Transcription factors regulating terpene synthases in tomato trichomes

Spyropoulou, E.

Citation for published version (APA):
Spyropoulou, E. (2012). Transcription factors regulating terpene synthases in tomato trichomes.
Chapter 5

RNA sequencing of *Solanum lycopersicum* trichomes identifies transcription factors that interact with terpene synthase promoters

Eleni A. Spyropoulou\(^1\), Raymond Hulzink\(^2\), Michel A. Haring\(^1\) and Robert C. Schuurink\(^1\)

\(^1\)Department of Plant Physiology, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

\(^2\)KeyGene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands
ABSTRACT

Glandular trichomes are production and storage organs of specialized metabolites such as terpenes, which play a role in the plant’s defense. The present study aimed to shed light on the regulation of terpene biosynthesis in *Solanum lycopersicum* by identification of transcription factors (TFs) that control the expression level of biosynthetic genes. A trichome transcriptome database was created with a total of 27,195 contigs, which contained 743 annotated transcription factors. Sixteen TFs were selected for further analysis, out of which two were shown to interact with terpene synthase promoters. For one, a MYC TF, binding to a monoterpenes synthase promoter was confirmed in yeast and stably transformed overexpressing lines have been constructed in order to shed further light into the role of this TF in the regulation of terpene biosynthesis.

INTRODUCTION

Trichomes have interested scientists initially for the taxonomic classification of plant species since the seventeenth century. Over the decades the interest in these structures expanded to involve their biochemistry, differentiation and development as well as their physiological and ecological roles (Hallahan and Gray, 2000). Specialized glandular trichomes, produce and accumulate large quantities of terpenoids, phenylpropanoids, flavonoids and alkaloids, which they can also secrete (Schilmiller et al., 2008). Sequencing techniques in combination with metabolite profile analysis of glandular trichomes and proteomics techniques have shed light onto the biosynthesis of these specialized metabolites in the trichomes of various plant species (Tissier, 2012).
Functional evaluation of expressed sequence tags (ESTs) from oil gland secretory cells of peppermint (*Mentha x piperita*; Lamiaceae) revealed that 25% of the ESTs code for enzymes involved in secondary metabolism. Since in mint oil glands the mevalonate (MVA) pathway is blocked, only enzymes of the methylerthritol (MEP) pathway were identified, that lead to the production of terpenoids- a highly represented class of compounds in these glands. Lipid transfer protein (LTP) homologues were also very abundant, whereas enzymes involved in the flavonoid metabolism formed a quantitatively less significant group in mint trichomes (Lange *et al*., 2000). In basil (*Ocimum basilicum*; Lamiaceae), 13% of the ESTs of a leaf peltate gland database belonged to genes of the phenylpropanoid pathway. Genes of the terpenoid pathways as well as lipid transfer proteins represented only 1% of the ESTs each (Gang *et al*., 2001). EST and micro-array analysis of alfalfa (*Medicago truncatula*; Fabaceae) glandular trichomes revealed that these glands primarily produce flavonoid compounds, with the second largest group being proteins of lipid metabolism or transport. Transcripts from the terpenoid biosynthetic pathway were underrepresented, in agreement with the fact that no terpenes have been detected in alfalfa trichomes (Aziz *et al*., 2005). In hop (*Humulus lupulus*; Cannabaceae) lupulin glands, three major classes of secondary metabolites have been found: essential oils, prenylated polyketides and prenylflavonoids. In a hop gland cDNA library enzymes from all biosynthetic pathways were identified in high abundance. LTPs were also highly expressed (Wang *et al*., 2008). Finally in a tobacco (*Nicotiana tabacum*; Solanaceae) trichome cDNA library, secondary metabolism accounted for only ~3% of the total metabolism-related ESTs, which seemed much lower than in other plant species and was attributed to primary metabolism-related genes being much more abundant and thus limiting the number of selected clones from the secondary metabolism (Cui *et al*., 2011). This however is not in accordance with other plant species where primary metabolism enzymes accounted also for a large number of the EST clones (e.g. 48% in mint oil gland secretory cDNA library; Lange *et al*., 2000).
Although EST sequencing has been instrumental in the discovery of enzymes of trichome-specialized metabolism so far (Schilmiller et al., 2010), next generation sequencing (NGS), which is becoming widely available and cost-effective, can give a more in-depth picture of (trichome) transcriptomes.

Apart from the large volume of data obtained, a major advantage of NGS technologies is perhaps that the DNA fragments do not need to be subcloned in a suitable vector, such as was necessary for Sanger sequencing, but can be used directly (Wilhelm and Landry, 2009) in one of the major, commercially available sequencing platforms to date: Genome Sequencer (454 Life Sciences, Roche), Genome Analyzer (Illumina) and Sequencing by Oligo Ligation and Detection (SOLiD; Applied Biosystems). Template in all three technologies is fragmented double-stranded DNA (genomic or cDNA).

Roche’s 454 GS FLX Titanium system is based on emulsion PCR and pyrosequencing, producing reads of about 400bp with a sequencing accuracy of ~99% (http://www.454.com/index.asp). The Illumina Genome Analyzer (GA) system is based on solid-phase bridge PCR and a “sequencing by synthesis” approach and makes use of fluorescent dye-labelled reversible terminator nucleotides for imaging. Depending on the sequencing chemistry it can produce reads of 35-150 bases (http://www.illumina.com). Finally, the Applied Biosystems SOLiD technology is based on emulsion PCR in combination with sequencing by ligation with fluorescently labeled oligonucleotides, producing reads of 50-75 bases (http://www.appliedbiosystems.com).

Each of these technologies has its advantages and disadvantages (Metzker, 2010), but since 454 sequencing produces the longest reads, it has been opted for creating transcriptomic databases by researchers (Brautigam and Gowik, 2010). Contigs are assembled based on the overlapping sequence of the reads and the identity within the overlapping region and this task becomes easier and more accurate with longer reads. Depending on whether a reference transcriptome (or genome) is available (and if this is desired) the reads can be mapped to the reference database and then the overlapping reads assembled into contigs. Alternatively, the reads can be
directly assembled into contigs (de novo assembly; Garber et al., 2011). In quantification of transcripts and expression profiling, next generation sequencing overcomes limitations of micro-array experiments, such as probe cross-hybridization (Lister et al., 2009) and for such applications, technologies like that of Illumina or Applied Biosystems are perhaps better suited, as they produce a larger sequence output per run compared to the 454 Roche technology (18-30Gb compared to 0.45Gb per run; Metzker, 2010).

Deep sequencing of cDNA using NGS technologies (termed RNA sequencing) is already being used for characterization of trichome transcriptomes- for example from plants of medical importance like Artemisia annua (Asteraceae; Wang et al., 2009) or Huperzia serrata and Phlegmariurus carinatus (Huperziaceae; Luo et al., 2010) and/or for gene discovery: 454 sequencing in combination with shotgun proteomics and metabolite analysis of tomato (Solanum lycopersicum; Solanaceae) trichomes led to the discovery of the leaf-trichome-specific β-charyophellene/α-humulene synthase (CAHS; Schilmiller et al., 2010) and 454 sequencing of trichomes deriving from wild and cultivated tomato varieties led to the discovery and characterization of various sesquiterpene synthases, providing insight into the evolution of terpene synthases (Bleeker et al., 2011).

Terpene biosynthesis in tomato plants is of major interest as these compounds are believed to play a role in the plant’s defense (van Schie et al., 2007; Bleeker et al., 2009; Schilmiller et al., 2009; Kang et al., 2010) in this agronomically important crop. The sequencing of the cultivated tomato genome has enabled the characterization of its terpene synthase (TPS) gene family (Falara et al., 2011) but not much is known about the regulation of the terpenoid pathway. Transcriptional control of biosynthetic genes is a major mechanism by which secondary metabolite production is regulated (Vom Endt et al., 2002). Well-studied cases of transcription factors regulating plant secondary metabolic pathways involve the biosynthesis of flavonoids and terpenoid indole alkaloids (Broun et al., 2006).

Here, we used 454 pyrosequencing of tomato stem trichomes as a tool for gene discovery. First, a transcriptome database was created from normalized cDNA,
which was mined for transcription factors (TF). Then, in order to narrow down the number of TFs potentially involved in terpene biosynthesis, an expression profiling database was created using Illumina sequencing of trichome RNAs from plants treated with or without jasmonic acid (JA), since JA is known to regulate plant’s indirect defenses in response to herbivore feeding and wounding (Ballare, 2011). Two transcription factors were shown to transiently transactivate terpene synthase promoters in *Nicotiana benthamiana* leaves.

**MATERIALS AND METHODS**

**Hormone Treatment and RNA Isolation**

Tomato plants (*Solanum lycopersicum* cultivar Moneymaker) were grown in soil in a greenhouse with day/night temperatures of 23°C/18°C and a 16/8h light/dark regime for four weeks. They were then sprayed either with JA solution (1mM JA; Duchefa, NL, in tap water + 0,05% SilwetL-77; GE Silicones, VA, USA) or with control solution (0,05% SilwetL-77 in tap water). Stem pieces were collected 30min, 2h, 8h and 24h later for pyrosequencing or 24h later for expression analyses and trichomes were isolated by shaking the stems in liquid nitrogen. Total RNA was isolated using TRIzol (Invitrogen, Paisley, UK) according to the manufacturer’s instructions. Equal amount of trichome RNA from the different time points was pooled creating the control (C) and JA samples. RNA used for pyrosequencing was then purified on a RNeasy Plant column (Qiagen, Valencia, CA, USA).
Transcriptome Database Construction

RNA quality was determined with the Agilent RNA pico chip (Agilent Technologies, Waldbronn, Germany). Synthesis and amplification of cDNA was performed using the SMART PCR cDNA Synthesis and Advantage 2 PCR kits (Clontech Inc., CA, USA) according to the manufacturer’s instructions with some modifications of adapters to eliminate 3’ poly(A)-stretches prior to sequencing. cDNA quality was determined with the Agilent DNA 7500 chip (Agilent Technologies, Waldbronn, Germany) or on an 1% agarose/EtBr gel. Normalization of the cDNA was carried out using the Evrogen TRIMMER kit (Evrogen, Moscow, Russia) according to the manufacturer’s protocol with some modifications of the adapters. The normalization efficiency was determined both on an agarose/EtBr gel (1%) and with an Agilent DNA 7500 chip (Agilent Technologies, Waldbronn, Germany). Amplified cDNA was purified and concentrated using the Qiaquick PCR purification kit (Qiagen, Valencia, CA, USA). cDNA shearing and FLX Titanium library preparation was carried out using the Roche GS FLX Titanium General Library Preparation Method kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer’s protocol. The size range of the fragments was determined with an Agilent DNA 1000 chip (Agilent Technologies, Waldbronn, Germany). Exclusion of smaller-sized fragments was performed using the double SPRI method as described in the Roche GS FLX Titanium General Library Preparation protocol (Roche Diagnostics, Mannheim, Germany). End-polishing, small fragment removal, library immobilization, fill-in reaction and single-stranded library isolation was performed using the GS FLX Titanium General Library Preparation Method kit (454 Life Sciences, Roche Diagnostics, Mannheim, Germany) according to manufacturer’s instructions.
Expression Profiling Database Construction

Starting from the same total RNA samples (C and JA, see above), mRNA was amplified and purified using the MessageAmp II aRNA Amplification kit (Applied Biosystems/Ambion, CA, USA) according to manufacturer’s instructions. RNA quality was determined with the Agilent RNA pico chip (Agilent Technologies, Waldbronn, Germany). Synthesis of cDNA was performed using the MessageAmp II aRNA Amplification kit (Applied Biosystems/Ambion, CA, USA) according to manufacturer’s instructions with modifications of the adapters to enable sequencing of 3’ cDNA ends. cDNA was purified with the Qiaquick PCR purification kit (Qiagen, Valencia, CA, USA). cDNA quality was determined with the Agilent DNA 7500 chip (Agilent Technologies, Waldbronn, Germany) or on an 1% agarose/EtBr gel. Shearing and ligation was carried out using standard Illumina PE adapters containing a specific sample ID tag. Adapter-ligated cDNA fragments were column purified with the Qiaquick PCR purification kit (Qiagen, Valencia, CA, USA). The size range of the fragments was determined with an Agilent DNA 1000 chip (Agilent Technologies, Waldbronn, Germany). Exclusion of smaller-sized fragments was performed using a single SPRI procedure as described in the Agencourt Ampure PCR Purification protocol (Agencourt Bioscience Corporation, MA, USA). The size range of single-stranded fragments was determined with an Agilent RNA pico 6000 chip (Agilent Technologies, Waldbronn, Germany). Expression profiling was performed using the Illumina GA II System (Illumina, USA).

Databases Assembly, EST Annotation and Homology Searches

The 454 sequencing reads (Control and JA combined) were assembled into contigs de novo by Vertis Biotechnologie AG, Germany using the CLCbio software (http://www.clcbio.com). Nucleotide sequences of the contigs were then blasted
against the SGN tomato unigenes v2 database (ftp.solgenomics.net/unigene_builds/combined_species_assemblies/tomato_species) for annotation, using a local Eblast tool ($E$ value $1\text{e}^{-9}$). The GA II reads (Control and JA separately) were mapped to the annotated contigs of the 454 sequencing trichome database by Vertis Biotechnologie AG, Germany.

The resulting contigs were also imported in the bioinformatics tool Blast2GO v.2.5.0 (Conesa et al., 2005) and were compared against the National Center for Biotechnology Information (NCBI) non-redundant protein database BLASTX ($E$ value $1\text{e}^{-3}$). Further analyses with this tool included functional annotation by Gene Ontology (GO) terms and Enzyme Commission numbers (EC code), InterPro terms (InterProScan; Quevillon et al., 2005) and metabolic pathways (Kyoto Encyclopedia of Genes and Genomes, KEGG; Ogata et al., 1998).

cDNA Synthesis and Quantitative Real Time PCR

DNA was removed from RNA with DNase (Ambion, Huntingdon, UK) according to the manufacturer and cDNA was synthesized from 1.5\(\mu\)g RNA using M-MuLV H\(^{-}\) Reverse Transcriptase (Fermentas, St. Leon-Rot, Germany) in 20\(\mu\)l volume. For Q-RT-PCR, cDNA equivalent to 100ng total RNA was used as template in 20\(\mu\)l volume and reactions were performed in the ABI 7500 Real-Time PCR System (Applied Biosystems) using the Platinum SYBR Green qPCR SuperMix-UDG kit (Invitrogen, Paisley, UK) with the following cycling program: 2min 50°C, 7min 95°C, 45 cycles of 15sec at 95°C and 1min at 60°C, followed by a melting curve analysis. Primer pairs were tested for amplification kinetics and linearity with a standard cDNA dilution curve and new primers were designed if necessary. Expression levels were normalized using ACTIN (SGN-U579547) mRNA levels. Effect of JA in gene expression was analyzed in three biological replicates by T-test using PASW Statistics 17.0 (http://www.spss.com). The homogeneity of variance was tested by Levene’s test.
**Cloning and construct design**

TFs *SlMYC1*, *SlWRKY3* and *SlWRKY4* (sequence of the full-length ORFs obtained from the 454 trichome database) were cloned between restriction sites *Ncol* (at the ATG) and *SacI* (at the 3’ end of the sequence) in vector pKG1662 (KeyGene, Wageningen, NL; for a map of the vector see patent nr US2011/0113512A1) driven by the CaMV 35S promoter. TF *SlWRKY1* (sequence of the full-length ORF obtained from the 454 trichome database) was cloned downstream of the CaMV 35S promoter in vector pJVII, a pMON999 vector (Monsanto, St. Louis, MO) with a modified multiple cloning site, between restriction sites *XbaI* (at the ATG) and *BsrGI* (at the 3’ end of the sequence). The sequences of these four TFs are provided in Figure S1. All constructs were verified by sequencing and then the expression cassettes containing 35S promoter, cDNA of interest and *nos* terminator were transferred to the MCS of the binary vector pBINplus (van Engelen et al., 1995) between *HindIII* and *SmaI* restriction sites. The final constructs were transformed to *Agrobacterium tumefaciens* GV3101 (pMP90). The promoter:GUS constructs used in the transient transactivation assay have been described earlier (Chapter 4).

**Transient transactivation assay in *Nicotiana benthamiana* leaves**

*Agrobacterium tumefaciens* GV3101 (pMP90) cultures were grown overnight from a single colony and diluted in infiltration buffer (50mM MES pH 5.8, 0.5% glucose, 2mM NaH₂PO₄, 100μM acetosyringone; Sigma-Aldrich) to OD₆₀₀ of 0.3. Leaves of five week old *Nicotiana benthamiana* plants were then infiltrated with mixtures carrying various promoter:GUS reporter constructs and the 35S:EOT1 effector construct in a 1:1 ratio. In order to normalize for transformation efficiency and protein extraction efficiency, a construct containing *Photinus pyralis* (firefly) luciferase (*LUC*) driven by the CaMV 35S promoter was co-infiltrated with the
reporter/effector mix in a 2:5 ratio (LUC: reporter/effector mix). Two leaves of three plants were infiltrated for each combination. Two days later leaf disks from the infiltrated areas were collected, frozen in liquid nitrogen and crude extracts were prepared in extraction buffer containing 25 mM Tris phosphate pH 7.8, 2 mM DTT, 2 mM CDTA pH 7.8, 10% glycerol and 1% Triton X-100. The enzymatic GUS activity was determined spectrophotometrically using 4-methylumbelliferyl \( \beta \) -D-glucuronide (MUG) as a substrate according to Jefferson et al. (1987). The luciferase assay was performed using the same extraction buffer, according to van Leeuwen et al. (2000) and measured in a FluoroCount Microplate Fluorometer (Packerd BioScience Company) using a 560 nm emission filter. Enzymatic GUS activity was normalized for luciferase activity for each sample. Significant differences between samples were tested using PASW Statistics 17.0 (http://www.spss.com) by ANOVA followed by a Tuckey’s B posthoc test. The homogeneity of variance was tested by Levene’s test and values were log transformed before the analysis if necessary. The experiments were repeated at least twice with similar results.

**Yeast-One-Hybrid Screen**

The full-length *SlMT51* promoter or promoter fragments -1254 to -1047 bp, -1046 to -807 bp, -806 to -613 bp, -612 to -409 bp, -408 to -208 bp and -207 to -1 bp were cloned between restriction sites EcoRI (at the 5’ end of the sequence) and XbaI (at the 3’ end of the sequence) in the MCS of vector pHISi (Clontech, Mountain View, CA, USA) upstream of the minimal promoter of the *HIS3* locus driving *HIS3*. The *HIS3* gene of pHISi was used as selectable marker for integration into the nonfunctional his3 locus of yeast strain PJ69-4A. The vector was linearized at the XhoI restriction site before integration. Leaky *HIS3* expression enabled enough colony growth, making it possible to use as a selectable marker. Background growth was controlled during library screening by the addition of 5 mM 3-amino-
1,2,4-triazole (3-AT) in the selective medium (synthetic dextrose, SD). Transformations were performed using 1µg of the library plasmid pAD-GAL 4-2.1 carrying the ORF of *SlMYC1* cloned between restriction sites *NheI* (at the ATG) and *XhoI* (after the stop codon) in the MCS of the vector.

**List of primers used**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5' -&gt; 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIMYC1_QF2</td>
<td>TGCTGAATCGCGATGAATTATGTC</td>
</tr>
<tr>
<td>SIMYC1_QR2</td>
<td>GCCTCAACTCGAGATCTCTAGTA</td>
</tr>
<tr>
<td>SIWRKY1_QF</td>
<td>GGATCAGTCGTTGATCCTGTC</td>
</tr>
<tr>
<td>SIWRKY1_QR</td>
<td>CTTCTGACCTTTTGGTACCCAG</td>
</tr>
<tr>
<td>SIWRKY3_QF</td>
<td>CAACAAATCAAGGCTCCGATATAC</td>
</tr>
<tr>
<td>SIWRKY3_QR</td>
<td>TAGTCAATGCTACAAGAAACTTTGAC</td>
</tr>
<tr>
<td>SIWRKY4_QF</td>
<td>GAAGTTACTGAGCCTTCCAG</td>
</tr>
<tr>
<td>SIWRKY4_QR</td>
<td>AACCTGAATAAAGCTGATACAG</td>
</tr>
<tr>
<td>SIACT_QF</td>
<td>TCAGCACATTCCAGAGATGT</td>
</tr>
<tr>
<td>SIACT_QR</td>
<td>AACAGCACAGGACACTCGCAG</td>
</tr>
</tbody>
</table>

**RESULTS**

**Assembly of 454 sequencing data and Genome Analyzer II transcript profiling**

We have created a tomato trichome EST database by sequencing a mixture of glandular and non-glandular trichome RNAs, deriving from stems of *Solanum lycopersicum* cv. Moneymaker plants. The resulting cDNA was normalized before being used as input for 454 GS FLX Titanium pyrosequencing. A full plate was sequenced consisting of two halves: one with cDNAs originating from control plants and the other half with cDNAs originating from plants treated with JA. In total we obtained 979,076 high-quality reads with an average length of 337bp. The reads from control and JA samples combined were assembled *de novo* resulting in
27,195 contigs with an average length of 931bp, leaving 24,187 reads unmatched (singletons), with an average length of 241bp. Nucleotide sequences of the contigs were blasted against the Solanaceae Genomics Network (SGN) tomato database for annotation, using a local E-Blast tool. 3,295 contigs were not annotated.

For creating the transcript profiling databases with Genome Analyzer II, the same RNA material as for the 454 sequencing was used, but this time the cDNA derived from Control and JA-treated stem trichomes was not normalized before being used as input. From the Control sample 5,631,975 3’ sequences were obtained and from the JA sample 5,882,547, corresponding to one plate lane. 4,840,738 and 5,169,891 reads from the Control and JA-samples, respectively, were mapped uniquely to the 27,195 contigs of the trichome database. In addition, 38,699 (C) and 45,375 (JA) reads were mapped non-specifically and 791,237 (C) and 712,656 (JA) remained unmapped.

**Homology searches, gene ontology and protein function**

In order to dissect the *S. lycopersicum* stem trichome transcriptome the unique ESTs (27,195 contigs) were submitted to homology searches after translation (BLASTX) and were compared against the National Center for Biotechnology Information (NCBI) non-redundant protein database using Blast2GO (www.blast2go.com). 4,733 sequences did not return a BLASTX hit. The majority of the top hits were to protein sequences of *Vitis vinifera*, followed by *Populus trichocarpa*, *Ricinus communis* and *Solanum lycopersicum*.

Next, gene ontology (GO) and enzyme classifications (EC) were performed in order to classify the ESTs. It must be noted that one sequence could be assigned to more than one GO term. An overview of the GO annotations obtained (at level 2 of 5) is presented in Figure 1 (level 1 giving a general description whereas level 5 a more specific description of a given gene therefore list coverage decreases at higher levels, whereas term specificity increases). The highest percentage of
molecular function GO terms were involved in binding and catalytic activity (Fig.1c). In the biological processes, the majority of the GO terms was grouped into two categories- those of metabolic and cellular process (Fig.1b). Furthermore, for the cellular component class the assignments were mostly given to cell and organelle (Fig.1a). Finally, within the predicted ECs, the prevailing categories of enzymes were transferases and oxidoreductases (Fig.1d).

Figure 1. Gene ontology (GO) and enzyme classifications (EC) for S.lycopersicum stem trichome transcriptome at level 2. (a) Cellular component GO terms, (b) biological process GO terms, (c) molecular function GO terms and (d) general EC terms.
The search of additional databases for protein families, domains, regions and sites was performed remotely from Blast2GO via the InterPro EBI web server. The 30 top InterPro entries obtained are presented in Table 1. The most dominant class of enzymes was protein kinases. Abundantly represented were also cytochrome P450s.

Finally, within Blast2GO, the EC numbers were classified in KEGG pathways, enabling the presentation of enzymatic functions in the context of the metabolic pathways in which they are part of (Blast2GO Tutorial; Conesa and Goetz, 2009). Among the pathways identified, the ones related to secondary metabolism are shown in Table 2. Lipid transfer proteins represented 0.19% of the tomato stem trichome transcripts.

**Table 1.** Summary of the 30 most common InterPro entries found in the *S.lycopersicum* stem trichome transcriptome.

<table>
<thead>
<tr>
<th>InterPro</th>
<th>Frequency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR011009</td>
<td>571</td>
<td>Protein kinase-like domain</td>
</tr>
<tr>
<td>IPR000719</td>
<td>521</td>
<td>Protein kinase, catalytic domain</td>
</tr>
<tr>
<td>IPR022890</td>
<td>336</td>
<td>Serine/threonine-/ dual-specificity protein kinase, catalytic domain</td>
</tr>
<tr>
<td>IPR008271</td>
<td>286</td>
<td>Serine/threonine-protein kinase, active site</td>
</tr>
<tr>
<td>IPR020635</td>
<td>283</td>
<td>Tyrosine-protein kinase, catalytic domain</td>
</tr>
<tr>
<td>IPR016040</td>
<td>263</td>
<td>NAD(P)-binding domain</td>
</tr>
<tr>
<td>IPR013083</td>
<td>239</td>
<td>Serine/threonine-protein kinase, active site</td>
</tr>
<tr>
<td>IPR017441</td>
<td>187</td>
<td>Protein kinase, ATP binding site</td>
</tr>
<tr>
<td>IPR022885</td>
<td>181</td>
<td>Pentatricopeptide repeat</td>
</tr>
<tr>
<td>IPR015943</td>
<td>175</td>
<td>WD40/YYTN repeat-like-containing domain</td>
</tr>
<tr>
<td>IPR01841</td>
<td>172</td>
<td>Zinc finger, RING-type</td>
</tr>
<tr>
<td>IPR012677</td>
<td>172</td>
<td>Nucleotide-binding, alpha-beta plait</td>
</tr>
<tr>
<td>IPR016204</td>
<td>166</td>
<td>Armadillo-type fold</td>
</tr>
<tr>
<td>IPR001245</td>
<td>164</td>
<td>Serine/threonine/tyrosine-protein kinase catalytic domain</td>
</tr>
<tr>
<td>IPR000504</td>
<td>158</td>
<td>RNA recognition motif domain</td>
</tr>
<tr>
<td>IPR001680</td>
<td>153</td>
<td>WD40 repeat</td>
</tr>
<tr>
<td>IPR011046</td>
<td>152</td>
<td>WD40 repeat-like-containing domain</td>
</tr>
<tr>
<td>IPR011990</td>
<td>147</td>
<td>Tetratricopeptide-like helical</td>
</tr>
<tr>
<td>IPR019889</td>
<td>145</td>
<td>Armadillo-like helical</td>
</tr>
<tr>
<td>IPR001128</td>
<td>141</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>IPR017986</td>
<td>133</td>
<td>WD40-repeat-containing domain</td>
</tr>
<tr>
<td>IPR017853</td>
<td>130</td>
<td>Glycoside hydrolase, superfamily</td>
</tr>
<tr>
<td>IPR012287</td>
<td>119</td>
<td>Homeodomain-related</td>
</tr>
<tr>
<td>IPR001611</td>
<td>115</td>
<td>Leucine-rich repeat</td>
</tr>
<tr>
<td>IPR009057</td>
<td>115</td>
<td>Homeodomain-like</td>
</tr>
<tr>
<td>IPR016196</td>
<td>112</td>
<td>Major facilitator superfamily domain, general substrate transporter</td>
</tr>
<tr>
<td>IPR012336</td>
<td>108</td>
<td>Thioredoxin-like fold</td>
</tr>
<tr>
<td>IPR013781</td>
<td>107</td>
<td>Glycoside hydrolase, subgroup, catalytic domain</td>
</tr>
<tr>
<td>IPR002213</td>
<td>102</td>
<td>UDP-glucuronosyl/UDP-glucosyltransferase</td>
</tr>
<tr>
<td>IPR002401</td>
<td>95</td>
<td>Cytochrome P450, E-class</td>
</tr>
</tbody>
</table>
A closer look was taken at the terpene biosynthesis pathway in order to see if the precursor pathways were up-regulated by JA treatment. As shown in Figure 2a, expression of some precursor genes in tomato was induced by JA although not strongly (max induction ~2.5-fold for HDS). Like in other plants (Tholl and Lee, 2011), genes encoding enzymes of the precursor pathways are not always single copy and it appears that expression levels and JA inducibility can vary. Transcript abundance of precursor genes is presented in Figure 2b for comparison reasons with the expression levels of the terpene synthases (TPSs) found in stem trichomes. Transcripts for enzymes involved in the jasmonic acid biosynthesis and signaling pathway were also identified in the database. Obtained data for some of these enzymes are presented in Table 3.
Figure 2. (a) Enzymes involved in the biosynthesis of mono-(C10), sesqui-(C15) and di-(C20) terpenes and (b) terpene synthases found in stem trichomes. An explanation of the abbreviations used in the pathways and the GAII reads for each enzyme are shown in the tables. Genes upregulated (> 1.7x) and downregulated (< 0.6x) by JA treatment are shown in red and green respectively. Chr; chromosome.
Table 3. List of selected enzymes involved in the jasmonic acid biosynthesis and signaling pathway (see references).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Contig nr</th>
<th>SGN Nr</th>
<th>Annotation</th>
<th>Transcript length (bases)</th>
<th>Reads JA</th>
<th>Reads C</th>
<th>Fold JA-induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOXA</td>
<td>6402</td>
<td>SGN-U592535</td>
<td>Lipoxigenase A</td>
<td>2837</td>
<td>207</td>
<td>9</td>
<td>23,00</td>
</tr>
<tr>
<td>AOC</td>
<td>24817</td>
<td>SGN-U562649</td>
<td>Allene oxide cyclase</td>
<td>1645</td>
<td>11170</td>
<td>2352</td>
<td>4,75</td>
</tr>
<tr>
<td>JAZ1</td>
<td>6863</td>
<td>SGN-U579837</td>
<td>Jasmonate ZIM-domain 1</td>
<td>1156</td>
<td>72</td>
<td>4</td>
<td>18,00</td>
</tr>
<tr>
<td>JAZ3</td>
<td>20751</td>
<td>SGN-U564446</td>
<td>Jasmonate ZIM-domain 3</td>
<td>986</td>
<td>111</td>
<td>13</td>
<td>8,54</td>
</tr>
<tr>
<td>COI1</td>
<td>2260</td>
<td>SGN-U568988</td>
<td>Coronatine-insensitive 1</td>
<td>2260</td>
<td>65</td>
<td>48</td>
<td>1,35</td>
</tr>
</tbody>
</table>

a; Li et al., 2004: LOXA (U09026), AOC (AW624058), COI1 (NM_001247535), b; Thines et al., 2007: JAZ1 (EF591123), c; Katsir et al., 2008: JAZ3 (EU194561)

Selection of candidate transcription factors potentially involved in terpene biosynthesis

Based on the annotated contigs 743 transcription factors of different classes were found: 69 WRKY, 151 MYB, 8 MYC, 52 bZIP, 9 ARF, 71 ERF, 17 ZnF, 28 bHLH, 12 MADS, 1 NAC and 325 of unknown function/class. Out of those, 325 were up-regulated (>1.1x) by the treatment with JA, 208 were down-regulated (<0.9x) and expression of 210 TFs remained unaltered (1.1x - 0.9x). Since jasmonic acid is known to play a role in the plant’s direct and indirect defenses (Ballare, 2011) we were interested in those transcription factors that are induced by JA and could therefore potentially be involved in up-regulating terpene biosynthesis. 56 of the TFs that were up-regulated by JA showed an induction higher than 2-fold. The sequence of these 56 TFs was blasted against the tomato genomic sequence (Solanaceae Genomics Network, SGN) and complete ORFs were constructed when possible (if not provided by the 454 sequencing). These sequences were submitted to homology search after translation (BLASTX) against the National Center for Biotechnology Information (NCBI) database for identifying conserved domains. Out of this analysis 16 TFs were selected according to their class to be further investigated (Table 4) based on the published data presented below.

There are only few transcription factors that are known to be involved in regulation of the terpenoid pathway. The first evidence came from Catharanthus roseus cells overexpressing ORCA3, a jasmonate-responsive APETALA2 (AP2)-domain
transcription factor, that showed increased accumulation of terpenoid indole alkaloids (van der Fits and Memelink, 2000). In 2004 Xu et al. identified a methyljasmonate-inducible WRKY transcription factor from *Gossypium arboreum* that regulates the sesquiterpene synthase (+)-δ-cadiene synthase A in cotton fibers. In 2009 Ma et al. demonstrated that a methyl-jasmonate-inducible WRKY transcription factor from *Artemisia annua* is involved in the regulation of artemisinin biosynthesis. Finally, CrWRKY1 was recently identified as being involved in the root-specific accumulation of serpentine in *C. roseus* plants and as being induced by phytohormones including JA (Suttipanta et al., 2011). Based on this knowledge eleven transcription factors of the AP2 class and four of the WRKY class were selected for further investigation of their potential involvement in terpene biosynthesis.

Lastly, a sixteenth gene was added to the list, a MYC transcription factor, since in *Arabidopsis*, AtMYC2, the first transcription factor shown to be involved in the transcriptional regulation of the JA signaling pathway, is known to regulate pathogen and wound response genes (Dombrecht et al., 2007).

### Table 4. List of selected transcription factors (TF) potentially involved in terpene biosynthesis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Contig nr</th>
<th>SGN nr</th>
<th>Annotation</th>
<th>Transcript length (bases)</th>
<th>Reads JA</th>
<th>Reads C</th>
<th>Fold JA-induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>SlAP2_9</td>
<td>83</td>
<td>SGN-US72361</td>
<td>ERF (ethylene response factor) subfamily B-3 of ERF/AP2 TF family</td>
<td>371</td>
<td>57</td>
<td>23</td>
<td>2.5</td>
</tr>
<tr>
<td>SlAP2_2</td>
<td>1719</td>
<td>SGN-US96590</td>
<td>DREB subfamily A-1 of ERF/AP2 TF family</td>
<td>422</td>
<td>80</td>
<td>11</td>
<td>7.3</td>
</tr>
<tr>
<td>SlAP2_6</td>
<td>5289</td>
<td>SGN-US63871</td>
<td>AP2 domain-containing TF</td>
<td>933</td>
<td>267</td>
<td>72</td>
<td>3.7</td>
</tr>
<tr>
<td>SlAP2_3</td>
<td>7031</td>
<td>SGN-US63215</td>
<td>DREB subfamily A-1 of ERF/AP2 TF family</td>
<td>981</td>
<td>35</td>
<td>8</td>
<td>4.4</td>
</tr>
<tr>
<td>SlAP2_7</td>
<td>7965</td>
<td>SGN-US87768</td>
<td>DREB subfamily A-4 of ERF/AP2 TF family</td>
<td>775</td>
<td>13</td>
<td>4</td>
<td>3.3</td>
</tr>
<tr>
<td>SlAP2_4</td>
<td>10714</td>
<td>SGN-US86437</td>
<td>ERF (ethylene response factor) subfamily B-3 of ERF/AP2 TF family</td>
<td>611</td>
<td>16</td>
<td>4</td>
<td>4.0</td>
</tr>
<tr>
<td>SlAP2_10</td>
<td>14672</td>
<td>SGN-US85539</td>
<td>AP2 domain-containing TF</td>
<td>297</td>
<td>111</td>
<td>45</td>
<td>2.5</td>
</tr>
<tr>
<td>SlAP2_5</td>
<td>16204</td>
<td>SGN-US77088</td>
<td>ERF (ethylene response factor) subfamily B-4 of ERF/AP2 TF family</td>
<td>523</td>
<td>4</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>SlAP2_11</td>
<td>25582</td>
<td>SGN-US84756</td>
<td>ERF (ethylene response factor) subfamily B-2 of ERF/AP2 TF family</td>
<td>660</td>
<td>7</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>SlAP2_11</td>
<td>25585</td>
<td>SGN-US86438</td>
<td>ERF (ethylene response factor) subfamily B-3 of ERF/AP2 TF family</td>
<td>942</td>
<td>96</td>
<td>3</td>
<td>32.0</td>
</tr>
<tr>
<td>SlAP2_8</td>
<td>26452</td>
<td>SGN-US81852</td>
<td>Ethylene-responsive element-binding factor 4 homolog</td>
<td>788</td>
<td>346</td>
<td>137</td>
<td>2.5</td>
</tr>
<tr>
<td>SWRKY2</td>
<td>9827</td>
<td>SGN-US65154</td>
<td>WRKY family TF</td>
<td>1366</td>
<td>4</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>SWRKY3</td>
<td>10561</td>
<td>SGN-US84367</td>
<td>WRKY family TF</td>
<td>1326</td>
<td>3</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>SWRKY1</td>
<td>13300</td>
<td>SGN-US65157</td>
<td>WRKY family TF</td>
<td>1099</td>
<td>7</td>
<td>1</td>
<td>7.0</td>
</tr>
<tr>
<td>SWRKY4</td>
<td>20918</td>
<td>SGN-US71278</td>
<td>WRKY family TF</td>
<td>1453</td>
<td>3</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>SlMYC1</td>
<td>24332</td>
<td>SGN-US76396</td>
<td>MYC TF</td>
<td>2174</td>
<td>191</td>
<td>128</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Tissue specificity and JA responsiveness of selected transcription factors

The candidate transcription factors (TFs) were selected from a sequencing trichome database as potential regulators of terpene biosynthesis. Ideally these TFs would be trichome-specific genes, induced by jasmonic acid. In order to investigate the expression pattern of these genes, cDNA was synthesized from different *S. lycopersicum* cv. Moneymaker tissues: leaves, stems, isolated stem trichomes and roots from 4-week-old plants, as well as flowers and fruit of mature plants. In Figure 3 expression data are presented for four of the sixteen selected transcription factors. For the other candidate TFs expression in the trichomes was much lower than that in other tissues (data not shown) and these were thus discarded from further analysis. TF *SIMYC1* was predominately expressed in trichomes, but also in leaves and flowers (Fig.3a). *SIWRKY1* was expressed in leaves, trichomes, roots and flowers (Fig.3b). *SIWRKY3* was a trichome-specific gene (Fig.3c) and *SIWRKY4* was expressed in trichomes, roots and fruit (Fig.3d). None of the selected transcription factors was significantly induced by JA according to the Q-RT-PCR data (Fig.3). *SIWRKY4* expression appeared to be roughly 1.7-fold reduced in JA treated plants (p=0.07).
Figure 3. Tissue specific expression and JA induction of selected TFs. Relative transcript levels for (a) *SlMYC1* (b) *SIWRKY1*, (c) *SIWRKY3* and (d) *SIWRKY4* as determined by Q-RT-PCR. Mean values (+SE) of 3 biological replicas are shown, normalized for *Actin* expression. L; leaf, WS; whole stem, BS; bald stem, T; stem trichomes, R; root, Fr; fruit, Fl; flower; C; control and JA; jasmonic acid induced stem trichomes. ns; not significant according to T-test.

*SlMYC1* and *SIWRKY4* can transactivate terpene synthase promoters in *Nicotiana benthamiana* leaves

In order to investigate whether these TFs can interact with terpene synthase promoters, a transient assay was used in *Nicotiana benthamiana* leaves, which has
been previously shown to work for the interaction between SlEOT1 and the SlMTS1 promoter (Chapter 4). In the reporter construct, expression of β-glucuronidase (GUS) was driven by the trichome specific promoter of SlMTS1 and therefore GUS activity was not detected in the leaf. However, co-infiltration with the 35S:SlEOT1 effector construct, which was expressed in leaves, transactivated the SlMTS1 promoter, leading to GUS expression in this heterologous system. As negative control a 35S:RFP construct was used. Various other reporter constructs with promoters of terpene synthases SlTPS3, SlTPS7 and SlTPS8 driving expression of β-glucuronidase (GUS) or GUSsYFP1 fusion (SlTPS9; Chapter 2) were used as well.

As shown in Figure 4, SlWRKY4 could transactivate the SlMTS1 promoter albeit to a lower extent than SlEOT1. The SlTPS3 and SlTPS7 promoters were weakly transactivated by SlWRKY4.

Transactivation of all tested promoters by SlWRKY1 and SlWRKY3 was similar to that of the negative control (35S:RFP) (data not shown).

![Graph](image)

**Figure 4.** Transactivation of terpene synthase promoters by SlWRKY4 in N. benthamiana leaves. Normalized GUS activity after co-infiltration with *A. tumefaciens* harboring the 35S:SlEOT1 (positive control) or the 35S:SlWRKY4 effector construct and various promoter:GUS reporter constructs. The 35S:RFP effector construct was used as negative control. The bars represent the obtained mean values and the error bars the standard error (n=3). RFP; red fluorescent protein. Bars with the same letters represent groups with the same mean according to Tuckey’s B posthoc test. Representative results from two experiments are shown.
The other TF investigated, SIMYC1, could transactivate all terpene synthase promoters tested except *SitPS8* and interaction with the trichome specific *SIMTS1* and *SitPS3* promoters was the strongest (Fig. 5). However it should be noted that GUS activity of a promoter driving the GUSsYFP1 fusion was lower than when the same promoter driving GUS alone was transactivated by an effector construct (data not shown), possibly because the fusion protein was less stable or because it was more difficult to be synthesized. Therefore, interaction of SIMYC1 with the trichome-specific *SitPS9* promoter is potentially stronger than that shown here.

![Graph showing normalized GUS activity](image)

**Figure 5.** Transactivation of terpene synthase promoters by SIMYC1 in *N. benthamiana* leaves. Normalized GUS activity after co-infiltration with *A. tumefaciens* harboring the 35S::*EOT1* (positive control) or the 35S::SIMYC1 effector construct and various promoter::GUS reporter constructs. The 35S::RFP effector construct was used as negative control. The bars represent the obtained mean values and the error bars the standard error (n=3). *RFP*: red fluorescent protein. Bars with the same letters represent groups with the same mean according to Tuckey’s B posthoc test. Representative results from three experiments are shown.

Furthermore it was investigated whether this transcription factor could interact with individual promoter fragments of the model *SIMTS1* promoter in yeast. As shown in Figure 6, SIMYC1 could interact only with the full-length promoter and not with the various promoter fragments. These data suggest either the existence of two
binding sites necessary for activity or that the binding site was interrupted in the promoter deletion fragments.

Figure 6. Interaction of SlMYC1 with the SlMTS1 promoter. Yeast cells with the full-length (FL) or shorter SlMTS1 promoter fragments A (-1254 to -1047bp), B (-1046 to -807bp), C (-806 to -613bp), D (-612 to -409bp), E (-408 to -208bp) and F (-207 to -1bp) integrated into their genome were transformed with pAD-GAL4-2.1_SlMYC1 and grown on SD medium with 5mM 3-AT.

Since SlEOT1, SlMYC1 and SlWRKY4 were shown in independent experiments to be able to transactivate the SlMTS1 promoter, we investigated what effect a combination of these transcription factors would have on the transactivation of SlMTS1 promoter. To this end, Agrobacterium cultures carrying the CaMV 35S-driven effector constructs were mixed in pairs or all three together and combined with the SlMTS1p:GUS reporter construct in N. benthamiana leaves (Fig.7). Interestingly, co-expression of SlEOT1 and SlMYC1 almost tripled the transactivation of SlMTS1 promoter compared to the effect of each TF alone. Adding SlWRKY4 did not have an additional effect, but rather seemed to have a negative effect on the combinatorial action of the other two TFs exerted on the SlMTS1 promoter, although not at a statistically significant level.
Figure 7. Transactivation of SlMTS1 promoter by SlEOT1, SlMYC1, SlWRKY4 or combination thereof in N. benthamiana leaves. Normalized GUS activity after co-infiltration with A. tumefaciens harboring the 35S:SlEOT1 (positive control), 35S:MYC1, 35S:SlWRKY4 effector constructs or combination thereof and the SlMTS1p:GUS reporter construct. The 35S:RFP effector construct was used as negative control. The bars represent the obtained mean values and the error bars the standard error (n=4). RFP; red fluorescent protein. Bars with the same letters represent groups with the same mean according to Tuckey’s B posthoc test. Representative results from two experiments are shown.

DISCUSSION

Although trichomes constitute a small fraction of a plant’s total tissue, they received distinctive attention for their ability to synthesize, store and secrete specialized metabolites. Through the production of mainly EST libraries, as well as micro-arrays and more recently high-throughput sequencing of (glandular) trichomes, research has focused on the expression of genes involved in the terpenoid, phenylpropanoid, alkaloid and flavonoid biosynthesis in various plant species, including tomato (McDowell et al., 2011; Schilmiller et al., 2010; Besser et al., 2009), sweet basil (Xie et al., 2008; Gang et al., 2001), tobacco (Cui et al., 2011; Harada et al., 2010), mint (Lange et al., 2000), alfalfa (Aziz et al., 2005), Artemisia annua (Wang et al., 2009) and hop (Wang et al., 2008).
Here we used massive parallel pyrosequencing on the 454 GS FLX Titanium platform to sequence *S.lycopersicum* stem trichome RNAs with the ultimate goal to identify transcription factors that are involved in terpene biosynthesis. In order to obtain a broad transcriptomic database optimal for gene discovery, it was deemed necessary to use normalized cDNA, so as to maximize representation of low abundant transcripts and reduce representation of highly abundant transcripts. Initial attempts to map the obtained reads to publicly available EST databases led to a high percentage of unmapped reads and assignment of the same reads to multiple unigenes and therefore the reads were in the end assembled *de novo*. 2.47% of the reads could not be matched and were not used in further analysis. 87.9% of the resulting contigs were subsequently annotated after blasting against the SGN tomato database using a local E-Blast tool.

First, in order to obtain a general overview of the transcripts that can be found in the tomato stem trichomes we used the functional annotation workstation Blast2GO. When blasting against the NCBI database through this automated software 82.6% of the contigs were annotated. Blast2GO provides a collection of tools that enable the assignment of Gene Ontology terms as well as Enzyme Classification codes with KEGG maps and InterPro motifs to the submitted sequences. However, for the data mining and the interpretation of the statistical results it must be kept in mind that one sequence can be assigned to more than one GO term. According to the InterPro entries (Table 1) by far the highest represented in the *S.lycopersicum* stem trichomes are protein kinases, which is not surprising since protein phosphorylation is a process implicated in responses to various signals and many regulatory enzymes are controlled by reversible phosphorylation (Stone and Walker, 1995). In addition, there are many proteins that contain WD40 repeats. Such proteins could have various roles in the trichomes as they are known to regulate diverse cellular functions, like cell division and trans-membrane signaling (Neer *et al*., 1994) and some could participate in the trichome formation, as in *Arabidopsis*, TRANSPARENT TESTA GLABRA1 (TTG1), a WD40-repeat protein, is one of the partners in the complex responsible for trichome initiation.
Moreover, several tomato stem trichome proteins appear to have zinc-finger domains. Since zinc finger-containing proteins are involved in a plethora of processes such as DNA recognition, transcriptional activation, protein folding and assembly and lipid binding (Laity et al., 2001) it is hard to speculate the particular function of these proteins in the trichomes. Furthermore, proteins with Armadillo-type repeats are also highly represented in the tomato stem trichomes. The Armadillo proteins are known to be involved in intra-cellular signaling and in Arabidopsis they have been shown to promote lateral root development (Coates et al., 2006). Finally, among the highly represented proteins are also cytochrome P450s, enzymes that are important for the biosynthesis of several compounds, such as hormones and defensive compounds involved in plant-insect interaction (Schuler, 2010).

In order to investigate whether in stem trichomes of tomato Moneymaker plants, jasmonic acid (JA) regulation of terpene biosynthesis is also on the precursor level except on the level of individual TPSs (Falara et al., 2011; Chapter 2), the quantitative database was mined for the genes encoding enzymes of the precursor pathways. The copy number of these genes varies between different plant species (Tholl and Lee, 2011) and as shown in Figure 2 different family members can vary in their expression levels and/or JA-inducibility. For example, although the methylerthritol (MEP) pathway enzymes are generally single copy (Phillips et al., 2008), 1-deoxy-d-xylulose 5-phosphate synthase (DXS) in contrast with Arabidopsis, which contains a single functional gene, has diversified into two isogenes in other plant species among which is tomato (Walter et al., 2002). Interestingly, while SlDXS1 is ubiquitously expressed, SlDXS2 is expressed only in a few tissues and in leaf trichomes its transcript abundance is much higher than that of SlDXS1 (Paetzold et al., 2010). Furthermore, SlDXS2 is moderately induced by wounding in cultivar Moneymaker (Paetzold et al., 2010), which correlates with the observed moderate induction of SlDXS2 by JA (1.7-fold, Fig.2). However, in cultivar Castlemart Manduca sexta larvae feeding up-regulates SlDXS2 expression approximately threefold (Sanchez-Hernandez et al., 2006). The regulation of
precursor enzymes of the MEP pathway by wounding, hormones or elicitors has been demonstrated in various plant species (Sanchez-Hernandez et al., 2006; Okada et al., 2007; Oudin et al., 2007; Arimura et al., 2008; Kim et al., 2009; Paetzold et al., 2010). Evidence for the regulation of precursor enzymes of the mevalonate (MVA) pathway is also abundant (Yang et al., 1991; Choi et al., 1994; Ha et al., 2003; Hui et al., 2003; Kondo et al., 2003; Bede et al., 2006). For example, HMGR enzyme activity and protein level was shown to be increased by fungal infection in potato tubers and sweet potato root (Kondo et al., 2003).

Furthermore, HMGRI transcripts were induced by treatment with methyl jasmonate also in potato, whereas HMGRI2 transcripts were reduced (Choi et al., 1994). However, in response to caterpillar herbivory, transcripts of alfalfa HMGRI1 were reduced (Bede et al., 2006). In tomato stem trichomes HMGRI1 and HMGRI3 were induced by JA treatment approximately 2-fold, whereas expression of HMGRI2 remained unaltered (Fig.2). None of the prenyl diphosphate synthases was induced in tomato trichomes by JA treatment, whereas two seemed to be downregulated (SGN-U578686; FPS and SGN-U573348; GGPS; Fig.2). However, expression of GGPS1 (SGN-U574849; transcripts not found in the stem trichome database) in tomato cv. Castlemart leaves was induced by spider mite feeding and JA (Ament et al., 2006) and in pepper roots FPS was induced by fungal infection (Ha et al., 2003). The emerging picture indicates that intermediates of the MEP and MVA pathways are subject to differential regulation between species and it would be interesting to confirm the obtained data for the Moneymaker stem trichomes by Q-RT-PCR.

As mentioned above, the primary goal of this sequencing run was to identify transcription factors that regulate terpene biosynthesis. Based on the annotated contigs (blasted against the SGN database) 2.7% of the transcripts in the tomato stem trichomes encode transcription factors. For comparison, in Arabidopsis thaliana ~6% of the genes in all tissues encode TFs (TAIR10 genome release, http://arabidopsis.org). In order to narrow down the number of transcription factors
to those that could potentially be involved in plant defense, a quantitative transcript profiling database was created by sequencing trichome RNAs from plants treated with JA or control solution. Terpene production and/or release are part of the plant’s indirect defenses, which involve emission of herbivore-induced volatiles that can be recognized by carnivorous and parasitoid insects. These defense responses to chewing arthropods are coordinated mainly by JA (Ballare, 2011) and therefore it was hypothesized that TFs involved in the regulation of terpene biosynthesis could also be JA-inducible genes. The transcription factors known to be involved in regulation of the terpenoid pathway (CrORCA3; van der Fits and Memelink, 2000; GaWRKY1; Xu et al., 2004; AaWRKY1; Ma et al., 2009; CrWRKY1; Suttipanta et al., 2011) are all jasmonate-inducible genes. However, from the field of JA signaling research in Arabidopsis, it was recently discovered that two previously unidentified MYC transcription factors (AtMYC3 and AtMYC4) are also direct targets of JAZ repressors and act additively with AtMYC2 in the activation of JA responses. Interestingly, in contrast to AtMYC2, AtMYC3 and AtMYC4 are only marginally induced by JA treatment (Fernandez-Calvo et al., 2011). Based on all the above, the initial selection of transcription factors to be analyzed from our quantitative stem trichome database was limited to TFs of the WRKY and AP2 class that showed a 2-fold or higher induction by JA treatment (2.3-fold was the induction rate of control gene SlMTS1). None of the MYC transcription factors of our database showed induction higher than 2, so for further analysis the closest homolog of AtMYC2 was selected. After discarding TFs that were not trichome-specific or at least showed highest expression in trichomes, the list was narrowed down to four candidate transcription factors. According to the Q-RT-PCR data however, none of these TFs was significantly induced by JA treatment (Fig.3). This could be explained in two ways. First, since the read count of these genes is very low both in the Control and JA samples (Table 4), perhaps the fold induction is over-estimated as 1 read versus 3 or 7 need not necessarily translate to a significant induction. Alternatively, since the RNA material used in the GAII sequencing run originated from pooled trichomes collected 30min, 2h, 8h
and 24h after JA treatment whereas the material for Q-RT-PCR was collected 24h after treatment, it is possible that these TFs show higher induction at an earlier time point, which remains to be tested.

A further and more specific indication of whether these TFs are involved in the terpene biosynthesis pathway would be to observe an interaction between the transcription factor and a terpene synthase. In transient transactivation assays in *N.benthamiana* leaves two of the four selected transcription factors were able to transactivate at least one terpene synthase promoter. SIWRKY4 showed strongest interaction with *SlMTS1* and in lesser extent with *SlTPS3* and *SlTPS7*. Although this TF is expressed highly in roots (Fig.3), it could not interact with the root-specific *SlTPS8*. It could also not interact with the trichome-specific sesquiterpene synthase *SlTPS9* so it is possible that SIWRKY4 can interact only with monoterpene synthase promoters or at least not the sesquiterpene synthase tested here (Fig.4). As shown in Table 5 this TF and the respective TPSs that it can transactivate are co-expressed in various tissues where the regulation could take place in the plant.

SlMYC1 showed strongest interaction with *SlMTS1* and *SlTPS3* and to a lesser extent with *SlTPS7* and *SlTPS9* but no interaction with *SlTPS8* (Fig.5), although this TF is also expressed in the root albeit not strongly (Fig.3). As shown in Table 5 SlMYC1 is expressed (at different levels) in every plant tissue and is able to transactivate all the terpene synthase promoters tested except one, so it seems to be a regulator of multiple TPSs, in contrast to SlEOT1 that is only expressed in the trichomes and can specifically interact with the SlMTS1 promoter and none of the other TPS promoters tested.
Table 5. Expression patterns and interaction overview of TFs and TPSs. Positive interaction of the transcription factors with the respective SlTPS promoters in N. benthamiana leaves is indicated by a colored box that represents the tissue in which they are co-expressed. Expression in the various tissues is indicated by +++, ++, +, +/- and -- according to Q-RT-PCR values. Darker shaded boxes indicate a stronger transient interaction between the TF and TPS.

<table>
<thead>
<tr>
<th></th>
<th>Leaf</th>
<th>Stem</th>
<th>Trichomes</th>
<th>Root</th>
<th>Fruit</th>
<th>Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>SlWRKY4</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td>SlMTS1</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS3</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
</tr>
<tr>
<td>SlTPS7</td>
<td>++</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
<td>--</td>
<td>++</td>
</tr>
<tr>
<td>SlTPS8</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS9</td>
<td>--</td>
<td>--</td>
<td>+++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Leaf</th>
<th>Stem</th>
<th>Trichomes</th>
<th>Root</th>
<th>Fruit</th>
<th>Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>SimYC1</td>
<td>+</td>
<td>+/-</td>
<td>++</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
</tr>
<tr>
<td>SimTS1</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS3</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
</tr>
<tr>
<td>SlTPS7</td>
<td>+</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
<td>--</td>
<td>++</td>
</tr>
<tr>
<td>SlTPS8</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS9</td>
<td>--</td>
<td>--</td>
<td>+++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Leaf</th>
<th>Stem</th>
<th>Trichomes</th>
<th>Root</th>
<th>Fruit</th>
<th>Flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>SlEOT1</td>
<td>--</td>
<td>--</td>
<td>+</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SimTS1</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS3</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
</tr>
<tr>
<td>SlTPS7</td>
<td>+</td>
<td>+/-</td>
<td>--</td>
<td>+/-</td>
<td>--</td>
<td>++</td>
</tr>
<tr>
<td>SlTPS8</td>
<td>--</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SlTPS9</td>
<td>--</td>
<td>--</td>
<td>+++</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

The interaction of SimYC1 with the promoter of SimTS1 in planta was confirmed also in yeast (Fig.5 and 6), however the binding region of this TF could not be identified. In the promoter sequence of SimTS1 (Fig.1, Chapter 3) there are two G-box-like elements (CACATG instead of the canonical CACGTG), one T/G-box element (AACGTG) and one T/G-box-like element (TACGTG), which could potentially be the binding site(s) of SimYC1. However even though the promoter deletion fragments used in yeast to test binding of SimYC1 did not interrupt any of
these motifs (Fig.6) no interaction was observed, suggesting that perhaps this TF has a binding site that has not been described before. The promoter of *SlTPS3*, with which *SlMYC1* interacts less strongly, contains one G-box-like element at position -1315 bases upstream of the ATG start codon and one T/G-box element at position -147 bases. The *SlTPS7* promoter, with which it can also interact, albeit even less strongly, contains one T/G-box at position -260 bases upstream of the ATG start codon. The *SlTPS9* promoter however, does not contain any of these elements, further indicating the existence of an uncharacterized motif on which *SlMYC1* binds. When using the naïve motif search program MEME (http://meme.sdsc.edu/meme/intro.html) for DNA sequences that are present in all four promoters which *SlMYC1* can transactivate transiently in *N.benthamiana* leaves, one 8bp motif was identified in the plus or minus (for *SlTPS9*) orientation: CTAGG(T/A)(A/G)G. The validation of this putative regulatory element or of the G-box-like or T/G-box motifs as the binding site(s) for *SlMYC1* in the respective promoters would require further experimentation.

Most strikingly, it was shown that *SlEOT1* and *SlMYC1* act synergistically in the activation of the *SlMTS1* promoter (Fig.7). Combinatorial control of transcriptional regulation is commonly found in plants and other eukaryotes (Singh, 1998). For example, in abscisic acid (ABA) signaling, the 67bp promoter region of the dehydration-responsive gene *rd22* contains a MYC and a MYB recognition site, where AtMYC2 and AtMYB2 bind. In *Arabidopsis* leaf protoplasts it was shown that these TFs could individually activate transcription of β-glucuronidase driven by this 67bp promoter region of *rd22* and that the transient transactivation was stronger when AtMYC2 and AtMYB2 were combined (Abe *et al*., 1997). Transgenic plants overexpressing these TFs each showed ABA hypersensitivity but the effect was more profound in plants overexpressing both cDNAs (Abe *et al*., 2003). In order to investigate in-depth the role of *SlMYC1* in terpene biosynthesis, transgenic overexpressing plants have been created under the control of the strong, trichome-specific *SlTPS9* (Chapter 2) promoter and are awaiting further analysis. **T0** plants will be crossed with **T0** individuals from the *SlEOT1* overexpressing
plants (Chapter 4) so as to explore the combinatorial effect of these transcription factors in tomato plants.

ACKNOWLEDGEMENTS

We thank Mattijs Bliek (Vrije Universiteit, Amsterdam) for kindly providing the yeast strain PJ69-4A, Carlos Villarroel Figueroa for help with the bioinformatics, Juan M. Alba Cano for help with the statistical analyses and Harold Lemeris, Ludeck Tikovsky and Thijs Hendrix for the excellent care of the tomato plants.

REFERENCES


Chapter 5


SUPPLEMENTAL DATA

**Figure S1.** Nucleotide sequence of transcription factors *SlWRKY1* (SGN-U565157), *SlWRKY3* (SGN-U584367), *SlWRKY4* (SGN-U571278) and *SlMYC1* (SGN-U576396).

**SlWRKY1**

```
atgGAAGATAGTTCTACATAAAATAATCTATTTTTTCATAAACAAAGAAGATTTCCACCACCGGAACTCCACCCGA
    TAATGCTGCTGATTTCTGTTTCTCCGTGAGTGAAGCGGCGGAAGTTAGCATGCCATCACCTAGAAAA
    AGTAAAGAGAGAGAAAAAAGAAGTAATTTTCTACGACCAATATTTGAGGATGAGAAGATACAAAGA
    AAAAGGGGAAGTTTATCCACCACACAGATTTCTGCTTGGGAAATATGGAACACCAATCTAAAAG
    GATCACCCTTATCCCGAGGGATATTATCGAGATGACTCATAAAAGAGGCTGCTCCAGGAAACAAGTAATC
    CGAAGCCGAGCCTGCTGACCTTCCACACATGCTTCTCCGCTCGTATACTCCGAGAATATGTGCTTAGA
    AAATCTCCCCCAAAGACCCAGATCTTGACATCTTGGGTAAATGTTGCGGTTATGCAGGAGGAAGATGTC
    GTGTACGGTCTGACGGTGAAAGGAACTCCTCGTCAAGGTCTGCTTGTGTTTTCCGGCGGTACAGTGGAATC
```

**SlWRKY3**

```
atgTCTGATAATCCTTTTTATATAGGAAATTTCCCGGAACTCCACCCGAACTCCACCCGAACTCCACCCGAACTCC
    TAATGCTGCTGATTTCTGTTTCTCCGTGAGTGAAGCGGCGGAAGTTAGCATGCCATCACCTAGAAAA
    AGTAAAGAGAGAGAAAAAAGAAGTAATTTTCTACGACCAATATTTGAGGATGAGAAGATACAAAGA
    AAAAGGGGAAGTTTATCCACCACACAGATTTCTGCTTGGGAAATATGGAACACCAATCTAAAAG
    GATCACCCTTATCCCGAGGGATATTATCGAGATGACTCATAAAAGAGGCTGCTCCAGGAAACAAGTAATC
    CGAAGCCGAGCCTGCTGACCTTCCACACATGCTTCTCCGCTCGTATACTCCGAGAATATGTGCTTAGA
    AAATCTCCCCCAAAGACCCAGATCTTGACATCTTGGGTAAATGTTGCGGTTATGCAGGAGGAAGATGTC
    GTGTACGGTCTGACGGTGAAAGGAACTCCTCGTCAAGGTCTGCTTGTGTTTTCCGGCGGTACAGTGGAATC
```

**SlWRKY4**

```
atgGAAGAGTAGGGAAGGAAAATACTAACAAACTAACAAACTAACAAACTAACAAACTAACAAACTAACAAACTAACAA
    TAATGCTGCTGATTTCTGTTTCTCCGTGAGTGAAGCGGCGGAAGTTAGCATGCCATCACCTAGAAAA
    AGTAAAGAGAGAGAAAAAAGAAGTAATTTTCTACGACCAATATTTGAGGATGAGAAGATACAAAGA
    AAAAGGGGAAGTTTATCCACCACACAGATTTCTGCTTGGGAAATATGGAACACCAATCTAAAAG
    GATCACCCTTATCCCGAGGGATATTATCGAGATGACTCATAAAAGAGGCTGCTCCAGGAAACAAGTAATC
    CGAAGCCGAGCCTGCTGACCTTCCACACATGCTTCTCCGCTCGTATACTCCGAGAATATGTGCTTAGA
    AAATCTCCCCCAAAGACCCAGATCTTGACATCTTGGGTAAATGTTGCGGTTATGCAGGAGGAAGATGTC
    GTGTACGGTCTGACGGTGAAAGGAACTCCTCGTCAAGGTCTGCTTGTGTTTTCCGGCGGTACAGTGGAATC
```
TCCAAATTAATTTTGAGCAAAAAAGAGAATGTGAGAGGAGAGAAGGAACTACCTAACCTGCTAGCTTAA
GATTGAAGGAGGACGAAAGAAAGGAACTGTACAAAAGTGGCACCTCACAACAGATCCTTGAGAAAG
GAATGGAAGGAAATGATGGGAGCACAAACTACCCAACTACACTACAAAGGAAAAGTTTCCGTCAGAGTTA
GATGGAACACCCAAACTACAGTGAACCTCAAAATGGAAGAAATGGAAGAAATACAAAGTTCAAGAG
GAACACCATGTCCTAGAGCTTTACTGTTGCACGTTAAGCTATGTTGTTCT
CTGTAATAACTCAAACACACACCTTTATACACCCACACTACATCTCCACTCCATTTTCTTCACACACTCCATCTC
AGTACCTTCATAAATACCTCTCTTTGATTTTAAACATAAATATCTGGGCTACTCTAAGTCTTATTACAC
AGACAAAGAATTCGAGTGTCAAAACAATAACATAGCAGATCGGAAATTATGCAAGAAAATCTGGGCAAGAT
ATGAAGGTCTACTGAGCCATTTTCAGTCTTTGTAGCTCCATCCACCTCCTCTCGTTGATCAGATAAA
TTACACA

SI"MYC1"

tagACGGGACTATAAGTTATGGAGTAATACCAATACACTAAAATACATATGTGATGATACTATGATGATGGAA
TTCTTTTTTATCTCCGATCCTCCTCTCTTTTTGAGGGCTCTCACCCTCCTAACCTGGCAGCTTGAAC
CGGAGTCGGAAGAACACGAGTTTTTCAAATACAACTGCAAGGAAGAGAAGAAGAAGAAGGAGGAAA
CGGAGTGGCACGTGTATGAGGTGGATACATCAGCAGTCTTATAGTTTTCGAGAGTTCT
CTGTAATAACTCAAACACACACCTTTATACACCCACACTACATCTCCACTCCATTTTCTTCACACACTCCATCTC
AGTACCTTCATAAATACCTCTCTTTGATTTTAAACATAAATATCTGGGCTACTCTAAGTCTTATTACAC
AGACAAAGAATTCGAGTGTCAAAACAATAACATAGCAGATCGGAAATTATGCAAGAAAATCTGGGCAAGAT
ATGAAGGTCTACTGAGCCATTTTCAGTCTTTGTAGCTCCATCCACCTCCTCTCGTTGATCAGATAAA
TTACACA

ATACCGGATCCTAGATGAGGAGATGATTGAATAGTTAATAACTCAGCGCTCAGAGCCGTAGTCCCAAATGTGTCTAAAATGGATAAGGCATCACTTC
TTGGAGATGCAATTGCATACATCAATGAGTTGAAATCAAAAGTTCAAAATTCAGATTTAGATAAAGA
GGAGTTGAGGAGACCAGCCAGGCTCAGGCAAGTTTAAAGGAAAGGAAATAACACCAACAGGCCAGTACAA
AGGCCACTCCCTCATTGAAATCAAGATGTCAAGATTGCTATAGTGAGCATAGTTAGCTTAAAGTGGATATTGA
TGAGGATGCTACTTGTATTCTGATTACAAATGATGATGAAAAAAGACACATCTGAGCGCGCTACTTGGAAGCGGAGCC
TCAAGGAGTCTATTGACGTAGTCACCCCGCTACGGTGTGGATAGTTTATGATCAGATACAAC
AGCCACAGTCAAATTTGAGGAGGCCATTTTATGCTCAAGAACAGCTTATGGAGACCTTAGCATACCAA
AAATGGGAATACGCGAtga