

Next generation sequencing unravels the biosynthetic ability of Spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics

Jin *et al.*

RESEARCH ARTICLE

Open Access

# Next generation sequencing unravels the biosynthetic ability of Spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics

Jingjing Jin<sup>1,2,3</sup>, Deepa Panicker<sup>1</sup>, Qian Wang<sup>1</sup>, Mi Jung Kim<sup>1</sup>, Jun Liu<sup>3</sup>, Jun-Lin Yin<sup>1</sup>, Limsoon Wong<sup>2</sup>, In-Cheol Jang<sup>1,4</sup>, Nam-Hai Chua<sup>3</sup> and Rajani Sarojam<sup>1\*</sup>

## Abstract

**Background:** Plant glandular trichomes are chemical factories with specialized metabolic capabilities to produce diverse compounds. Aromatic mint plants produce valuable essential oil in specialised glandular trichomes known as peltate glandular trichomes (PGT). Here, we performed next generation transcriptome sequencing of different tissues of *Mentha spicata* (spearmint) to identify differentially expressed transcripts specific to PGT. Our results provide a comprehensive overview of PGT's dynamic metabolic activities which will help towards pathway engineering.

**Results:** Spearmint RNAs from 3 different tissues: PGT, leaf and leaf stripped of PGTs (leaf-PGT) were sequenced by Illumina paired end sequencing. The sequences were assembled *de novo* into 40,587 non-redundant unigenes; spanning a total of 101 Mb. Functions could be assigned to 27,025 (67%) unigenes and among these 3,919 unigenes were differentially expressed in PGT relative to leaf - PGT. Lack of photosynthetic transcripts in PGT transcriptome indicated the high levels of purity of isolated PGT, as mint PGT are non-photosynthetic. A significant number of these unigenes remained unannotated or encoded hypothetical proteins. We found 16 terpene synthases (TPS), 18 cytochrome P450s, 5 lipid transfer proteins and several transcription factors that were preferentially expressed in PGT. Among the 16 TPSs, two were characterized biochemically and found to be sesquiterpene synthases.

**Conclusions:** The extensive transcriptome data set renders a complete description of genes differentially expressed in spearmint PGT. This will facilitate the metabolic engineering of mint terpene pathway to increase yield and also enable the development of strategies for sustainable production of novel or altered valuable compounds in mint.

**Keywords:** Spearmint, Next generation sequencing, Transcriptome, Glandular trichomes, Terpenes, Carvone, Terpene synthases

## Background

Plants produce an enormous variety of specialised metabolites among which terpenes are the largest and most structurally diverse class of natural products. They are the main components of plant essential oils. Many of these terpenes are produced and stored in specialised secretory structures called glandular trichomes [1,2].

These terpenes provide protection for plants against a variety of herbivores and pathogens [3] and are also commercially quite valuable. Therefore, the processes by which they are synthesised and stored in plants are main target for genetic manipulation for increased yield. But our knowledge about the development of secretory glandular trichomes and terpene production and its regulation is very limited making it difficult to engineer these metabolic pathways [4,5].

Aromatic essential oil produced by *Mentha* species is the source of the best known monoterpenes, menthol

\* Correspondence: rajanis@tll.org.sg

<sup>1</sup>Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore

Full list of author information is available at the end of the article

and carvone, which form the principal components of mint oil. They are extensively used in flavour and fragrance industries, pharmaceuticals and cosmetic products [6]. Peppermint variety mostly produces menthol whereas in spearmint variety carvone dominates [7,8]. From the PGT of peppermint variety (*Mentha X piperita*), 1,316 randomly selected cDNA clones, or expressed sequence tags (ESTs) were produced, which led to the identification of many genes, enzymes and substrates involved in the main menthol essential oil biosynthetic pathway [9,10]. Given the technical limitations at their time of study, an EST approach would possibly identify only cDNAs which are abundant in PGT. A recent proteomic analysis of spearmint PGT identified 1,666 proteins of which 57 were predicted to be involved in secondary metabolism [11]. But generation of sufficient genomic information with deep coverage is required to gain insights into the regulatory mechanism of terpene metabolism and glandular trichome development. This will promote successful engineering for improved yields or to develop mint as a platform for production of novel/altered terpenes. Mint is a well-suited plant for this as it is able to produce and store large amount of oils within PGT instead of exuding it on to the leaf surface. Storage within the PGTs also reduces the loss of volatile oils by emission into the atmosphere.

High-throughput RNA sequencing (RNA-Seq) has increasingly become the technology of choice to generate a comprehensive and quantitative profile of the gene transcription pattern of a tissue. Here, we report comparative analysis of RNA-seq transcriptome of different tissues of spearmint—namely PGT, leaf minus PGT (leaf-PGT) and leaf. The transcriptome data provided a genome-wide insight into the metabolic ability of PGT. Comparison of PGT and leaf-PGT showed that 3,919 unigenes were differentially expressed in PGT (minimum 4 times more in PGT when compared to leaf -PGT). Many of these were related to terpene production and other secondary metabolite pathways. From the various terpene synthases (TPS) transcripts identified, we functionally characterized 2 of these previously uncharacterized TPSs from mint and found them to be sesquiterpene synthases. Key pathway unigene transcripts were verified by qRT-PCR. Our results show the molecular specialisation of PGT for the production of different classes of metabolites.

## Results and discussion

### Spearmint PGT and their development

Spearmint leaves produce three different types of trichomes on their surfaces: non-glandular multicellular hair like, capitate glandular trichomes with a single secretory head cell and PGTs whose secretory head is composed of eight-cells with a single stalk and basal cell (Figure 1A). These PGT glands possess a large subcuticular storage

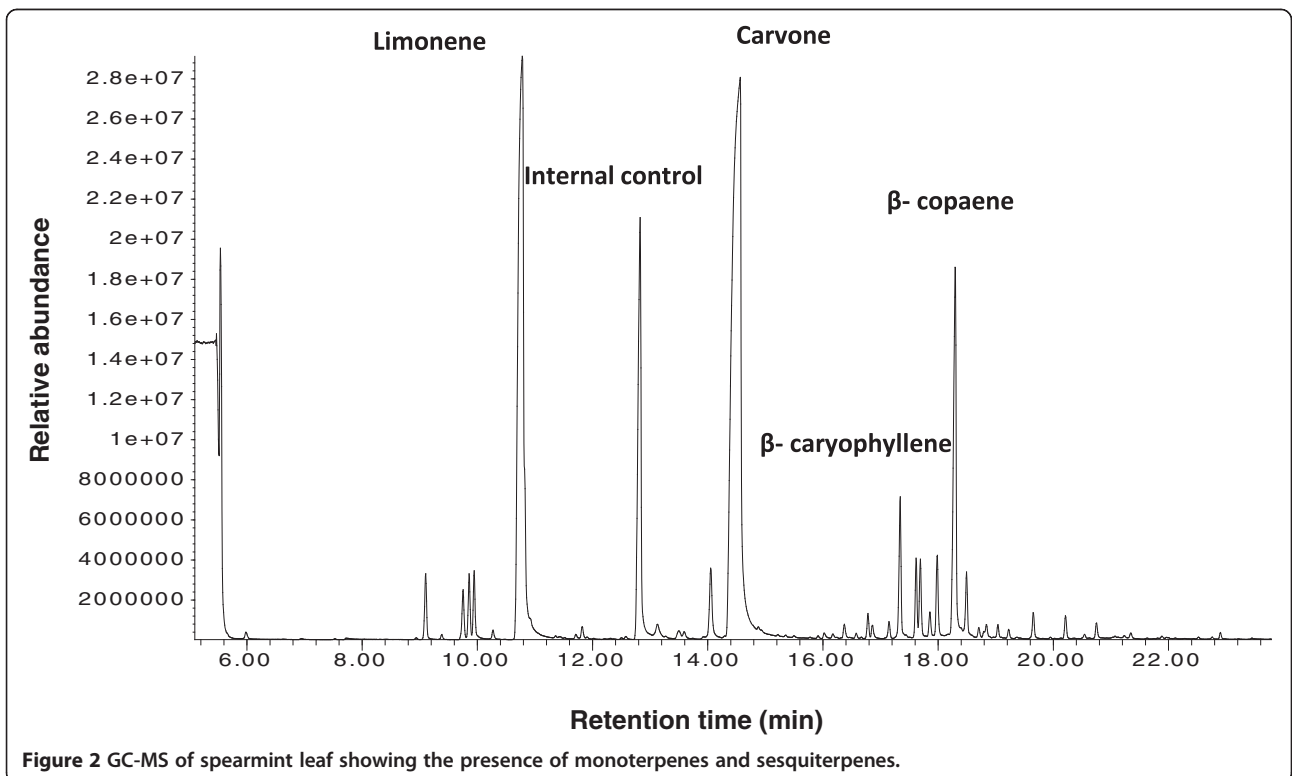
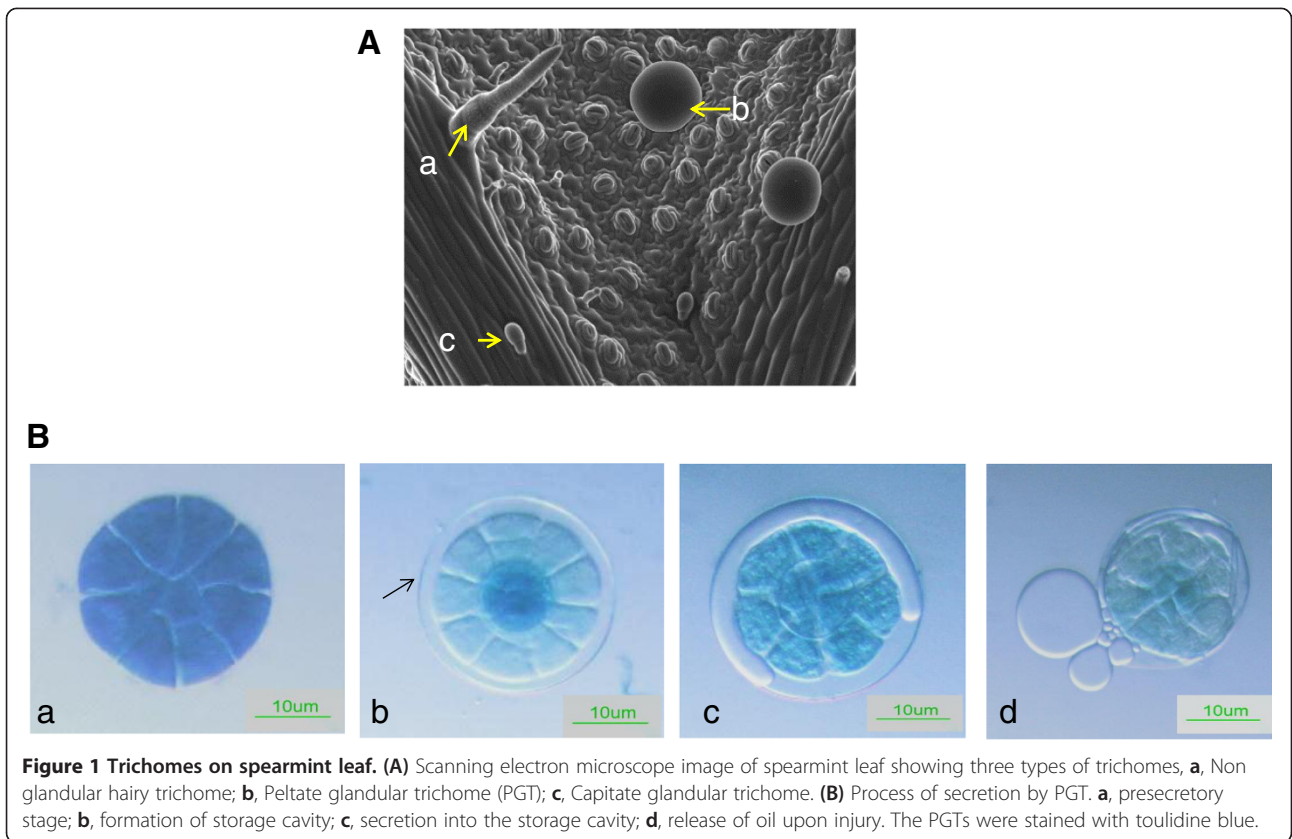
space that is formed by the separation of the cuticle from the apical cells and the essential oil is secreted into this cavity [12] (Figure 1B). It is known that new glands keep initiating on the leaf till expansion ceases and the monoterpene content and compositions change with the age of the leaf [13-16]. Different studies have indicated that monoterpene biosynthesis is most active in young 12–20 day old leaves of peppermint after which the rate of synthesis slowly declines [17-19]. We performed gas chromatography–mass spectrometry (GC-MS) analysis on young spearmint leaves (about 1–2 cm in length) and found abundance of both limonene and carvone monoterpenes (Figure 2). Limonene is the first committed step towards carvone pathway. In addition to these monoterpenes, the presence of sesquiterpenes was also observed. This indicated the dynamic terpene biosynthetic activity of leaves at this stage of development. PGT were purified from leaves of this stage and RNA isolated. The leaves of the same stage were brushed to remove all trichomes and RNA extracted from them as controls (Additional file 1).

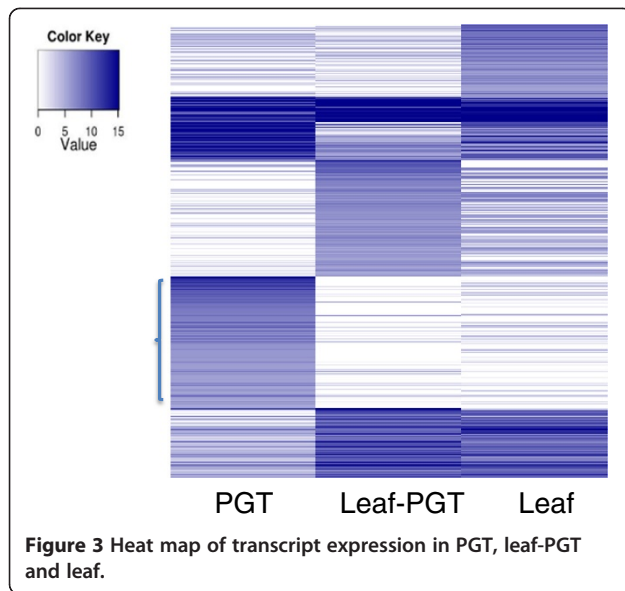
### Sequencing, *de novo* assembly and annotation of transcriptome

Three RNA libraries were prepared and sequenced by Illumina technology. More than 100 million high quality reads of 101 base pairs (bp) were generated from PGT, leaf-PGT and leaf (Additional file 2). Using the Trinity method [20] the sequence reads were finally assembled into 40,587 non-redundant unigenes, spanning a total of 101 Mb of sequence with a GC content of 43.14%. All unigenes were longer than 200 bp. The N50 of the final assembled transcripts was 1,774 bp. The unigenes were annotated by performing BLASTX search against various protein databases. Among the 40,587 non-redundant unigenes, 27,025 (67%) had at least one hit in BLASTX search with E-value  $\leq 1e-3$ . Functional classifications of Gene Ontology (GO) term of all unigenes were performed using Trinotate [20]. In order to calculate the expression level for assembled transcripts, we first mapped reads onto them using bowtie [21]. RSEM (RNA-seq by Expectation-Maximization) was used to estimate the abundance of assembled transcripts and to measure the expression level [22].

### Overview of expression profile of spearmint PGT

From the RNA seq data about 25,000 unigenes were observed to be expressed in spearmint PGT. The heat map in Figure 3 exhibits some specific expression patterns to PGT. Among this specific pattern for PGT we found transcripts for terpene biosynthesis, lipid transfer proteins and interesting transcription factors like MYBs and WRKYs. Comparison of PGT and leaf-PGT showed that 3,919 unigenes were differentially expressed in PGT (Additional file 3). These unigenes showed a minimum





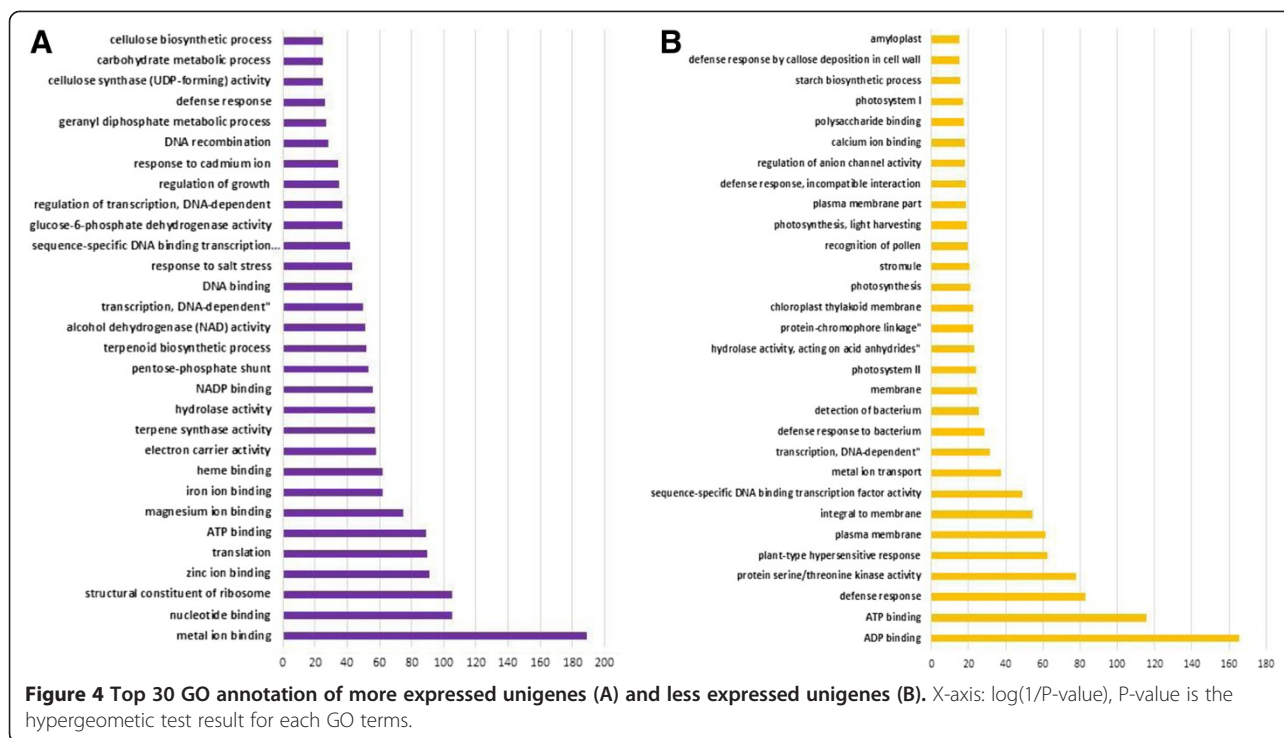
of 4 times increase in expression level in PGT as compared to leaf-PGT. About 30% of these unigenes encoded either hypothetical proteins or remained unannotated. Many of these unannotated unigenes showed none or minimal expression in leaf-PGT. They might represent novel genes that are unique to PGTs development and divergent from other plants, whose genomes have been sequenced. Data from proteomic analysis of spearmint PGT also showed that the largest functional category of the identified proteins was “unclear classification” and included proteins with unknown functions [11]. The absence or low levels of some PGT-specific transcripts from leaf RNA seq data indicated the dilution of PGT-specific RNAs among the total leaf RNAs and reaffirms the importance of isolating these organs for analysis.

Among the top 1000 differentially expressed unigenes, we identified 16 TPSS, 18 cytochrome P450s, 5 lipid transfer proteins (LTPs), 20 transcription factors, 2 ATP-binding cassette (ABC) transporters and several transcripts associated with cell wall. Cytochrome P450s are involved in the hydroxylation of terpenes [23] and LTPs have been suggested to be involved in intracellular transport and secretion of lipids and terpenes [9,24]. LTPs were among the most abundant unigenes in PGT and were confirmed by qRT-PCR (Additional file 4). The abundance of these LTPs suggests their importance in PGTs metabolic function and development. ABC transporters are also proposed to be involved in the active transport of secondary metabolites [11,25]. The spearmint PGT is presumed to undergo cell wall modification to form subcuticular storage space [12]. Among the differentially expressed unigenes there were few that were related to cell wall synthesis or modifications and a subset of these were confirmed by qRT-PCR (Additional files 4 and 5). Whether

they play a role in modification of cell wall layers to form the storage space remains to be investigated.

To characterize biological processes specific for PGT, GO term was determined for all differentially expressed unigenes. Additionally, we identified unigenes whose expression was reduced in PGT by comparing leaf-PGT and PGT. These unigenes showed a minimum of 4 times reduction in expression level in PGT when compared to leaf-PGT. GO term was determined for them as well. Figure 4 shows the top 30 GO terms for the more abundant and less abundant unigenes. GO terms associated with ribosome biogenesis, ribosome structural genes and translation are highly represented in PGT, which could reflect the high protein biosynthetic activity of PGTs. Other terms included terpene metabolism and most of the primary energy producing terms like glycolysis and tricarboxylic acid cycle. Furthermore, pentose phosphate related term (oxidative) was also enriched in PGT. This term provides NADPH for biosynthetic processes such as fatty acid synthesis, cytochrome P450 mediated hydroxylation and the assimilation of inorganic nitrogen [26]. These results indicate that the GO functions that provide energy equivalents and redox cofactors are very active in PGT. Secretory trichomes are biosynthetically very active, and hence, there is a high energy requirement in these cells. Unigenes from the GO terms of photosynthesis, chlorophyll biosynthesis and starch biosynthesis were among the less abundant ones. This shows that our PGT sample preparation was pure and not contaminated with leaf tissues as mint PGTs are non-photosynthetic.

Mint PGTs being non-photosynthetic and metabolically very active would presumably rely on exogenous supply of sucrose from underlying leaf tissues to use as carbon source for energy production. We found several transcripts encoding enzymes for sucrose catabolism expressed more in PGT like sucrose synthase and neutral/alkaline invertases that are important for channelling carbon from sucrose in non-photosynthetic tissues [27]. These enzymes convert sucrose to hexose phosphates. Plastids in the PGTs are the main sites of secondary metabolism. In contrast to chloroplasts, plastids of heterotrophic tissues have to rely on the import of ATP and carbon to drive their metabolic processes. We checked our set of differentially expressed unigenes to see if any known transporters are present. In most plants glucose 6-phosphate seems to be the preferred hexose phosphate taken up by nongreen plastids. The transporter proteins responsible for this import of carbon into plastids are known as Glc6P-phosphate translocator (GPT) and transcript similar to GPT was seen enriched in PGT (about 30 times more in PGT). This carbon can be used for starch biosynthesis or for the oxidative pentose phosphate pathway in plastids [28]. GO terms for oxidative pentose phosphate pathways were seen enriched in PGT. Additionally transcript similar to plastidic Phosphoenolpyruvate/



phosphate translocator was found to be expressed more in PGT. They are involved in the transport of phosphoenolpyruvate, an energy rich glycolytic intermediate from the cytoplasm into the plastids [29]. Further, ATP generated either by glycolysis or by oxidative phosphorylation in mitochondria can be imported into non-green plastids by plastidic nucleotide transporter (NTT). Transcript similar to NTT was also observed to be more abundant in PGT [29].

#### Analysis of spearmint PGT transcription factors

Although the cloning and functional characterization of enzymes involved in terpene biosynthesis has been quite successful in various plants, knowledge about regulation of these secretory trichome specific pathways is very rudimentary. Studies in peppermint show a close association between enzyme activity and transcript abundance for all gene/enzyme pairs, suggesting that, essential oil biosynthesis is primarily influenced at the transcriptional level [12,17,18,30]. Hence, identification of transcription factors that globally control metabolic pathway will provide an attractive strategy for engineering terpene production. Similarly, knowledge about the development of secretory glandular trichomes, the so-called factories of important terpenoid production, is very limited. Most of the transcription factors involved with trichome development have been isolated from Arabidopsis, which lacks secretory trichomes. Studies in tobacco and tomato are beginning to show that multicellular secretory trichomes and unicellular trichomes of Arabidopsis are not homologous structures, and they likely develop under

different regulatory conditions [31]. Our analysis of transcriptome data of young leaves or PGT did not uncover any transcripts that matched those of the major known trichome initiating gene transcripts from Arabidopsis, like *TRANSPARENT TESTA GLABRA1*, *GLABRA1*, *GLABRA3* [32]. Either these genes are not expressed or are expressed at a different developmental stage of leaves or PGT than the stage used in this study. Table 1 shows the top 20 transcription factors that were significantly more abundant in PGT when compared to leaf-PGT.

#### The MEP (2-Cmethyl-D-erythritol-4-phosphate) pathway is more abundant in spearmint PGT than the MVA (mevalonate) pathway

The building blocks for all different classes of terpenes produced by plants are C5 units of isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP). They are generated either by plastidial MEP or cytoplasmic MVA pathway. The MEP pathway requires seven enzymes to synthesize IPP and DMAPP from pyruvate and glyceraldehyde 3 phosphates which feed the monoterpene pathway [33]. From peppermint EST studies it has been proposed that the active pathway for the formation of IPP/DMAPP in the PGT is the MEP pathway. This is consistent with our analysis too where MEP pathway transcripts were more abundant in PGT than MVA. High expression of MEP pathway transcripts correlates well with the production of monoterpenes in PGT. It has been reported that 1-deoxy-D-xylulose-5-phosphate synthase (DXS), the first enzyme of this pathway is important

**Table 1 Top 20 enriched TFs in PGT compared to leaf-PGT**

Name	Leaf	Leaf-PGT	PGT	FC	Arabidopsis ID	Description
comp26629_c1	4.86	1.10	7.57	6.48	AT1G48000	myb
comp29031_c0	7.08	4.51	10.93	6.42	AT2G26580	YABBY
comp37102_c0	2.04	0.00	6.26	6.26	AT2G46150	Late embryogenesis abundant
comp31772_c0	2.48	0.10	6.03	5.93	AT5G13080	WRKY
comp31929_c0	3.97	1.52	7.27	5.75	AT2G47460	myb
comp25071_c0	3.11	1.24	6.93	5.69	AT5G60910	MADS-box
comp34871_c0	2.51	0.00	5.34	5.34	AT3G56850	Basic region/leucine zipper motif
comp34759_c0	1.69	0.00	5.25	5.25	AT1G14600	Homeodomain-like
comp43497_c1	2.42	0.53	5.36	4.84	AT1G68150	WRKY
comp32776_c1	0.70	0.00	4.61	4.61	AT2G41690	Heat shock TF
comp27387_c0	0.00	0.00	3.91	3.91	AT5G07680	NAC
comp35105_c0	2.06	0.00	3.85	3.85	AT5G65790	myb
comp25286_c0	2.13	0.00	3.80	3.80	AT5G01380	Homeodomain like TF
comp30569_c1	2.32	1.08	4.85	3.77	AT5G66350	Lateral root primordium
comp26625_c0	2.15	1.37	5.04	3.67	AT1G75390	bzip domain
comp17332_c0	0.00	0.00	3.65	3.65	AT1G73230	btf3
comp32174_c0	3.56	2.22	5.70	3.48	AT5G13080	WRKY
comp28852_c0	0.95	0.00	3.32	3.32	AT2G17770	Basic region/leucine zipper motif
comp29336_c0	3.03	1.18	4.39	3.21	AT3G24860	Homeodomain-like
comp42078_c1	3.48	0.77	3.84	3.08	AT4G32730	Homeodomain

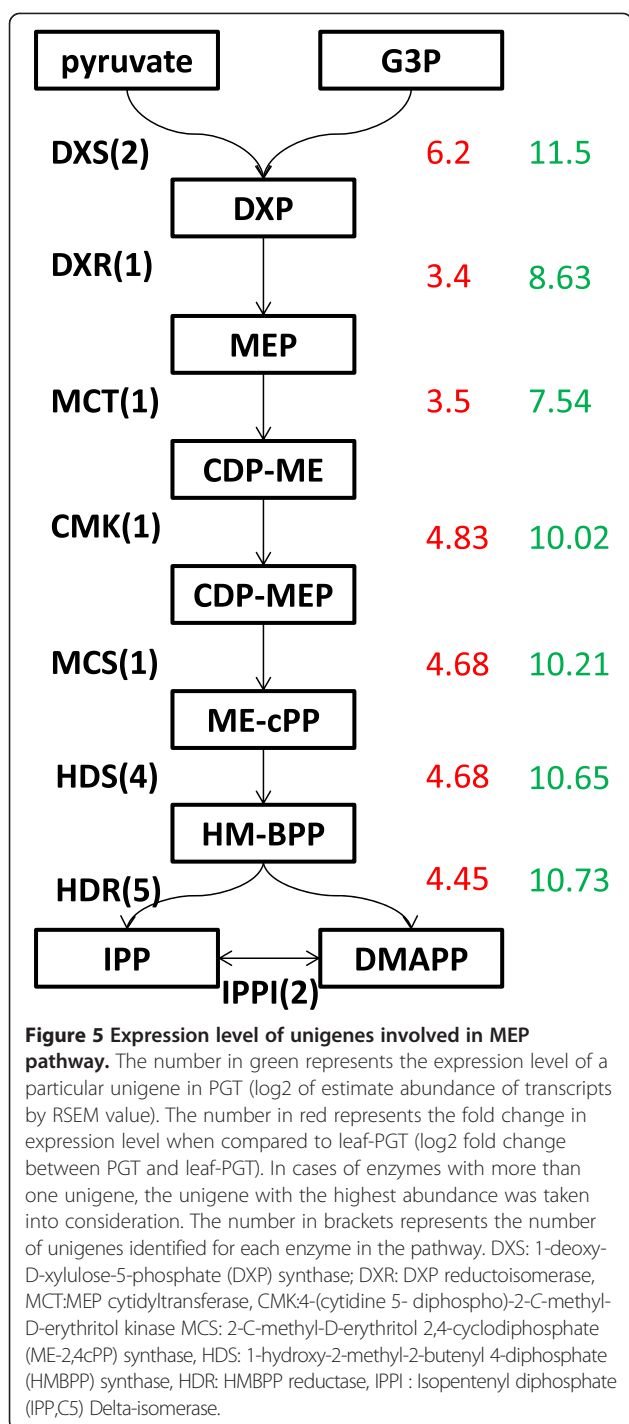
The value is log<sub>2</sub> RESM value for each assembled TFs. FC is the log<sub>2</sub> fold change in PGT when compared to leaf-PGT. Arabidopsis ID is the homolog ID in Arabidopsis protein database.

for the overall regulation of the pathway [34]. Multiple *DXS* genes have been found in plants like *Zea mays*, *Medicago truncatula*, *Oryza sativa*, *Ginkgo biloba* and *Pinus densiflora* and *Picea abies* [35-40]. In all these plants, two or three candidate *DXS* genes have been reported. From our data we were able to identify 2 different 1-deoxy-D-xylulose-5-phosphate synthase (*DXS*) unigenes showing different levels of abundance in PGT. The number of genes coding for each MEP pathway enzyme varies from plant to plant [33,41]. Presence of multiple genes with differential tissue-specific expression levels might contribute towards the regulation of the MEP pathway, in different organs of the plant. Figure 5 shows the number of unigenes identified for each enzyme of the MEP pathway and their RNA seq expression levels. In cases of enzymes with more than one unigene, the unigene with the highest abundance in PGT was taken into consideration. Their expression was further validated by qRT-PCR (Additional file 4). From our RNA seq data and qRT-PCR analysis, *DXR* and *MCT* transcript levels were low when compared to levels of other enzymes in MEP pathway. This might suggest that possibly these two enzymes are the rate limiting steps of this pathway. A possible option to explore in future will be to enhance the expression level of various rate limiting steps to enhance the production of terpenes.

In contrast to the MEP pathway, the transcript levels of MVA enzymes were very low. For RNA seq expression levels of unigenes involved in this pathway refer Additional file 6. Their expression was validated by qRT-PCR (Additional file 4). MVA pathway derived IPP is generally believed to be used for the production of cytosolic sesquiterpenes, triterpenes and mitochondrial terpenes. In addition to monoterpenes, mint produces a few sesquiterpenes although at much lower quantities than monoterpenes [9,42]. Lower level of MVA pathway can be one of the reasons. Interestingly, the transcripts of genes involved in the MEP pathway are also enriched in *Artemisia annua* trichomes, glandular trichomes of Hops and in Snapdragon flowers where sesquiterpene metabolism dominates [43-45] suggesting that the MEP pathway can also feed sesquiterpene production. Studies with labelled substrate predict exchange of metabolites between the MVA and MEP pathways [46,47]. How the IPP/DMAPP formed by MEP pathway is utilized to synthesize sesquiterpenes in mint remains to be investigated.

#### Monoterpene production is enriched in spearmint PGT

Subsequent condensation reactions between IPP and DMAPP are catalysed by GPP synthases (GPPS) that leads to the formation of geranyl diphosphate (GPP; C<sub>10</sub>) the



precursor for monoterpenes. The conversion of IPP to DMAPP and its equilibrium is maintained by IPP isomerase (IPPI). In most plant species this enzyme is encoded by a single gene whereas Arabidopsis has two IPPI genes [33]. We found 2 IPPI unigenes in spearmint both enriched in PGT. Peppermint GPPS (Mp GPPS) is a two-component heteromeric enzyme consisting of a large and a small subunit, and both the subunits are catalytically inactive by

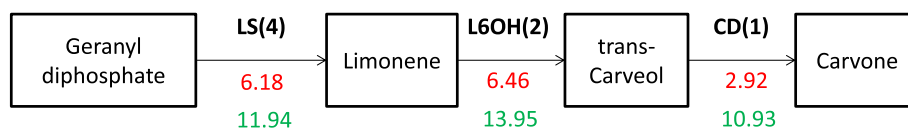
themselves [48,49]. In spearmint too, we found unigenes for both the small and large subunits of GPP synthase that showed high expression in PGT. The major constituent of spearmint essential oil is (-) carvone which is synthesised from GPP in a three step reaction. Transcripts for all the three enzymes involved in above reaction, Limonene synthase (LS), Limonene-6-hydroxylase (L6OH) and carveol dehydrogenase (CD) were highly expressed in PGT and verified by q-RT-PCR (Figure 6 and Additional file 4).

Interestingly, the precursor for menthol in peppermint and (-) carvone in spearmint is the same limonene. In peppermint it is oxygenated by (2)-4S-limonene-3-hydroxylase (L3OH) to form (2)-trans-isopiperitenol and it enters the menthol pathway whereas in spearmint limonene is oxygenated by (2)-4S-limonene-6-hydroxylase (L6OH) to form (2)-trans-carveol. Both these enzymes show a 70% identity at the amino acid level with major differences localized to the presumptive active sites [50]. The spearmint L6OH transcript is highly expressed in PGT as expected but the full set of downstream redox enzymes isopiperitenone reductase, (+)-pulegone reductase, and menthone reductase involved in menthol pathway were also found but poorly expressed in PGT. A previous study has shown that (-) carvone is not an efficient substrate for the initial double-bond reductase, therefore (-) carvone accumulates in spearmint even though the downstream redox enzymes are present [11,51]. Hence, the abundance of a single enzyme L6OH instead of L3OH changes the final monoterpene produced. This shows how simple changes in the production of a single intermediate can result in drastic changes in the metabolic profiles. When compared to heterodimeric GPP synthase, transcripts for farnesyl diphosphate synthase which is responsible for the formation of farnesyl diphosphate precursor for sesquiterpenes, is expressed around 4 times less in PGT. Apart from low MVA pathway, low levels of FPP synthase transcripts might also contribute to the reduced sesquiterpene production in mint PGT.

#### Functional characterization of Terpene Synthases (TPS) from spearmint

In plants, specific TPSs are responsible for the synthesis of various terpene molecules from the common precursors. Our transcriptome data provides a rich resource for identifying and functionally characterizing new TPSs from spearmint. From our enriched unigenes, we found 16 that were identified as terpene synthases; all of them were more than 1 kb and 10 of them were encoding full-length open reading frames (ORFs). We found TPSs annotated as limonene synthase, (E)-β-farnesene synthase, bicyclogermacrene synthase and cis muuroladiene synthase being preferentially expressed in PGT. However, the exact functional annotation of a new TPS requires activity characterization of the recombinant protein. In





**Figure 6 Carvone biosynthesis pathway unigene levels.** The number in green represents the expression level of a particular unigene in PGT ( $\log_2$  of estimate abundance of transcripts by RSEM value). The number in red represents the fold change in expression level when compared to leaf-PGT ( $\log_2$  fold change between PGT and leaf-PGT). In cases of enzymes with more than one unigene, the unigene with the highest abundance was taken into consideration. The number in brackets represents the number of unigenes identified for each enzyme in the pathway. LS: Limonene synthase, L6OH: Limonene-6-hydroxylase, CD: Carveol dehydrogenase.

mint species, to our knowledge limonene synthase from spearmint [52], (*E*)- $\beta$ -farnesene synthase from peppermint [42] and cis-muroladiene synthase from black peppermint [53] have been previously characterised with respect to their functions. From our RNA seq data we chose to characterize two unannotated full-length TPS. Phylogenetic comparison (Additional file 7) showed that both MsTPS1 and MsTPS2 belonged to the TPS-a subfamily of angiosperm sesquiterpene synthases. The main sesquiterpenes identified by GC-MS analysis in our spearmint variety were (*E*)- $\beta$ -farnesene,  $\beta$ -caryophyllene,  $\alpha$ -caryophyllene (Humulene), cis murola-3-5 diene,  $\beta$ -copaene, bicyclogermacrene and bicyclosquiphellandrene.

To determine MsTPS1 and MsTPS2 enzymatic activities, the full-length open reading frame encoding these enzymes were overexpressed in *E.coli*, purified and used for in vitro assays with GPP or FPP as substrate. In the presence of FPP MsTPS1 catalysed the formation of  $\beta$ -caryophyllene in vitro (Figure 7A) whereas MsTPS2 produced a peak from FPP that was identified as one of  $\beta$ -cubebene/Germacrene D/ $\beta$ -copaene by GCMS (Figure 7B).  $\beta$ -copaene was observed in our mint leaf GC-MS data suggesting that MsTPS2 is most likely to be  $\beta$ -copaene synthase. Both the TPSs failed to produce a peak with GPP as substrate (Figure 7A and B). Thus, our in vitro studies identified them to be sesquiterpene synthases. Furthermore, transient *Agrobacterium tumefaciens*-mediated plant expression [54] was used to investigate the terpenes produced by MsTPS1 and MsTPS2 in planta. Both MsTPS1 and MsTPS2 under the control of a 35S promoter were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi (days post-infiltration) by GC-MS. Both TPSs failed to form any new peak when observed by GC-MS. Studies have shown that overexpressing enzyme 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase (HMGR), a rate limiting step of the mevalonate pathway increases heterologous plant sesquiterpene production [55]. Accordingly, both the TPSs were coexpressed with HMGR in planta to observe the production of sesquiterpenes. MsTPS1 with HMGR produced  $\beta$ -caryophyllene as the major peak and  $\alpha$ -caryophyllene and caryophyllene oxide as minor peaks Additional file 8. MsTPS2 even with HMGR failed to produce any new peaks in planta suggesting that the compound formed by

MsTPS2 might be further metabolised endogenously by the plant (data not shown).

Sesquiterpene synthesis is thought to take place in the cytoplasm while monoterpene synthesis is believed to occur in plastids. Since our biochemical characterization shows that MsTPS1 and MsTPS2 are sesquiterpene synthases, we examined subcellular localization of these proteins in transient studies in tobacco leaves. YFP-tagged MsTPSs were transiently expressed in *N. benthamiana* leaf cells by *Agrobacterium*-mediated infiltration and visualized 3 dpi using the YFP channel of a confocal microscope. Both the MsTPSs showed cytoplasmic localization (Additional file 9) and their sequence analysis also showed lack of plastid targeting sequences which further affirms that these TPSs are sesquiterpene synthases. Therefore, either by sequence similarity or by functional characterization we were able to identify all the major terpene synthases that are responsible for the formation of major spearmint essential oil components.

#### Spearmint PGTs as plants chemical defense organs

Many of the secondary metabolites produced by glandular trichomes play a role in plant defence. Majority of them fall into the category of terpenes, phenylpropanes, flavonoids, methyl ketones, acyl sugars and defensive proteins. Apart from having a rich terpene pathway, spearmint PGT also shows presence of transcripts that are involved in the production of different secondary metabolites that may have a role in plant defense. Transcripts encoding enzymes for phenylalanine ammonia lyase, cinnamate 4-hydroxylase, and 4-coumarate CoA-ligase were seen expressed in PGT. These enzymes are involved in phenylpropanoid production. Presence of a variety of small molecular weight phenylpropanoids like caffeic, rosmarinic and ferulic acids has been detected in leaves of different mint germplasm [56]. Transcripts similar to Caffeate O-methyltransferase, an enzyme required for the conversion of caffeic acid into ferulic acid was preferentially expressed in PGT. Chalcone flavanone isomerase an enzyme in the flavonoid pathway in plants was also preferentially expressed in PGT. Unigenes encoding transcripts similar to plant invertase/pectin methylesterase inhibitors were highly expressed in PGT which are important to defend

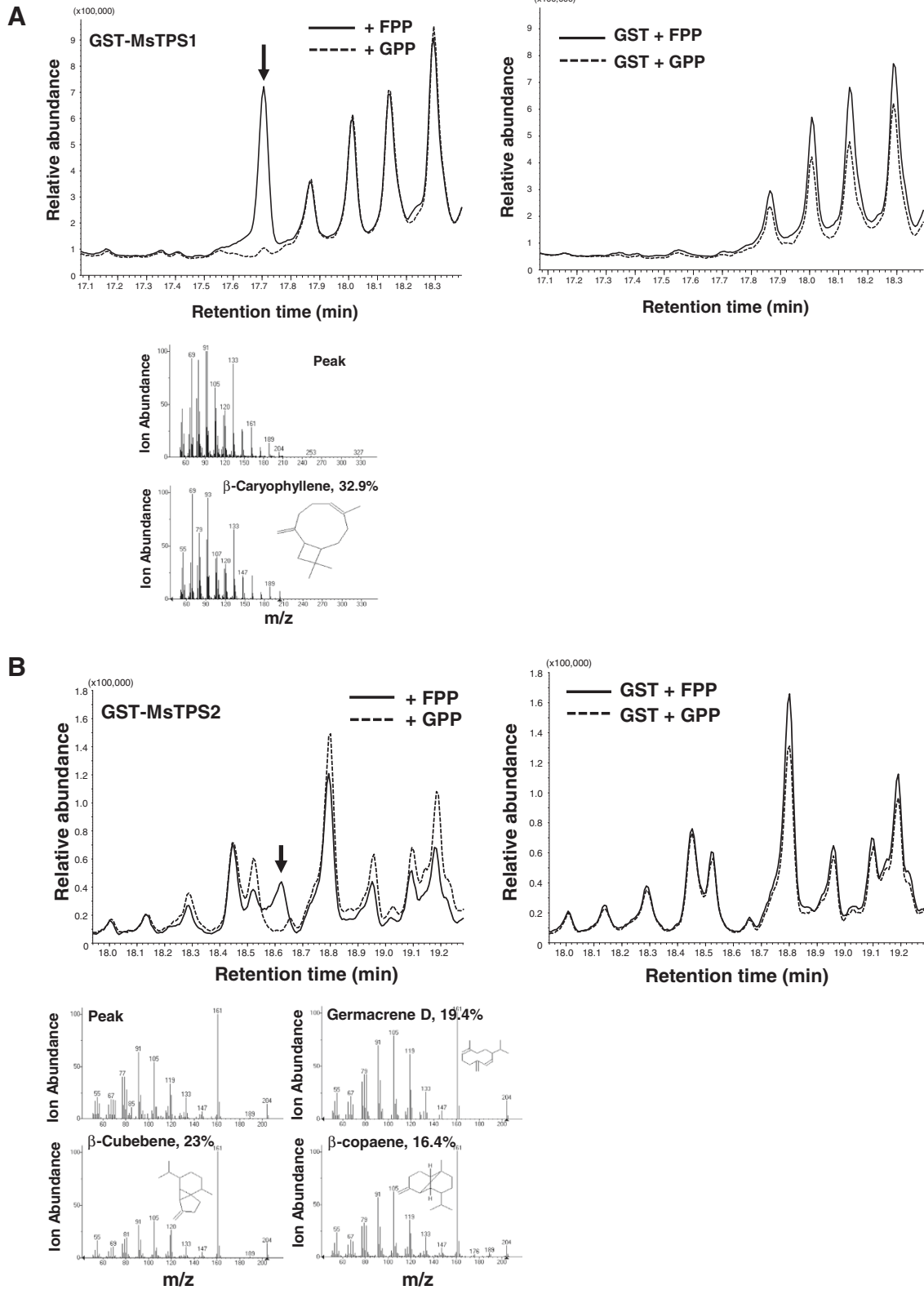


Figure 7 (See legend on next page.)

(See figure on previous page.)

**Figure 7 *In vitro* enzymatic assays of recombinant MsTPSs.** GST-tagged MsTPS recombinant enzymes were purified by glutathione-based affinity chromatography and used for *in vitro* assays with GPP or FPP as substrate. The final products were analysed by GC-MS. The peaks marked with an arrow in the GC traces were compared with the reference of the mass spectra library. Mass spectra for the peaks formed with FPP are shown at bottom of figure. m/z, mass-to-charge ratio, **(A)** Left panel,  $\beta$ -caryophyllene formation by GST-MsTPS1 with FPP; Right panel, the control GC-MS analyses of GST with GPP or FPP. **(B)** Left panel,  $\beta$ -cubebene/Germacrene D or  $\beta$ -copaene formation by GST-MsTPS2 with FPP; Right panel, the control GC-MS analyses of GST with GPP or FPP.

against plant pathogens [57,58]. All the above transcripts were verified by qRT-PCR (Additional file 4).

Oxylipins are large family of biologically active oxidized fatty acid-derivatives that are important for plant defense responses. They are produced mainly by the combined action of lipases, lipoxygenases (LOXs) and members of cytochrome P450 (CYP74) family specialized in metabolizing fatty acids hydroperoxides (HPs). HPs form the initial substrates for different branches of the oxylipin pathway [59-62]. We found enzymes involved in oxidation of fatty acids expressed more in PGTs. Transcripts similar to phospholipase A, lipoxygenase and allene oxide synthase and allene oxide cyclase were enriched in PGT. Allene oxide synthase and allene oxide cyclase are members of the cytochrome P450 (CYP74) family and are involved in the LOX pathways [63,64]. All the above mentioned transcripts showed a minimum of 8 times enrichment in PGT and were confirmed by qRT-PCR. However to understand the role of these enzymes in plant defence and exact nature of oxylipins formed by them, further characterization is essential, especially in terms of their positional specificity and substrate specificity. Plants possess both basal and inducible mechanisms to defend themselves against pathogens. The transcriptome data reflects the basal or constitutive state of defense mechanism existing in spearmint PGT. The changes in transcript abundance in all the above mentioned genes on stress induction will provide a better understanding of how PGT act as chemical defence organs.

## Conclusions

Availability of extensive genome resources in non-model plant *Mentha* is scarce. This is the first attempt at the de novo sequencing and assembly of transcriptomes derived from various tissues of *Mentha spicata* using NGS. Comparison of PGT and leaf-PGT led to the identification of 3,919 differentially expressed unigenes. Analysis of these unigenes provides an insight into the gene expression pattern and biological processes active in spearmint PGT. Further identification of various unknown PGT specific unigenes should facilitate new gene discovery. Our transcriptome data provides essential information for future genetic studies in spearmint. It will also help in developing strategies to engineer the terpene metabolic process that can be further extrapolated to other commercially important plants similar to mint where no genomic resources are available.

## Methods

### Plant material, PGT and RNA isolation

Commercial spearmint variety was grown in green house under natural light conditions. 1–2 cm leaves were collected in ice cold imbibition buffer similar to described in [9] in a 50 ml falcon tube. After 1 hour of soaking, the leaves (approximately 3 g) were transferred to a fresh falcon tube containing the extraction buffer as described in [9]. Trichomes were isolated by glass bead abrasion method. Glass beads (Sigma 425–600  $\mu$ m) were added to the falcon tube and vortexed for 30 sec twice with a 1 min resting period on ice. Then the mixture was carefully passed through cell strainer (100  $\mu$ m) to remove the cell wall debris and hairy trichomes and the flow through collected. This was again sieved using 40  $\mu$ m cell strainer to allow passage of capitate glandular trichome and collection of PGT above. The accumulated PGT on top of 40  $\mu$ m strainer was washed in isolation buffer few times and finally collected using RNAase free water into a 1.5 ml centrifuge tube. The collected PGT was then spun and excess water removed and immediately frozen in liquid nitrogen for further use. We randomly counted around 500 isolated PGT to get an estimate on sample purity and found 26 hair-like trichomes, mostly broken, and 4 CPT indicating a good purity level of isolated PGT. Same stage leaves were brushed in imbibition buffer to remove trichomes and checked under dissection microscope. Total RNA was extracted from PGT, leaf and leaf-PGT using the Spectrum™ Plant total RNA kit from Sigma. The quality of RNA was checked by measuring the ratio of OD<sub>260</sub> to OD<sub>280</sub> and the integrity was assessed by measuring the RNA Integrity Number (RIN) number using Agilent 2100 bioanalyser.

### Sequencing and assembly

The RNA libraries were prepared using the TruSeq RNA Sample Preparation Kits v2, set A (RS-122-2001, Illumina Inc.) according to manufacturer's instructions. The quality and size of cDNA libraries for sequencing were checked using the Agilent 2200 TapeStation system (Agilent Inc.). The libraries were run on single lanes for 100 cycles (paired-end) on HiSeq™ 2000 (illumine Inc.), individually. Raw reads were analysed by FastQC [65] for their quality and found to high quality reads with Q > 20. The Trinity method [20] was used for de novo assembly of the raw reads to generate unigenes. Functions of the unigenes

were annotated based on sequence similarities to sequences in the public nr database (National Centre for Biotechnology Information) and also the protein sequence databases from *Arabidopsis thaliana*, *Vitis vinifera* and *Oryza sativa*. The GO terms were retrieved by Trinote from the Gene Ontology database [20].

#### Preparation of recombinant proteins and *in vitro* enzyme assay

The full-length cDNAs of *MsTPS1* and *MsTPS2* were amplified with the following primer sets from PGT derived cDNA. *MsTPS1*: 5'-CACCATGGAAATTCCTGCACCGGTTTCGGCTTA-3' and 5'-AACTGTTAGGGGATCAACGAGTATGGATTTGATC-3'; and for *MsTPS2*: and 5'-5'CACCATGGCTGAAATCTGTGCGTCGGCTGCT-3' and 5' GTGCAGGGGATCTACGAGCACGGATTGAAT-3'. To construct the vectors for the production of recombinant GST-tagged proteins, the PCR-amplified *MsTPS1* and *MsTPS2* cDNAs were inserted into pGEX-4 T-1 (GE Healthcare Life Sciences) to generate GST-*MsTPS1* and GST-*MsTPS2*, respectively. Both these constructs were transformed into *E.coli* BL21-CodonPlus (DE3)-RIPL (Stratagene), and treated with 0.2 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside (IPTG) at 20°C for overnight to induce GST-tagged protein expression. The harvested cell pellets were resuspended in lysis buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, and protease inhibitors cocktail) and broken by sonication. The clarified lysate was collected by centrifugation and incubated with glutathione Sepharose 4B resin (GE Healthcare Life Sciences) at 4°C overnight. Proteins bound to glutathione Sepharose 4B resin were washed with the purification buffer, eluted from the column with 10 mM glutathione and dialyzed against a buffer containing 25 mM HEPES, pH 7.5, 100 mM KCl, 1 mM DTT, and 5% glycerol.

For *in vitro* enzyme assay for terpene synthase activity, 10  $\mu$ g of recombinant protein was used in final volume of 500  $\mu$ l of reaction buffer (25 mM HEPES, pH 7.5, 100 mM KCl, 7.5 mM MgCl<sub>2</sub>, 5 mM DTT, and 5% glycerol) with 10  $\mu$ g of substrate (farnesyl diphosphate, FPP or geranyl diphosphate, GPP; Sigma). Overlaid 250  $\mu$ l of hexane to trap volatile products and incubated at 30°C for 2 h. Extracts were analysed by GC-MS (Agilent).

#### *In vivo* characterization of *MsTPS1*

The full length cDNA encoding Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase (*AtHMGR1*, At1g76490) was amplified using two specific primers, *AtHMGR-F-XbaI* (5'-AACTCTAGAATGAAGAAAAA GCAAGCTGGTCCCCAACAGA-3') and *AtHMGR-R-AscI* primers (5'-AAAGGCGCGCCTGTTGTTGTTGT TGTCGTTGTCGTT-3'). The PCR-amplified product

was digested by *XbaI* and *AscI*, and cloned into pCAMBIA1300-3HA. *In vivo* characterization was done in *Nicotiana benthamiana* leaves by *Agrobacterium*-mediated infiltration. Initially *Agrobacterium* strain harbouring *MsTPS1* was used alone and then strain carrying *AtHMGR1* was mixed with *MSTPS1* strain and coinfiltrated. Overnight cultures of *Agrobacterium* grown at 28°C were harvested. It was resuspended to a final concentration to an absorbance of 1.0 at 600 nm in a solution containing 10 mM MgCl<sub>2</sub>, 10 mM MES pH 5.6 and 100  $\mu$ M acetosyringone. After 2 hour incubation at room temperature, the *Agrobacterium* mixture was injected into *N. benthamiana* leaves using a needleless syringe. Infiltrated plants were incubated in the growth chamber at 24°C for 3 days or 6 days. Three to four infiltrated leaves were taken for GC-MS analysis.

#### GC-MS analysis method

For extraction about 4–6 leaves were ground to a fine powder using liquid nitrogen and the powder homogenised in 500  $\mu$ l ethyl acetate including 1  $\mu$ l (10 mg/ml) of camphor as internal standard and incubated at least 2 h at room temperature with shaking. This mixture was centrifuged for 5 min at 12,000 rpm. The top organic layer was transferred to a new tube and dehydrated using anhydrous Na<sub>2</sub>SO<sub>4</sub>. The samples were analysed using GCMS (Agilent Technologies 7890A with 5975C inert Mass-spectro Detector with tripe axis detector). 2  $\mu$ l of samples were injected and separation was achieved with a temperature program of 50°C for 1 min and increased at a rate of 8°C/min to 300°C and held for 5 min, on a 30 m HP-5 MS column (Agilent Technologies).

#### Subcellular localization of TPSs-

The full-length cDNAs of *MsTPS1* and *MsTPS2* were amplified with the following primer sets from PGT derived cDNA. *MsTPS1*: 5'-CACCATGGAAATTCCTGCACCGGTTTCGGCTTA-3' and 5'-AACTGTTAGGGGATCAACGAGTATGGATTTGATC-3'; and for *MsTPS2*: and 5'-5'CACCATGGCTGAAATCTGTGCGTCGGCTGCT-3' and 5' GTGCAGGGGATCTACGAGCACGGA TTGAAT-3'. Both the cDNAs were cloned into pEN TR/D-TOPO vector (Invitrogen), and then transferred into pBA-DC-YFP [66] which contains the CaMV 35S promoter and C-terminal in frame YFP, to generate *MsTPS1*-YFP and *MsTPS2*-YFP, respectively. The *MsTPS1*-YFP or *MsTPS2*-YFP constructs were introduced into *A. tumefaciens* strain GV3101 by electroporation. YFP-tagged *MsTPSs* were co-expressed with CFP or alone to confirm the cellular localization of *MSTPS*-YFP. CFP expression was used as a cytoplasmic maker protein. Overnight cultures of *Agrobacterium* grown at 28°C were harvested. It was resuspended to a final concentration to an absorbance of 1.0 at 600 nm in a solution containing 10 mM MgCl<sub>2</sub>,

10 mM MES pH 5.6 and 100  $\mu$ M acetosyringone. After 2 hour incubation at room temperature, the *Agrobacterium* mixture was injected into *N. benthamiana* leaves using a needleless syringe. Infiltrated tobacco plants were placed in the growth chamber at 24°C for 3 days. Fluorescence signals were detected by a confocal scanning laser microscopy (Carl Zeiss LSM 5 Exciter) with a standard filter set.

#### Quantitative real time PCR (qRT-PCR)

The qRT-PCR was employed to validate gene expression pattern of transcriptome analysis. Approximately 1  $\mu$ g RNA was used for cDNA synthesis with iScript supermix (Bio-rad). The reaction was carried out by incubation for priming at 25°C for 5 min, followed by reverse transcription at 42°C for 30 min and inactivation at 85°C for 5 min. cDNA was stored at -20°C till further used. The qRT-PCR reactions were performed in 384-well PCR plate using ABI PRISM 900HT real-time PCR system and KAPA SYBR fast master mix (KAPA Biosystems). PCR reactions were performed using 0.3  $\mu$ l of the cDNA in a total of 5  $\mu$ l reaction volume and cycling profile was 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 60 s. After thermal cycles, the dissociation analysis (melting curve) was carried out to confirm specific amplification of PCR reaction. All reactions were performed in triplicate with three biological replicates, including non-template control. The threshold cycle ( $C_T$ ) value of gene is the cycle number required for SYBR Green fluorescence signal to reach the threshold level during the exponential phase for detecting the amount of accumulated nucleic acid [67]. In current study, elongation factor 1 (*efl*) was used as internal control, due to its stable expression in plant [68] and it also showed similar expression levels in all the tissues in our RNA seq data. Comparative delta  $C_T$  values of target genes to *efl* were taken as relative expression among different tissues. The amount of target gene, normalized to *efl* gene, was calculated by  $2^{-(C_T \text{ target} - C_T \text{ efl})}$ . Results were represented as mean  $\pm$  SD. The unigene names and sequence of primers used are listed in Additional file 10.

#### Availability of supporting data

The raw RNA seq data supporting the result of this article is available in the DDBJ: DNA Data Bank of Japan, with accession number is: DRA001856 (<http://trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA001856>).

#### Additional files

**Additional file 1:** Tissues from which RNA were isolated. a, isolated PGT; b, leaf-PGT; c, leaf.

**Additional file 2:** Quality of reads and statistics of sequencing.

**Additional file 3:** Differentially expressed genes in PGT.

**Additional file 4: qRT-PCR analyses of described unigenes.** DXS: 1-deoxy-D-xylulose-5-phosphate (DXP) synthase; DXR: DXP reductoisomerase, MCT:MEP cytidyltransferase, CMK4:(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase MCS: 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) synthase, HDS: 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMBPP) synthase, HDR: HMBPP reductase, IPP1: Isopentenyl diphosphate (IPP,C5) Delta-isomerase, GPPS: geranyl diphosphate synthase, LS: limonene synthase, L6OH: Limonene-6-hydroxylase, CD: carveol dehydrogenase, LTP: Lipid transfer protein.

**Additional file 5: Subset of cell wall related unigenes more abundant in PGT.**

**Additional file 6: Expression levels of transcripts involved in MVA pathway.** The number in green represents the expression level of a particular unigene in PGT ( $\log_2$  of estimate abundance of transcripts by RSEM value). The number in red represents the fold change in expression level when compared to leaf-PGT ( $\log_2$  fold change between PGT and leaf-PGT). In cases of enzymes with more than one unigene, the unigene with the highest abundance was taken into consideration. The number in brackets represents the number of unigenes identified for each enzyme in the pathway.

**Additional file 7: Phylogenetic analysis of full-length MstPS1 and MstPS2 to other plant terpene synthases.** The neighbour-joining tree was created using MEGA5.2 program from an alignment of *Mentha piperita* (*E*)- $\beta$ -farnesene synthase (AF024615), *Lycopersicon esculentum* germacrene C synthase (AF035630), *Salvia officinalis* 1,8-cineole synthase (AF051899), *Mentha spicata* 4S-limonene synthase (L13459), *Perilla citriodora* limonene synthase (AF241790), *Salvia officinalis* (+)-bornyl diphosphate synthase (AF051900), *Salvia officinalis* (+)-sabinene synthase (AF051901), *Perilla frutescens* linalool synthase (AF444798), *Citrus limon* (+)-limonene synthase 1 (AF514287), *Arabidopsis thaliana* myrcene/ocimene synthase (AF178535), *Artemisia annua* (3R)-linalool synthase (AF154124), *Antirrhinum majus* nerolidol/linalool synthase 1 (EF433761), *Antirrhinum majus* (*E*)- $\beta$ -ocimene synthase (AY195607), *Solanum lycopersicum* copalyl diphosphate synthase (AB015675), *Cucurbita maxima* copalyl diphosphate synthase (AF049905), *Zea mays* terpene synthase 1 (AF529266), *Cucurbita maxima* copalyl diphosphate synthase 1 (AF049905), *Abies grandis*  $\delta$ -selinene synthase (U92266), *Abies grandis* (-)-4S-limonene synthase (AF006193), *Abies grandis* pinene synthase (U87909), *Abies grandis* terpinolene synthase (AF139206), *Abies grandis* myrcene synthase (U87908), *Picea abies* *E*- $\alpha$ -bisabolene synthase (AY473619), *Cichorium intybus* germacrene A synthase short form (AF498000), *Lactuca sativa* germacrene A synthase LTC1 (AF489964), *Solidago canadensis* germacrene D synthase (AJ583447), *Solidago canadensis* germacrene A synthase (AJ304452), *Artemisia annua*  $\beta$ -caryophyllene synthase QHS1 (AF472361), *Artemisia annua* amorpho-4, 11-diene synthase (JQ319661), *Artemisia annua* (*E*)- $\beta$ -farnesene synthase (AY835398), *Arabidopsis thaliana* copalyl diphosphate synthase (NM\_116512), *Arabidopsis thaliana* ent-kaurene synthase GA2 (AF034774), *Stevia rebaudiana* copalyl pyrophosphate synthase (AF034545), *Stevia rebaudiana* kaurene synthase (AF097310). The scale bar indicates the number of amino acid substitutions per site. Five TPS subfamilies, a to e are based on the taxonomic distribution of the TPS families [69].

**Additional file 8: GC-MS data generated from *in vivo* characterization of MstPS1.** MstPS1 with or without HMGR was transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi (days post-infiltration) by GC-MS. Numbered peaks were identified by the reference of the mass spectra library and the mass spectra of compounds are shown at right side.

**Additional file 9: Subcellular localization of MstPS1 and MstPS2 in *N. benthamiana* leaf.** (A) YFP-tagged MstPSs were transiently expressed in *N. benthamiana* leaf cells by *Agrobacterium*-mediated infiltration and visualized 3 dpi using YFP channel of a confocal microscope. (B) YFP-tagged MstPSs were co-expressed with CFP to confirm the cytoplasmic localization of MstPS-YFP. CFP expression is used as a cytoplasmic marker protein. CFP: CFP channel image, YFP: YFP channel image, Auto: chlorophyll auto fluorescence, Light: light microscope image, Merged: merged image between Light and YFP.

**Additional file 10: Primer sequences used for qRT-PCR analysis.**

## Abbreviations

PGT: Peltate glandular trichomes; TPS: Terpene synthases; EST: Expressed sequence tags; GC-MS: Gas chromatography–mass spectrometry; RSEM: RNA-seq by Expectation-Maximization; LTP: Lipid transfer proteins; GPT: Glc6P–phosphate translocator; NTT: Nucleotide transporter; MVA: Mevalonate pathway; MEP: 2-C-methyl-D-erythritol-4-phosphate; IPP: Isopentenyl diphosphate; DMAPP: Dimethylallyl diphosphate; GPP: Geranyl diphosphate; FPP: Farnesyl diphosphate; GGPP: Geranylgeranyl diphosphate; DXS: 1-deoxy-D-xylulose-5-phosphate synthase; DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase; MCT: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK: 4-(cytidine 5-diphospho)-2-C-methyl-D-erythritol kinase; MCS: 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS: 1-hydroxy-2-methyl-2-butenyl 4-diphosphate synthase; HDR: 1-hydroxy-2-methyl-2-butenyl 4-diphosphate reductase; IPPi: Isopentenyl diphosphate isomerase; GPPS: Geranyl diphosphate synthase; LS: Limonene synthase; L6OH: Limonene-6-hydroxylase; CD: Carveol dehydrogenase; ORF: Open reading frame; dpi: Days post-infiltration; HMGR: Hydroxy-3-Methylglutaryl Coenzyme A Reductase; LOX: Lipoxygenases.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

JJ, JL and LSW performed the bioinformatics analysis of the RNA seq data, DNP did the trichome isolation and RNA preparation of spearmint tissues, QW performed the qRT-PCR analysis, MJK and ICJ conducted the TPS characterization; JLY did the GC-MS studies. NHC and RS developed the project and wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Hufeng Zhou for helping with bioinformatics analysis. We thank the Rockefeller University sequencing facility and Dr. Hui-Wen Wu for RNA seq. Prasanna Nori Venkatesh for maintaining the greenhouse. This research was funded by a grant from Singapore National Research Foundation (Competitive Research Programme Award No: NRF-CRP8-2011-02) to SR, ICJ and NHC and a grant from Singapore Millennium Foundation to NHC. JJ was partially supported by a Singapore Ministry of Education Tier-2 grant, MOE2009-T2-2-004. We also thank Temasek Life Sciences Laboratory central facilities.

## Author details

<sup>1</sup>Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore. <sup>2</sup>School of Computing, National University of Singapore, Singapore 117417, Singapore. <sup>3</sup>Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA. <sup>4</sup>Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore.

Received: 9 July 2014 Accepted: 16 October 2014

Published online: 01 November 2014

## References

- McCaskill D, Croteau R: Isopentenyl diphosphate is the terminal product of the deoxyxylulose-5-phosphate pathway for terpenoid biosynthesis in plants. *Tetrahedron Lett* 1999, **40**(4):653–656.
- Lange BM, Turner GW: Terpenoid biosynthesis in trichomes—current status and future opportunities. *Plant Biotechnol J* 2013, **11**(1):2–22.
- Langenheim JH: Higher plant terpenoids: a phytocentric overview of their ecological roles. *J Chem Ecol* 1994, **20**(6):1223–1280.
- Glas JJ, Schimmel BC, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR: Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *Int J Mol Sci* 2012, **13**(12):17077–17103.
- Tissier A: Glandular trichomes: what comes after expressed sequence tags? *Plant J* 2012, **70**(1):51–68.
- Lange BM, Mahmoud SS, Wildung MR, Turner GW, Davis EM, Lange I, Baker RC, Boydston RA, Croteau RB: Improving peppermint essential oil yield and composition by metabolic engineering. *Proc Natl Acad Sci U S A* 2011, **108**(41):16944–16949.
- Saharkhiz MJ, Motamedi M, Zomorodian K, Pakshir K, Miri R, Hemyari K: Chemical composition, antifungal and antibiofilm activities of the essential oil of mentha piperita L. *ISRN Phar* 2012, **2012**:718645.
- Chauhan RS, Kaul MK, Shahi AK, Kumar A, Ram G, Tawa A: Chemical composition of essential oils in *Mentha spicata* L. accession [IIIM(J)26] from North-West Himalayan region, India. *Ind Crops Prod* 2009, **29**(2–3):654–656.
- Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R: Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc Natl Acad Sci U S A* 2000, **97**(6):2934–2939.
- Croteau RB, Davis EM, Ringer KL, Wildung MR: (–)-Menthol biosynthesis and molecular genetics. *Naturwissenschaften* 2005, **92**(12):562–577.
- Champagne A, Boutry M: Proteomic snapshot of spearmint (*Mentha spicata* L.) leaf trichomes: a genuine terpenoid factory. *Proteomics* 2013, **13**(22):3327–3332.
- Turner GW, Gershenzon J, Croteau RB: Development of peltate glandular trichomes of peppermint. *Plant Physiol* 2000, **124**(2):665–680.
- Burbott AJ, Loomis WD: Evidence for metabolic turnover of monoterpenes in peppermint. *Plant Physiol* 1969, **44**(2):173–179.
- Croteau R, Martinkus C: Metabolism of monoterpenes: demonstration of (+)-neomenthyl-beta-d-glucoside as a major metabolite of (–)-menthone in peppermint (*mentha piperita*). *Plant Physiol* 1979, **64**(2):169–175.
- Maffei M, Gallino M, Sacco T: Glandular trichomes and essential oils of developing leaves in *mentha viridis* *lavanduliodora*. *Planta Med* 1986, **52**(03):187–193.
- Brun N, Colson M, Perrin A, Voirin B: Chemical and morphological studies of the effects of ageing on monoterpene composition in *Mentha x piperita* leaves. *Can J Bot* 1991, **69**(10):2271–2278.
- Gershenzon J, McConkey ME, Croteau RB: Regulation of monoterpene accumulation in leaves of peppermint. *Plant Physiol* 2000, **122**(1):205–214.
- McConkey ME, Gershenzon J, Croteau RB: Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint. *Plant Physiol* 2000, **122**(1):215–224.
- Kjonaas R, Croteau R: Demonstration that limonene is the first cyclic intermediate in the biosynthesis of oxygenated p-menthane monoterpenes in *Mentha piperita* and other *Mentha* species. *Arch Biochem Biophys* 1983, **220**(1):79–89.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren B, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011, **29**(7):644–652.
- Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, **9**(4):357–359.
- Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, **12**:323.
- Weitzel C, Simonsen H: Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. *Phytochem Rev* 2013, **1**–18.
- Choi YE, Lim S, Kim HJ, Han JY, Lee MH, Yang Y, Kim JA, Kim YS: Tobacco NtLTP1, a glandular-specific lipid transfer protein, is required for lipid secretion from glandular trichomes. *Plant J* 2012, **70**(3):480–491.
- Yazaki K: ABC transporters involved in the transport of plant secondary metabolites. *FEBS Lett* 2006, **580**(4):1183–1191.
- Kruger NJ, von Schaewen A: The oxidative pentose phosphate pathway: structure and organisation. *Curr Opin Plant Biol* 2003, **6**(3):236–246.
- Sturm A, Tang GQ: The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends Plant Sci* 1999, **4**(10):401–407.
- Fischer K, Weber A: Transport of carbon in non-green plastids. *Trends Plant Sci* 2002, **7**(8):345–351.
- Flugge UI, Hausler RE, Ludewig F, Gierth M: The role of transporters in supplying energy to plant plastids. *J Exp Bot* 2011, **62**(7):2381–2392.
- Turner GW, Croteau R: Organization of monoterpene biosynthesis in *Mentha*. Immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. *Plant Physiol* 2004, **136**(4):4215–4227.
- Yang C, Ye Z: Trichomes as models for studying plant cell differentiation. *Cell Mol Life Sci* 2013, **70**(11):1937–1948.

32. Ishida T, Kurata T, Okada K, Wada T: A genetic regulatory network in the development of trichomes and root hairs. *Annu Rev Plant Biol* 2008, **59**:365–386.
33. Vranova E, Coman D, Gruissem W: Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol* 2013, **64**:665–700.
34. Bruckner K, Tissier A: High-level diterpene production by transient expression in *Nicotiana benthamiana*. *Plant Methods* 2013, **9**(1):46.
35. Cordoba E, Porta H, Arroyo A, San Roman C, Medina L, Rodriguez-Concepcion M, Leon P: Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *J Exp Bot* 2011, **62**(6):2023–2038.
36. Walter MH, Hans J, Strack D: Two distantly related genes encoding 1-deoxy-d-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant J* 2002, **31**(3):243–254.
37. Kim BR, Kim SU, Chang YJ: Differential expression of three 1-deoxy-D-xylulose-5-phosphate synthase genes in rice. *Biotechnol Lett* 2005, **27**(14):997–1001.
38. Kim SM, Kuzuyama T, Chang YJ, Song KS, Kim SU: Identification of class 2 1-deoxy-D-xylulose 5-phosphate synthase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase genes from *Ginkgo biloba* and their transcription in embryo culture with respect to ginkgolide biosynthesis. *Planta Med* 2006, **72**(3):234–240.
39. Kim YB, Kim SM, Kang MK, Kuzuyama T, Lee JK, Park SC, Shin SC, Kim SU: Regulation of resin acid synthesis in *Pinus densiflora* by differential transcription of genes encoding multiple 1-deoxy-D-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase genes. *Tree Physiol* 2009, **29**(5):737–749.
40. Phillips MA, Walter MH, Ralph SG, Dabrowska P, Luck K, Uros EM, Boland W, Strack D, Rodriguez-Concepcion M, Bohlmann J, Gershenzon J: Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (*Picea abies*). *Plant Mol Biol* 2007, **65**(3):243–257.
41. Kim SM, Kuzuyama T, Kobayashi A, Sando T, Chang YJ, Kim SU: 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms *Ginkgo biloba* and *Pinus taeda*. *Planta* 2008, **227**(2):287–298.
42. Crock J, Wildung M, Croteau R: Isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (*Mentha x piperita*, L.) that produces the aphid alarm pheromone (E)-beta-farnesene. *Proc Natl Acad Sci U S A* 1997, **94**(24):12833–12838.
43. Wang W, Wang Y, Zhang Q, Qi Y, Guo D: Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics* 2009, **10**:465.
44. Wang G, Tian L, Aziz N, Broun P, Dai X, He J, King A, Zhao PX, Dixon RA: Terpene biosynthesis in glandular trichomes of hop. *Plant Physiol* 2008, **148**(3):1254–1266.
45. Dudareva N, Andersson S, Orlova I, Gatto N, Reichelt M, Rhodes D, Boland W, Gershenzon J: The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proc Natl Acad Sci U S A* 2005, **102**(3):933–938.
46. Arigoni D, Sagner S, Latzel C, Eisenreich W, Bacher A, Zenk MH: Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement. *Proc Natl Acad Sci U S A* 1997, **94**(20):10600–10605.
47. Hemmerlin A, Hoeffler JF, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ: Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J Biol Chem* 2003, **278**(29):26666–26676.
48. Burke C, Croteau R: Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. *J Biol Chem* 2002, **277**(5):3141–3149.
49. Chang TH, Hsieh FL, Ko TP, Teng KH, Liang PH, Wang AH: Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (*Mentha piperita*) reveals intersubunit regulation. *Plant Cell* 2010, **22**(2):454–467.
50. Lupien S, Karp F, Wildung M, Croteau R: Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (–)-4S-limonene-3-hydroxylase and (–)-4S-limonene-6-hydroxylase. *Arch Biochem Biophys* 1999, **368**(1):181–192.
51. Croteau R, Karp F, Wagschal KC, Satterwhite DM, Hyatt DC, Skotland CB: Biochemical characterization of a spearmint mutant that resembles peppermint in monoterpene content. *Plant Physiol* 1991, **96**(3):744–752.
52. Colby SM, Alonso WR, Katahira EJ, McGarvey DJ, Croteau R: 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. *J Biol Chem* 1993, **268**(31):23016–23024.
53. Prosser IM, Adams RJ, Beale MH, Hawkins ND, Phillips AL, Pickett JA, Field LM: Cloning and functional characterisation of a cis-muroladiene synthase from black peppermint (*Mentha piperita*) and direct evidence for a chemotype unable to synthesise farnesene. *Phytochemistry* 2006, **67**(15):1564–1571.
54. Hellens RP, Allan AC, Friel EN, Bolitho K, Grafton K, Templeton MD, Karunairetnam S, Gleave AP, Laing WA: Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 2005, **1**:13.
55. Song AA, Abdullah JO, Abdullah MP, Shafee N, Othman R, Tan EF, Noor NM, Raha AR: Overexpressing 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) in the lactococcal mevalonate pathway for heterologous plant sesquiterpene production. *PLoS One* 2012, **7**(12):e52444.
56. Tahira R, Naeemullah M, Akbar F, Masood MS: Major phenolic acids of local and exotic mint germplasm grown in Islamabad. *Pak J Bot* 2011, **43**(Special):151–154.
57. Lionetti V, Raiola A, Camardella L, Giovane A, Obel N, Pauly M, Favaron F, Cervone F, Bellincampi D: Overexpression of pectin methyltransferase inhibitors in *Arabidopsis* restricts fungal infection by *Botrytis cinerea*. *Plant Physiol* 2007, **143**(4):1871–1880.
58. An SH, Sohn KH, Choi HW, Hwang IS, Lee SC, Hwang BK: Pepper pectin methyltransferase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta* 2008, **228**(1):61–78.
59. Blee E: Impact of phyto-oxylipins in plant defense. *Trends Plant Sci* 2002, **7**(7):315–322.
60. Feussner I, Wasternack C: The lipoxygenase pathway. *Annu Rev Plant Biol* 2002, **53**:275–297.
61. Prost I, Dhondt S, Rothe G, Vicente J, Rodriguez MJ, Kift N, Carbonne F, Griffiths G, Esquerre-Tugay MT, Rosahl S, Castresana C, Hamberg M, Fournier J: Evaluation of the antimicrobial activities of plant oxylipins supports their involvement in defense against pathogens. *Plant Physiol* 2005, **139**(4):1902–1913.
62. Yang WY, Zheng Y, Bahn SC, Pan XQ, Li MY, Vu HS, Roth MR, Scheu B, Welti R, Hong YY, Wang XM: The patatin-containing phospholipase A pPLA1alpha modulates oxylipin formation and water loss in *Arabidopsis thaliana*. *Mol Plant* 2012, **5**(2):452–460.
63. Tijet N, Brash AR: Allene oxide synthases and allene oxides. *Prostaglandins Other Lipid Mediat* 2002, **68–69**:423–431.
64. Stenzel I, Otto M, Delker C, Kirmse N, Schmidt D, Miersch O, Hause B, Wasternack C: Allene Oxide Cyclase (AOC) gene family members of *Arabidopsis thaliana*: tissue- and organ-specific promoter activities and in vivo heteromerization. *J Exp Bot* 2012, **63**(17):6125–6138.
65. FastQC. [http://www.bioinformatics.babraham.ac.uk/projects/fastqc]
66. Zhang X, Garretton V, Chua NH: The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev* 2005, **19**(13):1532–1543.
67. Walker NJ: Tech.Sight. A technique whose time has come. *Science* 2002, **296**(5567):557–559.
68. Nicot N, Hausman JF, Hoffmann L, Evers D: Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 2005, **56**(421):2907–2914.
69. Chen F, Tholl D, Bohlmann J, Pichersky E: The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 2011, **66**(1):212–229.

doi:10.1186/s12870-014-0292-5

Cite this article as: Jin et al.: Next generation sequencing unravels the biosynthetic ability of Spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics. *BMC Plant Biology* 2014 **14**:292.