a list of candidates. As proteome analysis on small samples such as trichomes is difficult, the decision was made to work on the transcriptome level. New 2nd generation high-throughput sequencing platforms offer the opportunity to perform whole transcriptome analysis. In Chapter I, the basis of these sequencing technologies will be explained and an overview is given of several commercial platforms. As the Illumina HiSeq2000 was used in this project, most attention will be given to discuss the workflow of this system. To obtain good results with 2nd generation sequencing, the quality of the input RNA material is important. Procedures to collect good quality RNA were optimized as reported in Chapter II. In Chapter III, it was investigated whether the commercial Illumina protocol introduces bias if barcodes were used for multiplexing several samples.

With 2nd generation sequencing, also called next generation sequencing, millions of DNA or RNA fragments can be sequenced in parallel. This enables for example the comparison of the whole transcriptome in glandular and filamentous trichomes to find candidate genes for artemisinin production. At the genome level, it also facilitates for example genetic association studies to link genomic regions of *A. annua* cultivars with traits such as biomass production, density of trichomes and artemisinin production in each trichome [2]. In 2005, 454 sequencing (Roche Applied Science, Basel, Switzerland) was the first 2nd generation sequencing technology that became commercially available [3], followed by other platforms: Illumina or Solexa (Illumina, San Diego, California, USA), SOLiD and Ion Torrent (Life Technologies, Paisly, UK).

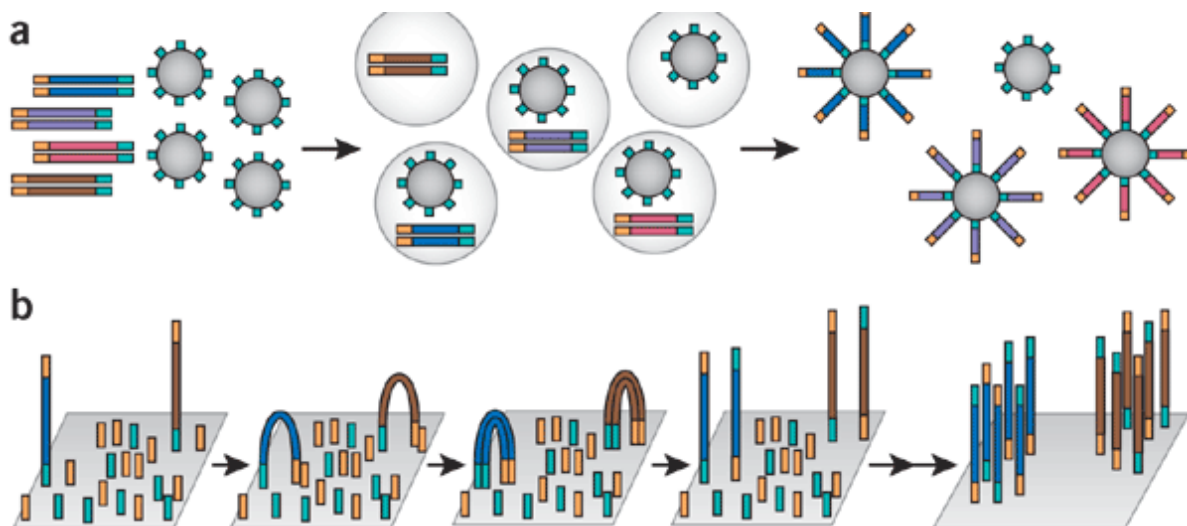


Figure 1: Clonal amplification for second generation sequencing [4].

A: Emulsion PCR in 454, SOLiD and Ion Torrent platforms. **B:** Bridge PCR in Illumina platform.

Although the sequencing biochemistry of these platforms is quite diverse, they are based on a similar concept. Depending on the platform, the maximum read length is between 100 and 1000 nucleotides and to obtain fragments with appropriate lengths, the starting material is randomly fragmented. To overcome detection problems during sequencing of single molecules, fragments are clonally amplified *in vitro* by bridge PCR or emulsion PCR as shown in Figure 1. In the case of bridge amplification, adaptors are ligated to the DNA fragments which bind to spatially isolated oligonucleotides attached on a flow cell. Around each DNA sequence, a cluster is formed with clonally amplified fragments (see Figure 1). Library preparation by bridge amplification is explained on page 51. This technology is adopted in the Illumina platform. In 454 sequencing, SOLiD and Ion Torrent, emulsion PCR

is performed. Adaptors are ligated to the DNA fragments. Within an oil phase, adaptors of each DNA molecule bind to a bead in an aqueous droplet (see Figure 1). Ideally, each droplet contains exactly one bead and only one DNA molecule. This aqueous droplet forms a mini-reactor and isolates the clonally amplified fragments. For sequencing, each clonal-group is spatially isolated on a plate (SOLiD) or in wells (454 and Ion Torrent) which enables sequencing in parallel. In SOLiD, fragments are sequenced by ligation whereas in the other platforms they are sequenced by synthesis. In the next paragraph, the sequencing biochemistry of these 2nd generation sequencing technologies is explained and advantages or disadvantages of the systems are given.

1 Comparison of sequencing platforms

The sequencing technology used in the 454 systems is called pyrosequencing. The basic principle is illustrated in Figure 2. All four types of nucleotides are sequentially added and incorporation of a complementary nucleotide releases pyrophosphate. In the presence of ATP sulfurylase, pyrophosphate reacts with adenosine 5' phosphosulfate (APS) and ATP is formed [5]. The amount of ATP correlates with the number of incorporated nucleotides. Luciferase converts ATP and luciferin to oxy luciferin and visible light is generated. Apyrase degrades unincorporated nucleotides and subsequently another nucleotide is added [5]. Instead of dATP, dATP α S is used because this is not recognised by luciferase [5]. The main advantages of 454 sequencing is the long read length up to 1000 bp with the GS FLX Titanium XL+ system (<http://454.com/products/gs-flx-system/index.asp>). Its speed is an additional advantage since it takes only 10 hours to complete the sequencing itself [3]. Another advantage is that library construction can be automated and emulsion PCR semi-automated which reduces manpower costs but on the other hand reagents are costly: \$12.56 per million bases [3]. However, an important disadvantage is that sequencing of homopolymers longer than 6 bp results in high error rates as it becomes very difficult to distinguish the light emitted by incorporation of e.g. 7 or 8 nucleotides [3].

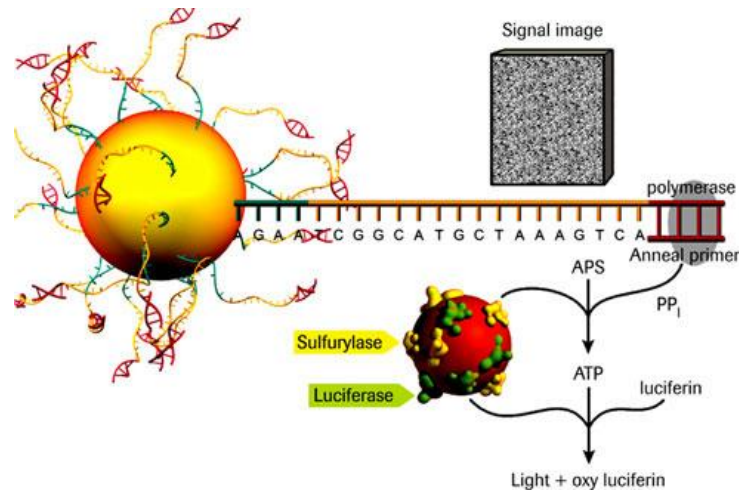


Figure 2: Pyrosequencing chemistry in 454 sequencing.

Incorporation of a nucleotide generates the formation of oxy luciferin and light.

(<http://454.com/products/technology.asp>)

Illumina sequencing is like 454 sequencing based on sequencing-by-synthesis but instead of pyroluminescence, it uses fluorescence. Each nucleotide is modified with a reversible chain terminator and labelled with a distinct fluorescent dye. Because of this, all 4 nucleotides can be added at once. After incorporation of 1 nucleotide in each strand, the fluorescent dye and chain terminator group are removed and the following nucleotide can be incorporated. In comparison with the older Illumina GA series, HiSeq2000 (available since 2010) gives 2- to 5-fold more data output. The enhanced output is reached by imaging the top and bottom surface of the flow cell. Speed is increased in the new system by simultaneous detection of four bases by using 4 cameras instead of one and imaging by line scanning mode instead of area imaging [6]. Run time is between 3 and 10 days, depending on the read length required and single and or paired-end option chosen. In comparison with 454 sequencing, SOLiD and Ion Torrent PGM, Illumina HiSeq2000 generates the largest data output (600 Gb/run) at the cheapest reagent cost: \$ 0.02 per million bases [3]. Bias observed in Illumina systems are increased coverage of regions with < 47% GC and most frequently A to C conversions [6]. Illumina is most prone to errors after GC-rich motifs or long homopolymers > 20 bases but total error rates (< 0.4%) are lower than in Ion Torrent and PacBio [7]. For our transcriptome analysis, Illumina HiSeq2000 sequencing was used. The procedure for Illumina sequencing will be explained in more detail in paragraph 2 Illumina sequencing.

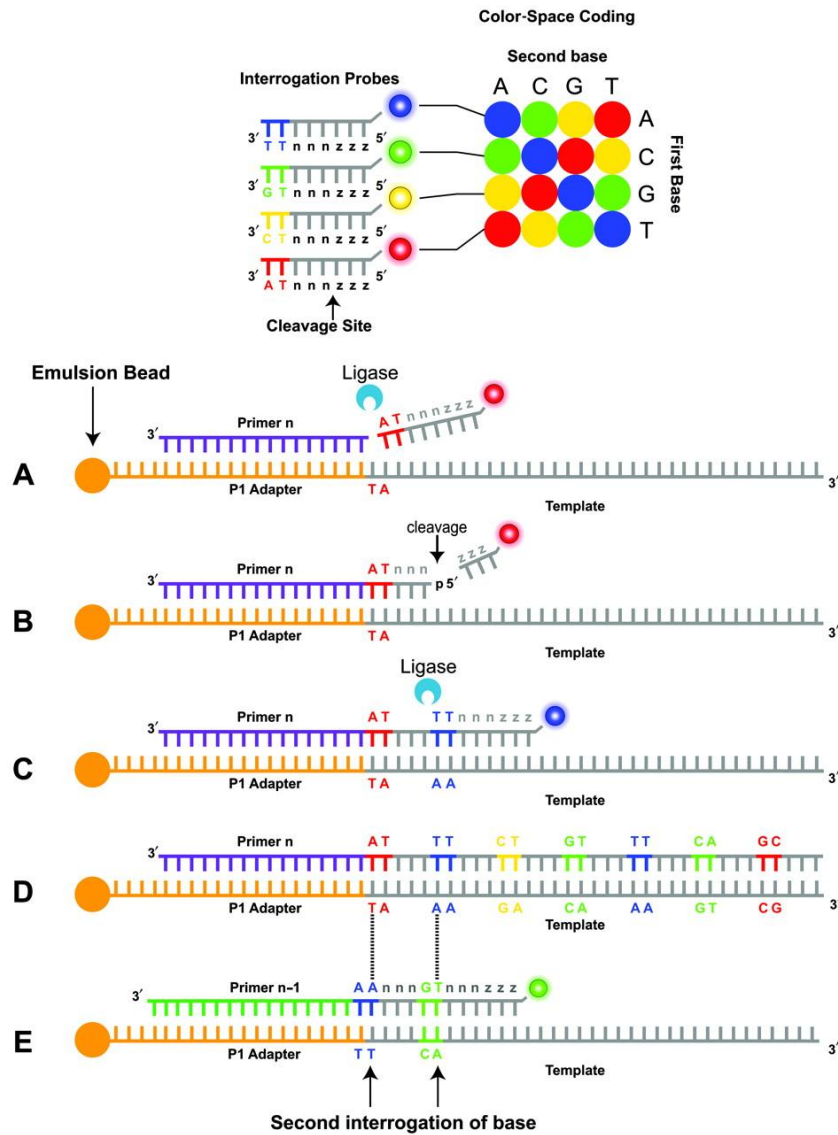


Figure 3: SOLiD sequencing [8].

Figure showing the colour-space coding from the first two nucleotides from the octamers. From **A-D**: starting from primer n, the first ligation cycle is finished; **E**: primer n-1 starts the second sequencing cycle out of 5.

SOLiD is the abbreviation for Sequencing by Oligo Ligation Detection [3]. The sequencing reaction by this system is complicated but the generated data is very accurate as the sequence of each nucleotide is deduced from two hybridizations. The principle of this technique is shown in Figure 3. At first, a primer (n) is annealed to the complementary template. In the next step, octamers are added with four fluorescent labels. Each fluorescent label is a colour-code for the combination of the first 2 nucleotides in its octamer. As shown in the colour-spacing code in Figure 3, the red label incorporated after the primer displays the possible combinations: AT, CG, GC or TA and codes for 4 out of 16 possible di-nucleotide

combinations. Nucleotides 3 to 8 of the octamer are not specific. After ligation, the octamer is cleaved between nucleotide 5 and 6 and the tail is removed. Subsequently another octamer is ligated to determine nucleotide 6 and 7 of the template and so on. At the end of this ligation cycle, no information is available from the in between sequences and the exact sequence of nucleotides 1 and 2 can still be one out of 4 possibilities as mentioned above (AT, CG, GC or TA). In order to attain this information, the ligation cycle is shifted by primer (n-1), one nucleotide away from the site where the first primer (n) annealed as shown in Figure 3 E. Adenosine hybridizes with the 3' end from the adaptor and the blue label shows that the next nucleotide is also adenosine. Therefore, the combination with the red label in Figure 3 A was AT. Each fragment is sequenced in five cycles, starting from primer n until primer n-4. SOLiD 5500xl has read lengths of only 85 bp which is its major shortcoming [3].

Ion Torrent PGM (Personal Genome Machine) was launched at the end of 2010 and is based on sequencing by synthesis and detection with semi-conductor technology [3]. An advantage of this technology is that no modified nucleotides (e.g. fluorescently labelled) nor expensive imaging technologies are needed [9]. During sequencing, the four nucleotides are sequentially introduced and if they are incorporated by DNA polymerase, hydrogen ions are released. Proportional to the number of nucleotides added, these hydrogen ions produce a shift in pH. Instead of scanning the whole chip, the semi-conductor detects protons directly and this speeds up the sequencing to a run time of 2 hours [9]. The read length is 200 bp [7]. A limitation of the system is the unreliable detection of the number of bases in homopolymers longer than 8 nucleotides [7]. IonTorrent PGM gives a low coverage in AT-rich regions and this bias is probably induced during amplification [7]. In the beginning of 2012, the Ion Proton sequencer (Life Technologies) was released. This system is based on the same principles but with a higher throughput (up to 10 Gb) than the Ion Torrent PGM (up to 1 Gb).

The 2nd generation sequencing platforms are based on clonal amplification of fragments since more fragments facilitate the detection of the sequencing signal. Despite this advantage, PCR amplification may introduce base sequencing errors or favour certain sequences over others, changing the relative abundance of DNA fragments [10]. This problem could be overcome by the introduction of very sensitive detectors enabling the sequencing of single DNA molecules. Sequencing of single DNA molecules is called 3rd generation sequencing.

In April 2011, a commercial 3rd generation sequencing system was released by Pacific Biosciences (Menlo Park, California, USA). PacBio RS enables single molecule real time sequencing (SMRT). This technology is based on two key inventions: zero-mode waveguides and nucleotides with a different phospho-linked fluorescent dye. The sequencing is executed in zero-mode waveguides, small nanophotonic visualisation chambers with a detection volume of 20×10^{-21} litres as shown in Figure 4 A [11]. In each chamber, a DNA polymerase with DNA strand are attached to the bottom. Nucleotides are diffusing in and out the detection chamber and during incorporation by DNA polymerase, the phosphate bond is cleaved and the fluorophore is held in place for tens of milliseconds (see Figure 4 B). This is much longer than the diffusion time of the non-incorporated nucleotides [11]. After fluorophore release, it rapidly diffuses out of the chamber and another nucleotide is incorporated. In this way, wash steps between the incorporation of each nucleotide are skipped which reduces the reagent costs and speeds up the process. Run times are approximately 2 hours [7]. PacBio has significantly higher (13%) error rates than Illumina (0.4%) and IonTorrent (1.78%) but errors were distributed evenly over the chromosome [7]. Another disadvantage for some applications is that PacBio needs input amounts of approximately 1 μg whereas 100 ng is enough for the other systems [7]. PacBio can be beneficial in de novo sequencing as the read lengths are substantially longer (1500 bp) and detection of alternative splicing can be facilitated [7]. The PacBio platform gives almost even coverage on GC and AT-rich regions [7].

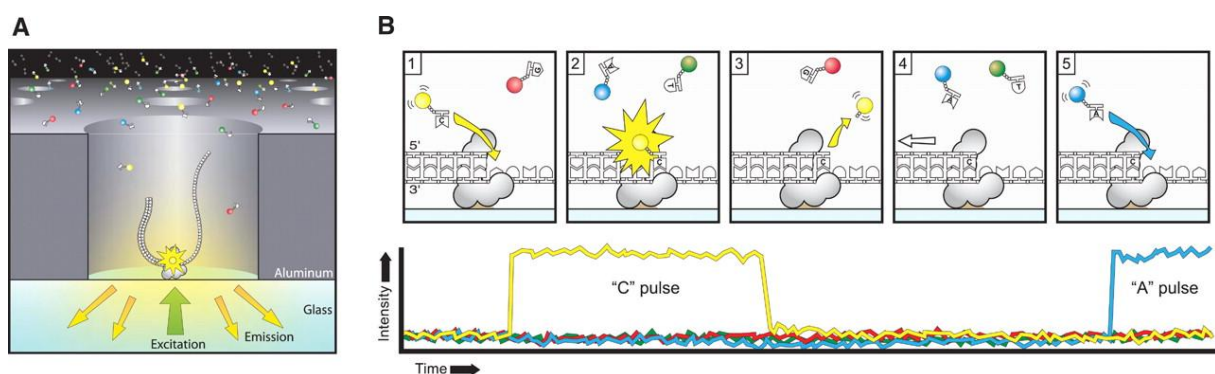


Figure 4: PacBio sequencing system [11].

Schematic representation of the PacBio sequencing system, **A**: nanophotonic visualisation chamber with DNA and DNA polymerase bound to the bottom, **B**: Incorporation of fluorescently labelled nucleotides gives a pulse of light.

Oxford Nanopore Technologies intended to release new 3rd generation sequencing platforms at the end of 2012: GridION and MinION. The MinION is a disposable sequencing device with the size of an USB memory stick and an expected price of less than \$900. GridION is a

high-throughput device with an output in the order of 10 Gb in 24 hours. More up-to date information can be found on: <http://www.nanoporetech.com/news/>.

2 Illumina sequencing

The Illumina system was used during this PhD project and will therefore be explained in more detail. The workflow comprises 3 steps: library preparation (Figure 5), cluster generation (Figure 6) and sequencing (Figure 7). To prepare the library, the ends of fragmented DNA are made blunt as shown in Figure 5. The 5' end is phosphorylated and the 3' end adenylated to enable the ligation of adaptors with a T-overhang. To both ends of the DNA fragment, a forked adaptor is ligated.

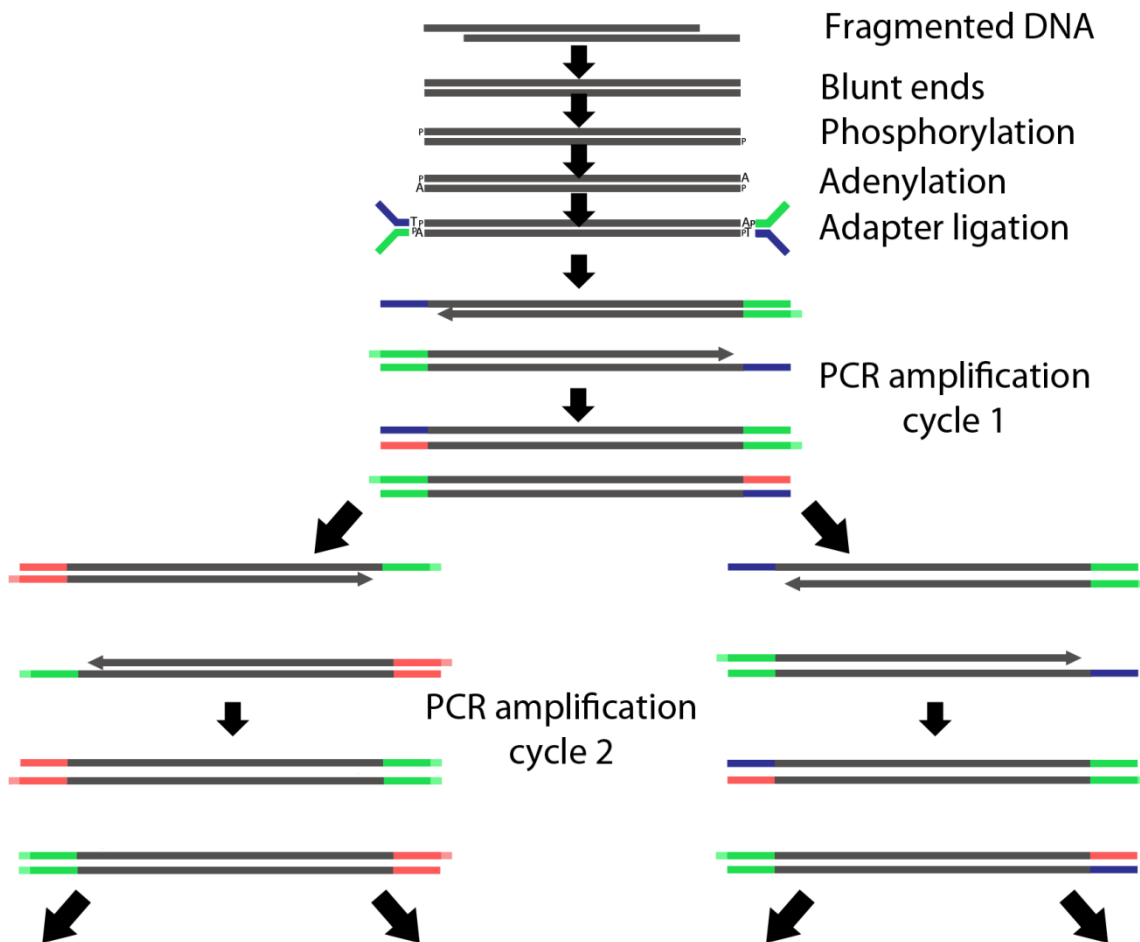


Figure 5: Illumina library preparation.

The ends of fragmented DNA are made blunt, 5' ends are phosphorylated and 3' ends adenylated. Forked adapters with a thymine-overhang are ligated and fragments with adapters at both ends are enriched by PCR amplification.

As illustrated in Figure 5, this adaptor structure ensures exponential amplification of fragments with an adaptor ligated to both ends since fragments without adaptors ligated or with one adaptor are not exponentially amplified. Adaptors and PCR primers incorporate sequences with 2 specific functions: enable hybridization to oligonucleotides on the flow cell and binding of the sequencing primer. It is possible to pool several samples in one flow cell. To do this, a specific sequence is added as barcode. In the Illumina library preparation protocols, this barcode is added just before or during PCR amplification. Adding an extra sequence during PCR amplification might create amplification bias. To investigate the influence of this, a new post amplification ligation-mediated barcoding protocol was developed and tested [12] and this will be discussed in Chapter III.

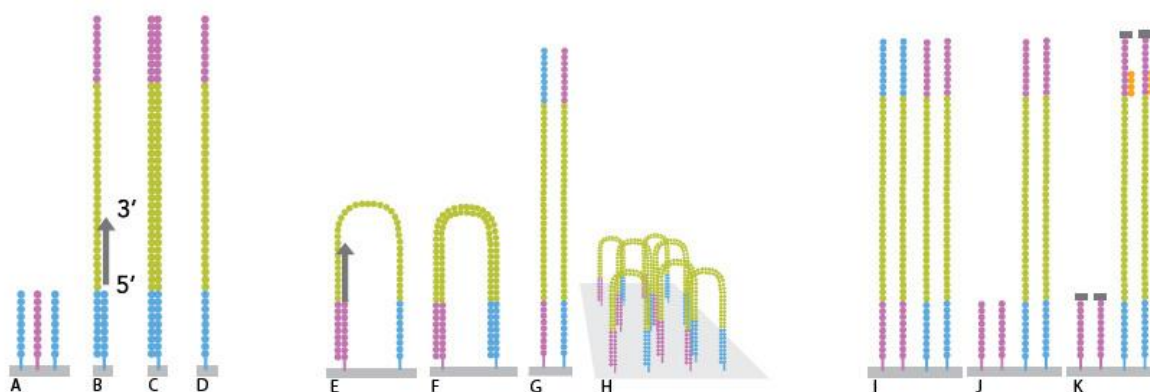


Figure 6: Cluster generation for Illumina sequencing.

Hybridization of template fragments to adaptors attached to the flow cell, bridge amplification and removal of reverse strands. Finally, 3' ends are blocked and sequencing primers are annealed.

(http://www.illumina.com/documents/products/datasheets/datasheet_cbot.pdf)

The cluster generation for Illumina is fully automated on the cBOT machine. As shown in Figure 6 A and B, library fragments are denatured and hybridized to oligonucleotides attached to the flow cell. These oligonucleotides act as a primer to start the synthesis of a new molecule that is covalently bound to the flow cell. The original template is washed away after denaturation with formamide (Figure 6 D) and the new molecule flips over and hybridizes with adjacent primers to form a bridge (Figure 6 E). Subsequently a second strand is synthesized and this forms a double stranded bridge (Figure 6 F). After denaturation, two copies of covalently bound single stranded templates are present (Figure 6 G). These are further amplified by bridge formation (Figure 6 H). At the end of the amplification, reverse strands are cleaved and removed (Figure 6 I and J). This results in a big cluster with only

forward strands. After blocking the 3' end, the sequencing primer is annealed and the flow cell is ready for sequencing (Figure 6 K).

The fragments are sequenced with reversible-terminated dNTP's labelled with different fluorescent dyes (see Figure 7). In each sequencing cycle all four dNTPs are presented to the flow cell but only one dNTP is incorporated per template. Fluorescent signals of each cluster are imaged and the terminator sequence with fluorescent dye is removed.

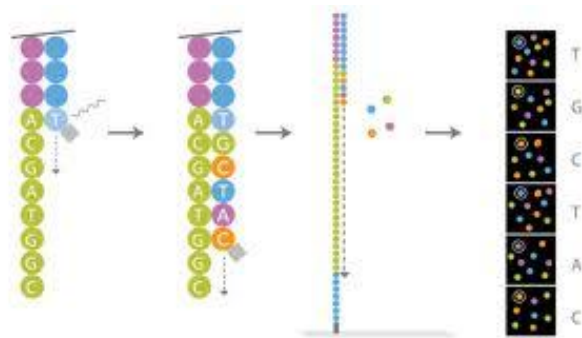
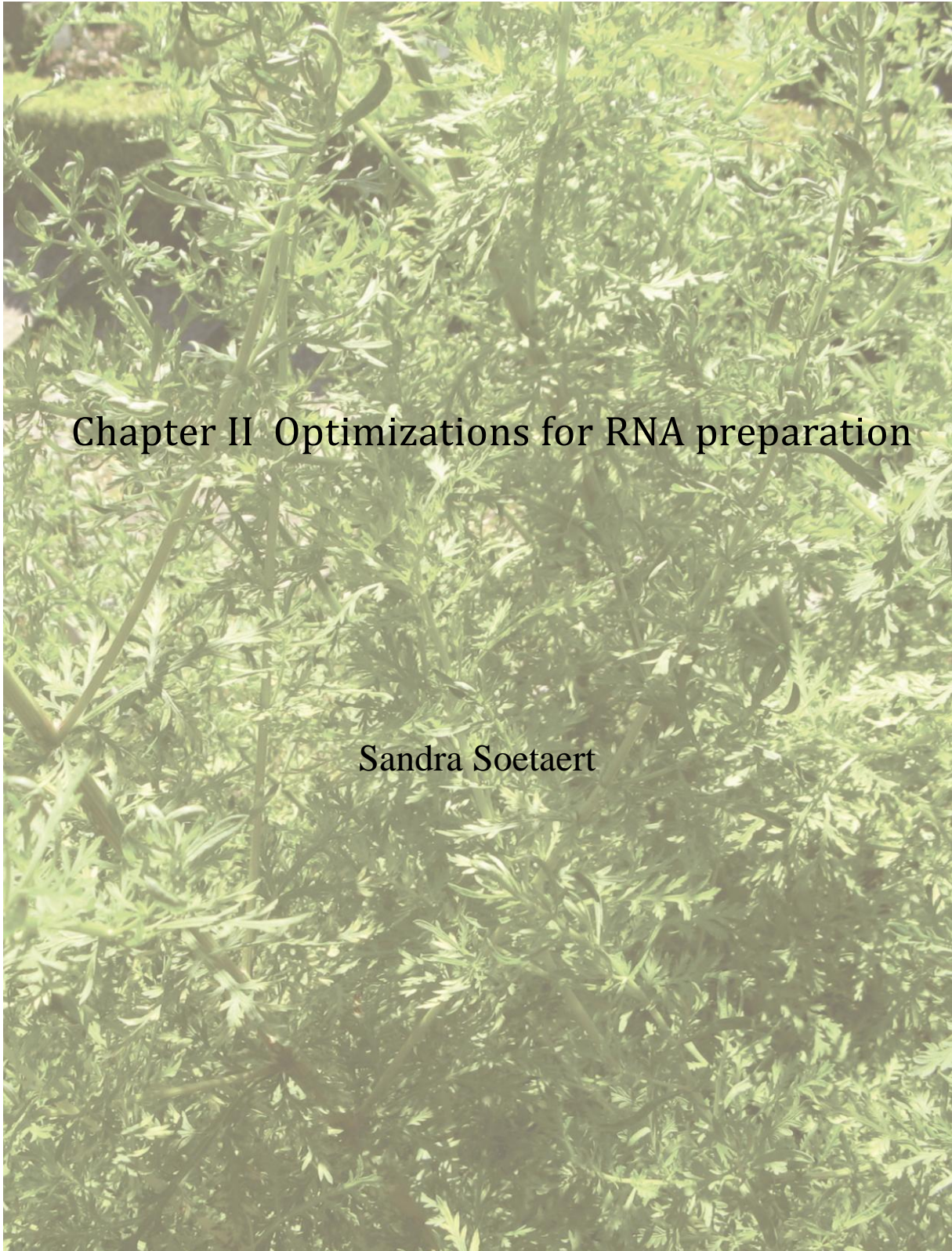


Figure 7: Illumina sequencing with fluorescently labelled nucleotides.

In each cycle, one dNTP with fluorescent label is incorporated in each cluster. Clusters are imaged and images from subsequent cycles are combined to determine the sequence of each fragment.

(http://www.illumina.com/documents/products/brochures/brochure_genome_analyzer.pdf)



Chapter II Optimizations for RNA preparation

Sandra Soetaert

To find candidate genes involved in the biosynthesis of artemisinin, glandular trichomes which produce artemisinin were compared at the transcriptome level with filamentous trichomes that do not synthesize artemisinin. As there were indications in literature that artemisinin is only synthesized in the apical cells on top of the glandular trichomes and not in the sub-apical cells [13], these cells were also compared. To sequence the transcriptome, RNASeq was performed on Illumina HiSeq2000.

To harvest the target sample-types as pure as possible and diminish the risk of contamination from other tissues, a microscope-guided collection technique was used. This technique is explained in Paragraph 1.1. Sample collection with laser capture microdissection is labour intensive and RNASeq is still an expensive technique which requires high quality input material. Because of this, attention was paid to optimize the collection and RNA extraction. As it was impossible to collect enough RNA with laser microdissection and laser pressure catapulting, an amplification step was introduced prior to RNASeq. An overview of these optimizations is given in this Chapter. Optimizations of the library preparation for RNASeq will be discussed in Chapter III.

1 Introduction and methods

1.1 Laser capture microdissection

For harvesting specific cells or tissues, a cutting laser is coupled with a microscope and a computer interface. The technique is called laser capture microdissection. Laser capture microdissection is a two-step process. In the first step named laser microdissection, the target is separated from surrounding tissues. For this, the laser is focused on the tissue, a trajectory is drawn with the computer (green in Figure 8 A) and the laser cuts while following this line (Figure 8 B). The second step is to collect the target sample as illustrated in Figure 8 (C-D) and for this, several technologies exist.

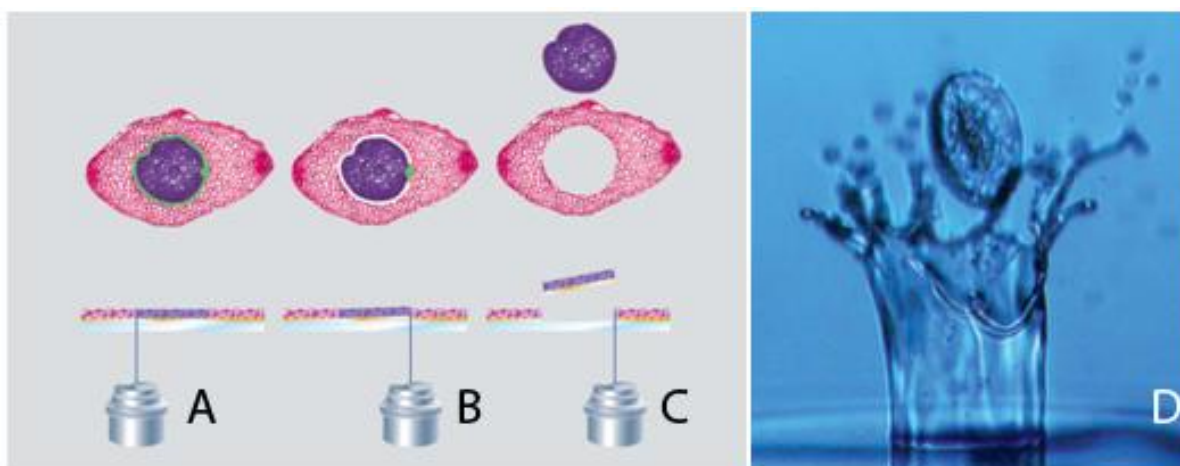


Figure 8: Laser microdissection and laser pressure catapulting with P.A.L.M. MicroLaser technology.

Pictures illustrating the operating procedure to collect tissue with laser microdissection and laser pressure catapulting, **A**: the area of interest is lined out, **B**: this trajectory is cut by laser microdissection, **C**: the target tissue is catapulted upwards by laser pressure catapulting, **D**: laser micro-dissected membrane catapulted out of liquid.

(A-C: [https://www.zeiss.com/C1256D18002CC306/0/D502284E9F21E27DC12574410058EBD0/\\$file/49-0010_e.pdf](https://www.zeiss.com/C1256D18002CC306/0/D502284E9F21E27DC12574410058EBD0/$file/49-0010_e.pdf), D: Image of Prof. A. Vogel, Lübeck, Germany

[https://www.zeiss.com/C1256D18002CC306/0/D502284E9F21E27DC12574410058EBD0/\\$file/49-0010_e.pdf](https://www.zeiss.com/C1256D18002CC306/0/D502284E9F21E27DC12574410058EBD0/$file/49-0010_e.pdf))

Two technologies are widely used: Arcturus (Life Technologies) and P.A.L.M. MicroLaser (Zeiss and P.A.L.M. Microlaser Technologies, München, Germany) as shown in Figure 9. The system of Arcturus places a cap with a thermoplastic film on the tissue sample and by IR-activation, a specific region of the film is selected to form a bridge between the cap and tissue.

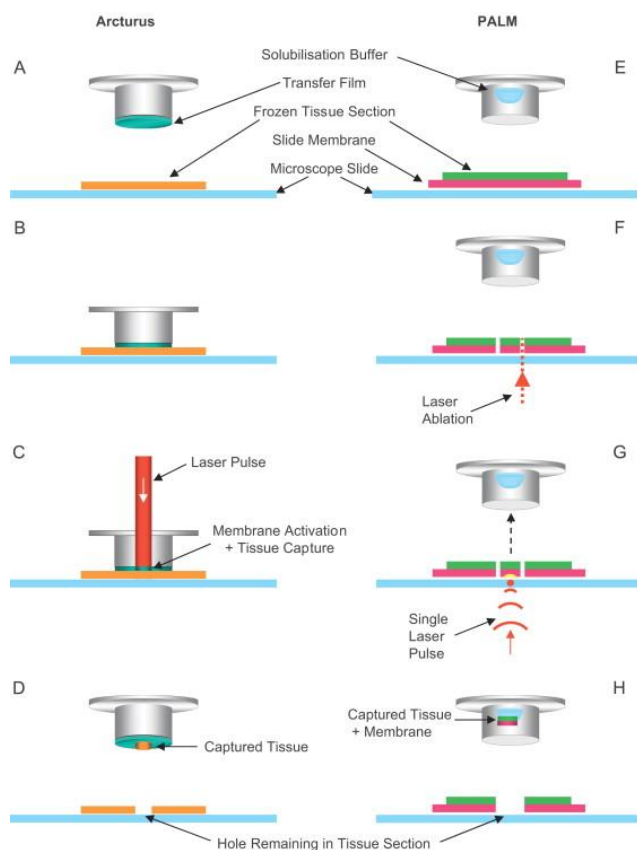


Figure 9: Mechanism of tissue capture with Arcturus and P.A.L.M. [1].

Arcturus captures samples on a transfer film whereas P.A.L.M. catapults samples in fluid.

showed that this is not harming or changing nucleic acids and other cellular macromolecules [15]. Therefore, laser capture microdissection can be applied to prepare samples for genomics, transcriptomics, proteomics and metabolomics studies [16].

Olsson *et al.* developed a protocol to use laser capture microdissection on glandular trichomes from *A. annua* [13] as shown in Figure 10. Instead of leaves, flower heads (Figure 11 C and D) were used since glandular trichomes protrude on the surface of floret buds (see Figure 11 B) and are sunken in the leaves (see Figure 11 A). Plant material was fixated under vacuum in formaldehyde and flower heads were chopped with a razor blade on a glass slide with formaldehyde. Thereafter, the P.A.L.M. MicroLaser technology fitted with a Robo-Mover to position the collection tube was used in combination with a microscope from Zeiss. In Figure 10, glandular trichomes were catapulted in the cap of a tube. For RNA extraction, cells were collected in 30 μ l of lysis buffer with β -mercaptoethanol from the Absolutely Nanoprep kit (Stratagene, La Jolla, CA, USA).

The film adheres to target cells and these cells are removed by lifting the cap (Figure 9 C-D). The P.A.L.M system isolates target cells by catapulting them with one laser pulse into the cap of a tube filled with fluid (Figure 9 G-H). The cap is positioned 1-3 mm above the slide. The catapulting strategy avoids sample contact and this is in contrast with the Arcturus system which uses adhesive films to tear away the target cells. Because of this, the Arcturus strategy is associated with a greater risk of contamination with non-target material. The narrower focusing width of the UV-A laser used in P.A.L.M. (0.5 μ m) enables more precise cutting than IR-lasers (7.5 μ m) [14]. There were some concerns over the impact of UV-A light on nucleic acids but recent studies

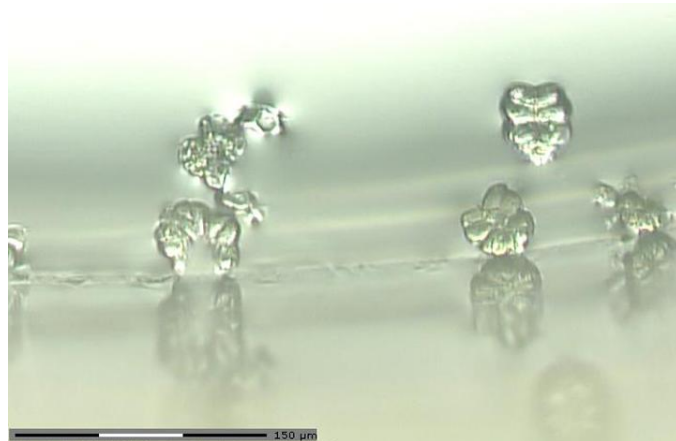


Figure 10: Glandular trichomes collected with P.A.L.M.

Picture of glandular trichomes from *A. annua* captured by laser pressure catapulting in the cap of a 500 μl tube.

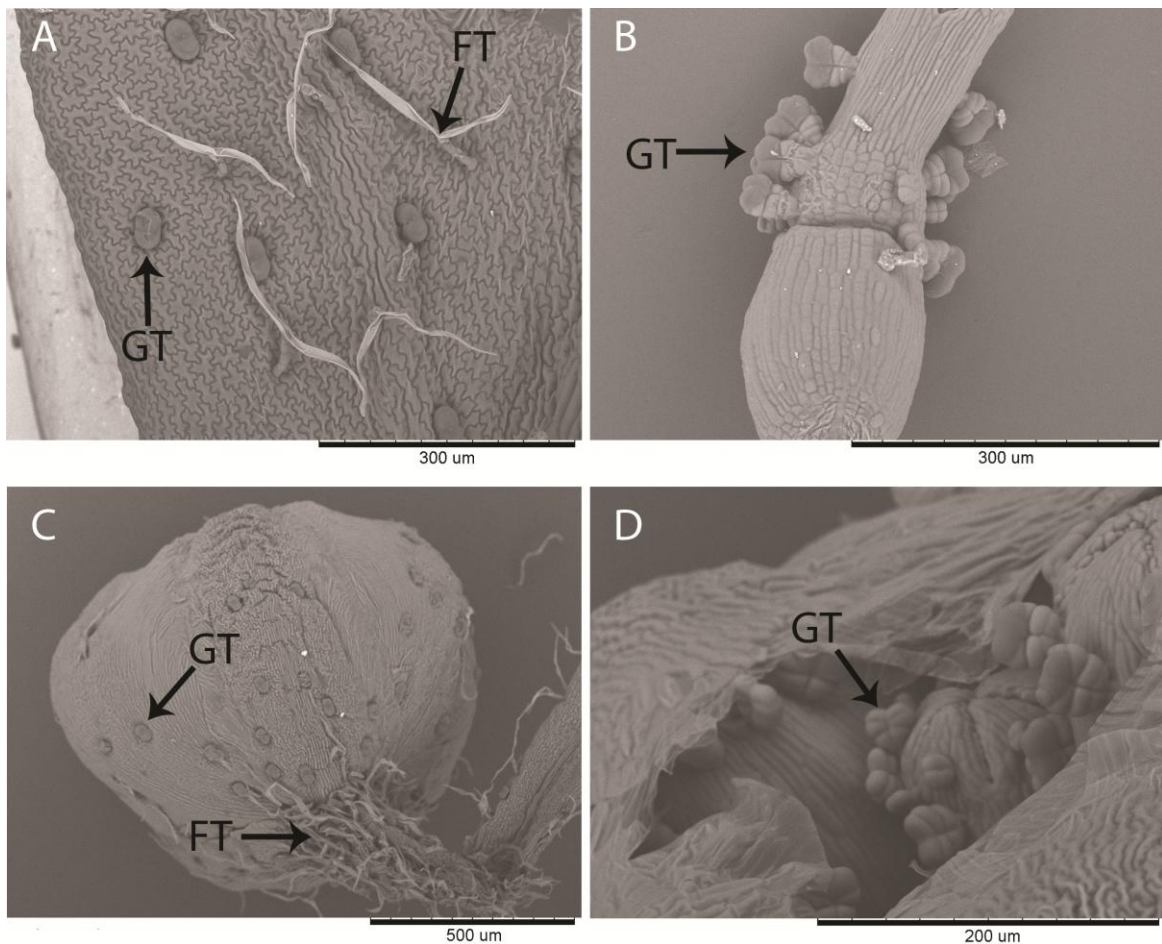


Figure 11: Glandular and filamentous trichomes from *Artemisia annua*.

Electron microscopy pictures of *A. annua* with glandular (GT) and filamentous (FT) trichomes. A: adaxial leaf surface with glandular trichomes sunken in the epidermis and filamentous trichomes; B: floret bud with protruding glandular trichomes; C: flower head (capitulum) with basal a lot of filamentous trichomes; D: floret buds in the involucrem of a flower head.

1.2 RNA extraction

After the collection of laser captured samples in lysis buffer with β -mercaptoethanol, RNA was extracted. The Absolutely RNA Nanoprep kit was chosen as this kit is optimized to extract RNA from extremely small numbers of cells ($1-10^4$ cells). Another advantage of this kit was that the big reaction volume in which the samples were collected, was easily reduced by pooling all the collection tubes on one filter. A DNase digestion was performed to remove DNA contamination and RNA was recovered in only 10 μ l.

To perform laser capture microdissection and RNA extraction, some factors were tested to estimate their influence on the final RNA quality such as the influence of fixation, the period of time that collected cells were kept in lysis buffer during sampling, the compatibility of lysis buffer with plant material from *A. annua* and the temperature at which the RNA was extracted. The amount of laser captured cells was too low to determine the RNA quality. Due to this, RNA preparation was optimized with more plant material. If possible for optimizing the procedure, the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) was used since this kit was better suited for bigger amounts of input material (in the range of 50 mg). These results were extrapolated to the laser captured samples.

Nanoprep and RNeasy RNA extraction kits are based on the same principles. To denature RNase enzymes, β -mercaptoethanol is added. The lysis buffer contains chaotropic salts which destabilize hydrogen bonds, Van der Waals forces and hydrophobic interactions. Also other proteins are denatured and the water shell from the negatively charged backbone of the nucleic acids is removed. This enables the binding of (ribo)-nucleic acids to a positively charged silica-filter [17, 18]. This binding can be improved with addition of ethanol [19] or sulfolane [20]. After washing, ribonucleic acids are eluted from the silica column with water or an elution buffer.

1.3 RNA quality control

The quality or integrity of RNA samples can be determined on a microfluidics-based electrophoresis system. Using such a sensitive system, it is easier to interpret the quality of the RNA than with normal gel electrophoresis. The two most frequently used systems are the Agilent bioanalyzer and the Experion system from Bio-Rad. In both systems, electrophoresis is performed on a chip with microchannels (Figure 12 A).

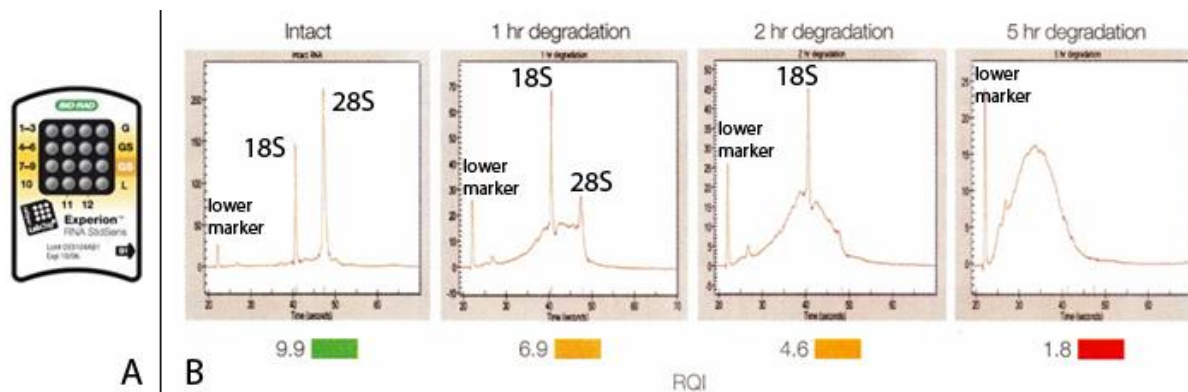


Figure 12: Experion electrophoresis.

RNA quality determination with Experion, **A:** Experion RNA StdSens electrophoresis chip.

B: electropherograms from RNA samples with their corresponding RQI score. X-axis: migration time in seconds, Y-axis: fluorescent intensity.

(adapted from http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5452B.pdf)

The RNA quality is assessed by an algorithm that calculates a quality score: RNA integrity number (RIN) (Agilent) or RNA quality indicator (RQI) score (Bio-Rad). As illustrated in Figure 12 B, scores are between 10 (highly intact RNA) and 1 (highly degraded RNA). To calculate these scores, several electrophoretic regions are taken into account as well as 18S and 28S peaks. The 18S and 28S rRNA are cleavage products from a common transcription unit. That means that exactly the same numbers of 18S and 28S molecules are present in each cell. In intact RNA, the peak intensity of 28S/18S has a ratio of two since the 28S fragment is approximately twice as long as 18S. Twice as many bases will bind twice as many dye molecules and this will lead to an increased signal detected during electrophoresis. For samples with a good quality, the RQI score is based on the ratio of 28S/18S whereas for samples with a quality that is less good, preferably the 18S and pre-18S regions are used. Instead of 28S, this ribosomal subunit is called 25S in plants. Each sample is loaded with a lower alignment marker as indicated in Figure 12 B. According to the Experion manual, at least 100 pg/ μ l of Total RNA is needed for quality determination with the Experion HighSens kit but preliminary tests in this project were not able to detect RNA from 150 laser captured glandular trichomes with Experion. Therefore, tests for optimizing the RNA quality were up-scaled with more material.

1.4 RNA amplification

For sequencing with Illumina, at least 0.1 μ g of total RNA is needed as input (Illumina TruSeq RNA Sample Preparation Guide). Since it was technically impossible to collect that

amount of RNA from laser captured trichome samples and apical as well as sub-apical cells, RNA was amplified. For amplification of RNA, several commercial kits are available and the choice was made to use a linear amplification kit since sequence-dependent bias and length-dependent bias may be amplified exponentially in the other kits [21]. Furthermore, highly variable amplification of low abundant transcripts might occur [21]. Another factor taken into consideration for selecting an amplification kit was the type of primers applied for amplification. Most kits use only oligo dT primers that bind on the 3' polyA tail of mRNA transcripts. The Ovation RNA-Seq system (NuGen, AC Bemmell, The Netherlands) uses oligo dT-based primers and random primers. With the use of random primers, it remains possible to sequence the 5' end if RNA shows signs of degradation. Because of the random primers and the linear amplification, the Ovation RNA-Seq system was chosen to amplify RNA from the laser captured trichome samples. A schematic overview of the Ovation RNASeq amplification protocol is given in Figure 13.

Input amounts of total RNA are preferably in the range of 500 pg to 100 ng and serve as template for first strand cDNA synthesis. For cDNA synthesis, chimeric primers are used. These primers have a DNA part which is an oligo dT or a random hexamer and an additional RNA fraction which does not hybridize across the transcripts but introduces a unique RNA sequence at the 5' end of the cDNA.

Within the cDNA/mRNA complex, fragmentation of the mRNA transcript occurs. This creates priming sites for DNA polymerase to synthesize a second strand DNA molecule. The newly synthesized strand contains an additional part complementary to the RNA sequence of the chimeric primers which creates a unique DNA/RNA heteroduplex at one end. This heteroduplex plays a key role during amplification.

RNaseH cleaves RNA in the DNA/RNA heteroduplex and makes the DNA part of the heteroduplex accessible for annealing a SPIA primer with RNA portion and replication is initiated by DNA polymerase. This forms the basis for single primer isothermal amplification (SPIA). The RNA sequence included in the newly synthesized strand is cleaved by RNaseH, another SPIA primer can bind and DNA polymerase can synthesize a new strand. With 500 pg total RNA starting material, an average amplification of 800-fold is observed. The amplification process is completed with a post-SPIA modification. In this step, the amplified

SPIA product is converted to double stranded cDNA appropriate for Illumina library preparation.

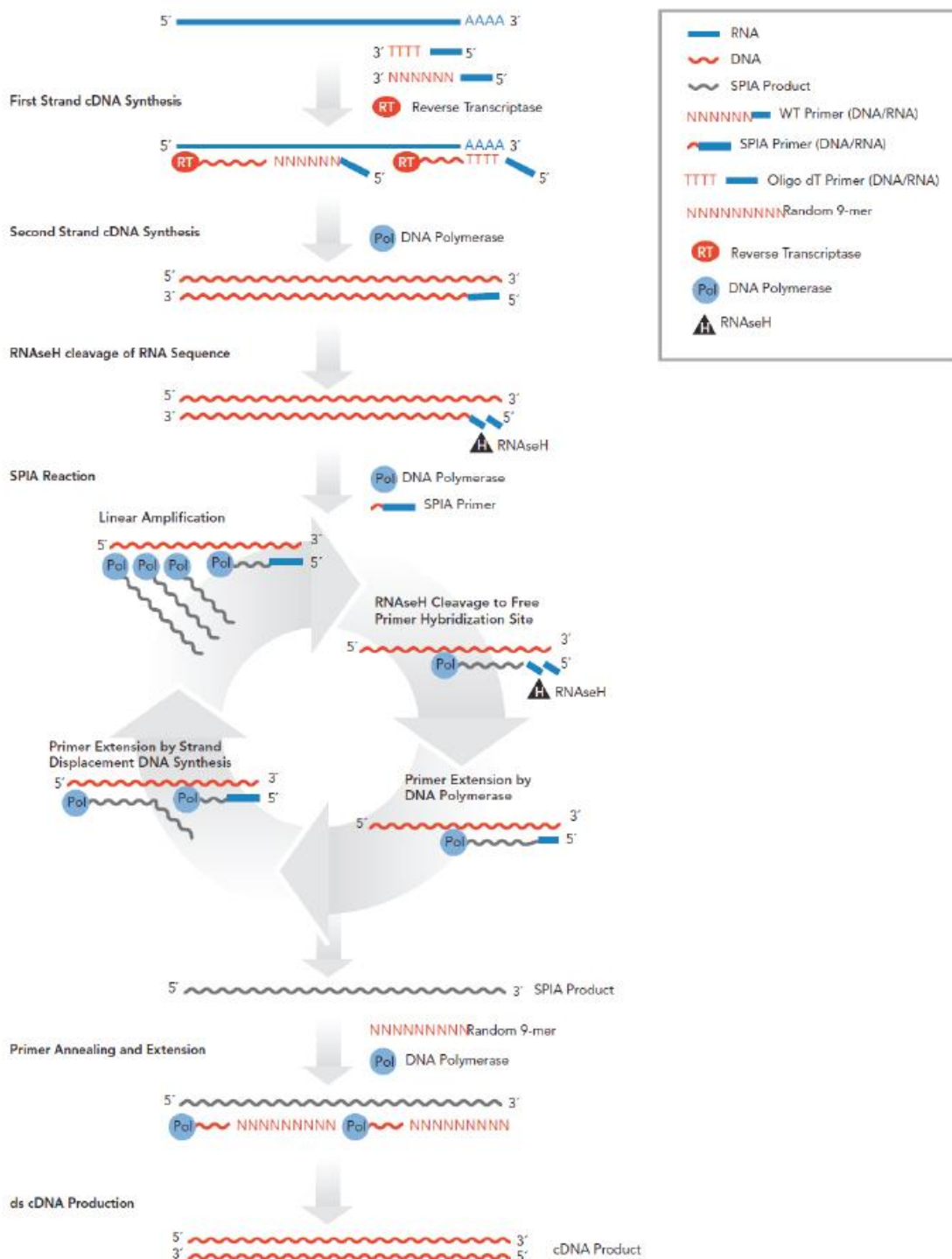


Figure 13: Ovation RNASeq amplification protocol (Nugen).

Starting from template RNA, cDNA is synthesized which contains an RNA/DNA heteroduplex. RNaseH cleaves the RNA portion and single primer isothermal amplification is performed. After amplification, SPIA product is converted to ds cDNA.

2 Optimization results and discussion

2.1 RNA quality

2.1.1 Optimization of sample preparation for laser capture microdissection

Olsson *et al.* developed a protocol with formaldehyde fixation of the plant material for laser capture microdissection (see page 61) [13]. To test the influence of formaldehyde fixation on the RNA quality, 70 mg of plant material was fixated on ice for 3 to 4 hours under vacuum with a 4% formaldehyde phosphate buffered saline solution. Another 70 mg of plant material was not fixated. RNA from both samples was extracted with the RNeasy Plant Mini Kit with RLT lysis buffer. The quality from these samples was analysed on an Agilent Pico chip. In the electropherogram of the non-fixated sample (10-fold diluted), ribosomal peaks 5.8S, 18S and 25S were clearly visible as well as additional peaks from the chloroplast rRNA. A RIN quality score of 8.7 was assigned whereas the fixated sample had a quality score of 2.5 and as shown in Figure 14, 18S and 25S were not detected in the undiluted fixated sample.

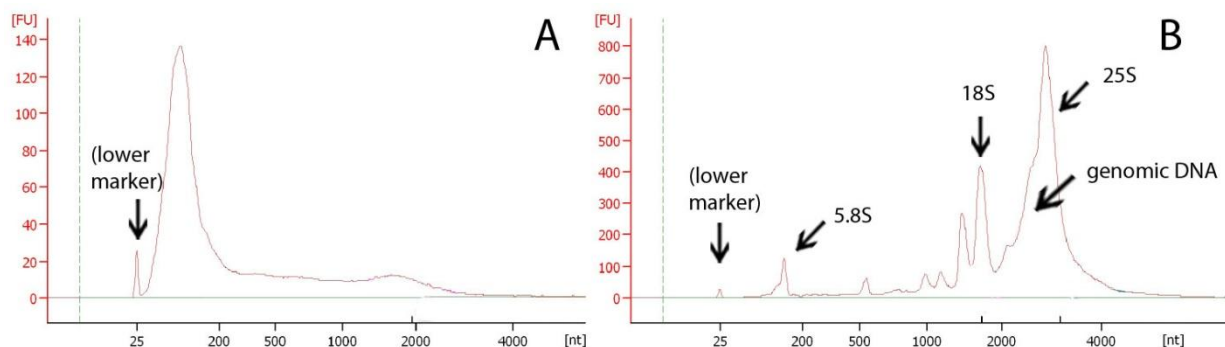


Figure 14: RNA extraction from fixated and non-fixated plant material.

Agilent traces showing the difference in quality of RNA extracted from **A:** fixated and **B:** non-fixated leaf material from *A. annua*. RNA from non-fixated material was 10-fold diluted.

Scale on X-axis is length in nucleotides and Y-axis is expressed in fluorescent units.

With formaldehyde fixation, RNA yield and quality was clearly deteriorated. The preparation procedure of tissues for laser microdissection balanced between morphological preservation and the recovery of molecules from the fixated tissue. Formaldehyde fixation cross-linked the cytoplasm and facilitated laser microdissection. This fixation made cells more rigid and during laser microdissection, cells were forced to separate as shown in Figure 15. Without fixation, cells were less robust and the risk of cutting through the cell walls increased.

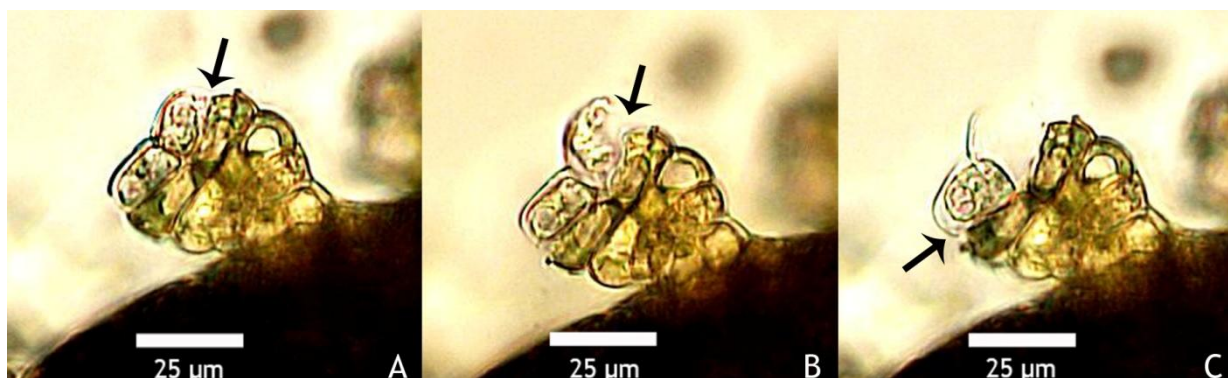


Figure 15: Laser capture microdissection of fixed tissue from *A. annua*.

Separation and isolation of apical cells from a glandular trichome from *A. annua* with P.A.L.M., **A:** laser microdissection diverges apical and sub-apical cells (arrow) of a formaldehyde fixed glandular trichome, **B:** apical and sub-apical cell are further separated (arrow) and **C:** apical cell is removed by laser pressure catapulting and the isolation of the 2nd apical cell is started (arrow).

The adverse effect of formaldehyde fixation on RNA quality was later also reported by Olofsson *et al.* [22]. They overcame this problem by using the following protocol: on a RNA free microscope slide, a closed flower head was cut in a drop of cold buffer with 25 mM MOPSO (pH6.3) (Sigma-Aldrich, Steinheim, Germany), 200 mM sorbitol (Alfa Aesar, Karlsruhe, Germany), 10 mM sucrose (Acros, Geel, Belgium), 5 mM thiourea (Sigma-Aldrich), 2 mM DTT (Fluka, Sigma-Aldrich), 5 mM MgCl₂ (Sigma-Aldrich) and 0.5 mM sodium-phosphate (Acros) [22]. For capturing apical and sub-apical cells, both cell types must be separated very precisely as shown in Figure 15. With the unfixed protocol from Olofsson *et al.* [22], laser microdissection to separate apical and sub-apical cells was difficult to perform and leakage occurred. Therefore, one experiment for RNASeq was performed with and another without fixation to collect apical and sub-apical cells.

To collect whole glandular and filamentous trichomes, fixation was less an issue. If trichomes were attached to a big piece of tissue, one laser shot was used to catapult a trichome. This was not the case if trichomes were attached to a small piece of tissue since the whole tissue would be catapulted upwards. In that case, trichomes were separated from the rest of the tissue by transecting the cells that connect the trichome to the floret or involucre. Subsequently, one laser pulse catapulted the trichome into the collection cap. Therefore, fixation was not needed to collect glandular and filamentous trichomes.

2.1.2 Compatibility of Nanoprep lysis buffer with *A. annua*

As described in the RNeasy Plant Mini Kit, some secondary metabolites in plant tissues can cause solidification of the sample, making extraction of RNA impossible. Therefore, the compatibility of the Nanoprep lysis buffer with plant material from *A. annua* was tested. Plant material (72 mg) of *A. annua* leaves was lysed in Nanoprep lysis buffer, homogenized on a QIAshredder column and further extracted with the RNeasy Plant Mini extraction kit. The extracted RNA had an RQI score of 9.9 and a 25S/18S ratio of 1.98 was automatically assigned. However it should be noted that this ratio was influenced by the genomic DNA present around the 25S peak but no visual degradation was observed (see Figure 16). Therefore, it was concluded that the lysis buffer of the Nanoprep kit is well suited for extraction of RNA from *A. annua*.

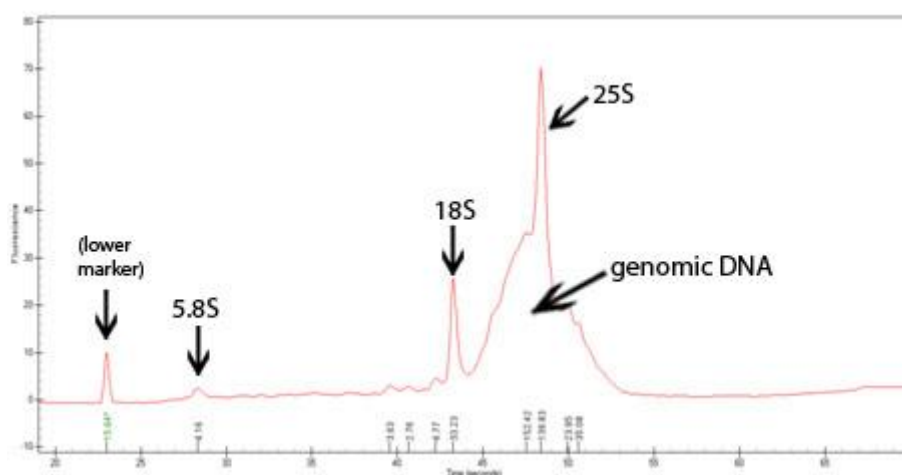


Figure 16: RNA extracted from *A. annua* with lysis buffer of the Nanoprep extraction kit.

Experiencing electropherogram from RNA extracted with Nanoprep lysis buffer. Ribosomal peaks are clearly visible. Y-axis: fluorescent intensity, scale of X-axis is migration time in seconds and other indications on X-axis are corrected peak areas.

2.1.3 Optimization of RNA extraction temperature

It was also tested at which temperature the RNA extraction with the Nanoprep kit is performed optimally since this information was not given in the manual. In a similar RNA extraction kit: RNeasy Plant Mini Kit, the manual recommends to work not colder than room temperature whereas working on ice is better to preserve RNA quality. Therefore, the Nanoprep kit was tested both at room temperature and on ice. For each test, two flower heads were crushed with a mortar in liquid nitrogen, homogenized on a QIAshredder column and the Nanoprep protocol was executed. For the test at room temperature, centrifuge was set at

20°C whereas for extraction on ice, centrifuge temperature was 4°C and samples were kept on ice during handling. Samples prepared on ice had a good RQI score of 8.6 whereas to RNA extracted at room temperature a moderate RQI score of 4.6 was assigned as shown in Table 1. Electrophoretic traces are shown in Figure 17.

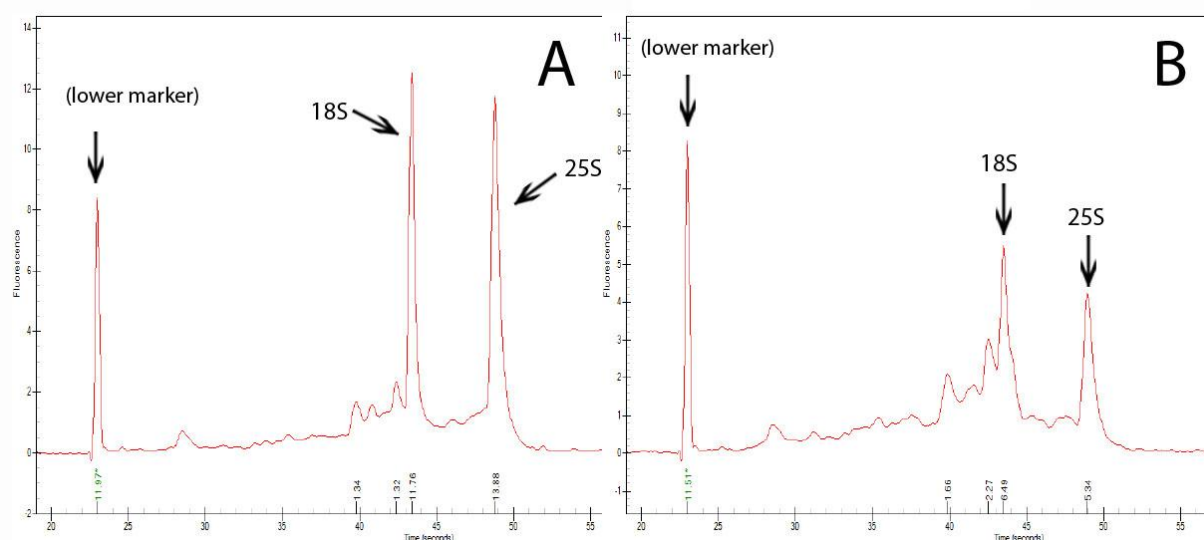


Figure 17: RNA extraction with the Nanoprep kit on ice and at room temperature.

Experion electropherogram from RNA extracted with Nanoprep lysis buffer **A**: on ice and **B**: at room temperature. RNA extracted at room temperature was more degraded. Y-axis: fluorescent intensity, scale of X-axis is migration time in seconds and other indications on X-axis are corrected peak areas.

Sample Name	RNA Area	RNA Concentration (ng/μl)	Ratio [28S:18S]	RQI	RQI Classification
Nanoprep on ice	75.00	27.06	1.18	8.6	■
Nanoprep at room temperature	62.95	22.72	1.11	4.6	■

Table 1: The influence on RNA quality of samples extracted with the Nanoprep kit on ice and at room temperature.

Experion result table in which RNA extracted on ice and room temperature were compared. The total RNA area and therefore also the RNA concentration in the electropherogram was slightly higher in the sample on ice. Ratios of rRNA were similar in both samples, Experion reports eukaryotic ribosomal ratios standard as 28S:18S but in plants the 28S is called 25S.

RQI score is higher with extraction on ice.

2.1.4 Collection time for laser capture microdissection

With laser capture microdissection, samples were catapulted directly in Nanoprep lysis buffer with β -mercaptoethanol at room temperature. It was more convenient to collect several trichomes in one collection tube. To simulate the influence of collecting samples in one cap for 2 hours, plant material from *A. annua* was directly extracted, incubated for 10 minutes or incubated for 2 hours in Nanoprep lysis buffer with β -mercaptoethanol at room temperature. Samples were subsequently extracted with the RNeasy Plant Mini Kit. In all tests, the RNA quality was good and no deterioration of RNA quality was observed after incubating the samples in lysis buffer with β -mercaptoethanol at room temperature for 2 hours (Figure 18).

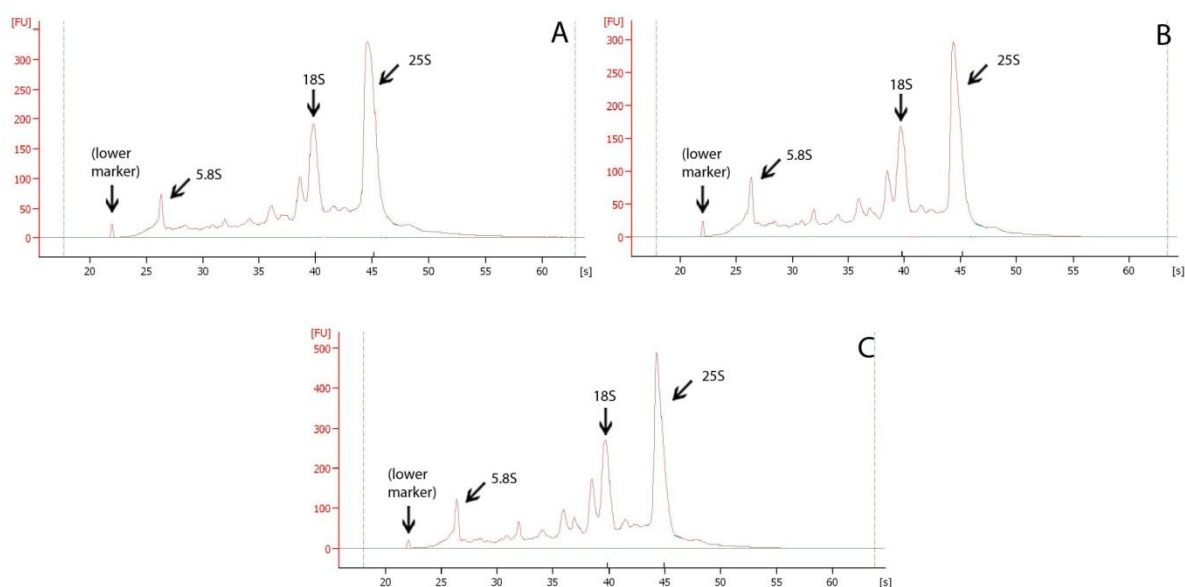


Figure 18: Incubation time in collection buffer and RNA quality.

Agilent electropherograms of **A**: direct RNA extraction with RNeasy Plant Mini Kit (RIN score: 8), **B**: 10 minutes incubation in lysis buffer and β -mercaptoethanol at room temperature before RNA extraction (RIN score: 7.6) and **C**: 2 hours incubation in lysis buffer and β -mercaptoethanol at room temperature before RNA extraction (RIN score: 7.8). Y-axis: fluorescent intensity, X-axis scale is migration time in seconds.

2.2 RNA quantification from laser captured trichomes

For the Ovation RNA-Seq amplification kit, it is best to have at least 500 pg RNA as input material. To estimate the yield from laser captured trichomes, glandular and filamentous trichomes were captured, extracted with the Nanoprep kit and the RNA amount was quantified with Quant-iT RiboGreen (Life Technologies).

To accelerate the collection of filamentous trichomes, not all stalk cells were included for each collected trichome. Extraction of 300 filamentous trichomes yielded approximately 600 pg RNA in 10 μ l. For the Ovation RNA-Seq amplification kit, only 5 μ l can be used which means that 1 pg input for amplification was obtained for each filamentous trichome. Therefore, preferably more than 500 filamentous trichomes were collected.

From glandular trichomes, 150 trichomes were captured with at least 6 intact secretory cells. This yielded approximately 2150 pg of RNA in 10 μ l elution buffer, which was for amplification 1075 pg or around 7 pg of RNA per glandular trichome. This means that at least 72 glandular trichomes had to be collected to obtain the recommended input amount of 500 pg.

Yields from glandular trichomes were extrapolated to estimate the number of apical and sub-apical cells needed. The amount of cells collected from 150 glandular trichomes was approximately 1040. This gave a yield of around 2 pg for each collected cell and 1 pg input for amplification. To obtain 500 pg, at least 500 apical or sub-apical cells needed to be collected but collecting this amount of cells was very laborious. As the technical support from Nugen recommended to go not below 300 pg of input material, it was decided to collect at least 300 cells and to have a two independent collection experiments.

3 Conclusions

From these optimization experiments, the conclusion was drawn that preferably non-fixated material is used for laser capture microdissection. During laser capture microdissection, samples can be collected for 2 hours in the same collection tube. The Nanoprep RNA extraction kit is suited to extract RNA from plant material of *A. annua* but it is better to perform the extraction procedure on ice. At least 72 glandular trichomes were collected for amplification whereas from filamentous trichomes around 500 were collected to have enough RNA. From apical as well as sub-apical cells, at least 300 cells were collected.



Chapter III Quantitative bias in Illumina TruSeq
and a Novel Post Amplification Barcoding Strategy
for Multiplexed DNA and Small RNA Deep
Sequencing [12]

Filip Van Nieuwerburgh¹, Sandra Soetaert¹, Katie
Podshivalova², Eileen Ay-Lin Wang², Lana Schaffer³,
Dieter Deforce¹, Daniel R. Salomon², Steven R. Head³, Phillip
Ordoukhanian³

¹ Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium,

² Department of Molecular and Experimental Medicine, The Scripps Research Institute,
La Jolla, California, United States of America,

³ Next Generation Sequencing Core, The Scripps Research Institute, La Jolla, California,
United States of America

With Ovation RNA-Seq amplification, RNA was converted to DNA for six sample-types collected with laser capture microdissection: glandular trichomes, glandular trichomes treated with the plant hormone jasmonic acid, filamentous trichomes, filamentous trichomes treated with jasmonic acid and apical as well as sub-apical cells from glandular trichomes. From each sample-type, 3 samples were sequenced. Sequencing these 18 samples in 3 lanes on an Illumina flow cell was enabled by tagging them with a unique barcode during the library preparation step.

In the TruSeq DNA library preparation protocol (Illumina), the barcode is introduced before the PCR step. During PCR amplification, several types and causes of bias can occur such as lower amplification of AT-rich regions [23], less efficient amplification of longer fragments [24] and formation of heteroduplexes in the annealing step in samples with a diverse population of nucleic acids [25]. Because of this, it is worth to investigate if the introduction of a unique barcode before PCR can cause sequence-dependent bias as this might influence the differential expression analysis.

A new barcoding protocol was developed which ligates barcodes after the PCR step, called Post Amplification Ligation Mediated (PALM) barcoding. Both PALM and TruSeq barcoding protocols were essayed on the same pool of reference DNA generated by converting *Saccharomyces cerevisiae* mRNA into double stranded DNA and results were compared. A similar PALM protocol was developed for miRNA.

After optimization on reference DNA, the PALM barcoding strategy was used in Part III for library preparation of samples from *A. annua*.

1 Introduction

Taking advantage of the increasing throughput achieved by second generation sequencing technologies, multiplexing several samples in one analysis can increase experimental throughput while reducing time and cost.

Several strategies have been described for barcoding sequencing libraries [26-30]. Vigneault *et al.* [29] published a miRNA barcoding protocol using ligation of 3' pre-adenylated barcoded adapter oligonucleotides as the first step of sequencing library preparation. Buermans *et al.* [30] published a miRNA sequencing protocol, introducing a barcode during PCR. Illumina recently released the TruSeq kits for multiplexed high-throughput sequencing. The Illumina TruSeq small RNA protocol introduces the barcode during the PCR step using differentially barcoded primers, while the TruSeq DNA (or messenger RNA converted to double stranded DNA) protocol introduces the barcode before the PCR step by ligation of differentially barcoded double stranded adaptors. All published methods place the barcodes within the adapters, downstream or within the PCR primer binding site or introduce the barcode during PCR. However, it is well established that multi-template PCR amplification can result in a sequence-dependent amplification bias, as some DNA species are amplified more efficiently than others [23, 24, 31, 32]. For this reason, introducing barcodes near a priming site might result in a barcode-specific quantitative bias. To our knowledge, no previous publication has provided in-depth data measuring PCR amplification bias resulting from the use of barcodes.

Our initial attempts to adapt previous barcoding strategies to multiplexed sequencing of small RNA used index sequences placed at the distal end of the 5' adapter in the Illumina small RNA library protocol. Despite a number of iterations of the design we consistently failed to avoid PCR amplification bias when identical samples with different barcodes were compared. Therefore, we designed a new strategy in which we ligate both the 3' and 5' adapters, perform the RT-PCR step and then ligate the barcode after the library PCR amplification, as the last step of the library preparation. We have called this strategy Post Amplification Ligation Mediated (PALM) barcoding. In the present study, we compared the de-multiplexed quantitative results of 12 differentially PALM barcoded miRNA samples, 12 TruSeq barcoded miRNA samples and 4 miRNA samples barcoded using our above-mentioned pre-

PCR barcoding strategy from the Human Brain Reference RNA (Ambion). Each pool was sequenced in a single lane on an Illumina GAIIx.

Parallel to PALM barcoding for small RNA, we also developed a PALM barcoding protocol for DNA samples or messenger RNA (mRNA) converted to double stranded DNA (dsDNA). The main difference compared to the PALM barcoding for small RNA, is the fact that double stranded adapters instead of single stranded adaptors need to be ligated before PCR. In the present study, we compared the de-multiplexed quantitative results of 12 differentially PALM barcoded DNA samples and 12 TruSeq barcoded DNA samples. Reference DNA was generated by converting *Saccharomyces cerevisiae* mRNA into double stranded DNA. Each pool was sequenced in a single lane on an Illumina HiSeq 2000.

2 Methods

2.1 PALM small RNA barcoding

The PALM miRNA barcoding protocol is similar to the Illumina Small RNA v1.5 Sample Preparation Guide. This protocol was modified to achieve a higher yield after the PCR amplification step using higher reaction volumes for the RT-PCR step. No extra cycles were added to the PCR reaction. The adapters used in the protocol were modified to allow for PALM barcoding and Illumina index sequencing with the Illumina multiplexing index read sequencing primer. The complete protocol, including the adapter sequences, is available in section 5.3. Figure S1 shows a typical Invitrogen 4% E-gel of a Human Brain Reference RNA (Ambion) library after PCR amplification and before barcode ligation. Figure 19 shows the necessary oligonucleotide components for PALM and how they are consecutively added to the miRNA sample. The key difference with respect to the current Illumina small RNA library protocol is the addition of the barcode to the library by ligation after PCR amplification. After ligation of the barcode, no further purification of the library is required. The library is quantified using analysis of area under the peaks with a BioAnalyzer 2100 (Agilent) to determine the correct loading concentration for subsequent sequencing.

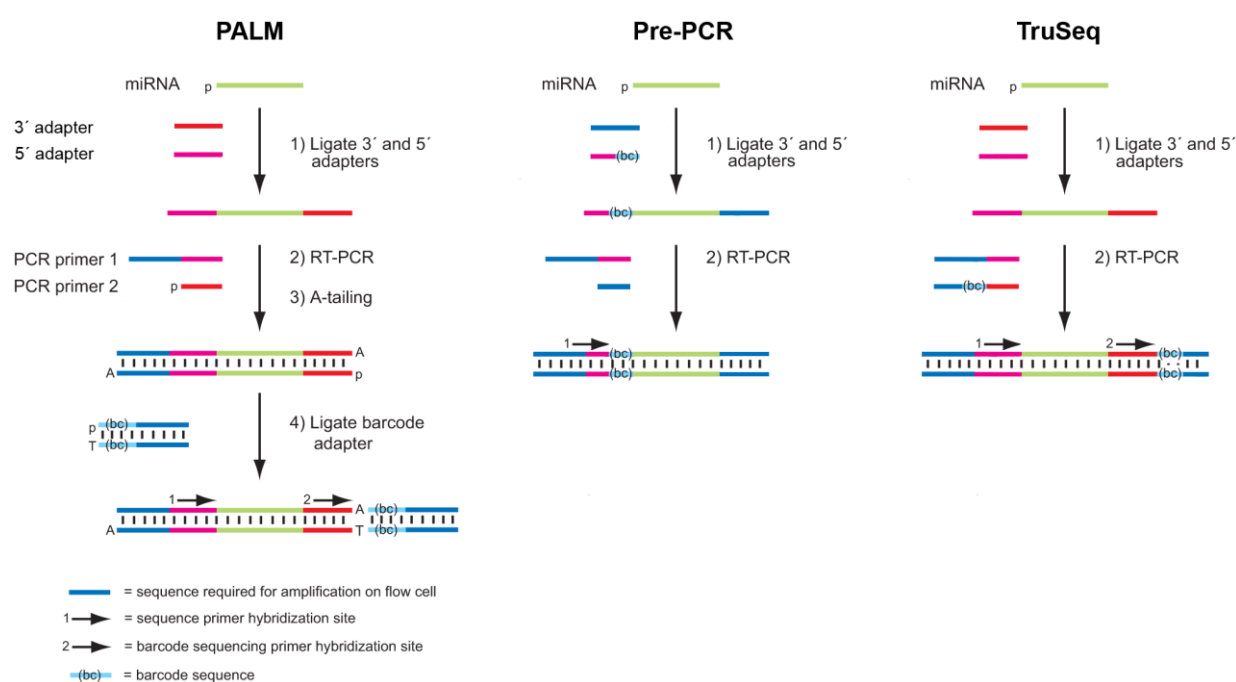


Figure 19: Comparative schematic of small RNA barcoding methods.

The three methods start with ligation of a 3' and 5' RNA adapter to generate a substrate for RT-PCR. In the pre-PCR barcoding method, the barcode is incorporated in the 5' adapter. In the TruSeq method, the barcode is incorporated in one of the RT-PCR primers. In the PALM barcoding method, the amplified RT-PCR product is A-tailed and ligated to a T-tailed barcoded adapter.

2.2 Pre-PCR barcoding of small RNA

The pre-PCR miRNA barcoding protocol is also similar to the Illumina Small RNA v1.5. Sample Preparation Guide. The adapters used in the protocol were modified to include a barcode and to allow for Illumina index sequencing with the Illumina multiplexing index read sequencing primer. The complete protocol, including the adapter sequences, is available in section 5.3.

2.3 Preparation of dsDNA from *S. cerevisiae* mRNA

Starting with poly A⁺ enriched RNA from *S. cerevisiae* (Clontech 636312), dsDNA was prepared with the NEBNext mRNA Sample Prep Reagent Set 1 (New England Biolabs E6100). During this procedure, RNA was fragmented with a fragmentation buffer and subsequently purified with the Qiagen RNeasy Minelute kit. After second strand cDNA synthesis, the dsDNA was purified with a Zymo DNA Clean and concentrator-5.

2.4 PALM DNA barcoding

The PALM DNA barcoding protocol is similar to the Illumina Genomic DNA Sample Preparation Guide. The adapters used in the protocol were modified to allow for PALM barcoding and Illumina index sequencing with the Illumina multiplexing index read sequencing primer. The complete protocol, including the adapter sequences, is available in section 5.3. The main difference compared to the current Illumina Genomic DNA library protocol is the addition of the barcode to the library by ligation after PCR amplification. After ligation of the barcode, no further purification of the library is required. The library is quantified using analysis of area under the peaks with a BioAnalyzer 2100 (Agilent) to determine the correct loading concentration for subsequent sequencing.

2.5 miRNA sequencing and data analysis

The pooled PALM and pre-PCR miRNA libraries were each sequenced in one lane on an Illumina Genome Analyzer Iix sequencer (40 bp single reads), using version 4 of cluster generation and sequencing kits. Sequencing of the pooled TruSeq miRNA libraries was done in one lane on an Illumina HiSeq 2000 sequencer (40 bp single reads), using version 4 of cluster generation and sequencing kits. Raw sequences were obtained from the Illumina GA Pipeline software CASAVA v1.7. The PALM barcoded sequences were demultiplexed using the Illumina pipeline and the pre-PCR barcodes using scripts written for this purpose. The pre-PCR barcodes cannot be demultiplexed using CASAVA because the pre-PCR barcode is not obtained with a separate read like the PALM and TruSeq barcode. The scripts allow for no mismatches in the barcode. Each barcode set was analyzed for small RNA using the Illumina pipeline add-on Flicker v2.7. Flicker trims the adaptor sequence from each read and does iterative alignment to the genome and to the miRNA database (miRBase v16) using the ELAND alignment strategy. The iterative alignment generates statistics of the number of reads aligning to the different classes of miRNA, as well as to individual miRNAs.

2.6 DNA/mRNA sequencing and data analysis

The pooled *S. cerevisiae* mRNA libraries were sequenced in one lane on an Illumina HiSeq 2000 sequencer (40 bp single reads), using version 4 of cluster generation and sequencing kits. The Illumina GA Pipeline software CASAVA v1.7. was used to obtain the reads and to demultiplex the PALM and TruSeq barcoded sequences. Each barcode read set was aligned

and annotated with CASAVA v1.7 using the *S. cerevisiae* S228C genome downloaded from the UCSC Genome website and the *S. cerevisiae* GTF exon and splice site annotation file downloaded from the Ensembl website. Reads that aligned to each exon and splice junction site were summed per gene.

2.7 TruSeq small RNA and DNA barcoding and sequencing

For the TruSeq sample preparation, the Illumina TruSeq Small RNA Sample Prep Kit (RS-200–0012) and the Illumina TruSeq DNA Sample Prep Kit (FC-121–1001) were used.

3 Results

3.1 Yields and quantification of libraries

The PALM barcode ligation step produces several DNA products but only the main product, library products with the barcode adapters ligated to both ends, are able to form clusters and generate sequencing data. For miRNA libraries, this product has a size of approximately 170 bp. For mRNA/DNA libraries this product has a size that is 102 bp longer than the size selected product before the PCR step. The other DNA products present in the library cannot form clusters or be sequenced: residual barcode adapters (~32 bp) can bind to the Illumina flow cell with one end, but will not produce clusters because bridge amplification only occurs when both ends of the DNA strand bind to the flow cell. Barcode-adapter dimers (~64 bp), can bind to the flow cell, but will not produce sequence because they lack a sequencing primer hybridization site. For this reason, no gel purification step is needed after the PALM barcode ligation step. When no final gel purification step is performed, quantification of the total quantity of DNA present in the library after barcode ligation would over-estimate the available material for optimal cluster generation and sequencing. Therefore, it is good practice to quantify the amounts of the desired products using an Agilent High Sensitivity DNA chip or an analogous gel- and microfluidics-based system to correctly load the flow cell.

For miRNA PALM barcoding, we optimized the yield of the Illumina small RNA library preparation protocol (version 1.5) for PALM barcoding by using higher reaction volumes for the RT-PCR step. No extra cycles were added to the PCR reaction. Starting from 1 μ g of Human Brain Reference total RNA, the protocol yields 11.14 ± 1.5 ng of gel purified PCR product. The PALM barcoding step worked well starting with between 2 and 20 ng of gel-

purified, PCR-amplified miRNA library. After PALM barcoding and AMPure XP bead purification, the final yield (in ng) of library with barcodes ligated to both ends, is approximately the same as the amount of PCR-amplified miRNA library used to start the PALM barcoding reaction.

The mRNA/DNA PALM barcoding protocol is based on the Illumina Genomic DNA Sample Preparation Guide. Starting from 5 ng of dsDNA, the PALM protocol yields ~200 ng PCR-amplified library (15 cycles) of which 100 ng was used in the PALM barcoding step. This generated >100 ng of library with barcodes ligated to both ends.

3.2 Deep Sequencing Results of Human Brain Reference RNA

We performed multiplexed miRNA deep sequencing on Human Brain Reference RNA using libraries prepared with three different protocols: PALM barcoded (12 barcodes), pre-PCR barcoded (4 barcodes) and TruSeq barcoded (12 barcodes). Sequencing of the brain RNA yielded 23,685,700 Illumina GAIIx pass-filter reads for the PALM barcoded pool, 24,171,696 Illumina GAIIx reads for the pre-PCR barcoded pool and 35,495,446 Illumina HiSeq 2000 reads for the TruSeq pool. Of the pass-filter reads from the PALM, pre-PCR and TruSeq barcoded libraries, 88%, 92% and 97% contained the barcode sequence respectively. Representation of the differentially barcoded libraries within the flow cell lanes was uniform and more than 50% of all the sequences mapped to mature miRNAs (Table S1).

3.3 Deep Sequencing Results of *S. cerevisiae* mRNA

We performed multiplexed mRNA deep sequencing on *S. cerevisiae* reference mRNA using libraries prepared with two different protocols: PALM barcoded (12 barcodes) and TruSeq barcoded (12 barcodes). Sequencing yielded 104,277,310 Illumina HiSeq 2000 pass-filter reads for the PALM barcoded pool and 115,419,701 Illumina HiSeq 2000 pass-filter reads for the TruSeq pool. Of the pass-filter reads from the PALM and TruSeq barcoded libraries, 94% and 97% contained the barcode sequence respectively. Representation of the differentially barcoded libraries within the flow cell lanes was uniform and more than 60% of all the sequences mapped to exons and splice junction sites (Table S2).

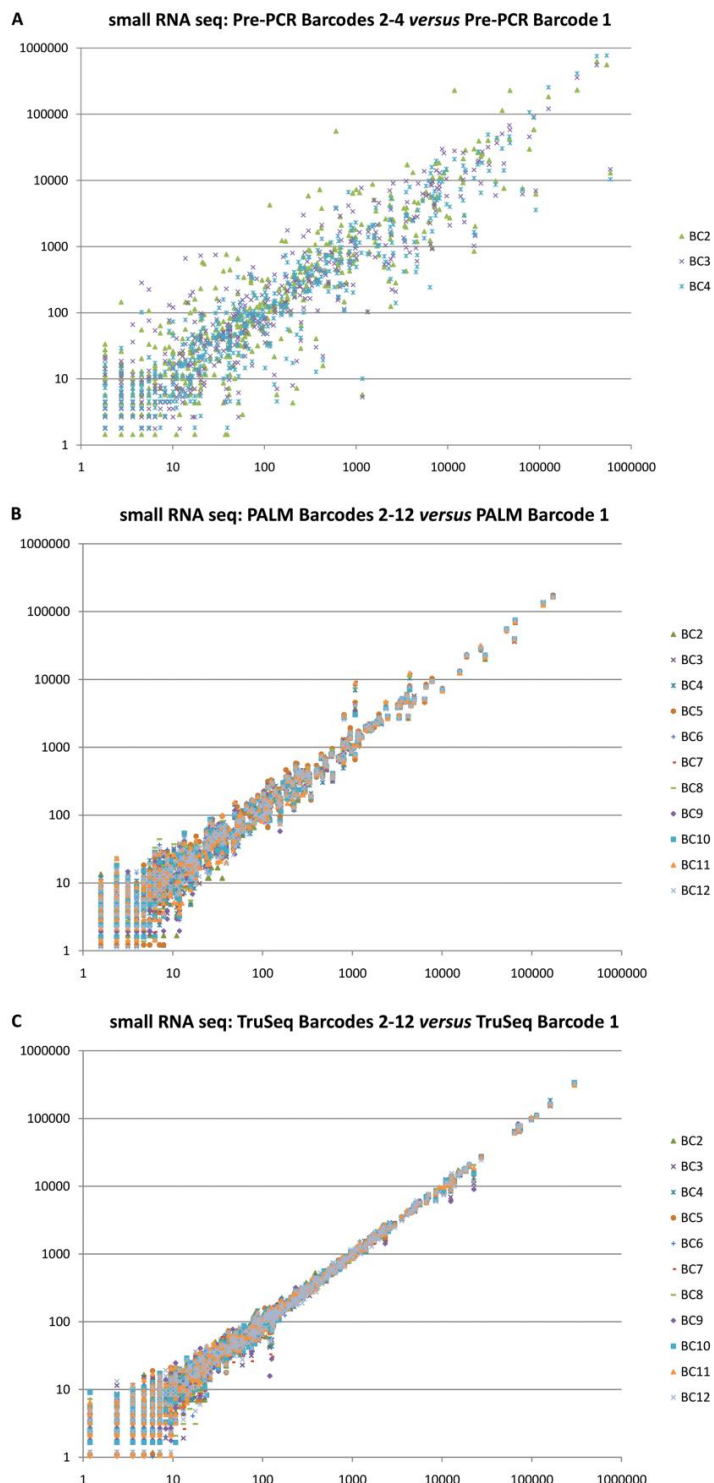


Figure 20: miRNA digital expression levels of all detected human brain reference sample miRNAs.

(a) in the pre-PCR barcoded library 1 versus their expression in the 3 other pre-PCR barcoded libraries, (b) in the PALM barcoded library 1 versus their expression in the 11 other PALM barcoded libraries, (c) in the TruSeq barcoded library 1 versus their expression in the 11 other TruSeq barcoded libraries.

3.4 Evaluation of bias for miRNA barcoding

We calculated the expression of each miRNA as its number of read counts normalized by the total number of reads for each library. The scatter plots in Figure 20 shows a side-by-side comparison of the miRNA expression profiles of the human brain reference libraries, barcoded using either the pre-PCR (A), PALM barcoding protocol (B) and the TruSeq barcoding protocol (C). This comparison reveals a very low variability in the miRNA expression profiles of the PALM and TruSeq barcoded samples but not for the pre-PCR barcoded samples, which is confirmed using a linear regression analysis on the miRNA with at least 10 counts (Table S1) for one of the barcodes: Barcode 1 against the other barcodes gives an $R^2 = 0.8197 \pm 0.1217$ for pre-PCR vs. $R^2 = 0.9930 \pm 0.0022$ for PALM vs. $R^2 = 0.9977 \pm 0.0016$ for TruSeq (See Table S3 for details). The bias introduced by the pre-PCR barcoding protocol precludes quantitative comparison of multiple samples using this strategy for multiplexing.

3.5 Evaluation of bias for mRNA/dsDNA barcoding

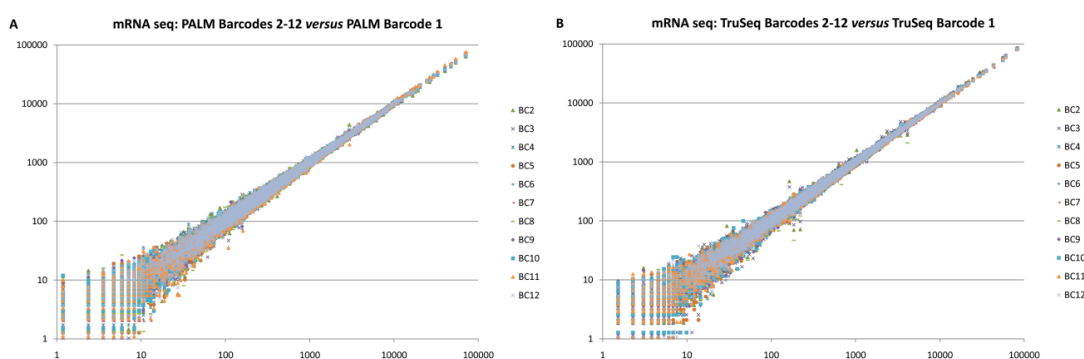


Figure 21: mRNA digital expression levels of all detected *S. cerevisiae* reference sample mRNAs.

A: in the PALM barcoded library 1 versus their expression in the 11 other PALM barcoded libraries, **B:** in the TruSeq barcoded library 1 versus their expression in the 11 other TruSeq barcoded.

We calculated the expression of each mRNA as its number of read counts normalized by the total number of reads for each library. The scatter plots in Figure 21 shows a side-by-side comparison of the mRNA expression profiles of the *S. cerevisiae* libraries, barcoded using either the PALM barcoding protocol (A) and the TruSeq barcoding protocol (B). This comparison reveals a very low variability in the mRNA expression profiles of the PALM and TruSeq barcoded samples, which is confirmed using a linear regression analysis on the mRNA with at least 10 counts (Table S2) for one of the barcodes: Barcode 1 against the other

barcodes gives an $R^2 = 0.9991 \pm 0.0005$ for PALM vs. TruSeq $R^2 = 0.9996 \pm 0.0003$ for TruSeq (See Table S3 for details).

4 Discussion

The constantly increasing throughput of next generation sequencers opens the possibility for multiplexed sequencing of samples. For example, sequencing one miRNA sample in one flow cell lane on an Illumina GAIIx generates an order of magnitude more read data than required: there are currently only 1037 known human miRNAs, representing a maximum of 25 kb of reference sequence [33]. Current Illumina technology provides >50 million reads from 1 flow cell lane. Thus, even multiplexing 12 different miRNA samples in one lane results in >2 million reads per sample. This coverage is still enough to accurately quantify all but the low abundant miRNA present in these samples.

A commonly used technique for multiplexing samples for deep sequencing is to incorporate a known, sample-specific nucleotide sequence in the DNA fragments during library preparation [26-30]. This sample-specific sequence (barcode) is sequenced together with the rest of the fragment. PCR amplification of a pool of DNA molecules with different nucleotide compositions, especially near priming sites, can however result in quantitative bias because some DNA species are amplified more efficiently than others [23, 24, 31, 32]. As we have shown here, introducing a barcode before the PCR step can result in a barcode-specific quantitative bias. Nonetheless, the currently published methods and commercial kits (i.e. Nugen and Bioo Scientific) introduce the barcode in the library before or during PCR-based library amplification. Unfortunately, none of these methods are provided with a quantitative analysis of the bias resulting from the use of barcodes. Thus, we reasoned that introduction of the barcode after library amplification would address this limitation by simply avoiding the problem and developed the PALM protocol. Illumina only recently introduced the TruSeq multiplexed sample preparation kits. The Illumina TruSeq small RNA strategy introduces the barcode during the PCR step using differentially barcoded primers, while the TruSeq DNA (or messenger RNA converted to double stranded DNA) strategy introduces the barcode before the PCR step by ligation of differentially barcoded adaptors. At the time of this publication, we are unaware of any published data demonstrating the impact of the TruSeq

protocols on the bias created by the combination of barcoding and PCR. For this reason, we compared the PALM barcoding strategy with the TruSeq barcoding strategy.

Our results describe a detailed quantitative analysis of PCR and barcoding bias obtained using the PALM and the TruSeq barcoding protocol. The PALM protocol demonstrates a robust and efficient multiplexing method for miRNA and mRNA expression profiling that is free of barcode-induced PCR bias. In contrast, our results for the same miRNA samples profiled with our pre-PCR barcoding protocol demonstrate significant barcode-specific bias. This bias is quite extreme, as the digital expression of the same miRNAs shows up to 100-fold differences in read counts for the top 200 most abundantly expressed miRNAs. Both the TruSeq miRNA and mRNA/dsDNA barcoding protocols show no bias. In the TruSeq miRNA protocol, the strategy of introducing the barcode during the PCR step using differentially barcoded primers does not result in bias. The TruSeq protocol for mRNA/dsDNA which introduces the barcode before the PCR step, surprisingly also produces results with no bias. It is unclear why our pre-PCR protocol for small RNA produces biased results, while the TruSeq protocol for mRNA/dsDNA produces unbiased results. Compared to our pre-PCR small RNA protocol which places the barcode only 3 bp away from the miRNA insert, the TruSeq mRNA/dsDNA protocol places the barcode 34 bp away from the mRNA/dsDNA. Another difference is that the mRNA/dsDNA protocol contains no reverse transcriptase step after barcoding and works with a typical insert size of 250 bp, instead of the miRNA insert size of approx. 22 bp. Because of this, the barcode sequence might have less impact on the quantitative results after PCR.

There are multiple sources of bias that can be introduced during sample purification and library preparation including ligation bias, secondary structures, PCR-bias created by amplification of differentially barcoded miRNAs and amplification bias introduced on the surface of the flow cell [34-36]. The important point in the context of the present work is that PALM and TruSeq barcoding, in contrast to the pre-PCR barcoding protocol we used, gives consistent and reproducible results allowing multiplexing and meaningful comparisons of differential miRNA and mRNA expression without the need for technical replicates with different barcodes. In addition, PALM is a transparent and adaptable alternative to commercial strategies with a limited number of barcodes. It allows the user to modify the

protocol and provides the flexibility to synthesize as many barcodes as needed in order to keep up with the ever-growing sequencing throughput.

5 Supporting information

5.1 Supplementary figures

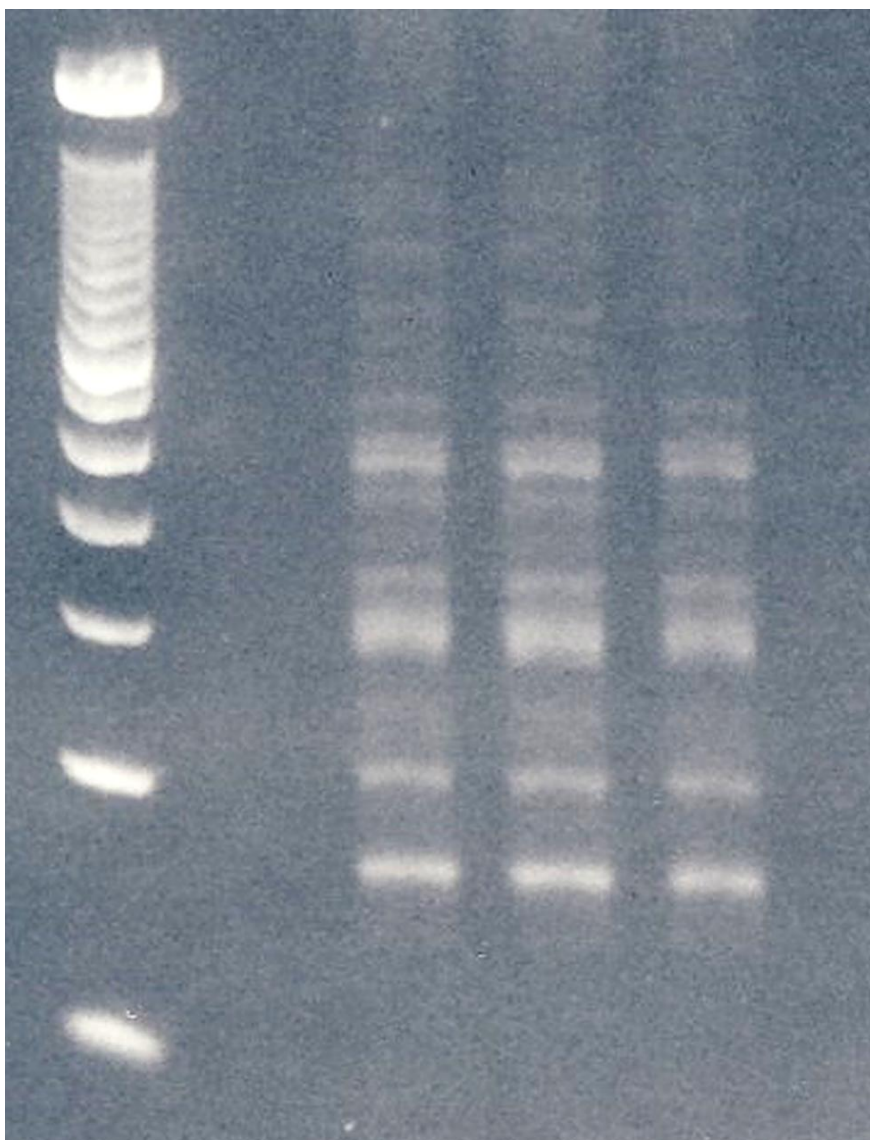


Figure S1: E-gel of a library after PCR amplification and before barcode ligation.

Typical Invitrogen 4% E-gel with 50 bp ladder of a Human Brain Reference RNA (Ambion) library after PCR amplification and before barcode ligation. The PCR product that needs to be purified from the gel is the band next to the 100 bp marker (second ladder band starting from the bottom of the picture). The bands closely above this PCR product should not be excised from the gel: doing so lowers the percentage of mature miRNA sequences in the sequencing results.

5.2 Supplementary tables

Table S1: : Human Brain RNA sequence quality statistics.

	Number of reads (mean \pm st. dev)	% (mean \pm st. dev)
12 PALM barcodes		
Total reads with barcodes	20,598,984	
Ave reads per barcoded sample	1,716,582 \pm 372,711	
Unaligned – too short	303,765 \pm 126,807	17.58 \pm 5.35
Unaligned – no match	123,341 \pm 24,814	7.22 \pm 0.53
Aligned - mature	904,344 \pm 207,458	52.69 \pm 4.74
Identified miRNA	491 \pm 19	
miRNA with >10 counts	325	
4 pre-PCR barcodes		
Total reads with barcodes	22,149,582	
Ave reads per barcoded sample	5,537,396 \pm 1,209,422	
Unaligned – too short	359,639 \pm 136,311	6.49 \pm 2.46
Unaligned – no match	261,967 \pm 82,974	4.73 \pm 1.50
Aligned - mature	3,359,695 \pm 698,009	60.67 \pm 12.61
Identified miRNA	594 \pm 4	
miRNA with >10 counts	434	
12 TruSeq barcodes		
Total reads with barcodes	34,590,185	
Ave reads per barcoded sample	2,882,515 \pm 386,221	
Unaligned – too short	63,515 \pm 9,874	2.2 \pm 0.34
Unaligned – no match	470,410 \pm 56,970	16.32 \pm 1.98
Aligned - mature	1,563,981 \pm 211,692	54.26 \pm 7.34
Identified miRNA	581 \pm 14	
miRNA with >10 counts	402	

Alignment statistics were generated by Flicker version 2.7. The pool of 12 PALM barcodes and the pool of 4 Pre-PCR barcodes were each sequenced in one Illumina GAIIx flow cell lane. The pool of 12 TruSeq barcodes was sequenced in one Illumina HiSeq 2000 lane. The table also shows the number of identified miRNA in each barcoded sample and the number of miRNA with at least 10 counts in at least one of the barcoded samples.

Table S2: *S. cerevisiae* mRNA sequence quality statistics.

	Number of reads	% (mean \pm st. dev)
12 PALM barcodes		
Total reads with barcodes	98,108,551	
Ave reads per barcoded sample	8,175,712 \pm 2,018,971	
Aligned	5,287,979 \pm 1,315,639	0,64 \pm 0,16
Unaligned	2,887,734 \pm 703,653	0,35 \pm 0,09
Identified mRNA	6482 \pm 24	
miRNA with >10 counts	6208	
12 TruSeq barcodes		
Total reads with barcodes	112,036,259	
Ave reads per barcoded sample	9,336,354 \pm 1,405,129	
Aligned	6,120,073 \pm 929,944	0,66 \pm 0,10
Unaligned	3,216,282 \pm 477,117	0,34 \pm 0,05
Identified mRNA	6496 \pm 14	
miRNA with >10 counts	6230	

Alignment statistics were generated with CASAVA v1.7. The pool of 12 PALM barcodes and the pool of 12 TruSeq barcodes were each sequenced in one Illumina HiSeq 2000 lane. The table also shows the number of identified mRNA for each barcoded sample and the number of miRNA with at least 10 counts in at least one of the barcoded samples.

Table S3: Matrix of correlations between differentially barcoded samples.

Only miRNA and mRNA with at least 10 counts in at least one of the barcoded samples were considered.

miRNA: Pre-PCR barcoding												
	BC 1	BC 2	BC 3	BC 4								
BC 1	1,0000	0,7334	0,7677	0,7774								
BC 2	0,7334	1,0000	0,8859	0,9363								
BC 3	0,7677	0,8859	1,0000	0,9479								
BC 4	0,7774	0,9363	0,9479	1,0000								
miRNA: PALM barcoding												
	BC 1	BC 2	BC 3	BC 4	BC 5	BC 6	BC 7	BC 8	BC 9	BC 10	BC 11	BC 12
BC 1	1,0000	0,9923	0,9920	0,9919	0,9928	0,9924	0,9919	0,9929	0,9930	0,9929	0,9921	0,9922
BC 2	0,9923	1,0000	0,9982	0,9986	0,9993	0,9995	0,9976	0,9994	0,9989	0,9983	0,9990	0,9968
BC 3	0,9920	0,9982	1,0000	0,9975	0,9995	0,9988	0,9992	0,9989	0,9997	0,9992	0,9983	0,9986
BC 4	0,9919	0,9986	0,9975	1,0000	0,9981	0,9990	0,9984	0,9986	0,9986	0,9989	0,9994	0,9982
BC 5	0,9928	0,9993	0,9995	0,9981	1,0000	0,9995	0,9988	0,9996	0,9996	0,9989	0,9989	0,9978
BC 6	0,9924	0,9995	0,9988	0,9990	0,9995	1,0000	0,9990	0,9999	0,9992	0,9986	0,9997	0,9975
BC 7	0,9919	0,9976	0,9992	0,9984	0,9988	0,9990	1,0000	0,9990	0,9993	0,9992	0,9992	0,9991
BC 8	0,9929	0,9994	0,9989	0,9986	0,9996	0,9999	0,9990	1,0000	0,9993	0,9986	0,9996	0,9975
BC 9	0,9930	0,9989	0,9997	0,9986	0,9996	0,9992	0,9993	0,9993	1,0000	0,9998	0,9990	0,9991
BC 10	0,9929	0,9983	0,9992	0,9989	0,9989	0,9986	0,9992	0,9986	0,9998	1,0000	0,9988	0,9997
BC 11	0,9921	0,9990	0,9983	0,9994	0,9989	0,9997	0,9992	0,9996	0,9990	0,9988	1,0000	0,9980
BC 12	0,9922	0,9968	0,9986	0,9982	0,9978	0,9975	0,9991	0,9975	0,9991	0,9997	0,9980	1,0000
miRNA: TruSeq barcoding												
	BC 1	BC 2	BC 3	BC 4	BC 5	BC 6	BC 7	BC 8	BC 9	BC 10	BC 11	BC 12
BC 1	1,0000	0,9988	0,9955	0,9977	0,9988	0,9991	0,9944	0,9975	0,9972	0,9970	0,9990	0,9976
BC 2	0,9988	1,0000	0,9971	0,9987	0,9996	0,9995	0,9963	0,9982	0,9991	0,9984	0,9995	0,9993
BC 3	0,9955	0,9971	1,0000	0,9956	0,9983	0,9984	0,9996	0,9992	0,9992	0,9995	0,9975	0,9984
BC 4	0,9977	0,9987	0,9956	1,0000	0,9980	0,9982	0,9957	0,9970	0,9972	0,9976	0,9983	0,9986
BC 5	0,9988	0,9996	0,9983	0,9980	1,0000	0,9998	0,9972	0,9990	0,9995	0,9989	0,9992	0,9994
BC 6	0,9991	0,9995	0,9984	0,9982	0,9998	1,0000	0,9977	0,9994	0,9993	0,9991	0,9996	0,9993
BC 7	0,9944	0,9963	0,9996	0,9957	0,9972	0,9977	1,0000	0,9988	0,9984	0,9992	0,9969	0,9980
BC 8	0,9975	0,9982	0,9992	0,9970	0,9990	0,9994	0,9988	1,0000	0,9990	0,9995	0,9989	0,9991
BC 9	0,9972	0,9991	0,9992	0,9972	0,9995	0,9993	0,9984	0,9990	1,0000	0,9994	0,9988	0,9994
BC 10	0,9970	0,9984	0,9995	0,9976	0,9989	0,9991	0,9992	0,9995	0,9994	1,0000	0,9988	0,9991
BC 11	0,9990	0,9995	0,9975	0,9983	0,9992	0,9996	0,9969	0,9989	0,9988	0,9988	1,0000	0,9990
BC 12	0,9976	0,9993	0,9984	0,9986	0,9994	0,9993	0,9980	0,9991	0,9994	0,9991	0,9990	1,0000
mRNA / dsDNA: PALM barcoding												
	BC 1	BC 2	BC 3	BC 4	BC 5	BC 6	BC 7	BC 8	BC 9	BC 10	BC 11	BC 12
BC 1	1,0000	0,9982	0,9995	0,9985	0,9991	0,9989	0,9994	0,9985	0,9993	0,9996	0,9989	0,9988
BC 2	0,9982	1,0000	0,9978	0,9995	0,9994	0,9995	0,9969	0,9997	0,9992	0,9991	0,9955	0,9995
BC 3	0,9995	0,9978	1,0000	0,9987	0,9989	0,9983	0,9988	0,9981	0,9988	0,9993	0,9991	0,9986
BC 4	0,9985	0,9995	0,9987	1,0000	0,9996	0,9993	0,9971	0,9995	0,9991	0,9993	0,9965	0,9996
BC 5	0,9991	0,9994	0,9989	0,9996	1,0000	0,9996	0,9981	0,9995	0,9995	0,9997	0,9973	0,9996
BC 6	0,9989	0,9995	0,9983	0,9993	0,9996	1,0000	0,9981	0,9996	0,9997	0,9995	0,9966	0,9996
BC 7	0,9994	0,9969	0,9988	0,9971	0,9981	0,9981	1,0000	0,9975	0,9987	0,9988	0,9991	0,9978
BC 8	0,9985	0,9997	0,9981	0,9995	0,9995	0,9996	0,9975	1,0000	0,9993	0,9992	0,9961	0,9996
BC 9	0,9993	0,9992	0,9988	0,9991	0,9995	0,9997	0,9987	0,9993	1,0000	0,9997	0,9974	0,9994
BC 10	0,9996	0,9991	0,9993	0,9993	0,9997	0,9995	0,9988	0,9992	0,9997	1,0000	0,9981	0,9995
BC 11	0,9989	0,9955	0,9991	0,9965	0,9973	0,9966	0,9991	0,9961	0,9974	0,9981	1,0000	0,9968
BC 12	0,9988	0,9995	0,9986	0,9996	0,9996	0,9996	0,9978	0,9996	0,9994	0,9995	0,9968	1,0000
mRNA / dsDNA: TruSeq barcoding												
	BC 1	BC 2	BC 3	BC 4	BC 5	BC 6	BC 7	BC 8	BC 9	BC 10	BC 11	BC 12
BC 1	1,0000	0,9994	0,9989	0,9998	0,9997	0,9996	0,9991	0,9994	0,9997	0,9997	0,9997	0,9996
BC 2	0,9994	1,0000	0,9994	0,9992	0,9991	0,9996	0,9989	0,9992	0,9992	0,9993	0,9990	0,9994
BC 3	0,9989	0,9994	1,0000	0,9985	0,9983	0,9989	0,9981	0,9981	0,9985	0,9993	0,9987	0,9994
BC 4	0,9998	0,9992	0,9985	1,0000	0,9998	0,9996	0,9991	0,9995	0,9998	0,9994	0,9995	0,9993
BC 5	0,9997	0,9991	0,9983	0,9998	1,0000	0,9996	0,9992	0,9995	0,9998	0,9992	0,9993	0,9991
BC 6	0,9996	0,9996	0,9989	0,9996	0,9996	1,0000	0,9990	0,9995	0,9996	0,9993	0,9991	0,9992
BC 7	0,9991	0,9989	0,9981	0,9991	0,9992	0,9990	1,0000	0,9994	0,9991	0,9987	0,9988	0,9988
BC 8	0,9994	0,9992	0,9981	0,9995	0,9995	0,9995	0,9994	1,0000	0,9995	0,9989	0,9991	0,9989
BC 9	0,9997	0,9992	0,9985	0,9998	0,9998	0,9996	0,9991	0,9995	1,0000	0,9994	0,9994	0,9993
BC 10	0,9997	0,9993	0,9993	0,9994	0,9992	0,9993	0,9987	0,9989	0,9994	1,0000	0,9996	0,9998
BC 11	0,9997	0,9990	0,9987	0,9995	0,9993	0,9991	0,9988	0,9991	0,9994	0,9996	1,0000	0,9996
BC 12	0,9996	0,9994	0,9994	0,9993	0,9991	0,9992	0,9988	0,9989	0,9993	0,9998	0,9996	1,0000

5.3 Supplementary Materials and Methods (Text S1)

5.3.1 PALM Barcoding of small RNA

Materials: All oligonucleotides were ordered HPLC purified from Integrated DNA Technologies Inc. (IDT). Enzymes were obtained from Invitrogen, Life Technologies, Inc., New England Biolabs, and Enzymatics Inc.

- PALM Adapters

5X PALM miRNA 3'adapter (10 μ M)
5' **rApp**-AGATCGGAAGAGCACACGTCT(**C3spacer**)

PALM miRNA 5'adapter (5 μ M)
5' GUUCAGAGUUCUACAGUCCGACGAUC

- 5X PALM Reverse Transcription primer

5X PALM RT primer (100 μ M)
5' GTGACTGGAGTTCAGACGTGTGCTCTCCGA

- PALM PCR primers

PALM miRNA PCR primer 1 (25 μ M)
5' AATGATACGGCGACCCAGCAGGTTTCAGAGTTCACAGTCCGA

PALM miRNA PCR primer 2 (25 μ M)
5' **P**-GTGACTGGAGTTCAGACGTGTGCTCTCCGA

- PALM Barcodes (50 μ M final concentration in annealed solution 50mM NaCl/10mM Tris pH 7.5)

barcode 1-P	5' P -G-ATCACG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 1-T	5' CAAGCAGAAGACGGGCATACGAGATCGTGATC*T
barcode 2-P	5' P -G-CGATGT-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 2-T	5' CAAGCAGAAGACGGGCATACGAGATACATCCG*T
barcode 3-P	5' P -G-TTAGGC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 3-T	5' CAAGCAGAAGACGGGCATACGAGATGCCTAAC*T
barcode 4-P	5' P -G-TGACCA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 4-T	5' CAAGCAGAAGACGGGCATACGAGATTGGTCAC*T
barcode 5-P	5' P -G-ACAGTG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 5-T	5' CAAGCAGAAGACGGGCATACGAGATCACTGTCT*T
barcode 6-P	5' P -G-GCCAA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 6-T	5' CAAGCAGAAGACGGGCATACGAGATATTGGCC*T
barcode 7-P	5' P -G-CAGATC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 7-T	5' CAAGCAGAAGACGGGCATACGAGATGATCTGC*T
barcode 8-P	5' P -G-ACTTGA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 8-T	5' CAAGCAGAAGACGGGCATACGAGATTCAGTC*T
barcode 9-P	5' P -G-GATCAG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 9-T	5' CAAGCAGAAGACGGGCATACGAGATCTGATCC*T
barcode 10-P	5' P -G-TAGCTT-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 10-T	5' CAAGCAGAAGACGGGCATACGAGATAAGCTAC*T
barcode 11-P	5' P -G-GGCTAC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 11-T	5' CAAGCAGAAGACGGGCATACGAGATGTAGCCC*T
barcode 12-P	5' P -G-CTTGTA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 12-T	5' CAAGCAGAAGACGGGCATACGAGATTACAAGC*T

(**P** = 5'phosphate, **3AmMO** = 3'amine, and * = phosphorothioate linkage)

- T4 RNA Ligase 2, truncated; T4 RNA Ligase; Finnzymes 2X Phusion HF master mix; and Taq DNA polymerase (New England Biolabs)
- RNaseOUT, SuperScript™ II Reverse Transcriptase, Qubit Fluorometer, dsDNA High Sensitivity assay, Magnetic Particle Concentrator, E-Gel EX 4% Agarose Gel, TrackIt 50 bp ladder, and 10 mM ATP (Invitrogen, Life Technologies, Inc.)
- Agarose Dissolving Buffer (ADB) and DNA Clean and concentrator -5 and -25 kits (Zymo Research Co.)
- T4 DNA Ligase (Rapid), using 2X Rapid Ligation Buffer (Enzymatics, Inc.)
- Molecular Biology Grade, DNase- and RNase-free MgCl₂ (Sigma-Aldrich, Co.)
- 2100 Bioanalyzer and High Sensitivity DNA kit (Agilent Technologies)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, Inc.)
- Dark Reader® transilluminator (Clare Chemical Research, Inc.)

Methods:

1) Ligation of 3'- and 5'- adapters to total RNA

1 μ g of total RNA sample was added to 10 pmol of PALM miRNA 3' adapter in a final volume of 6 μ L. The mixture was incubated at 70°C for 2 minutes, and then transferred to ice. Subsequently, the following reagents were added to the mixture: 1 μ L 10X T4 RNA Ligase 2 truncated reaction buffer; 0.8 μ L 100 mM MgCl₂; 1.5 μ L T4 RNA Ligase 2 truncated; and 0.5 μ L RNaseOUT, and the reaction was incubated at 22°C for 1 hour. Just prior to this reaction finishing, the PALM miRNA 5' adapter was denatured by heating it at 70°C for 2 minutes and transferring it to ice. Then, the following reagents were added to the reaction mixture: 1 μ L of 10 mM ATP; 1 μ L PALM miRNA 5' Adapter; and 1 μ L T4 RNA Ligase. The reaction was incubated at 20°C for 1 hour and then transferred to ice.

2) Reverse transcription of adapter ligated products.

12 μ L of the above RNA Ligation reaction was then taken and added directly to 60 pmol of PALM RT primer in a final volume of 15 μ L. The mixture was then heated to 70°C for 2 minutes, and then transferred to ice. In a separate, nuclease-free PCR tube the following reagents were premixed: 6 μ L 5X First Strand Buffer; 1.5

μL 12.5 mM dNTP mix; 3 μL 100 mM DTT; and 1.5 μL RNaseOUT. The 15 μL of Ligation products and PALM RT primer were then combined with the 12 μL of reagents in the PCR tube. The mixture was heated to 48°C for 3 minutes, and then 3 μL SuperScript™ II was added. The reaction was then incubated at 44°C for 1 hour, and then transferred to ice.

3) PCR Amplification.

In a separate microfuge tube the following reagents were combined and placed on ice: 50 μL 2X phusion master mix; 2 μL PALM miRNA PCR primer 1; 2 μL PALM miRNA PCR primer 2; 16 μL nuclease-free purified water. This mixture of reagents was then added to the 30 μL from the reverse transcription reaction in a PCR tube. The mixture was then placed in a thermocycler using the following program: 1 initial hold, 98°C, 30 seconds; cycled 12 times, 98°C, 10 seconds; 60°C, 30 seconds; and 72°C, 15 seconds; 1 final hold, 72°C, 10 minutes.

After the PCR reaction was complete, the PCR products were purified using a Zymo-25 DNA column clean up protocol and eluted from the column in 38.5 μL nuclease-free purified water.

4) 3' A-tailing of PCR products.

The eluted PCR products were mixed with 5 μL 10X Taq buffer, 5 μL 10 mM dATP and 1.5 μL Taq polymerase. The reaction was then incubated at 72°C for 30 minutes. The A-tailed PCR products were then isolated from the reaction using a Zymo-25 DNA column clean up protocol and eluted from the column in 60 μL nuclease-free purified water.

5) Purification of PCR products.

The desired PCR products were purified by loading onto a 4% Agarose E-gel, 20 μL per lane (3 lanes per sample). 5 μL of TrackIT 50bp Ladder was loaded into a single lane and used as a DNA size marker. The gel was electrophoresed for approximately 30 minutes, until the desired bands (at approximately 100bp \pm 10bp, see Supplementary Figure 1), were at the bottom of the gel. The gel was imaged using a Dark Reader® Transilluminator and the desired bands were excised using a clean razor blade. Approximately, 4 volumes of ADB buffer were added to the gel pieces and the mixture was incubated for 5-10 minutes at 37 °C, while mixing several times during the period. PCR products were isolated from the mixture using a Zymo-5 DNA column clean up protocol. The desired PCR

products were eluted from the column with 21 μL nuclease-free, purified water. 1 μL of the eluted DNA products was used for quantification with a Qubit Fluorometer.

6) Ligation of PALM barcodes to PCR products.

The amount of material going into ligation reaction varied but worked well at 2-20 ng. The eluted PCR products were mixed with 25 pmol of annealed PALM barcode in a final volume of 20 μL and added to 25 μL of 2X Rapid Ligation buffer with 5 μL T4 DNA Ligase (Rapid) and incubated at 20°C for 30 minutes. 1 μL of 0.5 M EDTA was added to stop the ligation reaction.

7) AMPure XP magnetic bead isolation of DNA products.

After the ligation reaction, 90 μL of AMPure XP beads was added, mixed, and allowed to incubate at room temperature for 5 minutes. Using a Magnetic Particle Concentrator (MPC), the beads were pelleted against the wall of the tube. The tube will remain in the MPC during all subsequent wash steps. The supernatant was removed, and the beads were washed twice with 500 μL of 70% ethanol (allowed to sit for 1 minute each time during wash). The AMPure XP beads were allowed to air dry for approximately 5 minutes. Then the tube was removed from the MPC and 30 μL of TE was added and vortexed to re-suspend the beads. After a 2 minute wait step, the tube was placed back in the MPC to pellet the beads against the wall of the tube and the eluate was removed and transferred to a fresh microfuge tube.

8) Agilent High Sensitivity DNA chip analysis and Illumina flow cell loading.

1 μL of the eluted DNA library was run on an Agilent High Sensitivity DNA chip to determine size and concentration of the desired barcoded products (at approximately 145bp and 180bp). The concentration of the desired products was determined from the Agilent 2100 Bioanalyzer electropherogram and used to prepare an equimolar pool of 9 pM solution for cluster generation on an Illumina flow cell.

5.3.2 Pre-PCR Barcoding of small RNA

Materials: All oligonucleotides were ordered HPLC purified from Integrated DNA Technologies Inc. (IDT). Enzymes were obtained from Invitrogen, Life

Technologies, Inc., New England Biolabs, and Enzymatics Inc.

- Adapters

5X miRNA 3'adapter (10 μ M)
5' **rApp**-ATCTCGTATGCCGTCTTCTGCTTG (**ddc**)

miRNA 5'adapter (5 μ M)
5' GUUCAGAGUUCUACAGUCC**GACG**AUC

Note: Bases in **red** are the barcode sequences. The 4 barcodes used in this study were:

BC1 5' GUUCAGAGUUCUACAGUCCGACGAUC **GACG** AUC

BC2 5' GUUCAGAGUUCUACAGUCCGACGAUC **AAGA** AUC

BC3 5' GUUCAGAGUUCUACAGUCCGACGAUC **ACUC** AUC

BC4 5' GUUCAGAGUUCUACAGUCCGACGAUC **CGCA** AUC

- 5X Reverse Transcription primer

5X miRNA RT primer (100 μ M)
5' CAAGCAGAAGACGGCATAACGA

- PCR primers

miRNA PCR primer 1 (25 μ M)
5' CAAGCAGAAGACGGCATAACGA

miRNA PCR primer 2 (25 μ M)
5'

AATGATACGGCGACCACCGACAGGTTCCAGAGTTCTACAGTCCGA

- T4 RNA Ligase 2, truncated; T4 RNA Ligase; and Finnzymes 2X Phusion HF master mix (New England Biolabs)
- RNaseOUT, SuperScript™ II Reverse Transcriptase, Qubit Fluorometer, dsDNA High Sensitivity assay, Magnetic Particle Concentrator, E-Gel EX 4% Agarose Gel, TrackIt 50 bp ladder and 10 mM ATP (Invitrogen, Life Technologies, Inc.)
- Agarose Dissolving Buffer (ADB) and DNA Clean and concentrator -5 and -25 kits (Zymo Research Co.)
- Molecular Biology Grade, DNase- and RNase-free MgCl₂ (Sigma-Aldrich, Co.)
- 2100 Bioanalyzer and High Sensitivity DNA kit (Agilent Technologies)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, Inc.)
- Dark Reader® transilluminator (Clare Chemical Research, Inc.)

Methods:

1) Ligation of 3'- and 5'- adapters to total RNA

1 μ g of total RNA sample was added to 10 pmol of miRNA 3' adapter in a final volume of 6 μ L. The mixture was incubated at 70°C for 2 minutes, and then transferred to ice. Subsequently, the following reagents were added to the mixture: 1 μ L 10X T4 RNA Ligase 2 truncated reaction buffer; 0.8 μ L 100 mM MgCl₂; 1.5 μ L T4 RNA Ligase 2 truncated; and 0.5 μ L RNaseOUT, and the reaction was incubated at 22°C for 1 hour. Just prior to this reaction finishing, the miRNA 5' adapter was denatured by heating it at 70°C for 2

minutes and transferring it to ice. Then, the following reagents were added to the reaction mixture: 1 μ L of 10 mM ATP; 1 μ L miRNA 5' Adapter; and 1 μ L T4 RNA Ligase. The reaction was incubated at 20°C for 1 hour and then transferred to ice.

2) Reverse transcription of adapter ligated products.

8 μ L of the above RNA Ligation reaction was added directly to 40 pmol of miRNA RT primer in a final volume of 10 μ L. The mixture was heated to 70°C for 2 minutes and then transferred to ice. In a separate, nuclease-free PCR tube the following reagents were premixed: 4 μ L 5X First Strand Buffer; 1 μ L 12.5 mM dNTP mix; 2 μ L 100 mM DTT; and 1 μ L RNaseOUT. The 10 μ L of ligation products and miRNA RT primer were then combined with the 8 μ L of reagents in the PCR tube. The mixture was heated to 48°C for 3 minutes and then 2 μ L SuperScript™ II was added. The reaction was then incubated at 44°C for 1 hour and then transferred to ice.

3) PCR Amplification.

In a separate microfuge tube the following reagents were combined and placed on ice: 25 μ L 2X phusion master mix; 1 μ L miRNA PCR primer 1; 1 μ L miRNA PCR primer 2 and 3 μ L nuclease-free purified water. This mixture of reagents was then added to the 20 μ L from the reverse transcription reaction in the PCR tube. The mixture was then placed in a thermocycler using the following program: 1 initial hold, 98°C, 30 seconds; cycled 12 times, 98°C, 10 seconds; 60°C, 30 seconds; and 72°C, 15 seconds; 1 final hold, 72°C, 10 minutes.

The PCR products were then isolated from the reaction using a Zymo-25 DNA column clean up protocol and eluted from the column in 60 μ L nuclease-free purified water.

4) Purification of PCR products.

The desired PCR products were purified by loading onto a 4% Agarose E-gel, 20 μ L per lane (3 lanes per sample). 5 μ L of TrackIT 50bp Ladder was loaded into a single lane and used as a DNA size marker. The gel was electrophoresed for approximately 30 minutes, until the desired bands (at approximately 100bp \pm 10bp, see Supplementary Figure 1), were at the bottom of the gel. The gel was imaged using a Dark Reader® Transilluminator and the desired bands were excised using a clean razor blade. Approximately, 4 volumes of ADB buffer were

added to the gel pieces and the mixture was incubated for 5-10 minutes at 37 °C while mixing several times during the period. PCR products were isolated from the mixture using a Zymo-25 DNA column clean up protocol. The desired PCR products were eluted from the column with 32 µL nuclease-free, purified water. 1 µL of the eluted DNA products was used for quantification with a Qubit Fluorometer and the dsDNA High Sensitivity assay.

5) Illumina flow cell loading.

The concentration of the desired products was determined from the Qubit Fluorometer and used to prepare an equimolar pool of 7 pM solution for cluster generation on an Illumina flow cell.

5.3.3 PALM Barcoding for DNA

Materials: All oligonucleotides were ordered HPLC purified from Integrated DNA Technologies Inc. (IDT). Enzymes were obtained from Invitrogen, Life Technologies, Inc., New England Biolabs, and Enzymatics Inc.

- PALM Adapters (50 µM mixture after annealing)

Adapter 1

5' P-GATCGGAAGAGCACACGTCT

Adapter 2

5' AACTCTTTCCCTACACGACGCTCTTCCGATC*T

- PALM PCR primers

PALM PCR primer 1 (25 µM)

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T

PALM PCR primer 2 (25 µM)

5' P-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

- PALM Barcodes (50 µM final concentration in annealed solution 50mM NaCl/10mM Tris pH 7.5)

barcode 1-P	5' P-G-ATCAG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 1-T	5' CAAGCAGAAGACGGCATACGAGATCGTGATC*T
barcode 2-P	5' P-G-CGATGT-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 2-T	5' CAAGCAGAAGACGGCATACGAGATACATCGC*T
barcode 3-P	5' P-G-TTAGGC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 3-T	5' CAAGCAGAAGACGGCATACGAGATGCCTAAC*T
barcode 4-P	5' P-G-TGACCA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 4-T	5' CAAGCAGAAGACGGCATACGAGATTGGTCAC*T
barcode 5-P	5' P-G-ACAGTG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 5-T	5' CAAGCAGAAGACGGCATACGAGATCACTGTC*T
barcode 6-P	5' P-G-GCCAA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 6-T	5' CAAGCAGAAGACGGCATACGAGATATTGGCC*T
barcode 7-P	5' P-G-CAGATC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 7-T	5' CAAGCAGAAGACGGCATACGAGATGATCTGC*T
barcode 8-P	5' P-G-ACTTGA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 8-T	5' CAAGCAGAAGACGGCATACGAGATTCAAGTC*T
barcode 9-P	5' P-G-GATCAG-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 9-T	5' CAAGCAGAAGACGGCATACGAGATCTGATCC*T

barcode 10-P	5' P-G-TAGCTT-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 10-T	5' CAAGCAGAAGACGGCATACGAGATAAGCTAC*T
barcode 11-P	5' P-G-GGCTAC-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 11-T	5' CAAGCAGAAGACGGCATACGAGATGTAGCCC*T
barcode 12-P	5' P-G-CTTGTA-ATCTCGTATGCCGTCTTCTGCTTG/3AmMO/
barcode 12-T	5' CAAGCAGAAGACGGCATACGAGATTACAAGC*T

(P = 5'phosphate, 3AmMO = 3'amine, and * = phosphorothioate linkage)

- 10X Phosphorylation buffer, dNTP mix, T4 DNA polymerase, *E. coli* DNA Polymerase I Large (Klenow) Fragment, T4 Polynucleotide Kinase, 10X NEBuffer2, 10 µL 1 mM dATP, Klenow Fragment (3'→5' exo-), 5X Phusion HF buffer, Phusion DNA Polymerase and Taq DNA polymerase, T4 DNA ligase, Quick Ligation Reaction buffer (New England Biolabs).
- Qubit Fluorometer, and dsDNA High Sensitivity assay, Magnetic Particle Concentrator, E-Gel EX 4% Agarose Gel, TrackIt 50 bp ladder, and 10 mM dATP, 0.5M EDTA pH8 (Invitrogen, Life Technologies, Inc.)
- Agarose Dissolving Buffer (ADB) and DNA Clean and concentrator -5 and -25 kits (Zymo Research Co.)
- T4 DNA Ligase (Rapid), using 2X Rapid Ligation Buffer (Enzymatics, Inc.)
- TE buffer pH8 (Sigma-Aldrich, Co.)
- 2100 Bioanalyzer and High Sensitivity DNA kit (Agilent Technologies)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, Inc.)
- Dark Reader® transilluminator (Clare Chemical Research, Inc.)

Methods:

1) End repair

5 ng of dsDNA was mixed with 10 µl of 10X Phosphorylation buffer, 4 µl of 10 mM dNTP mix, 5 µl of T4 DNA polymerase, 1 µl of *E. coli* DNA Polymerase I, Large (Klenow) Fragment and 5µl T4 Polynucleotide Kinase in a total reaction volume of 100µl and incubated at 20°C for 30 min. The end repaired products were then isolated from the reaction using a Zymo-5 DNA column clean up protocol and eluted from the column in 32 µL nuclease-free purified water.

2) 3' A-tailing of PCR products.

The eluted, end repaired DNA was mixed with 5 µl 10X NEBuffer2, 10 µL 1 mM dATP and 3 µL Klenow Fragment (3'→5' exo-). The reaction was then incubated at 37°C for 30 minutes. The A-tailed DNA was then isolated from the reaction using a Zymo-5 DNA column clean up protocol and eluted from the column in 23 µL nuclease-free purified water.

- 3) **Ligation of 3'- and 5'- adapters**

23 μ l of eluted, A-tailed DNA was mixed with 25 μ l 2X Quick Ligation Reaction Buffer, 1 μ l PALM adaptors mix (15 μ M each) and 1 μ l Quick T4 DNA Ligase. The mixture was incubated at room temperature for 15 minutes. The DNA was then isolated from the reaction using a Zymo-5 DNA column clean up protocol and eluted from the column in 20 μ L nuclease-free purified water.
- 4) **Gel purification**

The desired PCR products were purified by loading onto 4% Agarose EX-gel, 20 μ L per lane. 2 μ L of TrackIT 50bp Ladder (+ 18 μ l of H₂O) was loaded into a single lane and used as a DNA size marker. The gel was electrophoresed for approximately 20 minutes. The gel was imaged using a Dark Reader® Transilluminator and the desired product (200-250 bp) was excised using a clean razor blade. Approximately 4 volumes of Zymo ADB buffer were added to the gel pieces and the mixture was incubated for 10 minutes at 37 °C while mixing several times during the period. PCR products were isolated from the mixture using a Zymo-5 DNA column clean up protocol. The desired PCR products were eluted from the column in 29 μ L nuclease-free, purified water.
- 5) **PCR Amplification.**

The following reagents were combined in a PCR tube: 29 μ l of size selected DNA, 10 μ L 5X Phusion HF buffer, 1 μ L PALM DNA PCR primer 1, 1 μ L PALM DNA PCR primer 2, 1.5 μ l dNTP mix, 7 μ L nuclease-free purified water and 0.5 μ l Phusion DNA Polymerase. The mixture was then placed in a thermocycler using the following program: 1 initial hold, 98°C, 30 seconds; cycled 15 times, 98°C, 10 seconds; 60°C, 30 seconds; and 72°C, 30 seconds; 1 final hold, 72°C, 5 minutes. After the PCR reaction was complete, the PCR products were purified using a Zymo-25 DNA column clean up protocol and eluted from the column in 38.5 μ L nuclease-free purified water.
- 6) **3' A-tailing of PCR products.**

The eluted PCR products were mixed with 5 μ l 10X Taq buffer, 5 μ L 10 mM dATP and 1.5 μ L Taq polymerase. The reaction was then incubated at 72°C for 30 minutes. The A-tailed PCR products were then isolated from the reaction using a Zymo-5 DNA column clean up protocol and eluted from the column in 21 μ L nuclease-free purified water.
- 7) **Ligation of PALM barcodes to PCR products.**

100 ng of the eluted PCR products were mixed with 2 μ L of annealed PALM barcode (50 μ M) in a final volume of 23 μ L and added to 25 μ L of 2X Rapid Ligation buffer with 2 μ L T4 DNA Ligase (Rapid) and incubated at 20°C for 30 minutes. 1 μ l of 0.5 M EDTA was added to stop the ligation reaction.
- 8) **AMPure XP magnetic bead isolation of DNA products.**

After the ligation reaction, 90 μ l of AMPure XP beads was added, mixed, and allowed to incubate at room temperature for 5 minutes. Using a Magnetic Particle Concentrator (MPC), the beads were pelleted against the wall of the tube. The tube will remain in the MPC during all subsequent wash steps. The supernatant was removed, and the beads were washed twice with 500 μ L of 70% ethanol (allowed to sit for 1 minute each time during wash). The AMPure XP beads were allowed to air dry for approximately 5 minutes. Then the tube was removed from the MPC and 20 μ L of TE was added and vortexed to re-suspend the beads. After a 2 minute wait step, the tube was placed back in the MPC to pellet the beads against the wall of the tube and the eluate was removed and transferred to a fresh microfuge tube.
- 9) **Agilent High Sensitivity DNA chip analysis and Illumina flow cell loading.**

1 μ l of the eluted DNA library was run on an Agilent High Sensitivity DNA chip to determine size and concentration of the desired barcoded products. The concentration of the desired products was determined from the Agilent 2100 Bioanalyzer electropherogram and used to prepare an equimolar pool of 9 pM solution for cluster generation on an Illumina flow cell.
- 10) **Final (optional) gel purification**

The equimolar pool was purified by loading onto 4% Agarose EX-gel, 20 μ L per lane. 5 μ L of TrackIT 50bp Ladder was loaded into a single lane and used as a DNA size marker. The gel was electrophoresed for approximately 20 minutes. The gel was imaged using a Dark Reader® Transilluminator and the desired product (300-350 bp) were excised using a clean razor blade. Approximately, 4 volumes of Zymo ADB buffer were added to the gel pieces and the mixture was incubated for 5-10 minutes at 37 °C while mixing several times during the period. PCR products were isolated from the mixture using a Zymo-5 DNA

column clean up protocol. The desired PCR products were eluted from the column in 30 μ L nuclease-free, purified water. This eluate was used

to prepare an equimolar pool of 9 pM solution for cluster generation on an Illumina flow cell.

References

Articles and books:

1. Sluka P, O'Donnell L, McLachlan RI, Stanton PG: **Application of laser-capture microdissection to analysis of gene expression in the testis.** *Progress in Histochemistry and Cytochemistry* 2008, **42**(4):173-201.
2. Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM *et al*: **The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin.** *Science* 2010, **327**(5963):328-331.
3. Liu L, Li YH, Li SL, Hu N, He YM, Pong R, Lin DN, Lu LH, Law M: **Comparison of Next-Generation Sequencing Systems.** *Journal of Biomedicine and Biotechnology* 2012, **2012**(Article ID 251364):1-11.
4. Shendure J, Ji HL: **Next-generation DNA sequencing.** *Nature Biotechnology* 2008, **26**(10):1135-1145.
5. Ronaghi M: **Pyrosequencing sheds light on DNA sequencing.** *Genome Research* 2001, **11**(1):3-11.
6. Minoche AE, Dohm JC, Himmelbauer H: **Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems.** *Genome Biol* 2011, **12**(11 R112):1-15.
7. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y: **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.** *Bmc Genomics* 2012, **13**.
8. Voelkerding KV, Dames SA, Durtschi JD: **Next-Generation Sequencing: From Basic Research to Diagnostics.** *Clinical Chemistry* 2009, **55**(4):641-658.
9. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M *et al*: **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature* 2011, **475**(7356):348-352.
10. Pareek CS, Smoczynski R, Tretyn A: **Sequencing technologies and genome sequencing.** *J Appl Genetics* 2011, **52**(4):413-435.
11. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: **Real-Time DNA Sequencing from Single Polymerase Molecules.** *Science* 2009, **323**(5910):133-138.
12. Van Nieuwerburgh F, Soetaert S, Podshivalova K, Ay-Lin Wang E, Schaffer L, Deforce D, Salomon DR, Head SR, Ordoukhanian P: **Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing.** *PLoS One* 2011, **6**(10 e26969):1-6.
13. Olsson ME, Olofsson LM, Lindahl AL, Lundgren A, Brodelius M, Brodelius PE: **Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L.** *Phytochemistry* 2009, **70**(9):1123-1128.
14. Domazet B, MacLennan GT, Lopez-Beltran A, Montironi R, Cheng L: **Laser Capture Microdissection in the Genomic and Proteomic Era: Targeting the**

- Genetic Basis of Cancer.** *International Journal of Clinical and Experimental Pathology* 2008, **1**(6):475-488.
15. Ladanyi A, Sipos F, Szoke D, Galamb O, Molnar B, Tulassay Z: **Laser microdissection in translational and clinical research.** *Cytom Part A* 2006, **69A**(9):947-960.
 16. Nelson T, Tausta SL, Gandotra N, Liu T: **Laser microdissection of plant tissue: What you see is what you get.** *Annual Review of Plant Biology* 2006, **57**:181-201.
 17. Vogelstein B, Gillespie D: **Preparative and analytical purification of DNA from agarose.** *Proceedings of the National Academy of Sciences of the United States of America* 1979, **76**(2):615-619.
 18. Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheimvandillen PME, Vandernoordaa J: **Rapid and simple method for purification of nucleic-acids.** *J Clin Microbiol* 1990, **28**(3):495-503.
 19. Bruns DE, Ashwood ERMD, Burtis CA: **Fundamentals of Molecular Diagnostics:** Elsevier Health Sciences ISBN 9781416037378; 2007.
 20. Basehore SLB, C. Jeffrey; Novoradovskaya, Natalia: **Methods for the separation of biological molecules using sulfolane.** *Patent No. WO2008115708A1*, September 25th, 2008.
 21. Liu CL, Schreiber SL, Bernstein BE: **Development and validation of a T7 based linear amplification for genomic DNA.** *Bmc Genomics* 2003, **4**(1):19.
 22. Olofsson L, Lundgren A, Brodelius PE: **Trichome isolation with and without fixation using laser microdissection and pressure catapulting followed by RNA amplification: Expression of genes of terpene metabolism in apical and sub-apical trichome cells of *Artemisia annua* L.** *Plant Science* 2012, **183**(2012):9-13.
 23. Lopez-Barragan MJ, Quinones M, Cui KR, Lemieux J, Zhao KJ, Su XZ: **Effect of PCR extension temperature on high-throughput sequencing.** *Molecular and Biochemical Parasitology* 2011, **176**(1):64-67.
 24. Schutze T, Rubelt F, Repkow J, Greiner N, Erdmann VA, Lehrach H, Konthur Z, Glokler J: **A streamlined protocol for emulsion polymerase chain reaction and subsequent purification.** *Analytical Biochemistry* 2011, **410**(1):155-157.
 25. Schutze T, Arndt PF, Menger M, Wochner A, Vingron M, Erdmann VA, Lehrach H, Kaps C, Glokler J: **A calibrated diversity assay for nucleic acid libraries using DiStRO-a Diversity Standard of Random Oligonucleotides.** *Nucleic Acids Research* 2010, **38**(4 e23):1-5.
 26. Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA *et al*: **Identification of genetic variants using bar-coded multiplexed sequencing.** *Nature Methods* 2008, **5**(10):887-893.
 27. Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ: **A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing.** *Nucleic Acids Research* 2007, **35**(19 e130):1-9.
 28. Rigola D, van Oeveren J, Janssen A, Bonne A, Schneiders H, van der Poel HJA, van Orsouw NJ, Hogers RCJ, de Both MTJ, van Eijk MJT: **High-Throughput Detection of Induced Mutations and Natural Variation Using KeyPoint (TM) Technology.** *PLoS One* 2009, **4**(3 e4761):1-9.
 29. Vigneault F, Sismour AM, Church GM: **Efficient microRNA capture and bar-coding via enzymatic oligonucleotide adenylation.** *Nature Methods* 2008, **5**(9):777-779.

30. Buermans HPJ, Ariyurek Y, van Ommen G, den Dunnen JT, t Hoen PAC: **New methods for next generation sequencing based microRNA expression profiling.** *Bmc Genomics* 2010, **11**(716):1-16.
31. Meyer SU, Pfaffl MW, Ulbrich SE: **Normalization strategies for microRNA profiling experiments: a 'normal' way to a hidden layer of complexity?** *Biotechnology Letters* 2010, **32**(12):1777-1788.
32. Linsen SEV, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE *et al*: **Limitations and possibilities of small RNA digital gene expression profiling.** *Nature Methods* 2009, **6**(7):474-476.
33. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Research* 2008, **36**(Special issue):D154-D158.
34. Tian G, Yin XY, Luo H, Xu XH, Bolund L, Zhang XQ: **Sequencing bias: comparison of different protocols of MicroRNA library construction.** *Bmc Biotechnology* 2010, **10**(64):1-9.
35. Nelson PT, Wang WX, Wilfred BR, Tang GL: **Technical variables in high-throughput miRNA expression profiling: Much work remains to be done.** *Biochimica Et Biophysica Acta-Genes Regulatory Mechanisms* 2008, **1779**(11):758-765.
36. Romaniuk E, McLaughlin LW, Neilson T, Romaniuk PJ: **The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction.** *European Journal of Biochemistry* 1982, **125**(3):639-643.

Part III. Trichome analysis



Chapter I Transcriptome analysis of glandular and filamentous trichomes

Adapted from

Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua*

Sandra SA Soetaert¹, Christophe MF Van Neste¹, Mado L Vandewoestyne¹, Steven R Head³, Alain Goossens², Filip CW Van Nieuwerburgh¹, Dieter LD Deforce^{1§}

¹Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent University, Harelbekestraat 72, 9000 Ghent, Belgium

²Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium

³Next Generation Sequencing Core, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La Jolla, California, 92037, United States of America

In Part II Chapter II and III, technical optimizations were made to discover the best experimental setup to collect and prepare transcriptome material for 2nd generation sequencing. In this Chapter, these optimizations were used to perform a transcriptome analysis of glandular and filamentous trichomes. The experiment had two goals. The principal aim was to discover candidate genes that might be involved in artemisinin biosynthesis. These candidates show a decreased expression in filamentous trichomes and an enhanced expression in glandular trichomes and belong to the groups of cytochrome P450, peroxidases or dioxygenases that might be involved in the formation of the endoperoxide bridge in artemisinin. Little is known about the production of secondary metabolites in filamentous trichomes. They are assumed to form just a physical barrier by steric hindrance of herbivores. Therefore, tracking genes with an enhanced expression in filamentous trichomes was the second goal.

Since Maes *et al.* [1] showed the upregulation of artemisinin biosynthesis genes after treatment with jasmonic-acid, some plants were treated with this jasmonic-acid in our experiment to make transcriptome differences for artemisinin candidate genes more pronounced between glandular and filamentous trichomes.

1 Introduction

Artemisia annua L. (Sweet Wormwood) is a medicinal plant that produces artemisinin which is a sesquiterpene with anti-malarial properties. Since every year around 216 million people are infected with malaria [2], a high supply of artemisinin is needed at a reduced cost. Artemisinin production is being enhanced in *A. annua* by crossing high-producing plants [3]. Another strategy to increase production of artemisinin is synthesis of artemisinic-acid in engineered yeast and subsequent (photo)chemical conversion to artemisinin [4, 5]. A better insight in artemisinin biosynthesis could lead to a cheaper production method.

An important breakthrough unravelling artemisinin biosynthesis, was the localization of artemisinin production in the glandular trichomes [6]. Trichomes, named after the Greek word for hair are epidermal outgrowths covering plant organs. In *A. annua*, two types of trichomes are present: biserial peltate glandular trichomes (Figure 1 A) and filamentous trichomes composed of stalk cells and an elongated cell in a T-shape (Figure 1 B) [7]. Duke *et al.* compared a normal biotype of *A. annua* with both filamentous and glandular trichomes to a biotype with only filamentous trichomes. Only in the presence of glandular trichomes, artemisinin was detected [6].

This information was used to discover candidate genes for artemisinin biosynthesis with an expressed sequence tag (EST) approach [8]. Three EST libraries were constructed: glandular trichome, flower bud and glandular trichome-minus-flower-bud subtracted library. Several genes that were preferentially expressed in glandular trichomes are involved in artemisinin biosynthesis such as *CYP71AV1*, *ALDH1*, *DBR2* and *ADH1* [8-11].

Enzymes involved in artemisinin biosynthesis are known up to the formation of dihydroartemisinic acid. It is not yet clear whether the last step(s) from dihydroartemisinic acid to artemisinin involves a spontaneous auto-oxidation or is catalyzed by enzymes. Brown and Sy favour the theory of spontaneous chemistry because they see parallels between *in vitro* auto-oxidation and intermediates present *in vivo* [12]. Additionally, plants fed with labelled dihydroartemisinic acid and dried, contained the same proportion of labelled artemisinin as plants that were kept alive [12]. These are the main arguments for chemical conversion. On the other hand, while 70% of label incorporation was detected in the metabolites derived from

dihydroartemisinic acid; artemisinin and arteannuin H had only 5-15% label incorporation [12]. Therefore, it cannot be excluded that there is another more important pathway leading to artemisinin which was not accessible by the labelled precursor and that enzymes might be involved to catalyze this process in *A. annua*.

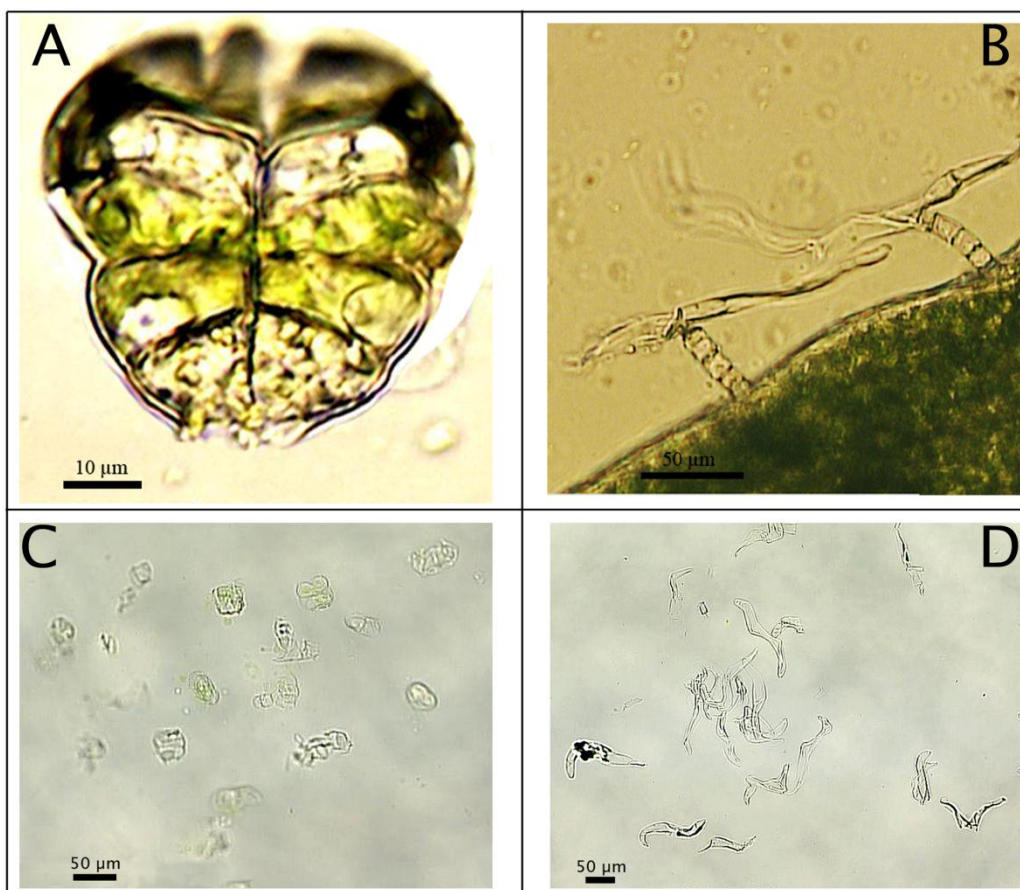


Figure 1: Glandular and filamentous trichomes of *A. annua*.

A: Glandular trichome with on top a pair of white apical cells and two pairs of green sub-apical cells surrounded by a sub-cuticular cavity (cells at the basis of the trichome are partially removed); B: Filamentous trichomes (T-shape) with stalk cells and elongated cell; C: Glandular trichomes captured by laser pressure catapulting in the cap of a tube; D: Filamentous trichomes captured by laser pressure catapulting in the cap of a tube.

To find candidate genes that catalyze the last step(s) to artemisinin, a more detailed analysis was needed of the transcriptome of glandular trichomes. Sequencing of the transcriptome from enriched glandular trichome preparations from *A. annua* was performed by Graham *et al.* on the Roche 454 platform to identify genes and markers for fast-track breeding [3]. In another study, Wang *et al.* performed a global transcriptome characterization of glandular

trichomes. To confirm the expression of some genes in glandular trichomes a semi-quantitative RT-PCR analysis was performed on filamentous and glandular trichomes. Three genes involved in terpene biosynthesis were tested: amorpha-4,11-diene synthase, a sesquiterpene cyclase and (3R)-linalool synthase. They found that these genes were expressed in glandular trichomes as well as in filamentous trichomes [13] and this raised the question whether filamentous trichomes are involved in the production of secondary metabolites.

Trichomes are in general classified as non-glandular (e.g. filamentous trichomes in *A. annua*) or glandular, based on their secretory capacity. In plants, glandular trichomes are most often production sites for multiple secondary metabolites which form a first-line defence at the surface of the plant through their capacity to entrap, deter or poison pathogens and herbivores [14]. Because of their interesting metabolite content as illustrated with artemisinin, a large number of studies have been devoted to glandular trichomes.

In contrast to the extensive literature describing glandular trichomes, less attention has been paid to non-glandular trichomes. Non-glandular trichomes are assumed to form a physical barrier by steric hindrance of herbivores [15, 16]. Non-glandular trichomes are mainly described for taxonomic and phylogenetic purposes [17-21] but little is known about their production of secondary metabolites [22].

To detect potential candidate genes for artemisinin biosynthesis and to investigate if filamentous trichomes produce important secondary metabolites, it is interesting to compare filamentous and glandular trichomes. Therefore, we performed a comparative transcriptome analysis of filamentous and glandular trichomes on the Illumina HiSeq platform. Several cytochromes, peroxidases and dioxygenases that are potentially involved in the biosynthesis of artemisinin and/or other terpenes were upregulated in glandular trichomes. Our transcriptome analysis confirms the established metabolic capacities of *A. annua* glandular trichomes but also points to specific metabolic activities in *A. annua* filamentous trichomes.

Additionally, two other transcriptome experiments were set up to discover potential candidate genes. First, the effect of jasmonic acid (JA) elicitation on glandular and filamentous trichomes was investigated as e.g. Maes *et al.* showed that artemisinin production can be stimulated by JA [1]. However, in our experimental setup, no influence of JA treatment was

detected. A possible explanation is that this is due to the use of plants in different developmental stages.

2 Methods

2.1 Overview of collected samples

In total 6 sample-types have been collected from capitula (flower heads) of *A. annua* (Table 1). Three independent repeats of glandular (Figure 1 C) and filamentous (Figure 1 D) trichomes were collected from the same mock- and JA-treated plants. In addition to this, two repeats of separated apical and sub-apical cells of glandular trichomes were collected from mock-treated plants. The RNA from one of the 2 repeats for each cell type was split in two and separately amplified, creating 3 repeats for sequencing. For the apical and sub-apical experiment, see Part III, Chapter II.

Sample-type	Apical cells		Sub-apical cells		Glandular trichomes		Filamentous trichomes	
	mock		mock		mock	JA	mock	JA
RNA extract 1	Unfixated		Unfixated		Unfixated Rep1	Unfixated Rep1	Unfixated Rep1	Unfixated Rep1
Amplification	ampl 1	ampl 2	ampl 1	ampl 2	ampl	ampl	ampl	ampl
RNA extract 2	Fixated		Fixated		Unfixated Rep2	Unfixated Rep2	Unfixated Rep2	Unfixated Rep2
Amplification	ampl		ampl		ampl	ampl	ampl	ampl
RNA extract 3					Unfixated Rep3	Unfixated Rep3	Unfixated Rep3	Unfixated Rep3
Amplification					ampl	ampl	ampl	ampl

Table 1: Overview of collected samples.

Table with an overview of collected samples from *A. annua*. Unfixated glandular and filamentous trichomes are collected from mock- and JA-treated samples in 3 repeats (Rep). For apical and sub-apical cells, one sample of unfixated and one sample of fixated cells are collected. The unfixated sample was amplified in 2 separate amplifications (ampl).

2.2 Plant preparation

Experiments were executed on *Artemisia annua* L. Anamed A3 cultivar (<http://www.anamed.net>). This cultivar contains up to 1,4% artemisinin (dry weight leaves) and is the result of cross breeding high-artemisinin producing plants by Mediplant Inc. (Conthey, Switzerland) [23]. Anamed A3 can grow well in tropical regions and is not as photosensitive as other breeds [24]. Under 8 hours day and 16 hours night, Anamed A3 starts flowering at the earliest after 6 months [see Additional File 1]. The length of this pre-flowering period is in line with observations under field conditions [25].

Seeds from *A. annua* Anamed were sterilized for 2 min in 70% EtOH (Merck, Darmstadt, Germany) and 10 min in a solution with 3.84 ml NaOCl (10-13% chlorine from Sigma-Aldrich, Steinheim, Germany), 5 µl Tween 20 (MP Biomedicals, Illkirch, France) and 6.16 ml sterile water. Subsequently, seeds were rinsed with sterile water and germinated on moist paper. After 1 or 2 weeks, shoots were transferred to soil and grown under a regime with 8 hours day, 16 hours night and a temperature of 20°C. After 6 months, flowers were appearing. For JA elicitation, plants were treated for 8 days before the start of the sampling procedure by spraying every 2 days with a solution of 100 µM JA (Duchefa, Haarlem, The Netherlands) containing 1.5 mM Tween20 (MP Biomedicals) and adding every 2 days 5 ml of 100 µM JA to the soil. During the sampling procedure, treatment was continued every 2 days. Control groups were treated with a mock (water) solution in a separate room.

2.3 Glandular and filamentous sample preparation

On the basal bracts and pedicel of the capitulum, filamentous trichomes are abundantly present [7]. Glandular trichomes are protruding on the corolla of the floret buds from the capitulum but are sunken in the capitulum bracts and in leaves [7]. An image of a flower head from *A. annua* Anamed was taken with Tabletop SEM (TM-1000, Hitachi, Tokyo, Japan) [see Additional File 2]. Sunken trichomes are difficult to collect with laser capture microdissection and therefore, capitula were used to collect glandular and filamentous trichomes. Trichomes were collected from mock and JA-treated plants in 3 independent biological repeats. For each repeat, trichomes were collected from a pool of 3 plants. The same plants were used for capturing 190 glandular and 670 filamentous trichomes.

On a RNA free microscope slide, a capitulum was cut in a drop of cold buffer with 25 mM MOPSO (pH6.3) (Sigma-Aldrich), 200 mM sorbitol (Alfa Aesar, Karlsruhe, Germany), 10 mM sucrose (Acros, Geel, Belgium), 5 mM thiourea (Sigma-Aldrich), 2 mM DTT (Sigma-Aldrich), 5 mM MgCl₂ (Sigma-Aldrich) and 0.5 mM sodium-phosphate (Acros) [26]. Trichomes were captured, using the Palm MicroBeam system (P.A.L.M Microlaser Technologies, München, Germany) with a nitrogen UV-A laser (wavelength 337 nm). The laser-beam was focussed on the tissue for laser microdissection, to separate trichomes or cells from the surrounding tissue. Free trichomes were thereafter captured with laser pressure catapulting by focussing the laser beam just below the tissue. Samples were collected in 30 µl lysis buffer with β-mercaptoethanol (Absolutely RNA Nanoprep Kit Stratagene, La Jolla, CA). An image of collected glandular and filamentous trichomes is shown in Figure 1 C and D. RNA was extracted with the Absolutely RNA Nanoprep protocol with DNase treatment. Samples were eluted in 10µl and half was used as starting material for RNA amplification.

2.4 RNA amplification and sequencing

Samples were amplified with a linear amplification system: the Ovation RNA-Seq System with 1 h 30 min Spia-amplification (NuGen, AC Bemmell, The Netherlands). In this amplification procedure, random primers and oligo dT primers are used during RNA amplification and consequently the 5' end of the mRNA is better amplified compared to when only oligo dT primers are used. The cDNA priming reaction used in the NuGEN amplification kits are designed to avoid amplification of rRNA sequences. After RNA amplification, barcoded Illumina sequencing libraries were made using a post amplification ligation-mediated strategy [27]. The 18 samples were sequenced with a read-length of 100 bp in 3 lanes of an Illumina HiSeq 2000 flowcell.

2.5 Quantitative real time (qRT)-PCR

Nugen amplified DNA from 3 independent mock-treated glandular and filamentous trichome samples was analyzed with qRT-PCR. As template, 2 ng DNA was used in 10 µl reactions containing 5µl iTaq SYBR Green Supermix with ROX (Bio-Rad, Watford, UK) and 400 nM primers. The qRT-PCR experiment was performed on a Light Cycler 480 (Roche) with hotstart at 95 °C for 2 min. and 42 cycles 95 °C (15 sec.), 52 °C (1 min.), including melting curve analysis. Each qRT-PCR reaction was executed in duplo and these technical repeats

were averaged prior to qbasePLUS (version 2.3) analysis with normalization to input DNA concentration [28]. To validate this normalization strategy, 3 genes were included that are expected to be similarly expressed in both trichome types: Actin 2 (homologue *AT3G18780*), protein phosphatase 2A subunit A3 (*PP2AA3*; homologue *AT1G13320*) and a pentatricopeptide repeat (PPR) superfamily protein (homologue *AT5G55840*). These genes were selected based on *Arabidopsis* data from Czechowski *et al.* [29] and expression of homologous transcripts in our RNASeq experiment. Other transcripts analyzed were artemisinin-synthesis and triterpene-synthesis related. Primers [see Additional File 3] were adopted from other manuscripts [1, 30] or designed with Primer-BLAST from NCBI [31].

2.6 Trinity and Blast2GO

A *de novo* transcriptome was assembled from all 18 samples. Trinity Release-2011-07-13 was used to perform an ALLPATHS error correction [32] on the reads prior to the *de novo* assembly which was made with Trinity Release-2011-08-20 [33]. The 150,288 Trinity contigs were annotated using BLAST and the NCBI non-redundant protein database. The best 5 BLAST hits were used to indicate a putative function. The Blast2GO suite [34] was used to generate gene ontology terms based on the BLAST output. Two sets of settings were sequentially used: strict and more permissive. In the more strict settings, a BLAST high-scoring segment pairs (hsp) length of 33 and a minimum hsp coverage of the query of 33 was required. The more permissive setting allowed for a shorter BLAST hsp length of 20 with no minimum hsp coverage of the query.

2.7 Bowtie and RSEM

Bowtie [35] and RSEM (RNASeq by Expectation Maximization) [36] were used for mapping the 289 million reads to the Trinity *de novo* assembly and counting the number of reads that matched to each contig. Standard options were used, but RSEM's polyA tail option was disabled.

2.8 edgeR

To perform the differential expression analysis, an R script was developed that makes use of the Bioconductor edgeR package [37]. All glandular and filamentous samples were normalized together. Normalization was performed by trimmed mean of M values (TMM).

TMM equates the overall expression levels of genes between samples under the assumption that the majority of them are not differentially expressed [38]. A p-value < 0.05 , adjusted for multiple testing, was used to determine which contigs could be differentially expressed. The edgeR parameter `prior.n` was set to 1.

2.9 MapMan

MapMan is a program to visualize various pathways and to indicate genes or contigs that are up or down-regulated [39]. Prior to MapMan, functional bins need to be assigned to contigs of the *de novo* assembly. To assign the bins, Mercator was used (default options, `Blast_cutoff: 50` and `IS_DNA`). Not all artemisinin biosynthesis genes were automatically assigned to a bin. For those genes, additional bins were added based on the definitions of the best 5 blast hits. The output of edgeR was visualized with MapMan to show which metabolic pathways are significantly differentially expressed.

3 Results and discussion

3.1 *De novo* assembly of the transcriptome

All 18 samples were sequenced in 3 lanes of an Illumina HiSeq 2000 flowcell. For each sample, on average 16 million single-end reads of 100 bp were generated. Around 60% of the reads had hits with a combined LSU and SSU rRNA reference database [40] and tRNA database with all known tRNAs [41]. The total of 289 million reads generated were processed using Trinity [33] to assemble a *de novo* transcriptome [see Additional File 4] containing 150,288 contigs in 108,400 homologous groups with an average contig-length of 412 bp and a minimum and maximum length of 201 and 7775 bp, respectively (Table 2).

To have a better estimate of the quality of the Trinity contigs, all 88 full-length *A. annua* mRNA sequences available in the NCBI non-redundant protein database were compared with the Trinity contigs. Of those, only 2 had no BLAST hits to the Trinity contigs. Each Trinity contig that showed a hit with one of the NCBI sequences, covered on average 58% of the length of the NCBI sequence. A combination of contigs with the same BLAST-hit could together cover on average 84% of the NCBI sequence. If the set of NCBI sequences is assumed to be representational of the real transcripts, one can expect that more than 80% of the length of a random transcript is present in our assembly.

Contig length (bp)	Number of contigs
< 500	118556
500-1000	24993
1000-1500	4658
1500-2000	1299
2000-3000	643
> 3000	133

Table 2: Length distribution of Trinity contigs.

3.2 Annotation and characterization of *de novo* transcripts

Eighty-four% of the 150,288 contigs showed at least one BLAST hit with the NCBI non-redundant protein database. The first five BLAST hits served as an indication for the putative function of a contig. The contigs were also annotated with the Blast2GO suite [34], which links BLAST hits to gene ontology (GO) terms: 46,711 contigs had a good connection to GO terms, using strict settings for Blast2GO; 5,596 contigs still had a connection to GO terms with more permissive settings of Blast2GO; the remaining 73,389 contigs with BLAST hits did not point to any GO terms. As *A. annua* is not a model organism and does not belong to a family of a model organism it is logical that a lot of contigs with BLAST hits were not annotated with GO terms. It is mostly for genes that show strong homologies with genes of model organism that one can expect a successful annotation using this strategy. Moreover, a gene that performs a different function but has a homology with a gene of a model organism, can be annotated wrongly.

To further characterize our *de novo* Illumina transcripts, they were also compared to the 454 glandular trichome *A. annua* contigs of Wang *et al.* [13]. Of their 42,678 contigs, 79% had a BLAST hit within our Trinity contigs. Vice versa, only 20% of our contigs had a BLAST hit within their contig set. This was not due to contamination of rRNA and tRNA reads that could have resulted from our method of mRNA amplification since only 0.6% of all our contigs showed a hit with rRNA or tRNA. Therefore the fact that the majority of our contigs are differential from those of Wang *et al.* is most likely the result from a greater coverage.

3.3 Influence of JA treatment

The transcriptome of mock and JA-treated trichome samples was compared and no significant differences (adjusted p-value < 0.05) were detected. Accordingly, at the metabolite level, equal amounts of artemisinin, arteannuin B, dihydroartemisinic acid and artemisinic acid were measured with HPLC-MS/MS as shown in additional data [see Additional File 5 and Additional File 6]. This indicates that the JA-treatment did not have a major influence in our experimental set-up. This was not expected since Maes *et al.* measured higher artemisinin levels after JA treatment [1]. A possible explanation is the use of plants in different developmental stages: Maes *et al.* used young seedlings while in our RNASeq experiment 6-months old plants with closed capitula were used. It has been shown that in flowers of *Arabidopsis thaliana*, JA levels increase 6.7-fold just before flower bud opening [42, 43]. Therefore it is plausible that endogenous JA-signalling already reached a maximum effect in our samples and that exogenous JA treatment did not trigger an additional response.

3.4 Glandular versus filamentous trichomes

Transcript levels from glandular trichomes and filamentous trichomes were compared to obtain a list of significantly differentially expressed contigs (adjusted p-value < 0.05). To make a robust statistical comparison, 3 samples of glandular trichomes with JA and 3 samples with mock-treatment were compared to the 6 samples (mock and JA) of filamentous trichomes. Of 150,288 contigs; 631 were significantly differentially expressed and all these contigs are listed in additional data [see Additional File 7]. From these, 204 contigs were more expressed in filamentous trichomes whereas, 427 contigs were more expressed in glandular trichomes. An overview with contigs discussed in this article and their normalized counts for each sample, log₂-fold changes and adjusted p-values are shown in additional data [see Additional File 8].

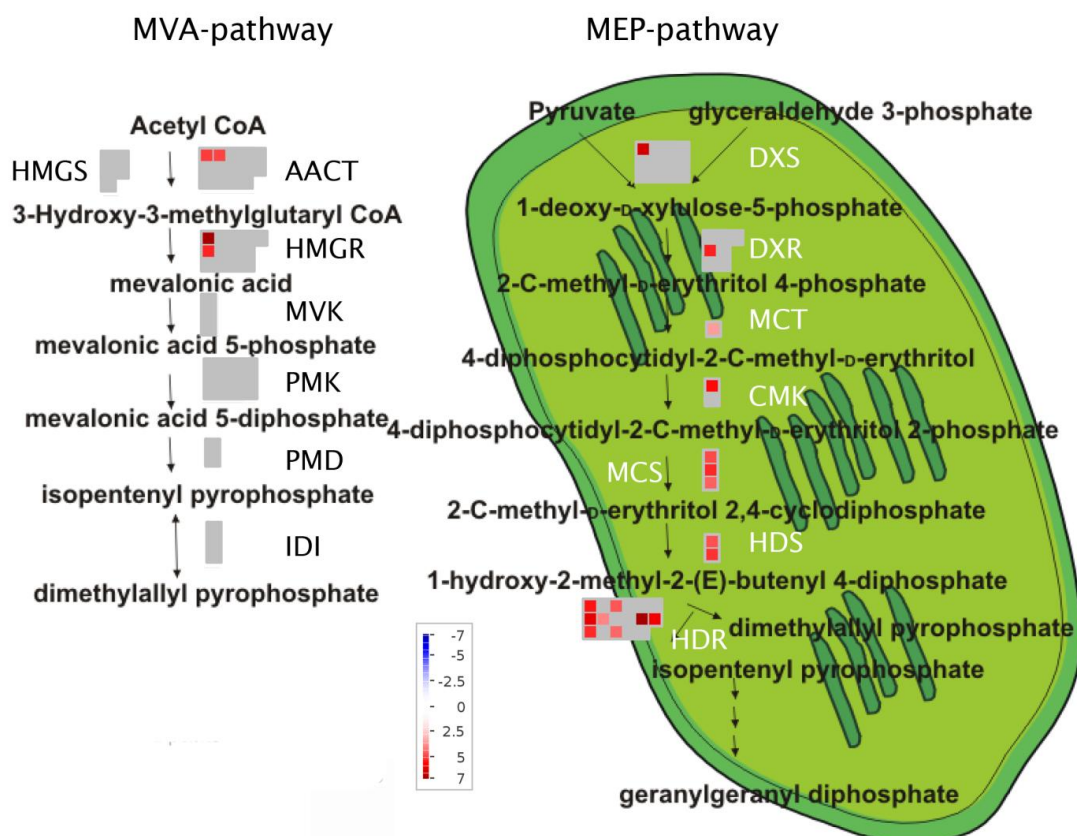


Figure 2: Differentially expressed MVA and MEP-pathway genes in glandular and filamentous trichomes.

MapMan figure (adapted from [44]) with comparison of MVA and MEP pathways in glandular and filamentous trichomes. Significantly more expressed contigs in glandular trichomes are shown in red with colour scale to indicate the \log_2 -fold changes. No contigs were significantly more expressed in filamentous trichomes. Grey represents contigs that were not significantly differentially expressed. **MVA pathway:** acetyl-CoA C-acetyltransferase (AACT), 3-hydroxy-3-methyl-glutaryl coenzyme A synthase (HMGS), 3-hydroxy-3-methyl-glutaryl coenzyme A reductase (HMGR), mevalonate kinase (MVK), phosphomevalonate kinase (PMK), diphosphomevalonate decarboxylase (PMD), isopentenyl diphosphate isomerase (IDI). **MEP pathway:** 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase (MCT), 4-cytidine 5'-diphospho-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS), hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (HDS) and hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR).

3.5 MVA and MEP pathway in glandular and filamentous trichomes

From the upregulated contigs, 5% in glandular and 0% in filamentous trichomes were involved in the mevalonate (MVA) or 2-C-methyl-D-erythritol 4-phosphate (MEP) pathways.

The MEP and MVA pathways produce isopentenyl diphosphate and its isomer dimethylallyl diphosphate, which are precursors for the production of terpenes.

MEP and MVA pathways were detected in glandular and filamentous trichomes. All transcripts coding for enzymes of the MEP pathway were significantly upregulated in glandular trichomes (Figure 2). From the MVA pathway, only *acetyl-CoA C-acetyltransferase (AACT)* and *3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR)* were significantly upregulated in glandular trichomes (Figure 2). Up-regulation of *HMGR* in glandular trichomes is important since it is shown that HMGR activity limits artemisinin biosynthesis [45, 46].

3.6 Artemisinin biosynthesis in glandular and filamentous trichomes

Terpene synthesis genes were accounting for 6% of the upregulated contigs in glandular trichomes and only 1% in filamentous trichomes. Starting from the MVA and MEP pathway, farnesyl diphosphate is synthesised by farnesyl diphosphate synthase (FDS) whose transcripts were not significantly upregulated in glandular trichomes (Figure 3 and Additional File 8). Subsequently, farnesyl diphosphate is converted to amorpha-4,11-diene which is the starting product for artemisinin biosynthesis. This reaction is catalyzed by amorpha-4,11-diene synthase (ADS) [47] and transcripts coding for this enzyme were detected more in glandular trichomes. By CYP71AV1, amorpha-4,11-diene is converted to artemisinic alcohol [4, 8]. Thereafter, artemisinic alcohol is oxidized by CYP71AV1 [4, 8] and alcohol dehydrogenase 1 (ADH1) [11] to artemisinic aldehyde. Artemisinic aldehyde is further oxidized by aldehyde dehydrogenase 1 (ALDH1) [9] and CYP71AV1 [4, 8] to artemisinic acid or reduced by artemisinic aldehyde Δ 11(13) double bond reductase (DBR2) [10] to form dihydroartemisinic aldehyde. A broad substrate oxidoreductase (RED1) can convert dihydroartemisinic aldehyde to dihydroartemisinic alcohol [48]. This reaction competes with ALDH1 using dihydroartemisinic aldehyde to form dihydroartemisinic acid [9]. Dihydroartemisinic acid is considered to be the precursor leading to artemisinin [10, 49, 50]. Transcripts corresponding to all these enzymes involved in the conversion of amorpha-4,11-diene to dihydroartemisinic acid were significantly upregulated in glandular trichomes, except for *RED1* (Figure 3).

These results confirm previous data pointing to glandular trichomes as the major artemisinin production site. A short dip in chloroform causes the collapse of the sub-cuticular space of

glandular trichomes and extracts almost all artemisinin [6, 51]. In addition to this, no artemisinin has been detected in a biotype of *A. annua* with only filamentous trichomes [6]. Nevertheless, in all samples of filamentous trichomes, transcripts corresponding to known artemisinin biosynthesis genes were detected, albeit at very low levels [see Additional File 8]. This is in agreement with previous reports, in which expression of *ADS* was also detected by RT-PCR in filamentous trichomes [13]. It should be noted though that in contrast, no staining of filamentous trichomes was observed in a promoter-GUS fusion study with the *ADS* promoter [52].

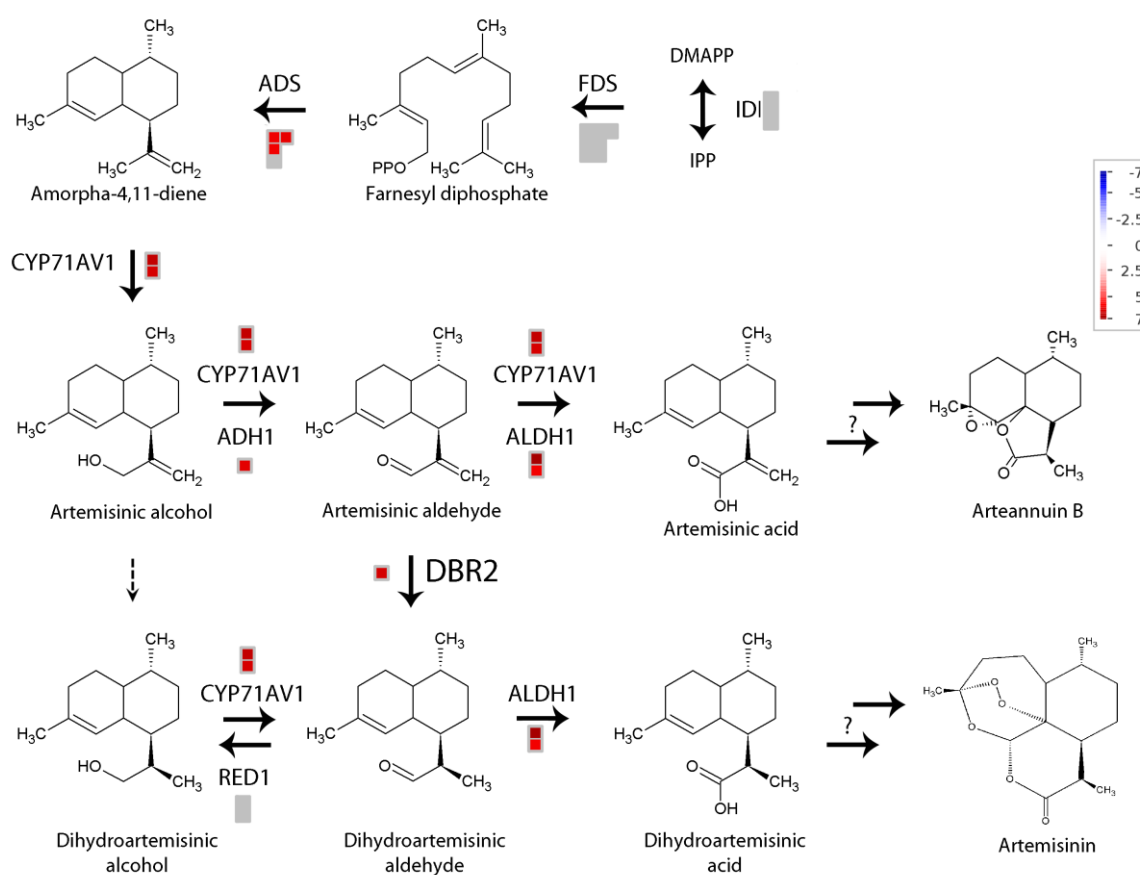


Figure 3: Differentially expressed artemisinin biosynthesis genes in glandular and filamentous trichomes. MapMan file with comparison of artemisinin biosynthesis pathways in glandular and filamentous trichomes. Significantly more expressed contigs in glandular trichomes are shown in red with colour scale to indicate the log₂-fold change. There were no contigs with higher expression in filamentous trichomes. Grey represents contigs that were not significantly differentially expressed.

3.7 Candidate genes for artemisinin biosynthesis

Contigs that are upregulated in glandular trichomes are possibly linked to artemisinin biosynthesis genes. The subset of these contigs which were annotated as cytochromes, peroxidases and dioxygenases might be possible candidates for the endoperoxide ring formation in artemisinin. An overview of these contigs is given in Table 3, more results per repeat are given in additional data [see Additional File 8]

Comp69 was detected 6 log₂-fold more in glandular compared to filamentous trichomes. The best BLAST-hit was with CYP81B1 [GenBank: CAA04117] from *Helianthus tuberosus*. This enzyme is functionally characterized to hydroxylate medium-chain saturated fatty acids [53]. In addition to this, CYP81B1 can be assumed to epoxidize these fatty acids [53, 54]. In glandular trichomes, comp548_c0_seq1 and comp548_c0_seq2 were detected 5 log₂-fold more than in filamentous trichomes, these contigs were annotated as P450 mono-oxygenase and showed homology with predicted sequences from the CYP82 family and CYP82C9v3 [GenBank: XP_002327091] from *Populus trichocarpa*. This CYP82 family is relatively uncharacterized [55]. A characterized enzyme from this family is CYP82G1 [GenBank: Q9LSF8] from *Arabidopsis thaliana* which is involved in homoterpene biosynthesis in which epoxidation might occur [56, 57]. But this characterized protein was not present in the list of best-BLAST hits with comp548. Another cytochrome P450 significantly more expressed in glandular trichomes was comp2774. The full length of this transcript was determined by Misra *et al.* and called CIM_CYP03 (CYP72A) [GenBank: GU318227] [58]. The activity of this enzyme was tested *in vitro* with artemisinic acid, dihydroartemisinic acid, arteannuin B and artemisinin as substrates but no activity was detected [58]. Other contigs significantly more expressed in glandular trichomes were comp15043_c0_seq2 annotated as CYP76B1 and comp3673 as cytochrome P450. Comp586_c0_seq3 and comp586_c0_seq4 were annotated as a cytochrome c-type. Some other cytochromes were significantly more expressed in filamentous trichomes as shown in additional data [see Additional File 8].

Comp252 (seq1, seq3, seq4 and seq5), 2084 and 6217_c0_seq4_1 are peroxidases detected respectively 6 log₂-fold, 4 log₂-fold and 30 log₂-fold more in glandular trichomes. Comp2084 was more specifically annotated as peroxidase 49 precursor and comp6217 was annotated as glutathione peroxidase. In filamentous trichomes, two contigs annotated as peroxidases were significantly more expressed than in glandular trichomes. These contigs are comp3274

annotated as peroxidase 1 from *A. annua* and comp16324 as alkaline leaf peroxidase from *Cyanara cardunculus*.

Annotation	Contig	Sum normalized counts of filamentous trichomes	Sum normalized counts of glandular trichomes	log ₂ Fold change	Adjusted p-value
Cytochromes					
CYP81	comp69 c0 seq1 1	1145	93262	5.9E+00	8.2E-05
CYP82	comp548 c0 seq1 1	91	5887	5.2E+00	3.5E-03
	comp548 c0 seq2 1	116	5247	4.8E+00	5.0E-02
	sum	207	11134		
CYP76B1	comp15043 c0 seq2	0	269	3.2E+01	4.6E-02
CYP72A	comp2774 c0 seq1 1	14	1797	6.2E+00	9.1E-03
CYP450	comp3673 c0 seq1 1	11	1656	6.4E+00	2.4E-02
cytochrome c-type	comp586 c0 seq1 1	27	3615	6.4E+00	1.3E-03
	comp586 c0 seq3 1	21	4979	7.4E+00	3.5E-05
	comp586 c0 seq4 1	0	157	3.1E+01	2.5E-02
	sum	48	8752		
Peroxidases					
Peroxidase	comp252 c0 seq1 1	79	6386	5.5E+00	9.1E-03
	comp252 c0 seq3 1	20	1411	5.5E+00	1.3E-02
	comp252 c0 seq4 1	73	6806	6.3E+00	4.5E-03
	comp252 c0 seq5 1	8	922	6.8E+00	1.4E-02
	sum	180	15525		
Peroxidase 49	comp2084 c0 seq1 1	134	3368	4.3E+00	3.0E-02
Gluthatione peroxidase	comp6217 c0 seq4 1	0	99	3.0E+01	4.5E-02
Dioxygenases					
Flavanone	comp225 c0 seq1 1	1352	26668	3.8E+00	1.0E-02
Flavanone	comp453 c0 seq1 1	657	20961	4.5E+00	1.3E-02

Table 3: Potential candidate genes for artemisinin biosynthesis.

Potential candidate cytochromes, peroxidases and dioxygenases with a significantly higher expression level in glandular trichomes. The normalized counts of 6 samples with filamentous trichomes are summed and compared with 6 glandular trichome samples. Log₂-fold changes are given as well as adjusted p-values.

Two dioxygenases were significantly more expressed in glandular trichomes: comp225_c0_seq1 and comp453 were both annotated as naringenin 2-oxoglutarate 3-dioxygenase (flavanone 3-hydroxylase).

3.8 Other terpene synthases in glandular and filamentous trichomes

Several contigs corresponding to enzymes involved in the biosynthesis of other sesquiterpenoids were significantly upregulated in glandular trichomes as well. These contigs are listed in the supplementary data [see Additional File 8]. Despite a higher *germacrene-A synthase* expression in glandular trichomes, germacrene-A was only detected in biotypes without glandular trichomes [59]. This can be the effect of up-regulation of germacrene-A oxidase which further oxidizes germacrene-A [60].

In filamentous trichomes, contig comp2645 corresponding to 8-*epi*-cedrol synthase, was significantly upregulated [see Additional File 8]. Low expression of 8-*epi*-cedrol synthase in glandular trichomes has been observed by qRT-PCR [26]. *In vitro*, recombinant 8-*epi*-cedrol synthase converts farnesyl diphosphate to 8-*epi*-cedrol, cedrol and minor amounts of α -cedrene and (E)- β -farnesene [61, 62]. Since a higher amount of 8-*epi*-cedrol synthase was detected in filamentous trichomes, these trichomes might synthesize the majority of these metabolites. Differences in (E)- β -farnesene concentration in glanded and glandless biotypes were estimated from Tellez *et al.* by correlating relative peak area to the oil content in fresh plant material [59]. Based on these estimates, the level of (E)- β -farnesene is approximately 1.8 times higher in glandless biotypes. This might correlate with the upregulated expression of 8-*epi*-cedrol synthase in filamentous trichomes. From α -cedrene, only trace amounts were measured in glanded and glandless biotypes [59] and the major product cedrol epimers were even not detected in extracts of *A. annua* [61, 62].

Regarding diterpenoid biosynthesis, only contigs annotated as *momilactone A synthase* were significantly more expressed in glandular trichomes [see Additional File 8]. Concerning monoterpenoid biosynthesis, many monoterpenoid synthases were significantly upregulated in glandular trichomes [see Additional File 8]. This is corroborated when comparing isoprenoid contents in glanded and glandless *A. annua* [59]. In oil from glanded biotypes, monoterpenes were predominant whereas in oil from glandless biotypes monoterpenes were almost absent.

β -amyrin synthase, an enzyme that converts 2,3-oxidosqualene to the triterpene saponin β -amyrin, has been characterized in *A. annua* [63]. This enzyme was represented in the *de novo* assembly by comp33386, comp59983, comp96251 and comp23239 and these contigs were not significantly differentially expressed. Contig comp7642_c0_seq2_1 shows homology with both

dammarenediol synthase and (β -)amyrin synthase and is a yet uncharacterized oxidosqualene cyclase (U_OSC). This contig was detected significantly more in filamentous trichomes [see Additional File 8].

3.9 Lipid biosynthesis in glandular and filamentous trichomes

In glandular trichomes, 15% of the significantly upregulated contigs were annotated to lipid biosynthesis. Transcripts and their corresponding significantly differentially expressed contigs are listed in supplementary data [see Additional File 8]. Acetyl-CoA carboxylase converts acetyl-CoA to malonyl-CoA and was significantly upregulated in glandular trichomes as shown in Figure 4. Subsequently, malonyl CoA-acyl carrier protein transacylase converts malonyl-CoA to malonyl-ACP [64]. This transcript was not significantly upregulated. Fatty acid biosynthesis is initiated by the condensation of malonyl-ACP with acetyl-CoA by β -ketoacyl-ACP synthase III (KAS). β -ketoacyl-ACP is reduced by β -ketoacyl-ACP reductase, dehydrated by β -hydroxyacyl-ACP dehydratase and reduced by enoyl-ACP reductase to yield butyryl-ACP. The latter transcripts except enoyl-ACP reductase were significantly more expressed in glandular trichomes as shown in Figure 4. The acyl-ACP end product has two carbons more than the original acetyl molecule [64]. Similar elongation cycles are continued with condensation of malonyl-ACP and acyl-ACP and the removal of the β -ketogroup. Three types of KAS were present with different acyl chain length specificities: KASIII (C_2 to C_4), KASI (C_4 to C_{16}) and KASII (C_{16} to C_{18}) [64]. All 3 types of KAS were upregulated in glandular trichomes.

Further extension of C_{16} and C_{18} to longer fatty acids requires their liberation from ACP by acyl-ACP thioesterase. Subsequently, fatty acids are exported out of the plastid to the endoplasmic reticulum [64]. The extension of fatty acids from long (C_{16} , C_{18}) to very long chains is catalyzed by β -ketoacyl-CoA synthase, β -ketoacyl-CoA reductase, β -hydroxyacyl-CoA dehydratase and enoyl-CoA reductase [65]. The rate-limiting step and specificity is determined by the β -ketoacyl-CoA synthase which was significantly more expressed in glandular trichomes [see Additional File 8].

Fatty acyl-CoA reductase 1 (TFARI) [1] was significantly more expressed in glandular trichomes. The encoded enzyme catalyzes the formation from acyl-CoA to fatty alcohols and is potentially involved in wax formation [1]. For the formation of unsaturated fatty acids,

omega-3 fatty acid desaturase is significantly upregulated in glandular trichomes. Contigs annotated as *cyclopropane-fatty-acyl phospholipid synthase* were highly expressed in glandular trichomes. This enzyme is forming a cyclopropane ring in unsaturated fatty acyl chains [66, 67]. Some contigs coding for lipid transfer proteins were significantly upregulated in glandular trichomes whereas other contigs coding for lipid transfer proteins were upregulated in filamentous trichomes [see Additional File 8].

The observed upregulation of lipid biosynthesis in glandular trichomes is in agreement with the results obtained by Tellez *et al.* [59], who measured that in glanded leaves 0.24% of fresh weight is oil and in glandless leaves only 0.06%.

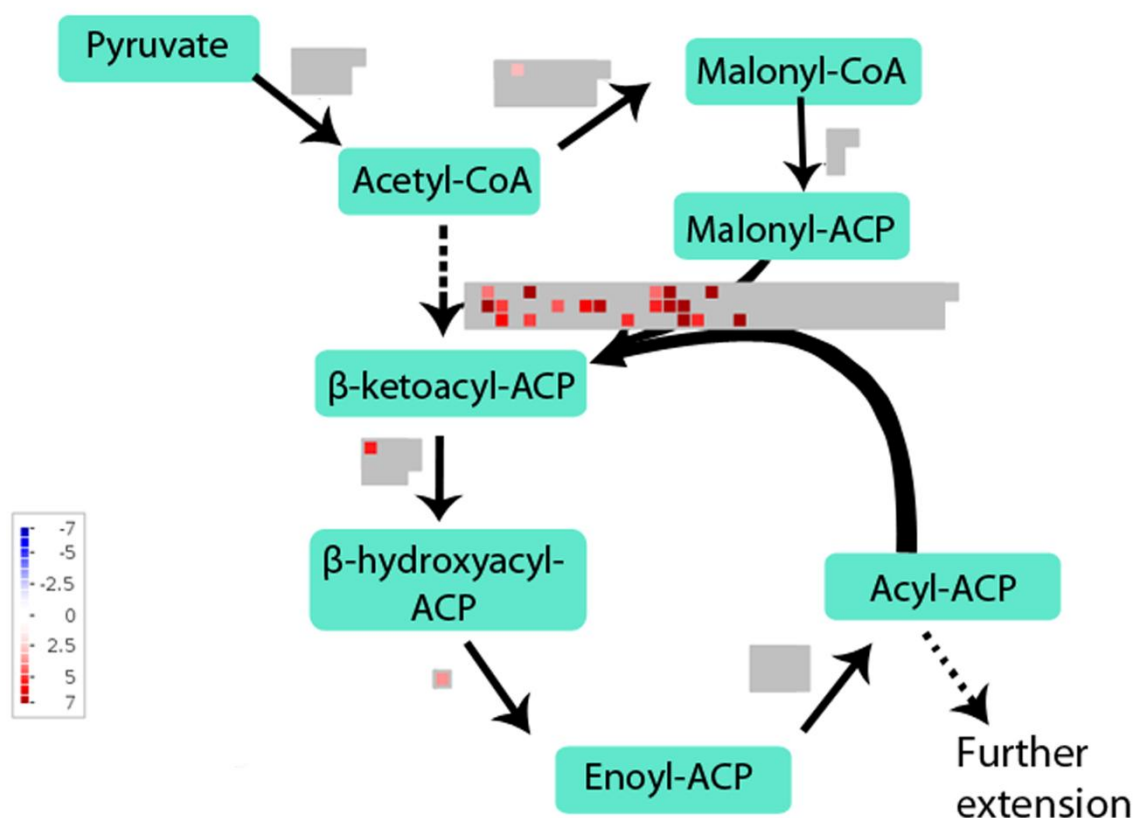


Figure 4: Differentially expressed lipid biosynthesis genes in glandular and filamentous trichomes.

MapMan file with comparison of lipid biosynthesis pathways in glandular and filamentous trichomes. Significantly more expressed contigs in glandular trichomes are shown in red with colour scale to indicate the log₂-fold change. There were no contigs with higher expression in filamentous trichomes. Grey represents contigs that were not significantly differentially expressed.

3.10 qRT-PCR

Enrichment for artemisinin-synthesis and triterpene-synthesis related transcripts in glandular and filamentous trichomes, respectively, was verified by qRT-PCR [see Additional File 9]. In mock-treated filamentous and glandular trichomes, similar expression levels were detected for *Actin2*, *PP2AA3* and *PPR protein* in both RNASeq and qRT-PCR. Similarly, with both techniques, no clear difference in *FDS* was observed between both trichome types. With qRT-PCR, the mean expression levels of *CYP71AV1*, *DBR2*, *Aldh1*, *Fatty acyl-CoA reductase 1 (TFAR1)* and 10 candidate genes for artemisinin biosynthesis were higher in glandular trichomes whereas the transcripts corresponding to uncharacterized oxidosqualene cyclase (U_OSC) and 8-epi-cedrol synthase were more abundant in filamentous trichomes. Hence, this qRT-PCR experiment confirmed the RNASeq data.

4 Conclusions

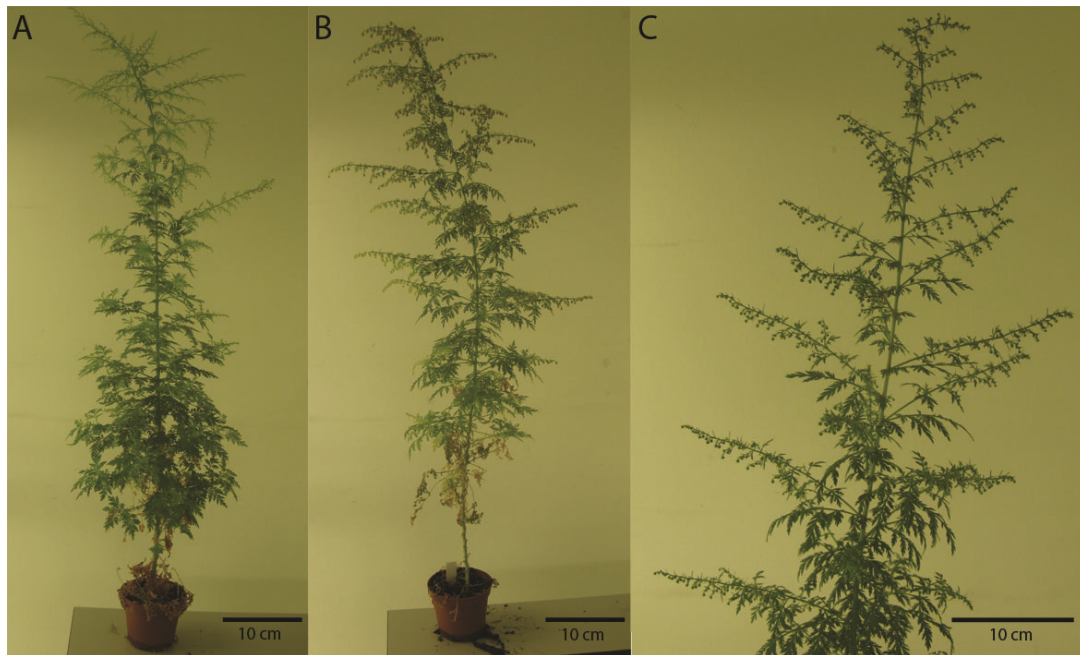
On the transcript level, MEP and MVA pathways were significantly upregulated in glandular trichomes in comparison with filamentous trichomes. In addition to this, transcripts coding for the artemisinin biosynthesis pathway, other sesquiterpene biosynthesis and monoterpene pathways were predominantly expressed in glandular trichomes. Novel cytochrome-, peroxidase- and dioxygenases-encoding genes highly expressed in glandular trichomes were detected and these might be potential candidate genes for the formation of the endoperoxide bridge in artemisinin. Lipid biosynthesis pathways were highly expressed in glandular trichomes and less in filamentous trichomes. In filamentous trichomes, some specific genes from sesquiterpenoid and triterpenoid pathways such as 8-epi-cedrol synthase and oxidosqualene cyclase were detected significantly more than in glandular trichomes. Between the transcriptome of apical and sub-apical cells from glandular trichomes, no differences could be observed in the expression of artemisinin biosynthetic enzymes.

This transcriptome analysis underscores the vast metabolic capacities of *A. annua* glandular trichomes and simultaneously points to the existence of specific terpene metabolic pathways in the filamentous trichomes. Therefore, it would be interesting to examine metabolic activities in filamentous trichomes of other plant species. Besides this, it would also be interesting to characterize the potential candidate genes for artemisinin biosynthesis. If they

are involved in the production of artemisinin, they can be used to produce artemisinin in yeast cells.

5 Supporting information

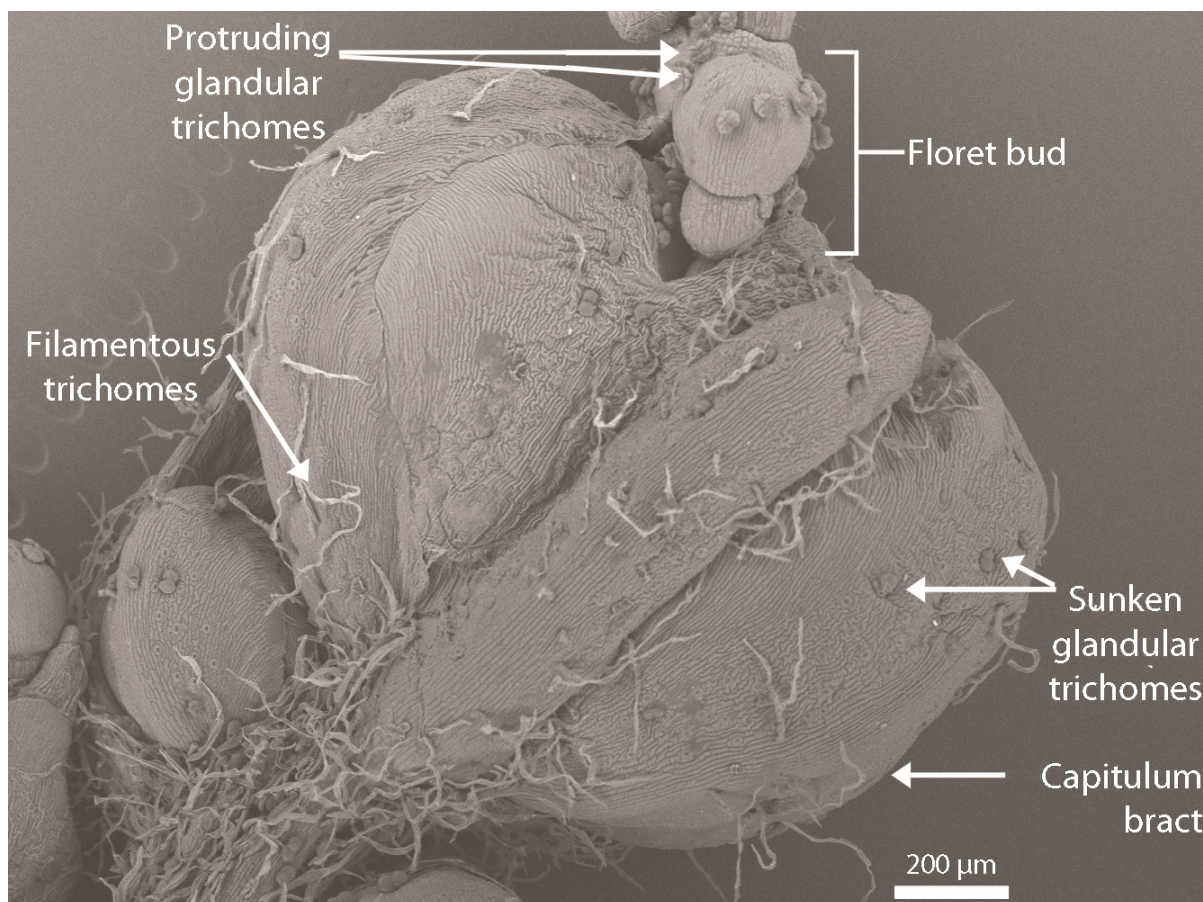
Additional File 1



***Artemisia annua* Anamed grown in 8h light, 16h night.**

Pictures of 7 months old *Artemisia annua* Anamed plants (A, B) grown in a growth room under 8h light, 16h night photoperiod. C: detail of the flower bud stage used to collect trichomes.

Additional File 2



Trichomes on a flower head of *A. annua*.

SEM picture of a flower head (capitulum) from *A. annua* Anamed with some bracts opened to show floret buds. On the floret buds, glandular trichomes are protruding whereas they are sunken in the capitulum bracts. Filamentous trichomes are abundantly present on the basal bracts.

Additional File 3

Transcripts	Primer sequences	
<i>Actin 2</i>	TCAGGCCGTGCTTTCTCTTT	AGATGGGCACTGTATGGGAC
<i>PP2AA3</i>	TACCATATACCGCACACGCC	GGGAAGTTGTTACAGCCCA
<i>PPR</i>	GGGAACTTAAGTTGCGGT	ACCATCCGAGAGTACCCAT
<i>FDS</i>	CCCGAGGTGATTGAAAGATTG ^[1]	CGAATACAGCCTGAAGATTGAGAG ^[1]
<i>CYP71AV1</i>	CGAGACTTTAACTGGTGAGATTGT ^[26]	CGAAGCGACTGAAATGACTTTACT ^[26]
<i>DBR2</i>	ATGGAAGTGAGGAGGAAG ^[1]	AAACAAGGTCAGGATTCG ^[1]
<i>Aldh1</i>	TCGGAGTAGTTGGTCACATC ^[1]	TCACGCCATCAGGAACAC ^[1]
<i>P450 comp69</i>	CTATTGCGCTTTGGGTCACG	GTAATCGAGGGCGGTTAGCA
<i>P450 comp2774</i>	TTTGATTGGGGGTTGCCAGA	CATGGTTAACCCCGCAAAA
<i>P450 comp 15043</i>	TTGTGGACTTGTTTGTGGCG	GTGTGCGGTTGCGTAGAA
<i>P450 comp 3673</i>	AGCGAGGATCACGTTGCATT	GCGGTCACCACATGAGGATT
<i>P450 comp 548</i>	GCACTCCAAGATTGCGGAT	AGTAATCCCGCGCTTTCAG
<i>Per comp2084</i>	TTCGTACAGGGCTGTGATGG	TGGGTGCGTTCTTTCTCCT
<i>Per comp252</i>	CAGGTCAGACCAGCAGTTGT	CCGCGAAATCCTCATGAAAC
<i>Per comp6217</i>	TTCTAGTCGACCGTGAGGGA	TTCTCGATGCTGAGTGACTG
<i>Diox comp225</i>	AGGAGAAGCGGTGCAAGATT	AACTGCCCTCCACTCTTTGG
<i>Diox comp453</i>	GGTGGCGAAGACTTGGTCTA	CATGTCGGTATGTGGGGCTA
<i>TFAR1</i>	CAAACCTTTTCAGTTCACCA ^[1]	ATGACAGCCTTTCATCCTTT ^[1]
<i>U_OSC</i>	CGGTCGAGCGTCAAGAAGTA	CGCATAAGCAAATCACCGCA
<i>epi-cedrol synthase</i>	TTGGTTCCCATAGGGCGAG	CGTAGGCTTGTGCTCGCTA

Primers for qRT-PCR.

Overview of primer sequences used for qRT-PCR.

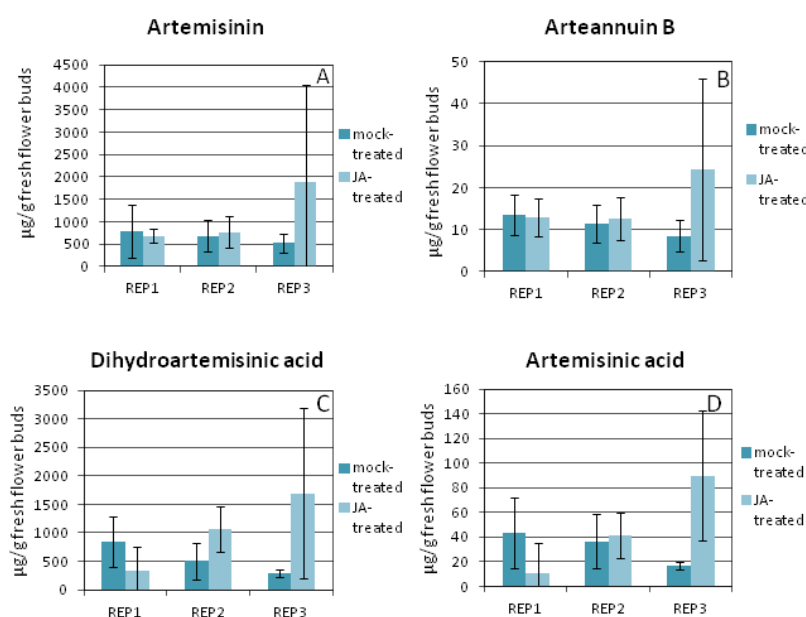
Additional File 4: Transcriptome assembly of trichomes from *Artemisia annua*.

All 150,288 *de novo* assembled sequences and their contig-numbers generated with Trinity. This list is provided in the supplementary CD.

Additional File 5: Metabolite concentrations in flower heads of mock- and JA-treated plants.

After sampling for RNA extraction, metabolites from remaining flower heads were separately extracted by a 1 min chloroform-dip [51]. Artemisinin, arteannuin B, dihydroartemisinic acid and artemisinic acid were quantified by means of a HPLC-MS/MS method, developed by Van Nieuwerburgh *et al.*, with following modifications [51]. The mobile phase of pump A consisted of ULC-MS pure water (Biosolve, Valkenswaard, the Netherlands) with 0.1% formic acid (Biosolve) and the mobile phase of pump B was 90% ULC-MS pure acetonitrile (Biosolve) and 10% water with 0.1% formic acid. Metabolites were separated in a run of 34 minutes (40% A and 60% B for 8 minutes, linear gradient of 9 minutes to 15% A and 85% B, 100% B for 5 minutes, 40% A and 60% B for 12 minutes). The capillary voltage of the ESI-source was 2.4 kV and the flow rate of nitrogen as desolvation gas was 400 l/h. Argon was used as a collision gas at 0.9 bar. The collision energy for dihydroartemisinic acid was set at 12 eV (m/z 237→163 + 191 + 201 + 219) and for arteannuin B the collision energy was 10

eV (m/z 249→185 + 189 + 231). The internal standard used was santonin with the collision energy set at 9 eV (m/z 247→173.2 + 201.2). For the other compounds, the same settings were used as in Van Nieuwerburgh *et al.* [51]. Standards santonin (I.S.) and artemisinin were supplied by Sigma-Aldrich, arteannuin B and artemisinic acid by the Walter Reed Army Institute of Research (Washington, U.S.A.) and dihydroartemisinic acid was donated by Patrick Covello (National Research Council Canada).



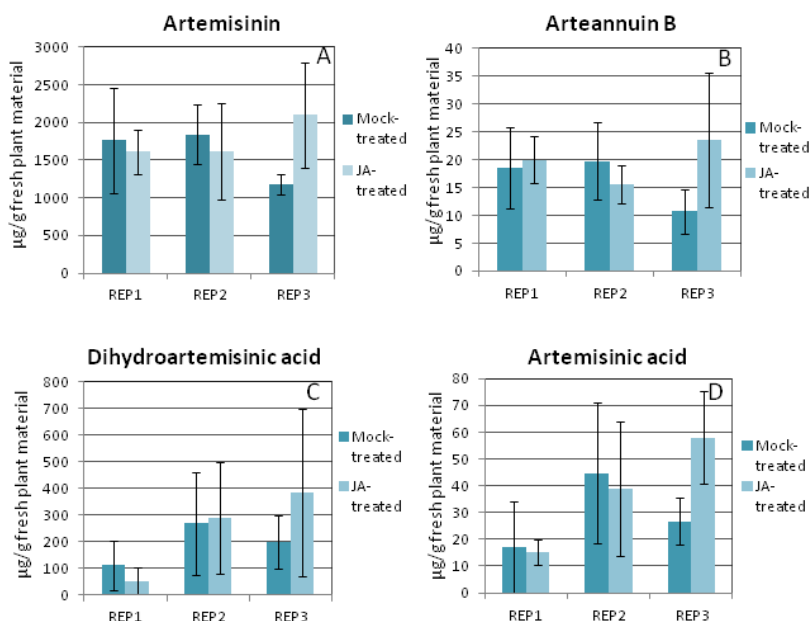
Metabolite concentrations in flower heads of mock- and JA-treated plants.

Metabolite concentrations are shown in $\mu\text{g/g}$ fresh flower head material of mock-treated and JA-treated plants. The results of all 3 repeats are shown separately. The concentration shown for each repeat is the average concentration of metabolites from 3 plants. The material for HPLC-MS/MS was derived from the same plants used for RNASeq. A: artemisinin, B: arteannuin B, C: dihydroartemisinic acid and D: artemisinic acid concentrations were measured. Equal quantities of artemisinin, arteannuin B, dihydroartemisinic acid and artemisinic acid were detected in mock- and JA-treated flower buds. Error bars represent standard deviations.

Additional File 6: Metabolite concentrations in entire plant material of mock- and JA-treated plants.

After sampling for RNA extraction, metabolites from remaining plant material were separately extracted by a 1 min chloroform-dip [51]. Artemisinin, arteannuin B, dihydroartemisinic acid and artemisinic acid were quantified by means of a HPLC-MS/MS method, developed by Van Nieuwerburgh *et al.*, with following modifications [51]. The mobile phase of pump A consisted of ULC-MS pure water (Biosolve, Valkenswaard, the Netherlands) with 0.1% formic acid (Biosolve) and the mobile phase of pump B was 90% ULC-MS pure acetonitrile (Biosolve) and 10% water with 0.1% formic acid. Metabolites were separated in a run of 34 minutes (40% A and 60% B for 8 minutes, linear gradient of 9 minutes to 15% A and 85% B, 100% B for 5 minutes, 40% A and 60% B for 12 minutes). The capillary voltage of the ESI-source was 2.4 kV and the flow rate of nitrogen as desolvation gas was 400 l/h. Argon was used as a collision gas at 0.9 bar. The collision energy for dihydroartemisinic acid was set at 12 eV (m/z 237→163 + 191 + 201 + 219) and for arteannuin B the collision energy was 10 eV (m/z 249→185 + 189 + 231). The internal

standard used was santonin with the collision energy set at 9 eV (m/z 247→173.2 + 201.2). For the other compounds, the same settings were used as in Van Nieuwerburgh *et al.* [51]. Standards santonin (I.S.) and artemisinin were supplied by Sigma-Aldrich, arteannuin B and artemisinic acid by the Walter Reed Army Institute of Research (Washington, U.S.A.) and dihydroartemisinic acid was donated by Patrick Covello (National Research Council Canada).



Metabolite concentrations in entire plant material of mock- and JA-treated plants.

Metabolite concentrations are shown in µg per g fresh plant material of mock-treated and JA-treated plants. The results of all 3 repeats are shown separately. The concentration shown for each repeat is the average concentration of metabolites from 3 plants. The material for HPLC-MS/MS was derived from the same plants used for RNASeq. A: artemisinin, B: arteannuin B, C: dihydroartemisinic acid and D: artemisinic acid concentrations were measured. Equal quantities of artemisinin, arteannuin B, dihydroartemisinic acid and artemisinic acid were detected in mock- and JA-treated flower buds. Error bars represent standard deviations.

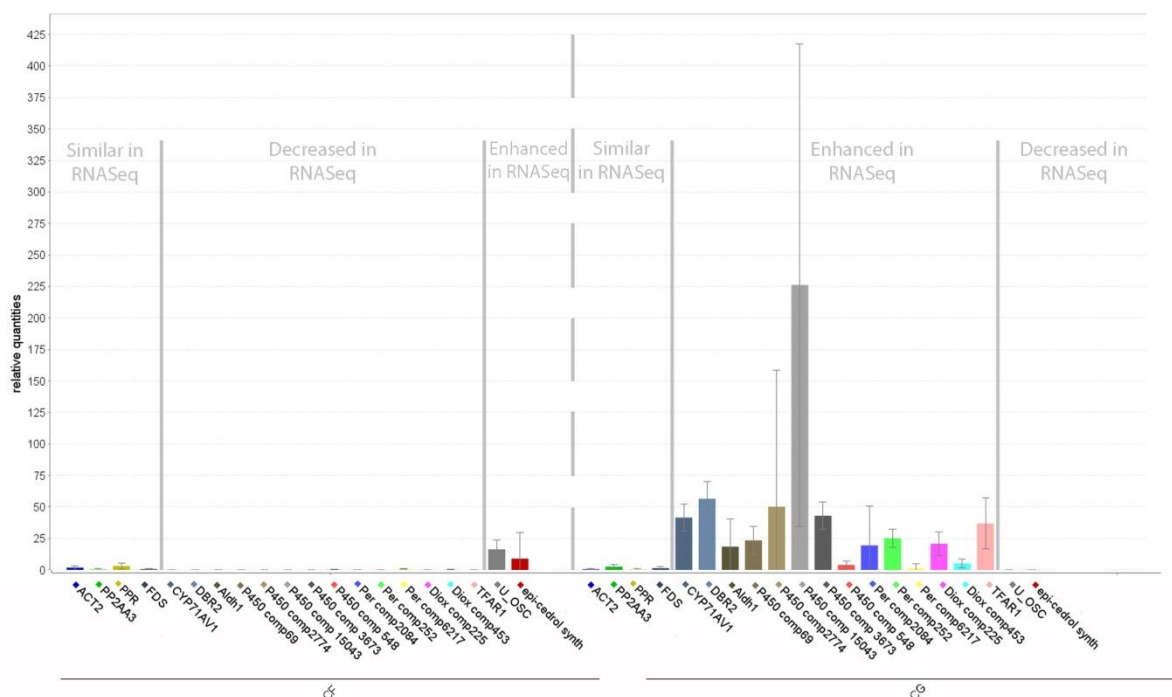
Additional File 7: Significant differences in transcriptome expression in glandular and filamentous trichomes.

List with all contigs that are significantly differentially expressed in glandular and filamentous trichomes. This list is provided in the supplementary CD.

Additional File 8: Overview from the commented differences between glandular and filamentous trichomes.

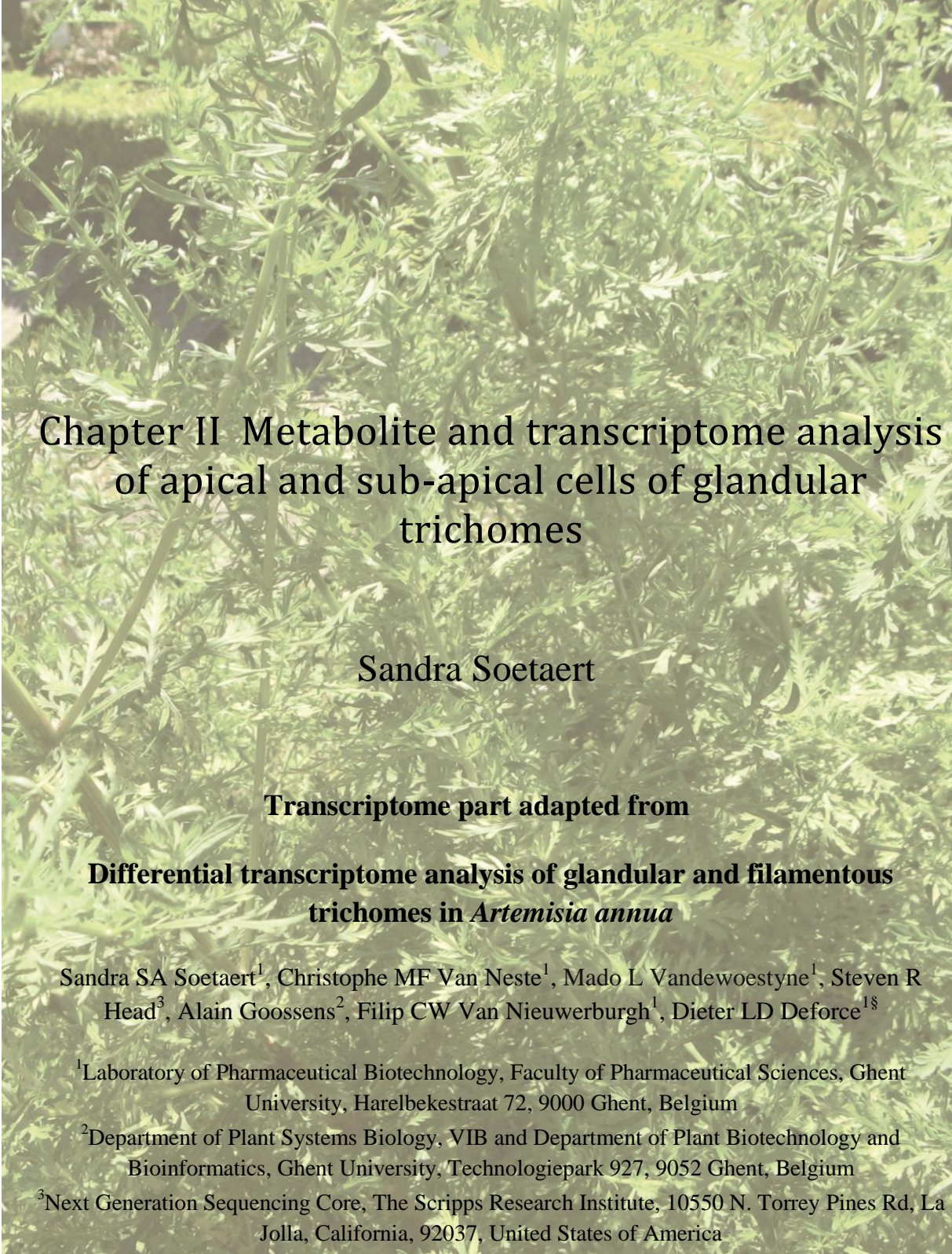
Overview of all contigs discussed in this article including calculation of \log_2 -fold changes and adjusted p-values. This list is provided in the supplementary CD.

Additional File 9: qRT-PCR Results.



qRT-PCR analysis on filamentous and glandular trichomes.

Bar chart showing qRT-PCR results on Nugen amplified material and on top the abundance of the corresponding transcripts in the RNASeq data (similar, decreased or enhanced abundance in filamentous and glandular trichomes). Of 3 biologic repeats, geometric averages from relative quantities were calculated with qbasePLUS and shown against a linear scale with as error bars the standard errors of the geometric mean. CF and CG represent mock-treated filamentous and glandular trichomes, respectively.



Chapter II Metabolite and transcriptome analysis of apical and sub-apical cells of glandular trichomes

Sandra Soetaert

Transcriptome part adapted from
**Differential transcriptome analysis of glandular and filamentous
trichomes in *Artemisia annua***

Sandra SA Soetaert¹, Christophe MF Van Neste¹, Mado L Vandewoestyne¹, Steven R
Head³, Alain Goossens², Filip CW Van Nieuwerburgh¹, Dieter LD Deforce^{1§}

¹Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, Ghent
University, Harelbekestraat 72, 9000 Ghent, Belgium

²Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and
Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium

³Next Generation Sequencing Core, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La
Jolla, California, 92037, United States of America

The discovery of new candidate genes more expressed in glandular trichomes and potentially involved in the production of artemisinin, was reported in Chapter I. In the current Chapter, HPLC-MS/MS experiments were performed to find the exact location of artemisinin and its bio-precursors inside the glandular trichomes. In literature, there were indications that the artemisinin precursors were synthesized on top of the glandular trichomes, in the apical cells. Therefore, apical and sub-apical cells were collected to determine the metabolite levels.

Besides the metabolite experiments, transcriptome experiments were executed on these cell types to investigate if all known artemisinin biosynthesis genes are expressed in apical and not in sub-apical cells and to see in which cell type the new candidate genes were expressed. The technical optimizations from Part II were used during the transcriptome analysis: RNA was collected as optimized in Part II Chapter II, amplified and the library was prepared according to the post amplification ligation-mediated strategy barcoding protocol from Part II Chapter III and sequenced on the Illumina HiSeq platform.

1 Introduction

Secretory cells from glandular trichomes are distinguishable in a pair of apical cells on top of the trichome and two pairs of sub-apical cells below. These cells are morphologically different: apical cells contain proplastids or leucoplasts instead of normal chloroplasts in sub-apical cells [68]. This raised the question whether these morphological differences imply a division of biosynthetic function [68]. For example, is artemisinin produced in only one cell type or in all secretory cells? To address this question, Olsson *et al.* collected apical and sub-apical cells with laser capture microdissection to determine which genes were expressed in both cell types [69]. In both cells, transcript expression of farnesyl diphosphate synthase was detected. This enzyme catalyzes the formation of farnesyl diphosphate which is a general precursor for sesquiterpenes and triterpenes. Transcriptional expression of ADS, CYP71AV1 and DBR2, three enzymes involved in artemisinin biosynthesis, was detected only in apical cells and not in sub-apical cells [69]. This indicated that artemisinic acid and dihydroartemisinic aldehyde are produced in the apical cells of the glandular trichomes. As the last steps of the artemisinin biosynthesis pathway are unknown, this conclusion cannot be drawn for artemisinin itself.

Therefore, in this PhD project, efforts were made to localize artemisinin itself by performing metabolite analysis on apical, sub-apical cells and the content of the sub-cuticula space from *A. annua* glandular trichomes. Laser capture microdissection was used to collect the cells and the sub-cuticular space was sampled with micropipetting. Artemisinin and its precursors were measured with a previously in-house developed HPLC-ESI-Q-TOF tandem mass spectrometry method [51].

Since Olsson *et al.* [69] detected transcript expression of artemisinin biosynthesis genes only in apical cells, a comparison of apical and sub-apical cells with RNASeq might lead to the discovery of other genes involved in artemisinin biosynthesis. Therefore, the transcriptome of apical and sub-apical cells was analysed with 2nd generation sequencing. For this, the experimental procedure used was similar to the RNASeq on glandular and filamentous trichomes in Part III Chapter I.

2 Materials and methods

2.1 Plant material

Seeds were sterilized as explained in Part III Chapter I Paragraph 2.2. For metabolite and transcriptome analysis, two different cultivars were used.

For metabolite analysis, seeds from *A. annua* cultivar Brazil (line 2/39) were provided by Patrick Covello (National Research Council Canada). These plants were grown in a growth room with a 12h day/12h night regime for a few months and then transferred to a greenhouse (16h day/8h night) for length growth. After a few months, the plants were transferred back to the growth room to initiate flowering.

For transcriptome analysis, *A. annua* cultivar Anamed (<http://www.anamed.net>) was used. These plants were grown under a regime with 8 hours day and 16 hours night. After 6 months, flower heads appeared.

2.2 Collection techniques

2.2.1 Laser capture microdissection for metabolite analysis

To measure the presence of artemisinin and its (proven and hypothetical) bio-precursors: 200 apical and 200 sub-apical cells were isolated by laser capture microdissection from fixated flower head buds. Fixation was carried out by subjecting the samples to a 4% formaldehyde (Merck, Hohenbrunn, Germany) in phosphate buffered saline (Gibco, Paisly, UK) solution for 3 to 4 hours in vacuum [69]. Glandular and filamentous trichomes were collected as well but for this, plant material was not fixated. Closed flower heads were cut finely in ULC-MS pure water (Biosolve, Valkenswaard, the Netherlands) or fixative and samples were microdissected and catapulted with a laser microscope in ULC-MS pure water.

2.2.2 Laser capture microdissection for transcriptome analysis

For the transcriptome experiment, 300 unfixated apical and sub-apical cells were collected. Flower head buds were cut into pieces in a cold buffer with 25 mM MOPSO (pH6.3) (Sigma-Aldrich), 200 mM sorbitol (Alfa Aesar, Karlsruhe, Germany), 10 mM sucrose (Acros, Geel, Belgium), 5 mM thiourea (Sigma-Aldrich), 2 mM DTT (Fluka, Sigma-Aldrich), 5 mM MgCl₂ (Sigma-Aldrich) and 0.5 mM sodium-phosphate (Acros) [26]. Cells were separated

with laser microdissection and captured with laser pressure catapulting (Figure 5). As explained in Part II Chapter II, formaldehyde deteriorates the yield and quality of extracted RNA but laser microdissection of unfixated apical and sub-apical cells was very difficult. Therefore, in another set of samples, 500 fixated cells of each cell-type were collected for transcriptome analysis. All cells for transcriptome analysis were collected in 30 μ l of lysis buffer with β -mercaptoethanol from the Absolutely Nanoprep kit (Stratagene, La Jolla, CA, USA).

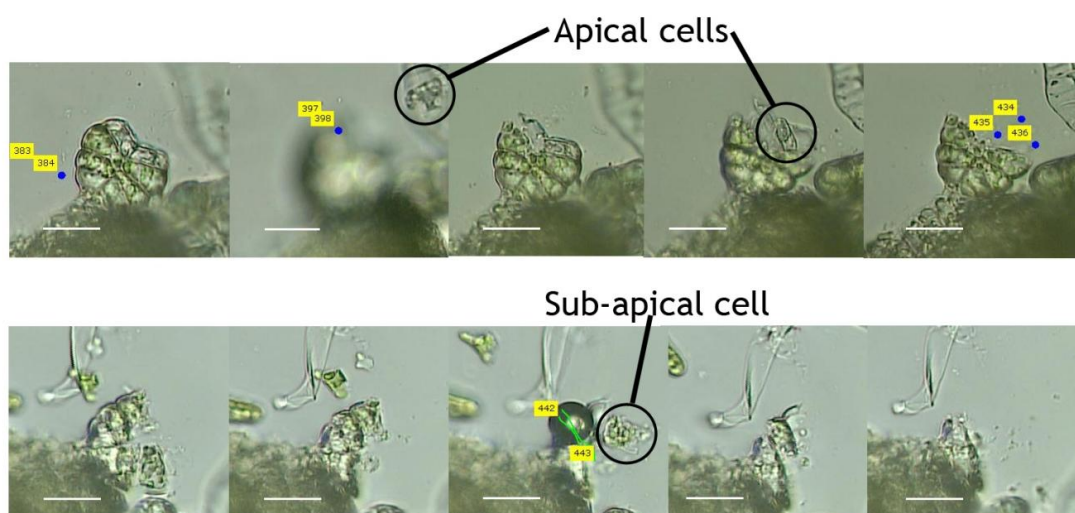


Figure 5: Laser capture microdissection of apical and sub-apical cells.

Laser capture microdissection with P.A.L.M of fixated apical and sub-apical cells from glandular trichomes of *A. annua* (scale bar is 25 μ m).

2.2.3 Micro-manipulation for metabolite analysis

To determine which metabolites are stored in the sub-cuticular cavity, it is impossible to use laser microdissection since the sub-cuticular cavity is not enclosed by one continuous membrane. Therefore, it was attempted to collect the fluid from the sub-cuticular cavity with a micro-pipette. This technique is called micro-manipulation (see Figure 6) and was executed with an Olympus IX70 microscope.

Closed flower heads were chopped up with a scalpel on a glass plate and then transferred to a petri dish. Ultra pure water was poured with care in the petri dish so the fragments remained on the bottom. To avoid cell debris which can obstruct the pipette, fragments with a lot of trichomes were selected and transferred to another petri dish and again ULC-MS pure water was gently poured over it. Pipettes made of micro capillaries (Drummond Scientific Co.) with

an average diameter of 200 μm (holding pipette) and 1 μm (sample collecting pipette) were filled with mineral oil (Western Family). The sample collecting pipette was filled at the top with dimethylpolysiloxane (viscosity 12,500 cSt 25°C supplied by Sigma-Aldrich). To enable visual control of the fluid in the collecting pipette, ultra pure water was aspirated first which visualized the boundary between dimethylpolysiloxane and the oily fluid aspirated from the sub-cuticular cavity.

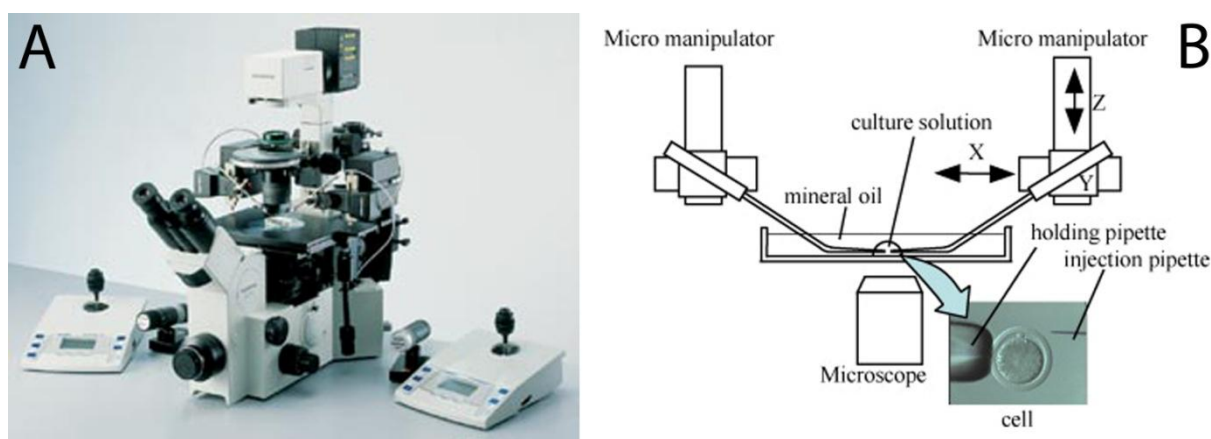


Figure 6: Micro-manipulation system.

The micro-manipulation system with a holding pipette and a collecting pipette, **A**: picture of an Olympus IX70 microscope for micro-manipulation, **B**: schematic overview of important components in a micro-manipulation device with an example of intra cytoplasmic sperm injection.

(A: http://www.sycos.co.kr/product/index.php?act=eppen&cate=inject_NK2, B: [70])

With a holding pipette, the plant material was fixed while the sample collecting pipette was inserted in the sub-cuticular cavity. To confirm the presence of the micropipette in the sub-cuticular cavity, the fluid was aspirated, respired and again aspirated for collection. This was microscopically observed as respectively shrinking, expansion and shrinking of the sub-cuticular cavity. The sample fluid was collected in a drop of ULC-MS pure water in another petri dish and transferred to a tube. The place where the drop of water was located in the petri dish was rinsed with another drop of water and also transferred to the tube. During fixation, crystals were formed in the sub-cuticular cavity as shown in Figure 7 and these crystals were not collected with micro-manipulation. Therefore, the experimental setup was changed to unfixated plant material to collect the content of the sub-cuticular cavities from 50 trichomes.

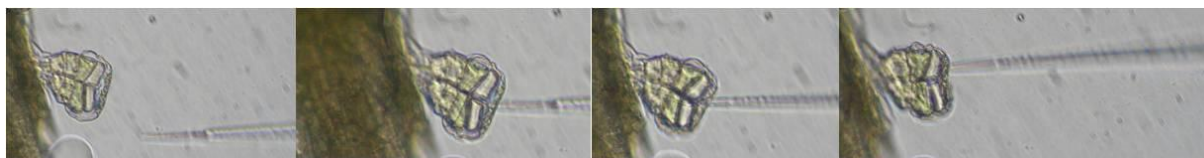


Figure 7: Micro-manipulation on fixated glandular trichomes.

The content of the sub-cuticular cavity of glandular trichomes from *A. annua* collected with micro-manipulation (scale bar not available but glandular trichomes are approximately 37 μ m in size).

2.3 Analysis methods

2.3.1 Metabolite analysis with HPLC-MS/MS

Artemisinin, arteannuin B, dihydroartemisinic alcohol, artemisinic alcohol, dihydroartemisinic acid, artemisinic acid, dihydroartemisinic aldehyde and artemisinic aldehyde were quantified by means of a HPLC-MS/MS method, adapted from Van Nieuwerburgh *et al.* [51]. The samples were pooled, sonicated (40 kHz) in an ultrasonic bath for 5 min, diluted to 50/50 H₂O/ACN with 0.1% formic acid and 0.1 ng/ μ l santonin as internal standard (I.S.) and 100 μ l was injected for analysis.

Chromatographic separations were performed on an Alltech Ultrasphere C18 IP 5 μ m column (150 x 2.1 mm). To protect the column, a Waters Xterra MS C18 5 μ m guard column (2.1 x 10 mm) was used. A Waters Alliance 2695 HPLC system was used to deliver the mobile phase at a flow rate of 0.2 ml/min. Pump A contained ULC-MS pure water (Biosolve, Valkenswaard, the Netherlands) with 0.1% formic acid (Biosolve), pump B 90% ULC-MS pure acetonitrile (Biosolve) and 10% water with 0.1% formic acid. Metabolites were separated in a run of 34 minutes: 40% A and 60% B for 8 minutes, linear gradient of 9 minutes to 15% A and 85% B, 100% B for 5 minutes and finally 40% A and 60% B for 12 minutes. Retention times were 4.21 minutes for santonin, 6.68 for arteannuin B, 8.78 for artemisinin, 15.55 for dihydroartemisinic acid, 16.36 for artemisinic acid, 16.99 for artemisinic alcohol, 17.94 minutes for dihydroartemisinic alcohol and 23.10 for artemisinic aldehyde. To divert $\frac{1}{4}$ to the LC/MS interface, a LC Packings ACUrateICP-04-20post-column splitter was used.

A Q-TOF Ultima mass spectrometer (Micromass, Manchester, UK) with an electrospray source in positive mode was used for the detection. The capillary voltage was optimized at 2.4

kV. Source and desolvation temperature were respectively 130 and 300 °C. As desolvation gas, N₂ was used with a flow rate of 400 l/h. Argon was used as collision gas at 0.9 bar. The collision energy for artemisinin was set at 7 eV (m/z 283.2→219.2 + 229.2 + 247.2 + 265.2), for arteannuin B 10 eV (m/z 249.2→185.2 + 189.2 + 231), dihydroartemisinic alcohol 9 eV (m/z 223.2→149.2 + 165.2 + 205.2), artemisinic alcohol 8 eV (m/z 221.2→203.2), dihydroartemisinic acid 12 eV (m/z 237.2→163.2 + 191.2 + 201.2 + 219.2), artemisinic acid 11 eV (m/z 235.2→189.2 + 199.2 + 217.2) and artemisinic aldehyde 11 eV (m/z 219.2→145.2 + 159.2 + 201.2). The internal standard used was santonin with the collision energy set at 9 eV (m/z 247→173.2 + 201.2). Standards santonin (I.S.) and artemisinin were supplied by Sigma-Aldrich, dihydroartemisinic acid, artemisinic alcohol, artemisinic aldehyde and dihydroartemisinic alcohol were donated by Patrick Covello (National Research Council Canada) and arteannuin B as well as artemisinic acid by the Walter Reed Army Institute of Research (Washington, U.S.A.). To ensure the specificity, only the sum of the fragments in MS/MS was used for quantification of the compounds.

2.3.2 Transcriptome analysis with Illumina RNASeq

RNA was extracted with the Absolutely RNA Nanoprep kit including a DNase treatment and the purified RNA was eluted in 10 µl. Of the RNA from the fixated samples, 5 µl was used as input for amplification. The 10 µl RNA from the unfixated samples was split in two and each subsample was separately amplified to create a third sample set for sequencing.

Samples were amplified with a linear amplification system: Ovation RNA-Seq with 1 h 30 min Spia-amplification (NuGen, AC Bemmell, The Netherlands). Illumina sequencing libraries were made by using a post amplification ligation-mediated strategy as explained in Part II Chapter III [27]. These samples were sequenced together with the glandular and filamentous trichomes with a read-length of 100 bp in 3 lanes of an Illumina HiSeq 2000 flowcell. Bioinformatics analysis was performed as described for glandular and filamentous trichomes in Chapter I.

3 Results and discussion

3.1 Metabolite analysis

The content of 50 unfixated sub-cuticular cavities was sampled with micro-manipulation and after HPLC-MS/MS, metabolite concentrations were recalculated to amounts per glandular trichome as shown in Figure 8. This facilitates the comparison with several types of trichome samples such as apical, sup-apical and complete glandular trichomes. HPLC-MS/MS revealed high levels of dihydroartemisinic acid and artemisinic acid which were above the limit of quantification. The peak of artemisinin showed a peak-to-noise ratio of 2.5 and the other metabolites were not detected.

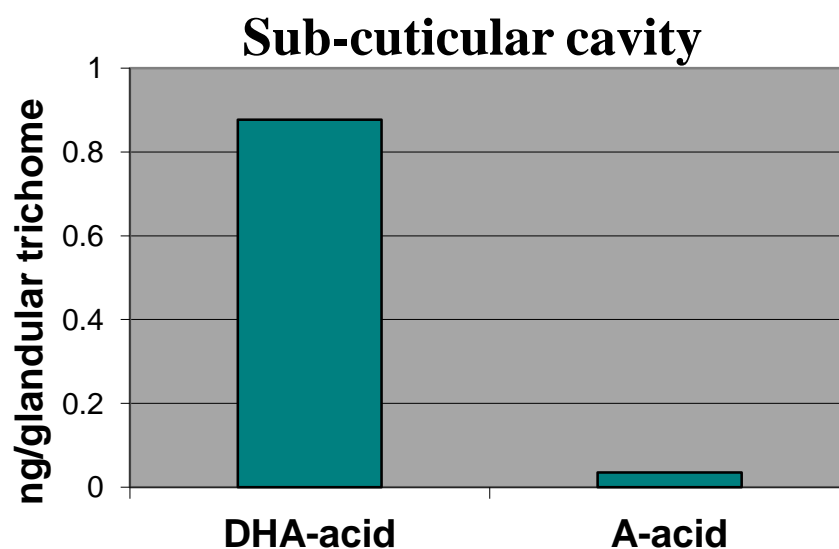


Figure 8: HPLC-MS/MS analysis of the content from the sub-cuticular cavity of glandular trichomes.

This graph represents the amounts of dihydroartemisinic acid (DHA-acid) and artemisinic acid (A-acid) detected in the sub-cuticular cavity of glandular trichomes with HPLC-MS/MS.

The amount is adjusted to the amount per trichome. Since only one measurement was performed, standard deviations are not available.

Another approach to find which metabolites are abundantly present in the sub-cuticular cavity was to collect unfixated glandular trichomes with intact sub-cuticular cavity and compare this to glandular trichomes from which the sub-cuticular cavity was removed. All trichomes were collected from one plant in three repeats. HPLC-MS/MS analysis of the same amount of trichome equivalents as in micro-manipulation, detected no artemisinin. As illustrated in Figure 9, dihydroartemisinic acid and artemisinic acid were detected in higher levels in

glandular trichomes with intact sub-cuticular cavities. This was consistent with the relatively high levels of dihydroartemisinic acid and artemisinic acid detected in fluid collected from these sub-cuticular cavities. The ratio of dihydroartemisinic acid to artemisinic acid was the same in fluid from sub-cuticular cavities, trichomes with and without sub-cuticular cavity.

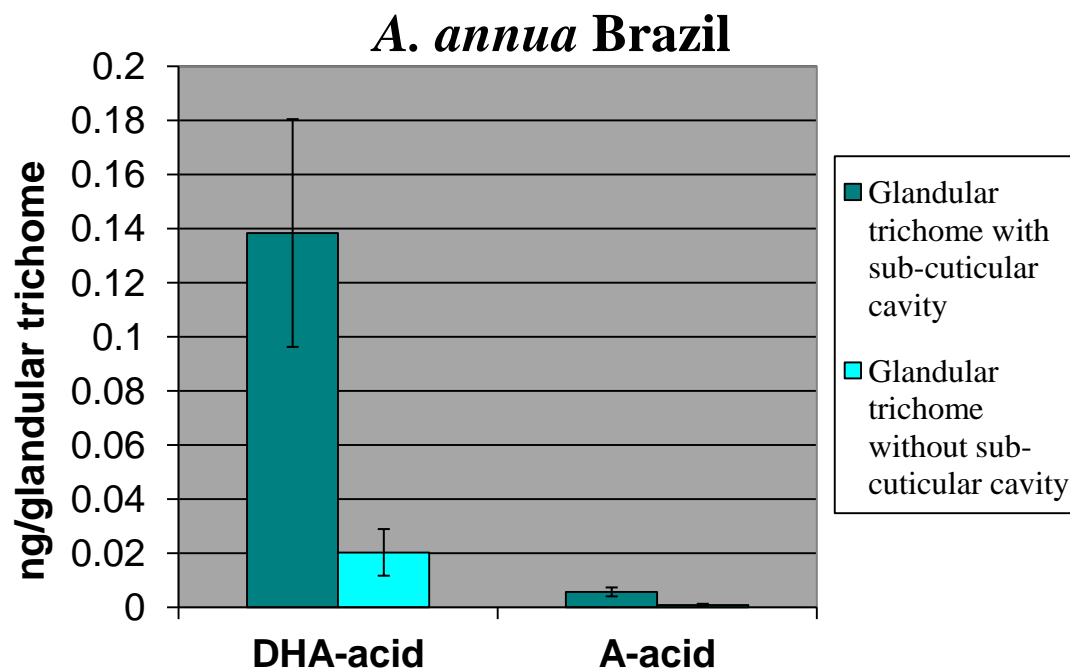


Figure 9: HPLS-MS/MS analysis of the glandular trichomes with sub-cuticular cavity and glandular trichomes with removed sub-cuticular cavity.

This graph represents the amounts of dihydroartemisinic acid and artemisinic acid detected in glandular trichomes with and without sub-cuticular cavity. The sub-cuticular cavity was removed by LCM and metabolite amounts are adjusted to the amount per trichome. Error bars represent standard deviations from 3 repeats on one plant of *A. annua*.

However, it should be noted that despite the same trends observed in both experiments, more dihydroartemisinic acid and artemisinic acid was detected in the sub-cuticular cavities alone than in the glandular trichome with intact sub-cuticular cavities (comparison of Figure 8 and Figure 9). Both tests were executed on another plant of the same cultivar and therefore, biological variation in artemisinin production might contribute to differences in metabolite levels. Another explanation might be that coeluting metabolites in the trichome cells caused ion suppression [71]. Sonication is well suited to extract metabolites from small samples [72] but it is possible that not all metabolites were extracted after 5 min. sonication of glandular trichomes in a sonication bath [73]. Another possible explanation for these discordant amounts of dihydroartemisinic acid and artemisinic acid detected after micro-manipulation

and laser capture microdissection might be the influence of the UV-A laser in the latter collection technique.

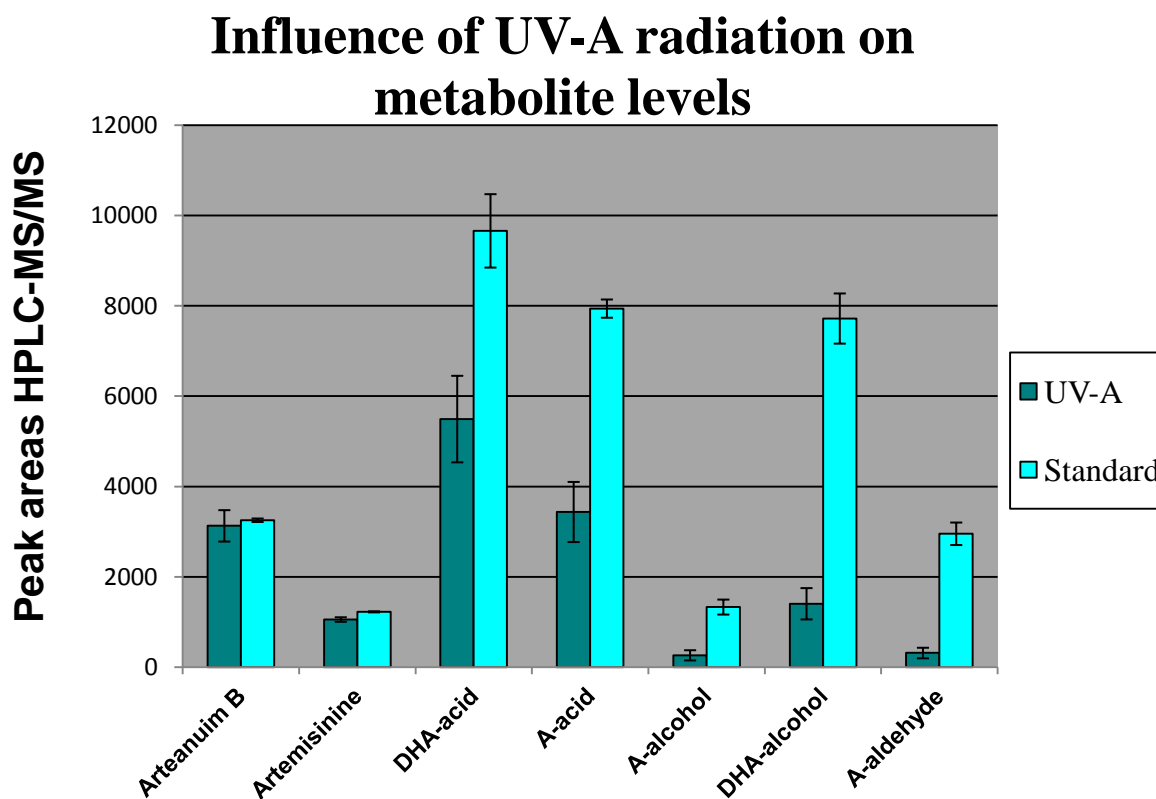


Figure 10: Influence of UV-A radiation on metabolite levels.

A standard of 0.8 ng/ μ l of each metabolite was irradiated for 1 hour with an UV-A laser and afterwards 8-fold diluted. Peak areas were compared to a standard of 0.1 ng/ μ l. Error bars represent standard deviation of 3 repeats.

As explained in Part I Chapter II Paragraph 4.2, dihydroartemisinic acid can be photochemically converted to artemisinin [5, 74] and therefore, the influence of UV-A light was investigated. A standard of 0.8 ng/ μ l of each standard metabolite (see page 143) in 50/50 H₂O/ACN with 0.1% formic acid was irradiated on a glass slide for one hour by continuous laser pulses. To rinse the glass slide, the sample was diluted 8-fold to 0.1 ng/ μ l (with correction for evaporation) and compared to a standard of 0.1 ng/ μ l. In Figure 10, average peak areas were given for three technical repeats. The levels of artemisinin and arteannuin B were relatively constant with or without UV-A radiation but the levels of the other metabolites were clearly reduced with UV-A radiation. Precaution should be taken to extrapolate these results in H₂O/ACN medium with 0.1% formic acid to metabolites in a cellular environment. It is possible that UV-A radiation contributed (partially) to the

discordant metabolite concentrations measured in the micro-manipulated and laser capture microdissection samples.

After analysing the content of the sub-cuticular cavity, experiments were focussed to cells of the glandular trichomes. Despite the possibly adverse effect of the UV-A laser, laser capture microdissection was the best suited method for separation and collection of specific trichome cells. Therefore, apical and sub-apical cells were collected and the equivalent of 90 cells in 50/50 H₂O/ACN with 0.1% formic acid was injected for HPLC-MS/MS. As shown in Figure 11, only the metabolites artemisinin, dihydroartemisinic acid and artemisinic acid were detected in apical cells in the first two experiments. In sub-apical cells, only dihydroartemisinic acid and artemisinic acid were detected but no artemisinin was observed. In a third test, artemisinin was not detected in apical and sub-apical cells (Figure 11). Additional experiments were performed but with inconsistent results. In some tests, artemisinin was not detected in apical cells and in another cultivar (Anamed), artemisinin was at the limit of detection in both apical and sub-apical cells.

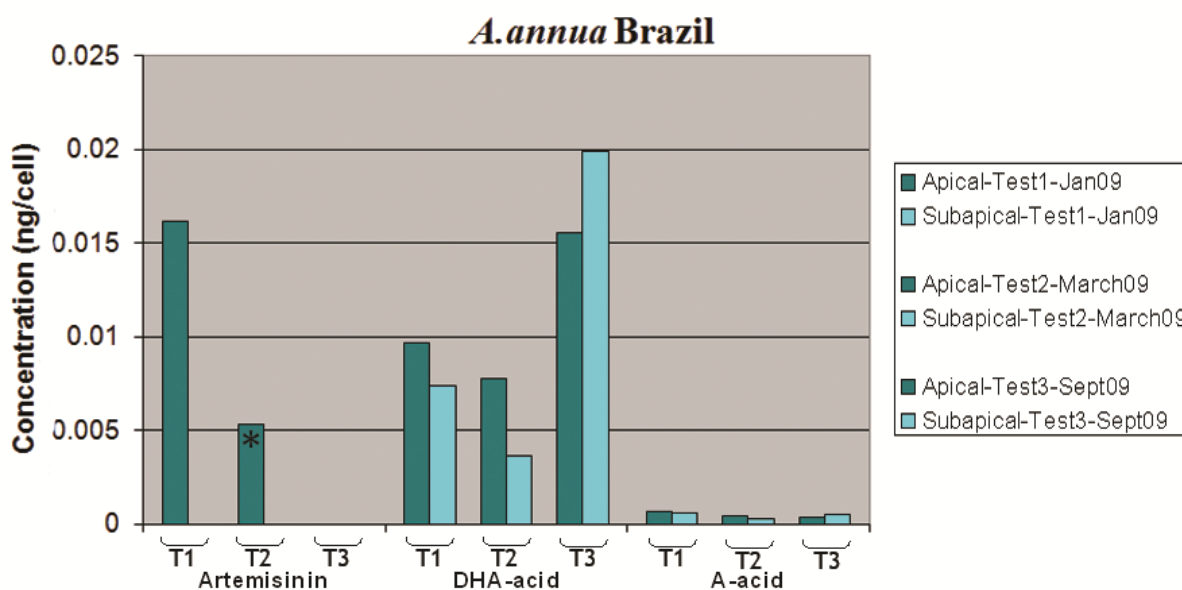


Figure 11: Metabolite analysis of apical and sub-apical cells.

Only artemisinin, dihydroartemisinic acid and artemisinic acid were detected. Results are given from 3 independent experiments. The bar with an * was above the detection limit but below the quantification limit. Artemisinin was never detected in sub-apical cells.

Therefore, as negative control, filamentous trichomes were collected and the equivalent of 83 trichomes was injected for HPLC-MS/MS. Artemisinin was detected with a peak-to-noise signal of 4 which was higher than the level detected in 90 apical cells of the same plant. This was not as expected since biotypes of *A. annua* with only filamentous trichomes lack artemisinin production [6]. Because of this, it was investigated if the detection of artemisinin in isolated filamentous trichomes might originate from contamination caused by the plant sample.

For laser capture microdissection, closed flower heads were immersed in ULC-MS pure water (or fixative) and cut up into pieces. Metabolites can be present in the wax layer surrounding the plant tissues or can be released during cutting. To test the order of magnitude of this potential contamination, 15 intact flower head buds were incubated for 30 min. in 300 μ l ULC-MS pure water, plant material sunked to the bottom and wash water was collected. After 12 washing steps, flower head buds were cut finely and washed in 3 subsequent steps. No trichomes were present in the washing water. Water from washing step 1 and 12 of intact flower head buds was analyzed with HPLC-MS/MS and to investigate if subsequently cutting the flower head buds increased the metabolite release, the washing water after cutting flower head buds was also analyzed. For HPLC-MS/MS, 50 μ l wash water was eluted to 200 μ l 50/50 H₂O/ACN with 0.1% formic acid. Three tests were performed on one plant. As illustrated on Figure 12, (undiluted) water from the first rinse of intact flower head buds contained between 0.75 and 1.2 ng/ μ l artemisinin. This level is relatively high (Figure 12) if compared to the amount measured in apical samples (Figure 11). Even in the 12th washing step, artemisinin was still present. After cutting these flower heads, additional artemisinin was released in the washing water. Artemisinic alcohol, artemisinic aldehyde and dihydroartemisinic alcohol were below the limit of detection.

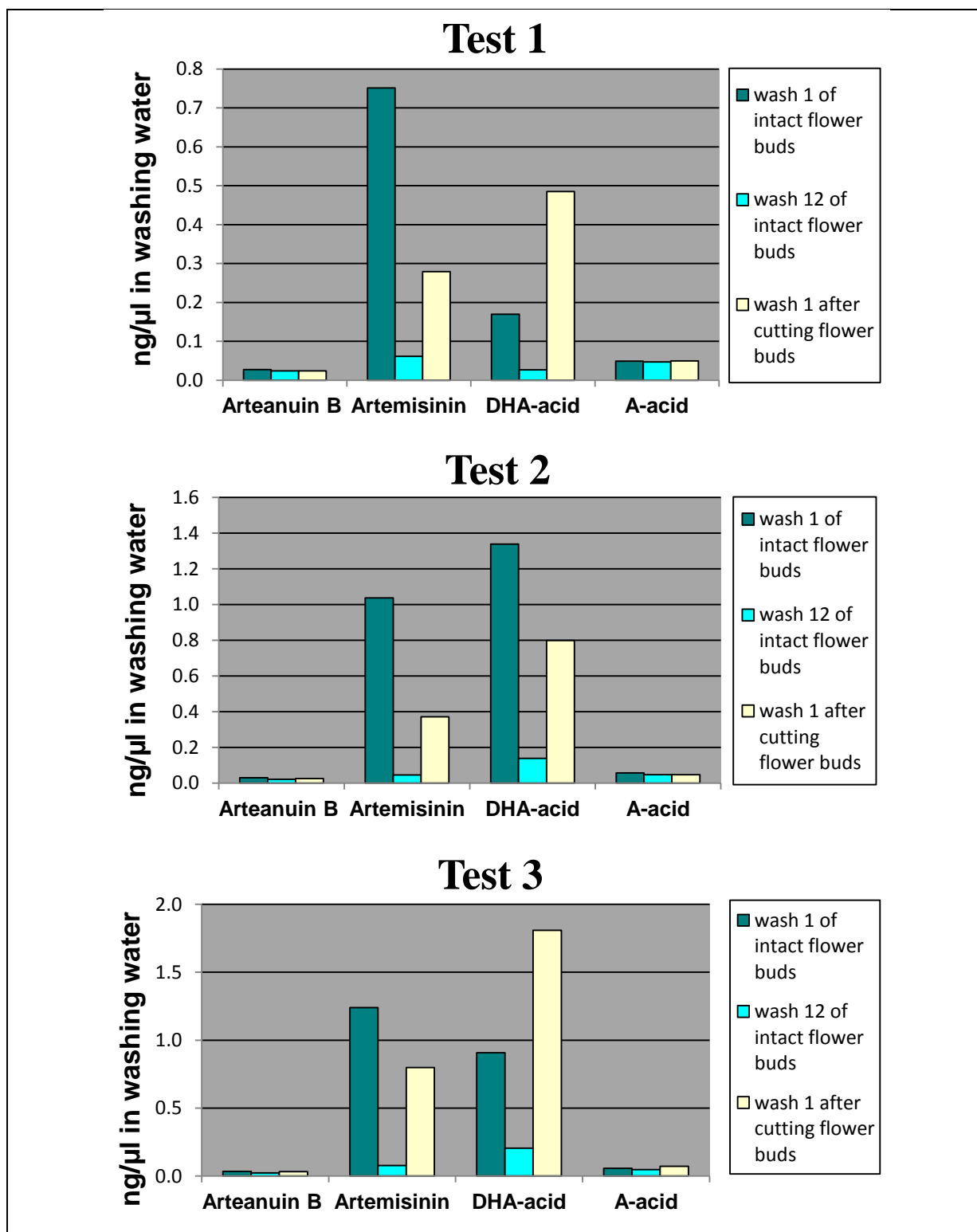


Figure 12: Metabolite analysis of wash water.

HPLC-MS/MS analysis of metabolites released in water if flower head buds were incubated for 30 min in 3 experiments. Wash 1 and wash 12 were analyzed and after wash 12, flower head buds were cutted and washed again. The other metabolites were below the limit of detection.

Therefore, the conclusion was made that arteannuin B, artemisinin, dihydroartemisinic acid and artemisinic acid were present in relatively high levels in the water on the glass slide for laser capture microdissection. This raised the question whether this can cause contamination in the collection tube? To simulate this effect, flower head buds of *A. annua* were cut into pieces and for 30 min. incubated in water as shown in Figure 13. The volume of water added per flower head bud and incubation time was the same as for laser capture microdissection of glandular, filamentous trichomes, apical and sub-apical cells. This washing water (without plant tissues) was pipetted on a glass slide, leaf material from mint was added and mint glandular trichomes were collected from this liquid.

Two tests were performed: one to enhance potential contamination: if the slide was drying, up to two times extra washing water was added to collect 55 mint trichomes and in another test to mimic the normal contamination level: if the slide was dried, a new slide was made to collect 45 mint trichomes. In the test that mimicked the normal contamination level, artemisinin was detected with a peak-to-noise ratio of 4 and dihydroartemisinic acid was 0.038 ng per collected glandular mint trichome. With the addition of extra washing water upon evaporation, the amount of artemisinin, dihydroartemisinic acid and artemisinic acid measured per glandular mint trichome was respectively 0.011, 0.038 and 0.026 ng. The other metabolites were not detected.

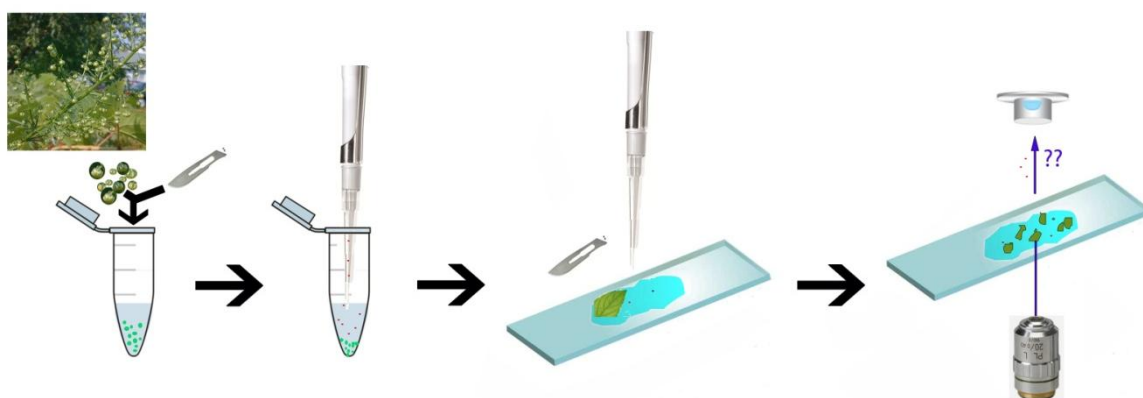


Figure 13: Test for contaminating metabolites during laser capture microdissection. Finely cut flower head buds of *A. annua* were incubated in water for 30 min. This water was collected and pipetted on a glass slide. A leaf of mint was finely cut and mint glandular trichomes were collected with laser capture microdissection. The red dots represent artemisinin-related metabolites dissolving in the incubation water.

As shown with mint trichomes, metabolites from the outer plant surface and from cutted cells of *A. annua*, dissolve in the aqueous prepartate and cause contamination of the laser capture microdissection samples. Therefore, it is required to put all data from paragraph 3.1 into perspective with this contamination problem and this makes it extremely difficult to draw conclusions. We can conclude that it was not possible to locate artemisinin in one or another cell type of the glandular trichomes and it seems that there is a lot of arteannuin B, artemisinin, dihydroartemisinic acid and artemisinic acid present on the outer surface of the flower head buds. These results were probably not caused by incubation in hypotonic ULC-MS pure water since similar results were obtained by incubation in phosphate buffered saline but it must be noted that the HPLC-MS/MS method was not validated for detecting natrium-adducts. The possible presence of these metabolites on the outer plant surface was also in line with the fact that during a 5 sec. chloroform dip, 97% from all artemisinin was extracted from leaf tissue [6].

3.2 Transcriptome analysis

For RNASeq of glandular and filamentous trichomes, more input material was available than for apical and sub-apical cells. The small number of micro-dissected apical and sub-apical cells yielded a suboptimal amount of input for the Ovation RNA-Seq amplification. Because of this, amplification was expected to be biased towards high abundant transcripts. Additionally, a high variation in the counts of low abundant transcripts was expected. Therefore less stringent settings were chosen (prior.n=estimated) by using a more tagwise dispersion. Out of 150,288 contigs, 195 contigs were reported as significantly differentially expressed (adj. p-value < 0.05). From these differences, 66 contigs were more expressed in apical cells, and 129 more in sub-apical cells. If these differences were examined, no clear connection was observed with functional metabolic pathways.

Normalized counts for artemisinin biosynthesis genes are shown in Figure 14 but none of the genes involved in artemisinin biosynthesis was significantly differentially expressed. This is contradictory to the results published in 2009 by Olsson *et al.* [69] who detected expression of artemisinin biosynthesis genes solely in apical cells. While our experiment was executed, the same research group presented in 2012 another manuscript in which they did more elaborate qRT-PCR tests [26]. Finally, they concluded that artemisinin biosynthesis genes are expressed in both apical and sub-apical cells and this is in agreement with our RNASeq data.

If the data set was analyzed with more strict settings as used for the analysis of glandular and filamentous trichomes: edgeR parameter `prior.n` set to 1, no significant differences were obtained with adjusted p-value < 0.05.

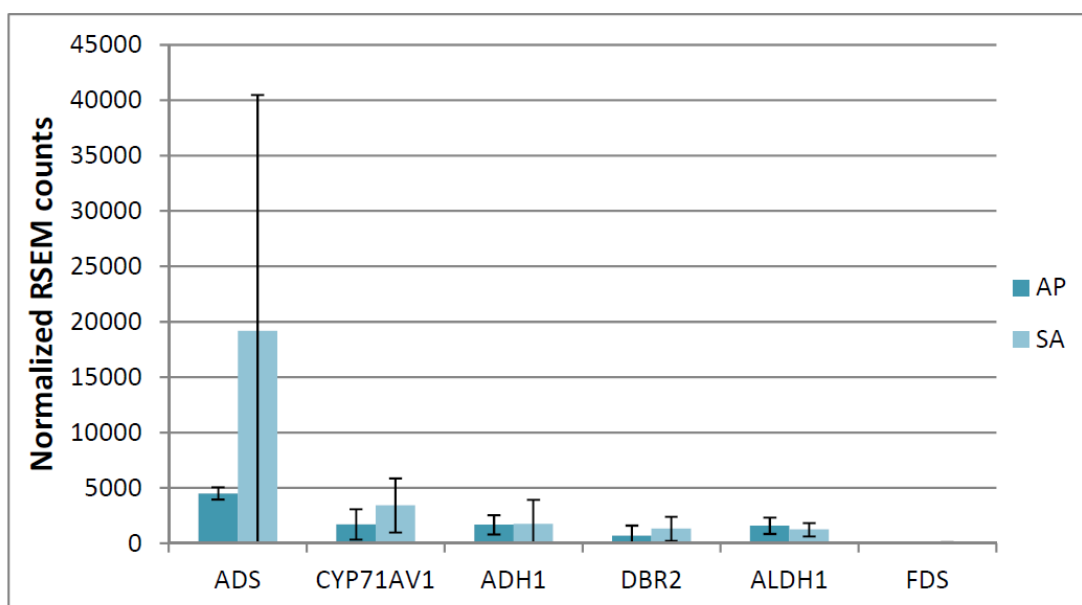


Figure 14: Expression of artemisinin biosynthesis genes in apical and sub-apical cells from glandular trichomes.

Normalized counts for artemisinin biosynthesis genes in apical (AP) and sub-apical (SA) cells. Error bars represent standard deviations.

4 Conclusions

Attempts to analyze metabolite levels of artemisinin and its bio-precursors in apical or sub-apical cells and glandular or filamentous trichomes were not successful. The reason for this was that the surface of the plant is covered with a wax layer which contains a variety of secondary metabolites including artemisinin. Therefore it was impossible to distinguish whether metabolites were originating from the cell content or from the wax layer on the plant surface which contaminates the aqueous microscopy slide.

The transcriptome comparison of apical and sub-apical cells revealed no significant differences in expression of artemisinin related genes. The 195 significantly differentially expressed genes were not clearly linked to functional pathways.

5 Supporting information

Additional File 10: Significant differences of transcriptome expression in apical and sub-apical cells.

List with all contigs that are significantly differentially expressed in apical and sub-apical cells. This list is provided in the supplementary CD.

References

Articles and books:

1. Maes L, Van Nieuwerburgh FCW, Zhang YS, Reed DW, Pollier J, Castele S, Inze D, Covello PS, Deforce DLD, Goossens A: **Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants.** *New Phytologist* 2011, **189**(1):176-189.
2. WHO: *World malaria report* Geneva; 2011.
3. Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM *et al*: **The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin.** *Science* 2010, **327**(5963):328-331.
4. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J *et al*: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440**(7086):940-943.
5. Levesque F, Seeberger PH: **Continuous-flow synthesis of the anti-malaria drug artemisinin.** *Angew Chem-Int Edit* 2012, **51**(7):1706-1709.
6. Duke MV, Paul RN, Elsohly HN, Sturtz G, Duke SO: **Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua* L.** *International Journal of Plant Sciences* 1994, **155**(3):365-372.
7. Ferreira JFS, Janick J: **Floral morphology of *Artemisia annua* with special reference to trichomes.** *International Journal of Plant Sciences* 1995, **156**(6):807-815.
8. Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS: ***Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin.** *Febs Letters* 2006, **580**(5):1411-1416.
9. Teoh KH, Polichuk DR, Reed DW, Covello PS: **Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*.** *Botany* 2009, **87**(6):635-642.
10. Zhang Y, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJH, Ross ARS, Covello PS: **The molecular cloning of artemisinic aldehyde Delta 11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*.** *Journal of Biological Chemistry* 2008, **283**(31):21501-21508.
11. Polichuk D; Teoh KH; Zhang Y; Ellens KW RD, Covello PS: **Nucleotide sequence encoding an alcohol dehydrogenase from *Artemisia annua* and uses thereof.** *Patent No. WO/2010/012074*, February 4th, 2010.
12. Brown GD, Sy LK: **In vivo transformations of dihydroartemisinic acid in *Artemisia annua* plants.** *Tetrahedron* 2004, **60**(5):1139-1159.
13. Wang W, Wang YJ, Zhang Q, Qi Y, Guo DJ: **Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing.** *Bmc Genomics* 2009, **10**(465):1-10.
14. Wagner GJ: **Secreting glandular trichomes: more than just hairs.** *Plant Physiology* 1991, **96**(3):675-679.

15. Levin DA: **The role of trichomes in plant defence.** *Quarterly Review of Plant Biology* 1973, **48**(1):3-15.
16. Mauricio R, Rausher MD: **Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense.** *Evolution* 1997, **51**(5):1435-1444.
17. Victorio CP, Moreira CB, Souza MD, Sato A, Arruda RDD: **Secretory Cavities and Volatiles of *Myrrhinium atropurpureum* Schott var. *atropurpureum* (Myrtaceae): An Endemic Species Collected in the Restingas of Rio de Janeiro, Brazil.** *Natural Product Communications* 2011, **6**(7):1045-1050.
18. Porto NM, De Figueiredo R, Oliveira AFM, Agra MD: **Leaf Epidermal Characteristics of *Cissampelos* L. (Menispermaceae) Species from Northeastern Brazil.** *Microscopy Research and Technique* 2011, **74**(4):370-376.
19. Bhatt A, Naidoo Y, Nicholas A: **The foliar trichomes of *Hypoestes aristata* (Vahl) Sol. ex Roem. & Schult var *aristata* (Acanthaceae) a widespread medicinal plant species in tropical sub-Saharan Africa: with comments on its possible phylogenetic significance.** *Biological Research* 2010, **43**(4):403-409.
20. Duarte MR, Lopes JF: **Leaf and stem morphoanatomy of *Petiveria alliacea*.** *Fitoterapia* 2005, **76**(7-8):599-607.
21. Bonzani NE, Barboza GE, Bugatti MA, Espinar LA: **Morpho-histological studies in the aromatic species of *Chenopodium* from Argentina.** *Fitoterapia* 2003, **74**(3):207-225.
22. Wagner GJ, Wang E, Shepherd RW: **New approaches for studying and exploiting an old protuberance, the plant trichome.** *Annals of Botany* 2004, **93**(1):3-11.
23. Delabays N, Simonnet X, Gaudin M: **The genetics of artemisinin content in *Artemisia annua* L. and the breeding of high yielding cultivars.** *Current Medicinal Chemistry* 2001, **8**(15):1795-1801.
24. Ferreira JFS, Laughlin JC, Delabays N, Magalhaes PM: **Cultivation and genetics of *Artemisia annua* L. for increased production of the antimalarial artemisinin.** *Plant Genetic Resources-Characterization and Utilization* 2005, **3**(2):206-229.
25. WHO: **Meeting on the production of artemisinin and artemisinin-based combination therapies** Tanzania, 2005-2006.
26. Olofsson L, Lundgren A, Brodelius PE: **Trichome isolation with and without fixation using laser microdissection and pressure catapulting followed by RNA amplification: Expression of genes of terpene metabolism in apical and sub-apical trichome cells of *Artemisia annua* L.** *Plant Science* 2012, **183**(2012):9-13.
27. Van Nieuwerburgh F, Soetaert S, Podshivalova K, Ay-Lin Wang E, Schaffer L, Deforce D, Salomon DR, Head SR, Ordoukhanian P: **Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing.** *PLoS One* 2011, **6**(10 e26969):1-6.
28. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J: **qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data.** *Genome Biology* 2007, **8**(2):R19.11-R19.14.
29. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR: **Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*.** *Plant Physiology* 2005, **139**(1):5-17.

30. Olofsson L, Engstrom A, Lundgren A, Brodelius PE: **Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L.** *BMC Plant Biol* 2011, **11**(45):1-12.
31. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL: **Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction.** *Bmc Bioinformatics* 2012, **13**(134):1-11.
32. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S *et al*: **High-quality draft assemblies of mammalian genomes from massively parallel sequence data.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(4):1513-1518.
33. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**(7):644-652.
34. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Research* 2008, **36**(10):3420-3435.
35. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biology* 2009, **10**(3):R25.21-R25.10.
36. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *Bmc Bioinformatics* 2011, **12**(323):1-16.
37. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**(1):139-140.
38. Robinson MD, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data.** *Genome Biol* 2010, **11**(3):R25.21-R25.29.
39. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M: **MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant Journal* 2004, **37**(6):914-939.
40. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glockner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.** *Nucleic Acids Research* 2007, **35**(21):7188-7196.
41. Chan PP, Lowe TM: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Research* 2009, **37**(D93-D97):D93-D97.
42. Reeves PH, Ellis CM, Ploense SE, Wu MF, Yadav V, Tholl D, Chetelat A, Haupt I, Kennerley BJ, Hodgens C *et al*: **A Regulatory Network for Coordinated Flower Maturation.** *PLoS Genet* 2012, **8**(2 e1002506):1-17.
43. Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, Guilfoyle TJ, Hagen G, Alonso JM, Cohen JD, Farmer EE *et al*: **Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation.** *Development* 2005, **132**(18):4107-4118.
44. Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D *et al*: **Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses.** *Plant Physiology* 2005, **138**(3):1195-1204.

45. Ram M, Khan MA, Jha P, Khan S, Kiran U, Ahmad MM, Javed S, Abdin MZ: **HMG-CoA reductase limits artemisinin biosynthesis and accumulation in *Artemisia annua* L. plants.** *Acta Physiologiae Plantarum* 2010, **32**(5):859-866.
46. Aquil S, Husaini AM, Abdin MZ, Rather GM: **Overexpression of the HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic *Artemisia annua* plants.** *Planta Medica* 2009, **75**(13):1453-1458.
47. Bouwmeester HJ, Wallaart TE, Janssen MHA, van Loo B, Jansen BJM, Posthumus MA, Schmidt CO, De Kraker JW, Konig WA, Franssen MCR: **Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis.** *Phytochemistry* 1999, **52**(5):843-854.
48. Ryden AM, Ruyter-Spira C, Quax WJ, Osada H, Muranaka T, Kayser O, Bouwmeester H: **The Molecular Cloning of Dihydroartemisinic Aldehyde Reductase and its Implication in Artemisinin Biosynthesis in *Artemisia annua*.** *Planta Medica* 2010, **76**(15):1778-1783.
49. Berteaux CM, Freije JR, van der Woude H, Verstappen FWA, Perk L, Marquez V, De Kraker JW, Posthumus MA, Jansen BJM, de Groot A *et al*: **Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*.** *Planta Medica* 2005, **71**(1):40-47.
50. Covello PS, Teoh KH, Polichuk DR, Reed DW, Nowak G: **Functional genomics and the biosynthesis of artemisinin.** In: *Annual Meeting of the Phytochemical-Society-of-North-American: 2006 2007; Oxford, MS: Pergamon-Elsevier Science Ltd; 2007: 1864-1871.*
51. Van Nieuwerburgh FCW, Castele SRV, Maes L, Goossens A, Inze D, Van Bocxlaer J, Deforce DLD: **Quantitation of artemisinin and its biosynthetic precursors in *Artemisia annua* L. by high performance liquid chromatography - electrospray quadrupole time-of-flight tandem mass spectrometry.** *Journal of Chromatography A* 2006, **1118**(2):180-187.
52. Wang H, Olofsson L, Lundgren A, Brodelius PE: **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.** *American Journal of Plant Sciences* 2011, **2**(4):619-628.
53. Cabello-Hurtado F, Batard Y, Salaun JP, Durst F, Pinot F, Werck-Reichhart D: **Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids.** *J Biol Chem* 1998, **273**(13):7260-7267.
54. Salaun JP, Weissbart D, Durst F, Pflieger P, Mioskowski C: **Epoxidation of cis-delta-9 and trans-delta-9-unsaturated lauric acids by a cytochrome P-450-dependent system from higher-plant microsomes.** *Febs Letters* 1989, **246**(1-2):120-126.
55. Tholl D, Sohrabi R, Huh JH, Lee S: **The biochemistry of homoterpenes - Common constituents of floral and herbivore-induced plant volatile bouquets.** *Phytochemistry* 2011, **72**(13):1635-1646.
56. Lee S, Badieyan S, Bevan DR, Herde M, Gatz C, Tholl D: **Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(49):21205-21210.
57. Donath J, Boland W: **Biosynthesis of acyclic homoterpenes in higher-plants parallels steroid-hormone metabolism.** *J Plant Physiol* 1994, **143**(4-5):473-478.

58. Misra A, Chanotiya CS, Gupta MM, Dwivedi UN, Shasany AK: **Characterization of cytochrome P450 monooxygenases isolated from trichome enriched fraction of *Artemisia annua* L. leaf.** *Gene* 2012, **510**(2):193-201.
59. Tellez MR, Canel C, Rimando AM, Duke SO: **Differential accumulation of isoprenoids in glanded and glandless *Artemisia annua* L.** *Phytochemistry* 1999, **52**(6):1035-1040.
60. Nguyen DT, Gopfert JC, Ikezawa N, MacNevin G, Kathiresan M, Conrad J, Spring O, Ro DK: **Biochemical Conservation and Evolution of Germacrene A Oxidase in Asteraceae.** *Journal of Biological Chemistry* 2010, **285**(22):16588-16598.
61. Hua L, Matsuda SPT: **The molecular cloning of 8-epicedrol synthase from *Artemisia annua*.** *Archives of Biochemistry and Biophysics* 1999, **369**(2):208-212.
62. Mercke P, Crock J, Croteau R, Brodelius PE: **Cloning, expression, and characterization of epi-cedrol synthase, a sesquiterpene cyclase from *Artemisia annua* L.** *Archives of Biochemistry and Biophysics* 1999, **369**(2):213-222.
63. Kirby J, Romanini DW, Paradise EM, Keasling JD: **Engineering triterpene production in *Saccharomyces cerevisiae*-beta-amyrin synthase from *Artemisia annua*.** *Febs J* 2008, **275**(8):1852-1859.
64. Samuels L, Kunst L, Jetter R: **Sealing plant surfaces: Cuticular wax formation by epidermal cells.** In: *Annual Review of Plant Biology*. vol. 59. Palo Alto: Annual Reviews; 2008: 683-707.
65. Kunst L, Samuels L: **Plant cuticles shine: advances in wax biosynthesis and export.** *Current Opinion in Plant Biology* 2009, **12**(6):721-727.
66. Bao XM, Katz S, Pollard M, Ohlrogge J: **Carbocyclic fatty acids in plants: Biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of *Sterculia foetida*.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(10):7172-7177.
67. Yu XH, Rawat R, Shanklin J: **Characterization and analysis of the cotton cyclopropane fatty acid synthase family and their contribution to cyclopropane fatty acid synthesis.** *BMC Plant Biol* 2011, **11**(97):1-10.
68. Duke SO, Paul RN: **Development and fine-structure of the glandular trichomes of *Artemisia annua* L.** *International Journal of Plant Sciences* 1993, **154**(1):107-118.
69. Olsson ME, Olofsson LM, Lindahl AL, Lundgren A, Brodelius M, Brodelius PE: **Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L.** *Phytochemistry* 2009, **70**(9):1123-1128.
70. Inoue K, Tanikawa T, Arai T: **Micro-manipulation system with a two-fingered micro-hand and its potential application in bioscience.** *Journal of Biotechnology* 2008, **133**(2):219-224.
71. Annesley TM: **Ion suppression in mass spectrometry.** *Clin Chem* 2003, **49**(7):1041-1044.
72. Pette D, Reichmann H: **A method for quantitative extraction of enzymes and metabolites from tissue samples in the milligram range.** *J Histochem Cytochem* 1982, **30**(4):401-402.
73. Wu J, Lin L, Chau FT: **Ultrasound-assisted extraction of ginseng saponins from ginseng roots and cultured ginseng cells.** *Ultrason Sonochem* 2001, **8**(4):347-352.
74. Sy LK, Brown GD: **The mechanism of the spontaneous autoxidation of dihydroartemisinic acid.** *Tetrahedron* 2002, **58**(5):897-908.

Websites:

1. http://www.sycos.co.kr/product/index.php?act=eppen&cate=inject_NK2 (03-12-2012)

Part IV. Overall conclusions

The first aim of this PhD project was to optimize a workflow to perform a whole transcriptome analysis of trichome-tissues with 2nd generation sequencing. In the following paragraphs, an overview of these optimizations will be given and this will be linked to the experimental setup of the final RNASeq experiments with whole trichome samples and apical and sub-apical cells. Thereafter, the biological results of these experiments will be discussed.

The first step in this whole optimization procedure was to collect trichome samples with laser capture microdissection. Laser microdissection was facilitated with formaldehyde fixation but this fixation clearly deteriorated the extracted RNA yield and quality. These results were in line with the observations of Olofsson *et al.* [1]. Therefore, if possible, formaldehyde fixation must be avoided. Filamentous and glandular trichomes were easily collected without fixation (Part III Chapter I) but this was not the case for apical and sub-apical cells. To collect the latter (Part III Chapter II), in the first experiment, the sample preparation protocol from Olofsson *et al.* [1] was used. In this protocol, formaldehyde fixation was omitted and microdissection of the plant material was performed in a buffer. Our practical experience was that in this way, laser microdissection to separate apical and sub-apical cells was difficult to perform and leakage occurred. Therefore, another experiment to collect apical and sub-apical cells for RNASeq was performed with fixation.

With laser pressure catapulting, trichome samples were collected in lysis buffer with β -mercaptoethanol. Switching collection tubes for each collected trichome was time consuming, therefore, RNA quality in lysis buffer with β -mercaptoethanol was monitored after 2 h. or 10 min. of incubation at room temperature. After 2 h of incubation, RNA quality was not deteriorated. Other optimization tests indicated that the RNA Nanoprep kit was appropriate to extract RNA from *A. annua* but that RNA with a better quality was obtained if the extraction was performed on ice. After extraction, RNA was linearly amplified to provide enough starting material for 2nd generation sequencing. In the linear amplification process with Ovation RNA-Seq system, oligo dT and random primers were used. An advantage was that with such a method, the coverage at the 5' RNA end is better [2]. Therefore, it was decided to use this amplification kit. In the results of our RNASeq experiment, it was seen that 60 % of the reads obtained were rRNA. That was the drawback of using these random primers and therefore, in the future, it should not be recommend to use the Ovation RNA-Seq system for plant tissues.

The following step after RNA amplification was library preparation for 2nd generation sequencing. Library preparation includes the introduction of a barcode if a group of samples e.g. glandular and filamentous trichomes is sequenced in one lane. After sequencing, the reads can be assigned to the appropriate sample by the barcodes. In the commercially available TruSeq DNA library preparation kit of Illumina, barcodes are ligated before PCR amplification. Introducing different barcodes prior to amplification might result in amplification bias [3-5] in which some products with one barcode might be more efficiently amplified than products with another barcode. To investigate if the multiplex barcodes from the TrueSeq cause amplification bias, a Post Amplification Ligation-Mediated (PALM) protocol was developed which adds the barcodes after the PCR amplification step from the library preparation [6]. From these experiments, the conclusion was drawn that virtually no bias was observed between the different barcodes in both the PALM and TruSeq protocol. Therefore, both protocols were appropriate for analyzing the multiplexed trichome samples and the decision was made to use the PALM protocol.

With the PALM protocol, a library was made of the trichome samples, clusters were generated and trichome samples were single-end 100 bp sequenced on the IlluminaHiSeq platform. A transcriptome assembly was made *de novo* which contained 150,288 contigs and from these contigs, 631 were significantly differentially expressed if glandular and filamentous trichomes were compared (Part III Chapter I). In filamentous trichomes, 204 contigs were significantly more expressed whereas 427 contigs were upregulated in glandular trichomes.

MEP and MVA pathways which produce precursors for the artemisinin biosynthesis were significantly more expressed in glandular trichomes if the comparison was made to filamentous trichomes. All genes were significantly upregulated in the MEP pathway whereas in the MVA pathway, only *acetyl-CoA C-acetyltransferase (AACT)* was upregulated and *3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR)*. This is important since the latter enzyme was reported to limit artemisinin biosynthesis [7, 8].

All known genes coding for enzymes involved in the artemisinin biosynthesis were significantly upregulated in glandular trichomes. These genes were *amorpha-4,11-diene synthase* which cyclises the general sesquiterpene precursor to amorpha-4,11-diene, oxidizing enzymes *CYP71AV1*, *alcohol dehydrogenase 1 (ADH1)*, *aldehyde dehydrogenase 1 (ALDH1)* and the reductase: *artemisinic aldehyde Δ 11(13) double bond reductase*. The enhanced

expression of these genes in glandular trichomes and decreased expression in filamentous trichomes was as expected since *A. annua* plants with only filamentous trichomes lack artemisinin production. As all known artemisinin biosynthesis genes were so clearly differentially expressed in this experiment, these results show that our differential expression analysis of glandular and filamentous trichomes is a good base to pick up artemisinin biosynthesis genes.

To find new candidates, the list of 427 upregulated contigs was screened for cytochrome P450, peroxidases and dioxygenases. This resulted in 5 cytochrome P450 genes significantly upregulated in the glandular trichomes. One cytochrome P450 from the CYP72A family (comp2774) in this list was very recently investigated by Misra *et al.* for its involvement in artemisinin biosynthesis but no *in vitro* activity was detected with dihydroartemisinic acid, artemisinin, artemisinic acid, arteannuinB as substrates [9]. Two other contigs: comp69 and comp548 were respectively annotated as CYP81B1 and CYP82. In both enzyme families, epoxidation reactions were reported in respectively fatty acids and terpenes. The best BLAST-hits of comp15043 were CYP76B1 (7-ethoxycoumarin O-deethylase) and geraniol 10-hydroxylase which hydroxylates the monoterpenoid geraniol [10]. Comp3673 was also in this list of significantly upregulated contigs. Three peroxidase-annotated sequences were upregulated in glandular trichomes: comp252, comp2084 and comp6217 and two dioxygenases: comp225 and comp453. To investigate in future experiments if one of these genes or a combination of these genes is involved in artemisinin biosynthesis, they can be expressed in a yeast strain which contains all artemisinin biosynthesis genes up to dihydroartemisinic acid and subsequently, it can be monitored if artemisinin is detected. Other possibilities are to make over-expression lines for these genes in *A. annua* plants or *in vitro* feeding experiments.

Some genes coding for lipid biosynthesis enzymes showed also a significantly enhanced expression in glandular trichomes for example *fatty acyl-CoA reductase*. This is corroborated by the results from Tellez *et al.* who measured a higher oil content in *A. annua* plants with glandular trichomes [11].

Because of their interesting metabolite content as illustrated with artemisinin, a large number of studies have been devoted to glandular trichomes. In contrast to the extensive literature describing glandular trichomes, less attention has been paid to non-glandular trichomes. Non-

glandular trichomes are assumed to form a physical barrier by steric hindrance of herbivores. To investigate if filamentous trichomes are involved in the production of interesting secondary metabolites, the contigs upregulated in filamentous trichomes were investigated (Part III Chapter I). Some specific genes from sesquiterpenoid and triterpenoid pathways such as 8-*epi*-cedrol synthase and an uncharacterized oxidosqualene cyclase were detected significantly more in filamentous than in glandular trichomes. Our results underscore the vast metabolic capacities of *A. annua* glandular trichomes but nonetheless point to the existence of specific terpene metabolic pathways in the filamentous trichomes.

In the RNASeq experiment, some plants were treated with jasmonic-acid (Part III Chapter I) to trigger the artemisinin biosynthesis genes and to make transcriptome differences for artemisinin candidate genes more pronounced between glandular and filamentous trichomes. Unfortunately, no difference was observed if plants were treated or not. This was not expected since Maes *et al.* measured higher artemisinin levels after jasmonic-acid treatment [12]. A possible explanation is the use of plants in different developmental stages: Maes *et al.* used young seedlings while in our RNASeq experiment 6-months old plants with closed capitula were used. Some authors report highest artemisinin levels in pre-flowering stages whereas others report the peak during flowering [13]. Another possible explanation is that distinct cultivars can react differently to jasmonic-acid treatment as previously reported by Maes *et al.* [12]. The added value from this experiment was that these jasmonic-acid samples gave extra coverage for the *de novo* assembly and these 3 additional repeats also gave a statistical surplus.

Another part of the RNASeq experiment (Part III Chapter II) was the comparison of apical versus sub-apical cells. Since Olsson *et al.* [14] detected transcriptional expression of ADS, CYP71AV1 and DBR2, three enzymes involved in artemisinin biosynthesis only in apical cells and not in sub-apical cells, genes with an enhanced expression in apical and not in sub-apical cells might also be candidate genes for artemisinin biosynthesis. If the transcriptome of apical and sub-apical cells was compared, significantly differentially expressed genes were observed but these differences were not clearly linked to functional pathways and none of these differences was known to be involved in artemisinin biosynthesis. Olofsson *et al.* [1] also detected the expression of artemisinin biosynthesis genes in both cell types which is in contrast with their previous results [14]. In this PhD project, attempts were also made to detect artemisinin metabolites in apical and sub-apical cells with HPLC-MS/MS (Part III

Chapter II), but it was impossible to localize artemisinin metabolites in a specific cell-type of glandular trichomes since a lot of metabolites were present on the outer plant surface.

References

Articles and books:

1. Olofsson L, Lundgren A, Brodelius PE: **Trichome isolation with and without fixation using laser microdissection and pressure catapulting followed by RNA amplification: Expression of genes of terpene metabolism in apical and sub-apical trichome cells of *Artemisia annua* L.** *Plant Science* 2012, **183**(2012):9-13.
2. Clement-Ziza M, Gentien D, Lyonnet S, Thiery JP, Besmond C, Decraene C: **Evaluation of methods for amplification of picogram amounts of total RNA for whole genome expression profiling.** *Bmc Genomics* 2009, **10**(246):1-15.
3. Lopez-Barragan MJ, Quinones M, Cui KR, Lemieux J, Zhao KJ, Su XZ: **Effect of PCR extension temperature on high-throughput sequencing.** *Molecular and Biochemical Parasitology* 2011, **176**(1):64-67.
4. Schutze T, Rubelt F, Repkow J, Greiner N, Erdmann VA, Lehrach H, Konthur Z, Glokler J: **A streamlined protocol for emulsion polymerase chain reaction and subsequent purification.** *Analytical Biochemistry* 2011, **410**(1):155-157.
5. Schutze T, Arndt PF, Menger M, Wochner A, Vingron M, Erdmann VA, Lehrach H, Kaps C, Glokler J: **A calibrated diversity assay for nucleic acid libraries using DiStRO-a Diversity Standard of Random Oligonucleotides.** *Nucleic Acids Research* 2010, **38**(4 e23):1-5.
6. Van Nieuwerburgh F, Soetaert S, Podshivalova K, Ay-Lin Wang E, Schaffer L, Deforce D, Salomon DR, Head SR, Ordoukhanian P: **Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing.** *PLoS One* 2011, **6**(10 e26969):1-6.
7. Ram M, Khan MA, Jha P, Khan S, Kiran U, Ahmad MM, Javed S, Abdin MZ: **HMG-CoA reductase limits artemisinin biosynthesis and accumulation in *Artemisia annua* L. plants.** *Acta Physiologiae Plantarum* 2010, **32**(5):859-866.
8. Aquil S, Husaini AM, Abdin MZ, Rather GM: **Overexpression of the HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic *Artemisia annua* plants.** *Planta Medica* 2009, **75**(13):1453-1458.
9. Misra A, Chanotiya CS, Gupta MM, Dwivedi UN, Shasany AK: **Characterization of cytochrome P450 monooxygenases isolated from trichome enriched fraction of *Artemisia annua* L. leaf.** *Gene* 2012, **510**(2):193-201.
10. Collu G, Unver N, Peltenburg-Looman AM, van der Heijden R, Verpoorte R, Memelink J: **Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis.** *FEBS Lett* 2001, **508**(2):215-220.
11. Tellez MR, Canel C, Rimando AM, Duke SO: **Differential accumulation of isoprenoids in glanded and glandless *Artemisia annua* L.** *Phytochemistry* 1999, **52**(6):1035-1040.
12. Maes L, Van Nieuwerburgh FCW, Zhang YS, Reed DW, Pollier J, Castele S, Inze D, Covello PS, Deforce DLD, Goossens A: **Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants.** *New Phytologist* 2011, **189**(1):176-189.
13. Ferreira JFS, Janick J: **Distribution of Artemisinin in *Artemisia annua*.** *Progress in new crops* 1996:579-584.

14. Olsson ME, Olofsson LM, Lindahl AL, Lundgren A, Brodelius M, Brodelius PE: **Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L.** *Phytochemistry* 2009, **70**(9):1123-1128.

Part V. Summary

In ancient Chinese medicine, extracts from *Artemisia annua* L. were used to cure people suffering from malaria. The active molecule in this herbal therapeutic was characterized as artemisinin, a sesquiterpene with an endoperoxide bridge. Due to the emerging resistance of *Plasmodium* species against chloroquine, mefloquine and sulfadoxine-pyrimethamine, the World Health Organization (WHO) recommends the use of artemisinin-based combination therapies. Since every year around 216 million people are infected with malaria, a high supply of artemisinin is needed at a reduced cost. To produce artemisinin in a more cost-effective way, several strategies are followed. Plants of *A. annua* produce low amounts of artemisinin (0.01-0.8% of dry weight) and by crossing high-producing plants, artemisinin production is enhanced. Another strategy is biosynthesis of artemisinin in heterologous hosts. Despite attempts made for engineering *E. coli* and *Saccharomyces cerevisiae* to produce artemisinin, only the production of precursors from artemisinin was achieved. This is due to an incomplete knowledge of the biosynthetic pathway of artemisinin. Therefore, the discovery of new candidate genes is the focus of this PhD project.

On leaves, stems and inflorescences of *A. annua*, artemisinin is produced in glandular trichomes which are specialized secretory hairs. Additionally, non-glandular hairs with a filamentous T-shape are present but these do not produce artemisinin. By comparison of gene expression in glandular and filamentous trichomes, new candidate genes were discovered. Glandular and filamentous trichomes were collected from flower heads with laser microdissection and laser pressure catapulting and RNA was isolated and amplified. The optimization of these procedures is discussed in Part II Chapter II.

Thereafter, the transcriptome was sequenced with 2nd generation sequencing. During library preparation, samples were prepared for cluster generation and sequencing by ligation of adaptors. If a group of samples e.g. glandular and filamentous trichomes is sequenced in one lane, library preparation includes the introduction of a barcode. These barcodes are short DNA fragments with a unique combination of nucleotides and they reflect which read is originating from which sample. In the commercially available TruSeq DNA library preparation kit of Illumina, barcodes are ligated before PCR amplification. Introducing different barcodes prior to amplification might result in amplification bias in which some products with one barcode might be more efficiently amplified than products with another barcode. To investigate if this type of bias is actually a problem and to prepare the trichome samples, a protocol was developed (Part II Chapter III) that adds barcodes after amplification

called Post Amplification Ligation-Mediated (PALM). In both the PALM and TruSeq protocol, the results showed virtually no bias between the different barcodes.

Trichome samples were single-end 100 bp sequenced on the IlluminaHiSeq platform. A transcriptome assembly was made *de novo* which contains 150,288 contigs. From these contigs, 631 were significantly differentially expressed if glandular and filamentous trichomes were compared (Part III Chapter I). In filamentous trichomes, 204 contigs were significantly more expressed whereas 427 contigs were upregulated in glandular trichomes. In glandular trichomes, as expected, all known genes involved in the biosynthesis pathway of artemisinin were upregulated. MEP and MVA pathways that produce precursors for the artemisinin biosynthesis, were also significantly upregulated in glandular trichomes. In addition to this, other sesquiterpene biosynthesis and monoterpene pathways as well as lipid biosynthesis pathways were predominantly expressed in glandular trichomes. Novel cytochrome P450-, peroxidase- and dioxygenases-encoding genes highly expressed in glandular trichomes were detected and these might be potential candidate genes for the formation of the endoperoxide bridge in artemisinin.

Because of their interesting metabolite content as illustrated with artemisinin, a large number of studies have been devoted to glandular trichomes. In contrast to the extensive literature describing glandular trichomes, less attention has been paid to non-glandular trichomes. Non-glandular trichomes are assumed to form a physical barrier by steric hindrance of herbivores. To investigate if filamentous trichomes are involved in the production of interesting secondary metabolites, the contigs upregulated in filamentous trichomes were investigated (Part III Chapter I). Some specific genes from sesquiterpenoid and triterpenoid pathways such as 8-epi-cedrol synthase and oxidosqualene cyclase were detected significantly more than in glandular trichomes. Our results underscore the vast metabolic capacities of *A. annua* glandular trichomes but nonetheless point to the existence of specific terpene metabolic pathways in the filamentous trichomes.

Glandular trichomes are composed of 10 cells from which 6 cells are surrounded by a subcuticular cavity in which metabolites are secreted. The 2 secretory cells on top (apical) of the glandular trichome are morphologically different from the 4 secretory cells below (sub-apical) and might have another metabolic function. Therefore, it is possible that only one of these cell-types is involved in artemisinin production. To test this hypothesis, apical and sub-apical

cells as well as complete glandular and filamentous trichomes were collected with laser microdissection and laser pressure catapulting for performing metabolite analysis (Part III Chapter II). Also an attempt was made to collect the content of the sub-cuticular cavity with micropipetting. Metabolites were measured with HPLC-MS/MS. From these experiments, it was concluded that these methods were not suitable for detecting metabolites present in the cells. For both collection techniques, an aqueous preparation was made by chopping closed flower heads. During this procedure, metabolites from the wax layer surrounding the plant and metabolites from chopped cells were dispersed in this aqueous medium and caused contamination of the samples. This was suspected since artemisinin was detected in filamentous trichomes. The final proof of this contamination was obtained by adding mint trichomes in this aqueous medium. After collecting these trichomes with the laser microscope, artemisinin was detected in these mint trichomes. Therefore, it was impossible to localize artemisinin metabolites in a specific cell-type of glandular trichomes.

Another attempt was made to look for differences in apical and sub-apical cells at the transcriptome level (Part III Chapter II). If the transcriptome of apical and sub-apical cells was compared, significantly differentially expressed genes were observed but these differences were not clearly linked to functional pathways and none of these differences was known to be involved in artemisinin biosynthesis.

Part VI. Samenvatting

De traditionele Chinese geneeskunde maakte gebruik van extracten van zoete alsem (*Artemisia annua* L.) voor de behandeling van malaria. Het actieve bestanddeel werd later gekarakteriseerd als artemisinine, een sesquiterpeen met endoperoxide brug. Door de opkomst van *Plasmodium* parasieten die resistent zijn aan chloroquine, mefloquine en sulfadoxine-pyrimethamine, beveelt de wereldgezondheidsorganisatie het gebruik aan van artemisinine-gebaseerde combinatietherapieën. Omdat elk jaar ongeveer 216 miljoen mensen geïnfecteerd worden met malaria is er nood aan veel artemisinine tegen een lage prijs. Om artemisinine op een goedkopere manier te produceren worden verschillende strategieën gebruikt. Zoete alsem bevat een relatief laag gehalte aan artemisinine (0.01-0.8% van het drooggewicht). Door planten met een hoog artemisinine gehalte te kruisen wordt een cultivar ontwikkeld die meer artemisinine aanmaakt. Een andere strategie die toegepast wordt is de productie van artemisinine in gist. Door gekende genen van de artemisinin biosynthese weg in gist tot expressie te brengen zijn wetenschappers erin geslaagd om precursoren van artemisinine aan te maken maar niet artemisinine zelf. Dit komt doordat nog niet alle enzymen gekend zijn die een rol spelen in de aanmaak van artemisinine. De zoektocht naar deze ontbrekende genen is de focus van dit doctoraatsproject.

Op de blaadjes, stengels en bloemknoppen van zoete alsem zijn een soort microscopisch kleine klierachtige haartjes aanwezig die ook wel glandulaire trichomen genoemd worden. Deze haartjes zijn verantwoordelijk voor de productie van artemisinine. Naast glandulaire trichomen zijn er ook T-vormige haartjes aanwezig maar die hebben geen zichtbare klierfunctie en maken geen artemisinine aan. Deze haartjes worden filamenteuze trichomen genoemd. Door het vergelijken van genexpressie in glandulaire en filamenteuze trichomen worden nieuwe kandidaat genen gezocht die een rol kunnen spelen in de productie van artemisinine. Tijdens dit doctoraatsproject werden glandulaire en filamenteuze trichomen van bloemhoofdjes verzameld met behulp van een laser microscoop door laser microdissectie en laser katapultatie. Van deze stalen werd het RNA geïsoleerd en geamplificeerd. De optimalisatie van deze procedures wordt besproken Part II Chapter II.

Vervolgens werd het transcriptoom gesequeneerd met 2^{de} generatie sequencing. Tijdens de library aanmaak worden de stalen voorbereid voor de sequencing door het ligeren van adaptoren. Als een groep stalen (vb. glandulaire en filamenteuze trichomen) wordt gesequeneerd in 1 baan, dan moeten barcodes toegevoegd worden tijdens de library aanmaak. Deze barcodes zijn korte DNA fragmentjes met een specifiek herkenbare combinatie aan

nucleotiden. Met behulp van die barcode kan achteraf teruggevonden worden welke data van welk staal afkomstig is. In de commercieel beschikbare DNA library voorbereidingskit TruSeq van Illumina worden barcodes geligeerd voor de PCR amplificatie stap. De introductie van specifieke barcodes per staal voor amplificatie zou kunnen resulteren in een betere amplificatie van sommige fragmenten met een bepaalde barcode en mindere amplificatie van fragmenten met een andere barcode. Om te onderzoeken of dit werkelijk een probleem geeft en om de trichoomstalen te analyseren werd een protocol ontwikkeld (Part II Chapter III) waarin barcodes na amplificatie worden toegevoegd. Als dit nieuw ontwikkelde protocol met het TruSeq protocol vergeleken werd, werd met beide methodes zo goed als geen afwijking waargenomen tussen de verschillende barcodes.

Het transcriptoom van de trichoomstalen werd enkelzijdig gesequeneerd in reads van 100 bp op het IlluminaHiSeq platform (Part III Chapter I). De reads werden geassembleerd tot *de novo* transcriptoom fragmenten. In totaal werden er 150,288 fragmenten of contigs gegenereerd. Van deze contigs kwamen er 631 significant verschillend tot expressie als glandulaire met filamenteuze trichomen vergeleken werden. In filamenteuze trichomen kwamen 204 contigs significant meer tot expressie en in glandulaire trichomen waren 427 contigs upgereguleerd. Zoals verwacht kwamen alle gekende artemisinine biosynthese genen meer tot expressie in glandulaire trichomen. Ook de mevalonate (MVA) en 2-C-methyl-D-erythritol 4-phosphate (MEP) biosynthese pathways kwamen meer tot expressie in de glandulaire trichomen. Deze pathways maken precursoren voor de aanmaak van artemisinine. Daarnaast kwamen andere sesquiterpeen- en monoterpeen- alsook de lipide-biosyntheseweg predominant voor in glandulaire trichomen. Ook nieuwe kandidaatgenen voor de artemisinin biosynthese werden opgepikt. Deze genen kwamen significant meer tot expressie in glandulaire trichomen, behoren tot de cytochrom P450, peroxidase of dioxygenase coderende genen en kunnen mogelijks betrokken zijn bij de endoperoxide brug formatie in artemisinine.

Veel interessante metabolieten worden geproduceerd in glandulaire trichomen en dit type trichomen werd dan ook uitvoerig bestudeerd in de literatuur. Aan niet glandulaire trichomen of filamenteuze trichomen werd minder aandacht besteed. Van niet glandulaire trichomen wordt aangenomen dat ze de plant beschermen door sterisch hinderen van herbivoren. Om uit te zoeken of filamenteuze trichomen ook specifiek interessante secundaire metabolieten produceren werd onderzocht welke contigs significant meer tot expressie kwamen in

filamenteuze trichomen (Part III Chapter I). Dit waren sommige genen van sesquiterpeen en triterpeen pathways zoals 8-epi-cedrol synthase en oxidosqualene cyclase. Deze resultaten bevestigen niet enkel het belang van glandulaire trichomen als productie plaats voor secundaire metabolieten maar geven ook een indicatie van de aanwezigheid van belangrijke specifieke terpeen biosynthese pathways in filamenteuze trichomen.

Glandulaire trichomen zijn opgebouwd uit 10 cellen waarvan 6 cellen omgeven zijn door een sub-cuticulaire blaas waarin metabolieten gesecreteerd worden. De 2 secretorische cellen aan de top (apicaal) van de glandulaire trichoom verschillen morfologisch van de 4 secretorische cellen die daaronder liggen (sub-apicaal). Deze morfologische verschillen zijn een indicatie dat deze celtypes mogelijk verschillende functies uitoefenen. Het is mogelijk dat maar 1 van beide celtypes betrokken is bij de productie van artemisinine. Daarom werden zowel apicale, sub-apicale als volledige glandulaire en filamenteuze trichomen voor een metabolietanalyse verzameld met laser microdissectie en laser katapultatie (Part III Chapter II). Ook werd geprobeerd om de inhoud van de sub-cuticulaire holte te verzamelen met een micropipet. De metaboliet gehalten werden bepaald met HPLC-MS/MS. Uit de resultaten kon geconcludeerd worden dat deze methode niet geschikt was om de aanwezigheid van metabolieten in de cellen te meten. Voor de collectie met laser katapultatie en micropipetteren werd een preparaat gemaakt door bloemknoppen te verhakken in water. Hierbij migreerden metabolieten afkomstig van de waslaag rondom de bloemknoppen en metabolieten uit de opengehakte cellen in het water en dit veroorzaakte contaminatie van de stalen. Dit werd aangetoond door munt trichomen in waswater toe te voegen: na collectie met de laser microscoop en analyse met HPLC-MS/MS werd artemisinine ook gedetecteerd in deze munt trichomen. Daaruit werd besloten dat het met deze methode niet mogelijk was om artemisinine te lokaliseren in een specifiek celtype van de glandulaire trichomen.

Vervolgens werden de problemen op metaboliet niveau omzeild door op transcriptoom niveau te kijken naar waar de gekende artemisinine biosynthese genen tot expressie komen in de glandulaire trichomen (Part III Chapter II). Als het transcriptoom van apicale en sub-apicale cellen vergeleken werd, werden significante verschillen in expressie niveaus gedetecteerd maar deze verschillen konden niet gelinkt worden aan functionele pathways en daarenboven was geen enkel van deze verschillen een gekend artemisinine biosynthese gen.

Part VII. Abbreviations

List of abbreviations

AACT: acetyl-CoA C-acetyltransferase

ACP: acyl carrier protein

ACT: artemisinin-based-combination therapy

ADS: amorpha-4,11-diene synthase

ADH1: alcohol dehydrogenase 1

ALDH1: aldehyde dehydrogenase 1

APS: adenosine 5' phosphosulfate

CASAVA: Consensus Assessment of Sequence and Variation

CMK: 4-cytidine 5'-diphospho-2-C-methyl-D-erythritol kinase

CNAP: Centre for Novel Agricultural Products

CYP71AV1: amorpha-4,11-diene monooxygenase

DBR2: artemisinic aldehyde Δ 11(13) double bond reductase

DMAPP: dimethylallyl pyrophosphate

DXP: 1-deoxy-D-xylulose-5-phosphate

DXS: 1-deoxy-D-xylulose-5-phosphate synthase

DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase

EST: expressed sequence tags

FDS: farnesyl diphosphate synthase

FT: filamentous trichome

GDP: geranyl diphosphate

GGDP: geranylgeranyl diphosphate

GO: gene ontology

GT: glandular trichome

HDS: hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase

HDR: hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase

HMGR: 3-hydroxy-3-methyl-glutaryl coenzyme A reductase

HMGS: 3-hydroxy-3-methyl-glutaryl coenzyme A synthase

HPLC-ESI-Q-TOF: high pressure liquid chromatography with electrospray ionisation and time of flight

HPLC-MS/MS: high pressure liquid chromatography with tandem mass spectrometry

Hsp: high-scoring segment pairs

IC50: half maximal inhibitory concentration

IDI: isopentenyl diphosphate isomerase

IPP: isopentenyl pyrophosphate

LSU: large subunit

JA: jasmonic acid

KAS: β -ketoacyl-ACP synthase

MCS: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase

MCT: 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase

MEP: 2-C-methyl-D-erythritol 4-phosphate

MVA: mevalonate

MVK: mevalonate kinase

NAT: nucleic acid-based tests

P.A.L.M.: photo ablation laser microdissection

PALM: post amplification ligation mediated barcoding

PCR: polymerase chain reaction

PGM: personal genome machine

PMK: phosphomevalonate kinase

PMD: diphosphomevalonate decarboxylase

PP2AA3: protein phosphatase 2A subunit A3

PPR: pentatricopeptide repeat superfamily protein

RDT: rapid diagnostic test

RED1: dihydroartemisinic aldehyde reductase

RIN: RNA integrity number

RNASeq: RNA sequencing

RQI: RNA quality indicator

RT-PCR: reverse transcriptase PCR

RSEM: RNASeq by Expectation Maximization

SEM: scanning electron microscopy

SERCAs: sarco/endoplasmic reticulum Ca²⁺ ATPase

SOLiD: sequencing by oligo ligation detection

SMRT: single molecule real time sequencing

SPIA: single primer isothermal amplification

SSU: small subunit

TFAR1: fatty acyl-CoA reductase 1

TMM: trimmed mean of M values

UCSC: University of California, Santa Cruz

U_OSC: uncharacterized oxidosqualene cyclase

WHO: World Health Organisation

Part VIII. Curriculum vitae

Personal Information

Name Sandra Soetaert
Address Begonialaan 9, 9140 Steendorp
Gsm 0496/05.89.68
Date of birth 14/08/1985
Nationality Belgian

Education and working experience

May 2013-current **Supervisor of chemical lab LQC2: quality control (Pfizer Puurs)**
Apr.-March 2013 **Project engineer (Pfizer Puurs)**
2008-2013 **PhD student (FWO fellowship)** in the lab of Pharmaceutical Biotechnology (Ghent University) and the lab of Secondary Metabolites (PSB/VIB).
Title: “Trichomes, the key to an increased production of artemisinin in *Artemisia annua*” (abstract in attachment).
2007-2008 **Master after Master Applied Pharmaceutical Sciences** obtained with great distinction at Ghent University.
Thesis: “The development and validation of a multiplex STR-PCR for forensic research”.
2005-2008 **Teacher training** obtained with distinction at Ghent University.
Teaching practice: Sint-Janscollege in Sint-Amandsberg.
2003-2007 **Master in Biology** obtained with great distinction at Ghent University.
Thesis: “Biological characterization and screening of East- en West-Flanders on the presence of the facultative parasitic nematode *Halicephalobus gingivalis*”.
1997-2003 Secondary school: Science and modern languages obtained with distinction at Edugo campus de Toren (Oostakker).

Practical experience

- **2nd generation sequencing**: experimental design, lab-work and bio-informatics analysis.
- **Laser microdissection and pressure catapulting**
- **HPLC-MS/MS**
- **Q-PCR**, multiplex PCR and nested PCR
- Sanger Sequencing

- Preparation of a **validation**-file for an in-house developed multiplex PCR for obtaining an extension of the Belac-accreditation and **implementation** in the forensic laboratory.

Additional Courses and Workshops	Organizer	Date
Project management	Ghent University	2012
Leading, Following and Collaborating	Ghent University	2011
Introduction to Linux for bioinformatics	BITS (VIB)	3 and 7/10/2011
Comparative & Regulatory Genomics in Plants	Ghent University	13-15/04/2011
qPCR experiment design and data analysis	Biogazelle	8-9/11/2010
Introduction to Bioinformatics	BITS (VIB)	4-5/11/2010
Advanced academic English: Writing Skills	Ghent University	2010

Scientific awards

2009 Presentation award Terpnet 2009 (Tokyo, Japan)

Teaching experience

2008-2012 Practicum Biotechnology
 2008-2012 Practicum Phytotherapy/Pharmacognosy

A1 Publications (Articles in Web of Science)

Filip Van Nieuwerburgh, Sandra Soetaert, Katie Podshivalova, Eileen Ay-Lin Wang, Lana Schaffer, Dieter Deforce, Daniel R. Salomon, Steven R. Head, Phillip Ordoukhanian: **Quantitative Bias in Illumina TruSeq and a Novel Post Amplification Barcoding Strategy for Multiplexed DNA and Small RNA Deep Sequencing**. *PLoS ONE* 2011, **6**(10): e26969 IF: 4.411

Sandra SA Soetaert, Christophe MF Van Neste, Mado Vandewoestyne, Steven R Head, Alain Goossens, Filip CW Van Nieuwerburgh, Dieter DL Deforce: **Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua***. Submitted to BMC Plant Biology.

Bart J.G. Broeckx, Frank Coopman, Geert E.C. Verhoeven, Wim Van Haeringen, Leanne van de Goor, Tim Bosmans, Ingrid Gielen, Jimmy H. Saunders, Sandra S.A. Soetaert, Henri Van Bree, Christophe Van Neste, Filip Van Nieuwerburgh, Bernadette Van Ryssen, Elien Verelst, Katleen Van Steendam, Dieter Deforce: **The prevalence of genetic disorders in a dog population from Belgium, the Netherlands and Germany**. Submitted to *PLoS ONE*.

Tessa Moses, Jacob Pollier, Sandra Soetaert, Marie-Laure Erffelinck, Robin Vanden Bossche, Johan M. Thevelein, Dieter Deforce and Alain Goossens: **Reconstitution of *Artemisia annua* triterpene biosynthetic pathway in yeast and tobacco**. Manuscript in preparation.

C3 Publications and Presentations

Poster Conference on Advances in Canine and Feline Genomics and Inherited Diseases 2012 (Visby, Sweden): **Relation between the FBN2 haplotype and phenotypical hip dysplasia**. Broeckx B., Coopman F., Verhoeven G., Bosmans T., Criel A., Gielen I., Saunders J., Soetaert S., van Bree H. and Van Haeringen W., Van Nieuwerburgh F., Van Ryssen B., Deforce D

Poster Terpnet 2011 (Kalmar, Sweden): **Transcriptome analysis of apical and sub-apical cells of *Artemisia annua* trichomes with next-generation-sequencing**. Sandra S.A. Soetaert, Filip C.W. Van Nieuwerburgh, Peter E. Brodelius, Alain Goossens, Dieter L.D. Deforce

Oral presentation Terpnet 2009 (Tokyo, Japan): **Metabolite analysis of laser captured apical and sub-apical cells of *Artemisia annua* trichomes**. Sandra S.A. Soetaert, Filip C.W. Van Nieuwerburgh, Sofie R.F. Vande Castele, Mikael E. Olsson, Mado Vandewoestyne, Peter E. Brodelius, Alain Goossens, Dieter L.D. Deforce

Poster Terpnet 2009 (Japan): **Phytohormonal regulation of gland formation and artemisinin biosynthesis in *Artemisia annua***. Filip Van Nieuwerburgh, Lies Maes, Sofie Vande Castele, Sandra Soetaert, Marc Van Montagu, Dirk Inzé, Dieter L.D. Deforce, Alain Goossens

Part IX. Dankwoord

Als laatste zou ik graag iedereen bedanken die rechtstreeks of onrechtstreeks geholpen heeft om mijn doctoraat tot een goed einde te brengen. Een doctoraat is een project met “ups” en “downs”, vandaar dat steun van collega’s, vrienden en familie zo belangrijk is om de motivatie te vinden om door te gaan. Andere doctoraatsstudenten wil ik graag de raad geven om nooit op te geven, de “ups” komen meestal pas op het einde...

Dieter, bedankt om mij als promotor onder je vleugels te nemen en mij te steunen dit doctoraat af te ronden. Je hebt mij de kans gegeven om heel zelfstandig te werken maar als ik ergens met een technisch probleem zat, dan kon ik altijd bij jou terecht en wist jij vaak heel creatieve oplossingen te bedenken.

Alain, als co-promotor was jij steeds beschikbaar om mij met raad en daad bij te staan. Je woorden “Je ziet dat het bij de rest ook niet altijd vlot verloopt.” zullen mij bij blijven. Je expertise op het gebied van plantenbiotechnologie hielp ook om de zaken vanuit een ander perspectief te bekijken.

Filip, als mede co-promotor wil ik je danken voor het schrijven van mijn FWO aanvraag, voor je hulp bij het uitwerken van het 2nd generation sequencing gedeelte en voor de begeleiding gedurende 4 jaar.

Christophe, als bioinformatica buddy, heb je met mij veel mailtjes heen en weer gestuurd om de RNASeq analyse tot een goed eind te brengen. Jouw programmeer skills hebben dit onderzoek in een hogere stroomversnelling gebracht. Toch kan ik nog niet geloven dat een kapotte harde schijf kan gemaakt worden door die in te vriezen... Het was leuk om met jou samen te werken.

Yens, jou wil ik ook graag bedankt voor de praktische hulp in het labo op het einde van mijn doctoraat. Op het einde was het echt rushen en ik was super blij dat ik kon rekenen op een extra paar handen. Jill, Liselotte, Kim en Anne-Marie, de thesisstudenten die geholpen hebben aan dit project, bedankt voor de assistentie!

Sofie, bedankt voor de opleiding tot HPLS-MS analist. Technische problemen worden vlot door jou opgelost, vandaar je bijnaam massa fluisteraar.

Maarten en Katleen, de *pater et mater familias* van “den bokaal” waarbij we altijd terecht konden voor goede raad. Marlies en Trees voor de gezellige babbels en bemoedigende woorden. Veerle en Liesbeth, het vaak komische duo aan den overkant van den bokaal die de dagen opvrolijkt. Pieter, merci voor de veelal filosofische discussies over projecten en Ellen, jouw enthousiasme is moeilijk te evenaren. Paulien, bedankt te luisteren als mijn project eens wat minder vlot ging. Bart, jij wist ook vaak de sfeer erin te brengen met de dierengeneeskundige weetjes van de week. Bert, bedankt voor je droge humor om de sfeer erin te houden en om de IT helpdesk te bemannen. Mado, veel geluk met je nieuwe job. De collega’s van het DNA labo (David, Saskia, Sylvie, Delphine, Petra, Sabine, Eveline, Evelien, Leen, Astrid en Nadine), bedankt voor de aangename babbels en toffe werksfeer. Ik heb ook heel goede herinneringen aan mijn manama thesis bij jullie! Inge, merci voor de hulp bij het regelen van “van alles en nog wat”. Super vlot, altijd met een glimlach en nooit is iets te veel gevraagd.

I would also like to thank the people from NXT-GNT and Mdx at the 3th Floor for the nice working environment and the technical support for transcriptomics analysis.

Tessa and Jacob, thanks for the nice congress company at Terpnet in Sweden and Tokyo. I also liked the in depth discussions about terpenoids, that helped to put things into another perspective. Thanks to the other metabol group members for their amusing BBQs, seminars (Blankenberge) and interesting group discussions.

Chen and Bjorn, thanks for the help with the micropipetting of the glandular cavities. This was a difficult piece of master work!

Mama en papa, Brigitte en Denis, Claudia en Simon, bedankt voor de ontspannen avondjes, weekendjes en wandelingen. Dat was een welkome afwisseling. Bedankt ook voor het begrip tijdens super drukke periodes en de steun bij het zoeken van een job.

Als laatste, bedankt Quinten om mijn steunpilaar te zijn gedurende die 4 jaar. Zonder jou zou ik dit niet volgehouden hebben. Je kon mij echt opbeuren als het eens wat minder ging en je stond steeds klaar voor een diepe wetenschappelijke discussie om een tweede opinie te geven.