Plant Physiology Preview. Published on October 5, 2015, as DOI:10.1104/pp.15.01353

Running head:

Diterpenoids biosynthesis in Stevia leaf tissues

Author for correspondence:

In-Cheol Jang

Temasek Life Sciences Laboratory and Department of Biological Sciences,

1 Research Link, National University of Singapore, Singapore 117604

Phone; (+65) 6872 7409

e-mail; jangi@tll.org.sg

Research area: Biochemistry and Metabolism

Comparative transcriptomics unravel biochemical specialization of leaf tissues of Stevia (*Stevia rebaudiana*) for diterpenoid production

Mi Jung Kim[1,5], Jingjing Jin[1,2,5], Junshi Zheng[1,3,5], Limsoon Wong[2], Nam-Hai Chua[4], and In-Cheol Jang[1,3,*]

[1]Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604

[2]School of Computing, National University of Singapore, Singapore 117417

[3]Department of Biological Sciences, National University of Singapore, Singapore 117543

[4]Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

[5]These authors contributed equally.

One-sentence summary

Metabolite-guided transcriptomics and biochemical enzyme characterization identified biosynthetic routes of specialized diterpenoids in leaf tissues of Stevia.

Footnotes:

This research was funded by grants from Temasek Life Sciences Laboratory and the Singapore National Research Foundation (Competitive Research Programme Award No: NRF-CRP8-2011-02) to ICJ.

Corresponding Author:

In-Cheol Jang; jangi@tll.org.sg

ABSTRACT

Stevia (*Stevia rebaudiana*) produces not only a group of diterpenoid glycosides known as steviol glycosides (SGs) but also other labdane-type diterpenoids that may be spatially separated from SGs. However, their biosynthetic routes and spatial distribution in leaf tissues have not yet been elucidated. Here we integrate metabolome and transcriptome analyses of Stevia to explore the biosynthetic capacity of leaf tissues for diterpenoid metabolism. Tissue-specific chemical analyses confirmed that SGs were accumulated in leaf cells but not in trichomes. On the other hand, Stevia leaf trichomes stored other labdane-type diterpenoids such as oxomanoyl oxide and agatholic acid. RNA-seq analyses from two different tissues of Stevia, provided a comprehensive overview of dynamic metabolic activities in trichomes and leaf without trichomes. These metabolite-guided transcriptomics and phylogenetic and gene expression analyses clearly identified specific gene members encoding enzymes involved in the 2-C-methyl-D-erythritol 4-phosphate pathway and the biosynthesis of steviol or other labdane-type diterpenoids. Additionally, our RNA-seq analysis uncovered *copalyl diphosphate synthase* (*SrCPS*) and *kaurene synthase 1* (*SrKS1*) homologs, *SrCPS2* and *KS-like* (*SrKSL*), which were specifically expressed in trichomes. *In vitro* and *in planta* assays showed that unlike SrCPS and SrKS1, SrCPS2 synthesized labda-13-en-8-ol diphosphate and successively catalyzed the formation of manoyl oxide and *epi*-manoyl oxide in combination with SrKSL. Our findings suggest that Stevia may have evolved to use distinct metabolic pathways to avoid metabolic interferences in leaf tissues for efficient production of diverse secondary metabolites.

INTRODUCTION

Stevia rebaudiana, a perennial shrub belonging to the Asteraceae family has been used for centuries in South America as sweetener for herbal teas and foods (De et al., 2013). In addition to being a sweetening agent, Stevia has been used as a cardiotonic for hypertension and heartburn and also to lower uric acid levels. For reasons of food and medicine, there has been wide-spread cultivation of this plant throughout Europe, Asia and North America (Philippe et al., 2014).

The sweetness of Stevia leaves is attributed to a group of diterpenoid derivatives known as steviol glycosides (SGs), which are 150-300 times sweeter than cane sugar. In contrast to artificial sweeteners, SGs are natural and can be used primarily as a non-calorie sweetener and/or flavor enhancer. Stevia accumulates SGs in its leaves to as high as 20% of the dry weight, which comprise of approximately 10 major components in varying levels (Geuns, 2003). All these compounds share a common backbone, a steviol aglycone, but with different number of sugar moieties (glucose, rhamnose and xylose) (Obtani and Yamasaki, 2002; Starratt et al., 2002). The quality and intensity of the taste are determined by the C-position (C-13 or C-19) of steviol being modified and the number of the sugar moieties (Obtani and Yamasaki, 2002).

As diterpenoid glycosides, SGs are synthesized from 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway in plastids, which produces two isoprene units, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) from pyruvate and glyceraldehyde-3-phosphate (G3P) through a series of 7 enzyme-catalyzed steps (Totté et al., 2000; Kumar et al., 2012). The MEP pathway is also important for the synthesis of many other terpenoids, in particular mono-, di- and tetraterpenes, since IPP and DMAPP are their common precursors (McGarvey and Croteau, 1995). Once the 5-carbon unit IPP is synthesized in plastids by the MEP

pathway, IPP isomerase (IDI) catalyzes the reversible isomerization of IPP to form DMAPP and this enzyme might play a regulatory role in determining cellular DMAPP levels (Adam et al., 2002; Hunter, 2007).

Among terpenoids, diterpenoids are classically defined by 20-carbon isoprenoids derived from the common precursor geranylgeranyldiphosphate (GGPP) which is catalyzed by GGPP synthase (GGPPS) inside plastids (Zi et al., 2014). Being required for the production of photosynthetic pigments and phytohormones (e.g. gibberellic acid and abscisic acid) diterpenoid biosynthesis is essential for plant growth and development. In addition, plants also produce over 10,000 different diterpenoids as secondary metabolites, which may function as direct defense compounds against herbivores and microorganisms (Zerbe et al., 2013; Zi et al., 2014). The most commonly found version of diterpenoids is labdane-type compounds such as gibberellic acids (GAs), phytoalexins, sclareol, forskolin, taxol and steviol, many of which are of significant value to human life as biobased phamaceuticals, flavors and fragrances (Zerbe and Bohlmann, 2015).

The structural diversity of different diterpenoids arises from divergent evolution of specialized diterpenoid biosynthetic pathways. These pathways share a common precursor GGPP and use a few types of enzymes including different classes of diterpene synthases (diTPSs), cytochrome P450-dependent monooxygenases (P450s), and various modifying transferases (e.g. acyl-, methyl-, or glycosyltranferases) which may be unique for a given plant species or family (Zerbe et al., 2013; Zi et al., 2014; Zerbe and Bohlmann, 2015). In angiosperms, specialized diterpenoids are characterized by their common biosynthetic origins in the initial dual cyclization and/or rearrangement reactions by a sequential pair of class II and class I diTPSs (Peters, 2010). Class II diTPSs contain a DxDD motif which is required for the protonation-initiated cationic cyclization and rearrangement of GGPP to distinct bicyclic diphosphate intermediates (Peters, 2010; Chen et al., 2011). Subsequently, class I

diTPSs possessing DDxxD and NSE/DTE motifs catalyze cleavage of the phosphate group of intermediates and further cyclize or rearrange the resulting carbocations (Peters, 2010; Chen et al., 2011).

In Stevia, SG biosynthesis shares three steps in common with GA biosynthesis pathway for the formation of kaurenoic acid, which requires two diTPSs, copalyl diphosphate synthase (CPS) and kaurene synthase (KS) and one member of the cytochrome P450 family, kaurene oxidase (KO) (Richman et al., 1999; Humphrey et al., 2006). Since GA is an important phytohormone for plant growth and elongation of cells, it is not surprising that these genes are conserved among higher plants (Hedden and Phillips, 2000). Moreover, an additional P450 monooxygenase, kaurenoic acid hydroxylase (KAH) is involved in the biosynthesis of steviol, which provides the backbone for all SGs (Brandle and Telmer, 2007). The last stage of SG biosynthesis is the glycosylation of the diterpenoid steviol by UDP-glycosyltransferases (UGTs) which transfer sugar residues from activated sugars to an aglycone acceptor (Richman et al., 2005; Brandle and Telmer, 2007).

The chemical complexity of lipophilic components produced in Stevia leaves has been previously examined by gas chromatography-mass spectrometry (GC-MS) showing the presence of diverse mono- and sesqui-terpenes and fatty acids (Markovic et al., 2008). Interestingly, apart from kaurene as a diterpenoid hydrocarbon precursor for GAs or steviol, other labdane-type diterpenoids such as manoyl oxide, sclareol, austroinulin and sterebins are also detected in Stevia leaves extracts, which may confer biological and pharmacological properties (Ibrahim et al., 2007; Markovic et al., 2008; Cho et al., 2013). However, the biosynthetic pathways or terpene synthases for these diterpenoids in Stevia have not yet been reported.

The biosynthesis and accumulation of many plant diterpenoids are restricted to specialized tissues or specific cell types, which enables plants to effectively synthesize specific natural chemicals and to avoid metabolic interference. For example, a diterpenoid forskolin accumulates in specialized root cork cells of *Coleus forskohlii* (Pateraki et al., 2014), diterpene resin acids in resin ducts of conifers (Zulak and Bohlmann, 2010) and Z-abienol in tobacco leaf trichomes (Sallaud et al., 2012). This cell/tissue-specific accumulation of diterpenoids may be associated with differential gene expression rather than their translocation and accumulation in specific cell/tissues. Although Stevia genes encoding CPS, KS and KO related to GA or steviol biosynthesis have been shown to express in leaf parenchyma (Humphrey et al., 2006), it is unclear where in the leaf tissues SGs are accumulated. In plants, specialized metabolites are stored in the place of their synthesis, or in some cases, they can be transported to other tissues (Alvarez, 2014).

The experimental approach of combining bioinformatics and metabolite analysis has facilitated the characterization of biosynthetic pathways and related genes for plant secondary metabolism (Bleeker et al., 2011; Zerbe et al., 2013; Jin et al., 2015; Zerbe and Bohlmann, 2015). As the Stevia genome has not yet been sequenced, only limited sources of gene information are available from randomly selected expressed sequence tags (ESTs) (Brandle et al., 2002). Recently, Chen et al., (2014) reported a RNA-seq dataset of three Stevia genotypes comparing expression patterns of previously reported genes involved in SG biosynthesis in these genotypes, but there were no additional new genes identified related to the MEP pathway, SG biosynthesis or the synthesis of other terpenoids. Detailed or enriched transcriptome datasets from different tissues of terpenoid-producing non-model plants not only provide opportunities for the discovery of new genes and enzymes, but also explain the biosynthetic ability of specialized terpene metabolism that may be spatially restricted to organs, tissues or cell types (Zerbe et al., 2013; Jin et al., 2014; Zerbe and Bohlmann, 2015).

Here, we performed comparative analysis of RNA-seq transcriptomes of two different tissues, namely, trichomes and leaves without trichomes (leaf-trichomes). Comparison of transcripts between these two tissues showed that more than 20% of the assembled unigenes were differentially expressed in trichomes, whereas genes specifically-expressed in the leaf-trichomes were less than 12%. Most genes involved in SG biosynthesis were predominantly expressed in leaf-trichomes where SGs accumulated. Moreover, we found that *SrCPS2* and *SrKSL*, *SrCPS* and *SrKS1* homologs, were specifically expressed in trichomes. Taken together, our results show that comparative transcriptomic analysis along with metabolite profiling provided useful information on the distinct biosynthetic pathways for specialized diterpenoids that preferentially accumulate in different tissues of the Stevia leaves.

172 **RESULTS**

173 **Metabolite profiles in Stevia leaves**

174 Steviol glycosides (SGs) are known to accumulate mainly in Stevia leaves (Brandle and Rosa,
175 1992; Kumar et al., 2012); however, it is unclear whether SGs are synthesized and stored in
176 specialized tissue such as trichomes or oil cells in Stevia leaf tissues. We first investigated
177 trichomes on the surface of Stevia leaves using scanning electron microscope (SEM). Fig. 1
178 shows that there were three different types of trichomes: glandular, long non-glandular and
179 short non-glandular. These three types of trichomes could be distinguished by their
180 differences in morphology, structure and size (Fig. 1). Among these, glandular trichomes in
181 other plants have been shown to produce, store and secrete large amounts of diverse
182 secondary metabolites such as terpenoids, phenylpropanoids and acyl sugars (Glas et al.,
183 2012). To examine if there is any differential distribution of terpenes in different cell/tissues
184 of Stevia leaves, we first purified trichomes (T) from young leaves using a standard glass
185 bead abrasion method (Lange et al., 2000). We also prepared leaf tissues stripped of
186 trichomes (leaf minus trichomes; L-T) by cold-brushing (Wang et al., 2001). Both SEM and
187 light microscopy confirmed the purity of the trichome preparation. Note that the preparation
188 still contained a mix of the three types of trichomes which are difficult to separate
189 (Supplemental Fig. S1).
190 Next, we used high-performance liquid chromatography (HPLC) and GC-MS to examine SG
191 content and other diterpenoids in T and L-T samples. Fig. 2A shows that most SGs were
192 exclusively found in L-T but not in T. In this HPLC analysis, each SG was identified by
193 comparison with the retention time of known standards. Whereas SGs were almost absent in
194 T, GC-MS analysis of T extract revealed a wider range of terpenes including mono-, sesqui-
195 and di-terpenes compared to the extract derived from L-T (Fig. 2B). Metabolites identified in
196 T are listed in Table I.

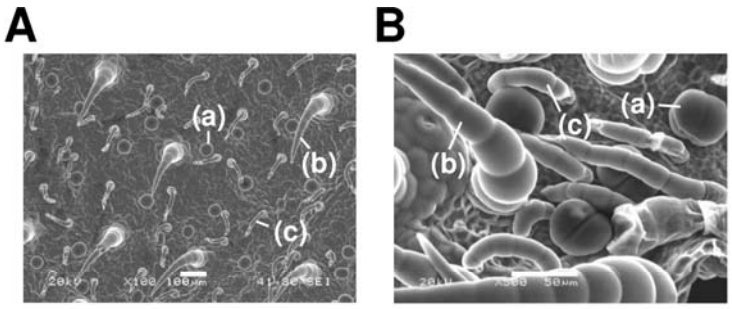


Figure 1. Glandular and filamentous trichomes of *S. rebaudiana*. A, Scanning electron microscopy (SEM) images of Stevia leaf showing three types of trichomes. (a) glandular trichome; (b) long non-glandular filamentous trichome; (c) short non-glandular filamentous trichome. B, A detailed view of the three types of trichomes of Stevia. Scale bars are included on all images.

197

198 **RNA-seq data of Stevia, *de novo* assembly and annotation of transcriptome**

199 To compare transcriptomes of different parts of Stevia leaves, we sequenced RNA-seq
200 libraries derived from T and L-T samples using Illumina HiSeq 2000. Illumina sequencing
201 runs generated more than 160 million high quality strand specific single-end reads of 101
202 base pairs (bp) from each of the two tissues. The quality of Illumina sequencing outputs was
203 high, as evaluated by FastQC (Supplemental Fig. S2). Our RNA-seq reads were *de novo*
204 assembled into 53,793 non-redundant unigenes, spanning a total of 95 Mb of sequence with a
205 GC content of 37.3%. The sequence length of the assembled unigenes was at least longer
206 than 200 bp. The resulting assembly had a N50 value of 1,482bp. As a result, the assemblies
207 produced 32,336 contigs for L-T and 45,918 contigs for T, giving a total of 53,793 unigenes
208 (Table II).

209 To predict accurate gene annotations and to maximize gene annotation percentages, the
210 assembled unigenes were blasted against the National Centre for Biotechnology Information
211 (NCBI) non-redundant protein database and protein databases from *Arabidopsis thaliana*,
212 *Vitis vinifera, Glycine max* and *Oryza sativa*. Amongst the 53,793 unigenes, 40,243 (74.8%)
213 unigenes were annotated through BLASTX search with E-value <= 1e-3. We mapped RNA-

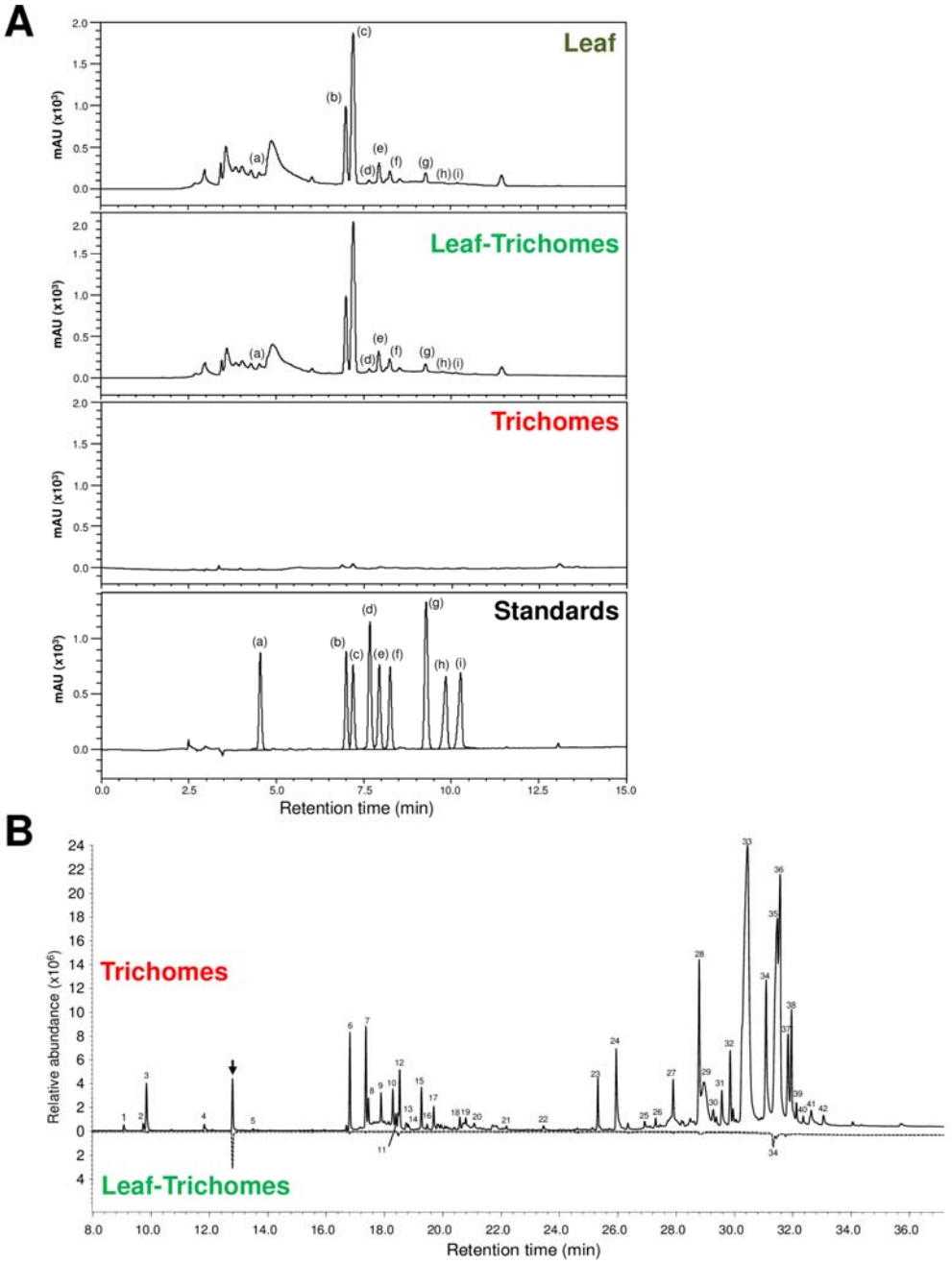


Figure 2. Distribution of diterpenoids in Stevia leaf tissues. Tissue-specific abundance of metabolites was evaluated by UHPLC (A) and GC-MS (B). A, Comparison of SGs content between leaf-trichomes and trichomes. Standards contain a mixture of nine known SGs. Peak (a) Reb D; peak (b) Reb A; peak (c) Stevioside; peak (d) Reb F; peak (e) Reb C; peak (f) Dulcoside A; peak (g) Rubusoside; peak (h) Reb B; peak (I) Steviobioside. B, GC traces showing the difference between leaf-trichomes and trichomes. The arrow indicates the peak of camphor (10 µg/µL), the internal standard used in the assay. The peaks numbered in GC traces were identical to those listed in Table I.

214 seq reads onto the assembled transcripts to calculate their expression levels using Bowtie

215 (Langmead and Salzberg, 2012). RSEM (RNA-seq by Expectation-Maximization) was used

to estimate the abundance of assembled transcripts and to determine the expression levels (Li and Dewey, 2011).

Differential gene expression between leaf without trichomes (L-T) and trichomes (T)

From the RNA-seq data, 32,336 and 45,918 unigenes were observed to be expressed in L-T and T, respectively. The two expression patterns were clearly distinct as shown by the heat map (Supplemental Fig. S3). Among the unigenes, 3,807 unigenes were preferentially expressed in L-T, whereas 10,529 unigenes were T-specific; the minimum differential expression level was at least 4 times between the two samples (Supplemental File S1). Note that more than 20% of the assembled unigenes were specifically expressed in T. Among the differential expressed genes, about 40% encoded either hypothetical proteins or were unannotated. These might be novel genes with sequences divergent from those of other plants but nevertheless essential for trichome development in Stevia and/or partly be a result of miss-assembled transcripts.

To further analyse possible functions of the differentially expressed unigenes from each sample we assessed their Gene Ontology (GO) classifications by Trinotate (Quevillon et al., 2005). Fig. 3 shows the GO terms for the more abundant unigenes in L-T or T. GO terms related to defense responses to fungus and virus and responses to salicylic acid stimulus were highly represented in T, reflecting the overall defence function of trichomes. However, unigenes from the GO terms of photosynthesis and chlorophyll biosynthesis were enriched in L-T but not detectable in T. Moreover, other terms for cellular components related to chloroplast envelope, thylakoid membrane and stroma were only found in L-T. These results were not unexpected since the L-T sample contained mesophyll cells where photosynthesis takes place. At the same time, our results showed that the T sample preparation was not

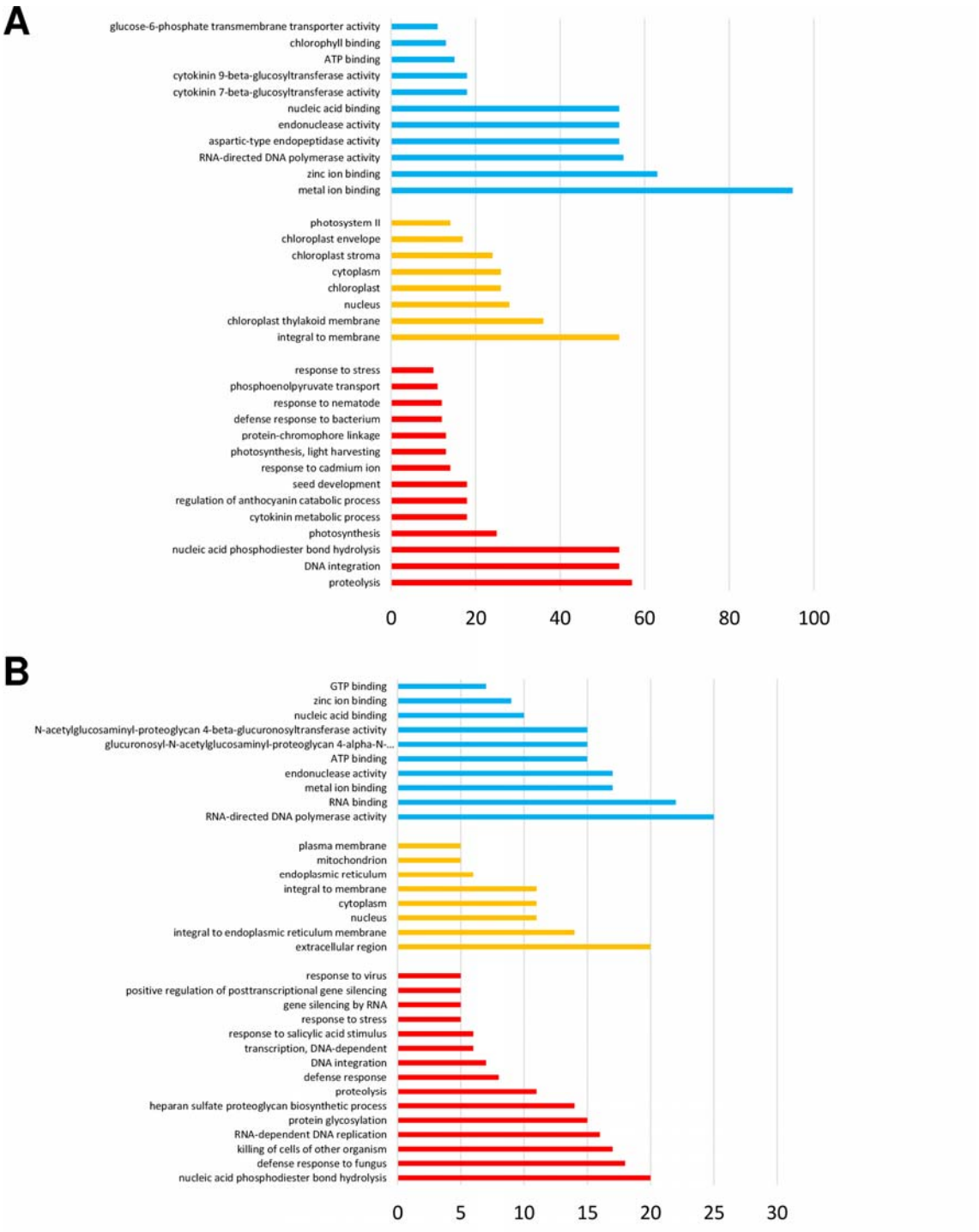


Figure 3. Gene Ontology (GO) analysis of more abundantly expressed unigenes in leaf-trichomes (A) or trichomes (B). X-axis, log(1/P-value). P-value is the hypergeometic test result for each GO terms.

240 contaminated by leaf tissues since Stevia trichomes are not implicated in photosynthesis by

241 GO annotation. Although genes encoding enzymes and proteins related to photosynthesis are

242 significantly expressed in glandular trichomes of tomato and tobacco (Harada et al., 2010;

243 Cui et al., 2011) they are not expressed in mint glandular trichomes (Lange et al., 2000)
244 indicating species differences between trichome types.

245 Since SGs were found in the L-T sample (Fig. 2A) we hypothesized that the 3,807 unigenes
246 that were preferentially expressed in this sample may provide an important resource for the
247 identification of novel genes involved in SG biosynthesis and its regulation.

248

249 **Expression and molecular analysis of MEP pathway and IPP isomerase (IDI) genes**

250 Fourteen genes encoding seven enzymes in the MEP pathway and IDI were identified from
251 our Stevia RNA-seq datasets, including genes for four DXSs (1-deoxy-D-xylulose 5-
252 phosphate synthase), two each of DXR (1-deoxy-D-xylulose 5-phosphate reductoisomerase),
253 HDR (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase) and IDI, one each of
254 MCT (2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase), CMK (4-(cytidine 5′-
255 diphospho)-2-C-methyl-d-erythritol kinase), MDS (2-C-methyl-d-erythritol 2,4-
256 cyclodiphosphate synthase) and HDS (4-hydroxy-3-methylbut-2-enyl diphosphate synthase)
257 (Supplemental Fig. S4). We found four *DXS* genes (designated as *SrDXS1-4*) containing the
258 complete open reading frame (ORF). Since *SrDXS* that is designated here as *SrDXS4* has
259 been previously reported (Totte et al., 2003, Accession number, AJ429232), three additional
260 *DXS* genes, *SrDXS1*, *SrDXS2* and *SrDXS3* are newly identified in this study. Among four
261 SrDXS proteins, SrDXS1 is the only member of the DXS1 clade, which is probably involved
262 in primary metabolism whereas SrDXS2 and SrDXS4 which belong to the DXS2 clade, may
263 be related to secondary metabolism (Fig. 4A; Supplemental Fig. S5; Walter et al., 2002;
264 Phillips et al., 2007). Some plant species have a third DXS protein cluster referred to as the
265 DXS3 clade, which has been proposed to participate in the synthesis of some products
266 essential for plant survival (Cordoba et al., 2009; 2011). We found that SrDXS3 was a

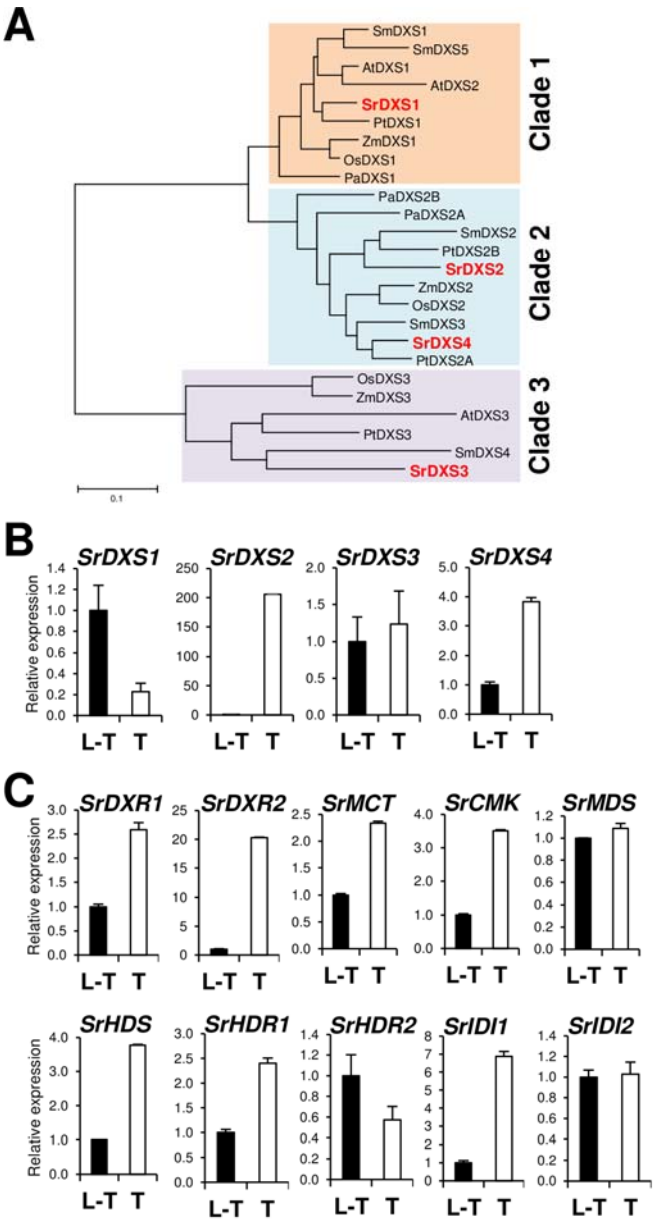


Figure 4. Analysis of genes involved in MEP pathway from *S. rebaudiana*. A, Phylogenetic analysis of DXSs from Stevia. The maximum likelihood tree was constructed by MEGA 6 program from an alignment of full-length SrDXSs with other plant DXSs. Abbreviations and accession numbers of proteins are listed in Supplemental Table S2. B, Relative expression levels of *SrDXS* genes in leaf-trichomes (L-T) and trichomes (T). C, Relative expression levels of *SrDXRs*, *SrMCT*, *SrCMK*, *SrMDS*, *SrHDS*, *SrHDRs* and *SrIDIs* in L-T and T. Transcript levels of genes expressed in L-T and T were measured by qRT-PCR. Amplification of *Actin* mRNA was used for normalization.

member of the DXS3 clade (Fig. 4A). We confirmed differential expression of *SrDXSs* by qRT-PCR using the L-T and T samples. Fig. 4B shows that *SrDXS1* transcripts were 4-5 times more abundant in L-T than T. On the other hand, *SrDXS4* which was previously

reported (Totte et al., 2003) was predominantly expressed in T. Interestingly, *SrDXS2* transcripts were rarely detectable in L-T, but were highly expressed more than 200 times in T.

From our RNA-seq data set, we were able to identify transcripts for all previously reported *DXR*, *MCT*, *CMK*, *MDS*, *HDS*, and *HDR* (Totte et al., 2003; Kumar et al., 2012). Moreover, we found transcripts for an additional copy of *DXR* and *HDR*, designated here as *DXR2* and *HDR2*, respectively. Note that *SrDXR1* (Accession number, AJ429233) and *SrHDR1* (Accession number, DQ269451) are previously reported genes (Totte et al., 2003; Kumar et al., 2012). Although SrDXR1 and SrDXR2 share 89.4% high amino acid identity (Supplemental Fig. S6), the expression pattern of *SrDXR1* and *SrDXR2* was clearly distinct in leaf tissues. Fig. 4C showed that *SrDXR1* expression was slightly higher in T compared to L-T by about 2.5-fold, whereas *SrDXR2* transcripts were 20-fold more abundant in T. The expression of other intermediate genes such as *SrMCT*, *SrCMK*, *SrMDS* and *SrHDS* was detected in both L-T and T samples (Fig. 4C). All genes except *SrMDS* were slightly more expressed in T by about 2-4 fold but *SrMDS* expression was comparable in both samples.

SrHDR2 showed 86.7% amino acid identity with SrHDR1 (Supplemental Fig. S7). However, genes for the two related proteins showed opposite expression patterns in L-T and T samples, but the differential expression level of these genes were not significant (Fig. 4C).

Transcripts for two *IDI* genes, *SrIDI1* and *SrIDI2* were found in our Stevia RNA-seq dataset. Although *SrIDI1* appeared to be the same as the previously reported *SrIDI* (Accession number, DQ989585) (Kumar et al., 2012) the protein encoded by *SrIDI1* is 64 amino acids longer than SrIDI at the N-terminus indicating that *SrIDI* might be a partial sequence (Supplemental Fig. S8). On the other hand, *SrIDI2* is a new sequence not reported before. Our gene expression analysis showed that *SrIDI1* was predominantly expressed more than 6 times in trichomes (Fig. 4C). By contrast, *SrIDI2* was expressed both in L-T and T. SrIDI1

294 and SrIDI2 share 68.4% identity at the amino acid level and both possess a transit peptide
295 sequence for plastidic targeting (Supplemental Fig. S8).

296

297 **Characterization of *GGPPS* gene family**

298 Analysis of our Stevia RNA-seq uncovered three *GGPPS* genes designated as *SrGGPPS1*,
299 *SrGGPPS2*, and *SrGGPPS3*. All three SrGGPPSs are clearly distinct from geranyl
300 diphosphate synthase (GPPS) or farnesyl diphosphate synthase (FPPS) family members and
301 have two conserved aspartate-rich domains, FARM (DDx_9RR) and SARM (DDxxD), which
302 are required for IPP and DMAPP substrate binding and catalysis (Fig. 5A and Supplemental
303 Fig. S9; Liang et al., 2002). *SrGGPPS1* transcripts were almost undetectable in the L-T
304 sample (Fig. 5B); rather, this gene was predominantly expressed more than 400 times in T.
305 By contrast, transcripts for *SrGGPPS2* and *SrGGPPS3* were more abundantly expressed in
306 the L-T sample. All three SrGGPPSs were predicted by the ChloroP program
307 (http://www.cbs.dtu.dk/services/ChloroP) to be plastid-localized. SrGGPPS1 and SrGGPPS3
308 had a putative N-terminal plastid transit peptide sequence of 33 and 45 amino acids,
309 respectively, whereas SrGGPPS2 was predicted to have only an extra 6 amino acids at the N-
310 terminus, which may be too short to function as a plastidic transit peptide sequence
311 (Supplemental Fig. S9). This was confirmed by subcellular localization of each SrGGPPS-
312 YFP fusion protein in *N.benthamiana* leaves using *Agrobacterium*-mediated infiltration. Fig.
313 5C shows that SrGGPPS1 and SrGGPPS3 were clearly localized in chloroplasts, whereas
314 SrGGPPS2 was distributed throughout the cytosol.

315

316 **Expression pattern of genes for diterpenoid steviol biosynthesis and their homologs**

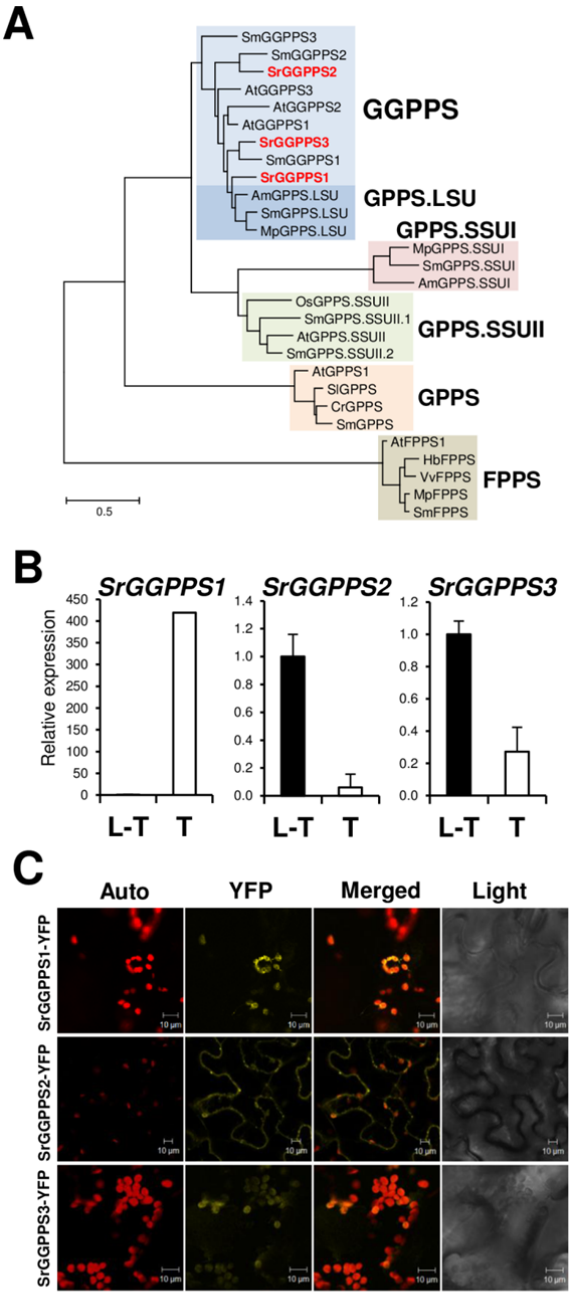


Figure 5. Analysis of GGPPSs from *S. rebaudiana*. A, Phylogenetic analysis of GGPPSs from Stevia. The maximum likelihood tree was constructed by MEGA 6 program from an alignment of full-length SrGGPPSs with other plant GGPPSs, GPPSs, GPPS.SSUIs, GPPS.SSUIIs and FPPSs. Abbreviations and accession numbers of proteins are listed in Supplemental Table S2. B, Relative expression levels of *SrGGPPS* genes in leaf-trichomes (L-T) and trichomes (T). Transcript levels of genes expressed in L-T and T were measured by qRT-PCR. Amplification of *Actin* mRNA was used for normalization. C, Subcellular localization of SrGGPPs. YFP-fused SrGGPPSs (SrGGPPS1-YFP, SrGGPPS2-YFP, and SrGGPPS3-YFP) were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration and visualized 3 dpi (days post-infiltration) using YFP channel of a confocal microscope. Auto, chlorophyll autofluorescence; YFP, YFP channel image; Light, light microscope image; Merged, merged image between Auto and YFP. Scale bars, 10 µm.

In addition to Stevia diTPSs, SrCPS and SrKS (Richman et al., 1999), analysis of our Stevia RNA-seq uncovered a homolog each for both SrCPS and SrKS, which were designated as SrCPS2 and KS-like (SrKSL), respectively. Because both *SrCPS2* and *SrKSL* were initially

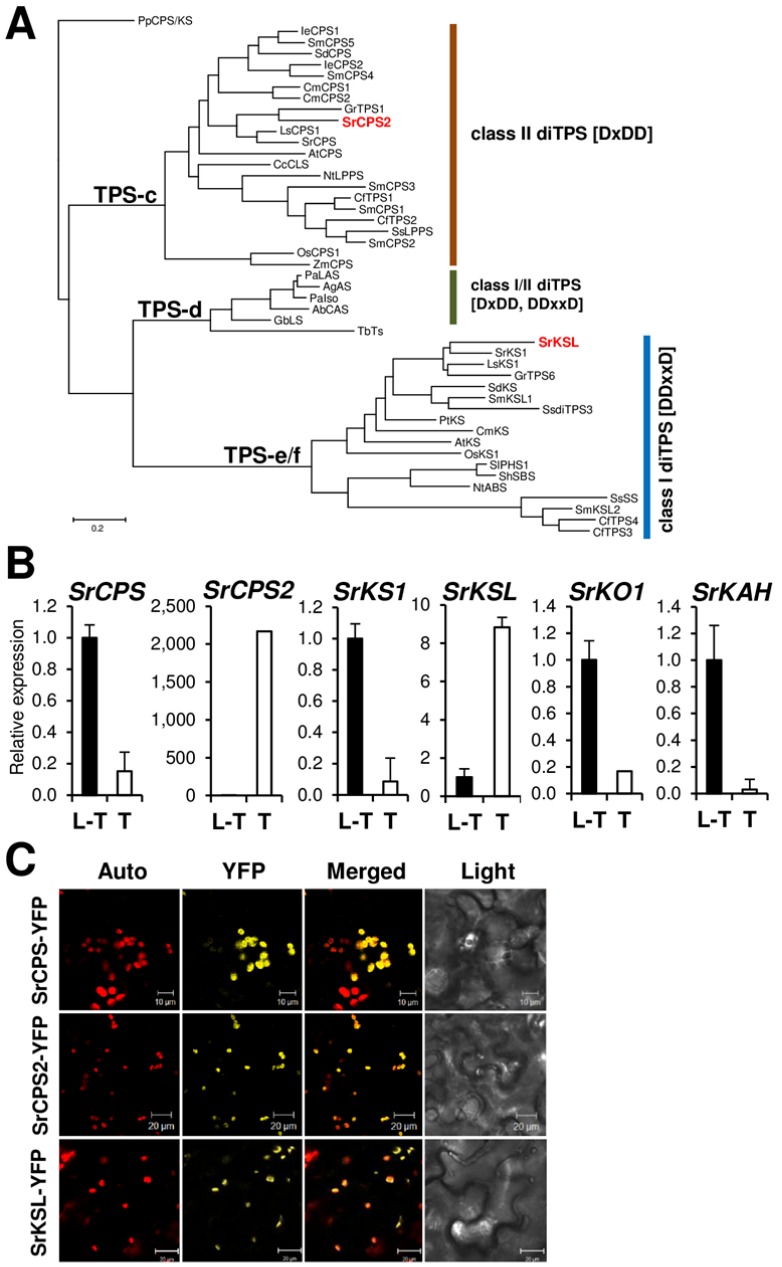


Figure 6. Analysis of diterpene synthases (diTPSs) from *S. rebaudiana*. A, Phylogenetic analysis of SrCPS2 and SrKSL from Stevia. The maximum likelihood tree was constructed by MEGA 6 program from an alignment of full-length SrCPS2 and SrKSL with other plant TPSs. Abbreviations and accession numbers of proteins are listed in Supplemental Table S2. B, Relative expression levels of Stevia diTPSs (*SrCPS*, *SrCPS2*, *SrKS1*, and *SrKSL*) and P450s (*SrKO1* and *SrKAH*) genes in leaf-trichomes (L-T) and trichomes (T). Transcript levels of genes expressed in L-T and T were measured by qRT-PCR. Amplification of *Actin* mRNA was used for normalization. C, Subcellular localization of Stevia diTPSs. YFP-fused SrCPS, SrCPS2 and SrKSL (SrCPS-YFP, SrCPS2-YFP, and SrKSL-YFP) were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration and visualized 3 dpi (days post-infiltration) using YFP channel of a confocal microscope. Auto, chlorophyll autofluorescence; YFP, YFP channel image; Light, light microscope image; Merged, merged image between Auto and YFP. Scale bars, 10 μm for SrCPS-YFP or 20 μm for SrCPS2-YFP and SrKSL-YFP.

identified as partial sequences from the RNA-seq dataset we performed 5'-RACE (Rapid Amplification of cDNA ends) to clone the full-length cDNAs. Phylogenetic comparison of the sequences of the full-length cDNAs with other related sequences assigned SrCPS2 and

SrKSL to the TPS-c and TPS-e/f subfamilies, respectively (Fig. 6A). These two families contain general or specialized diTPSs (Caniard et al., 2012; Chen et al., 2011).

SrCPS2 contains 777 amino acids which is 10 amino acids shorter than SrCPS and shares 56.1% amino acid identity with the latter (Supplemental Fig. S10). Like SrCPS, SrCPS2 possesses a DxDD motif for protonation-initiated cyclization of GGPP; this motif is a characteristic of class II TPSs (Supplemental Fig. S10). *SrCPS* transcripts were approximately 5 times more abundant in L-T than in T. By contrast, *SrCPS2* was only detectable in T suggesting that it may function differently in Stevia (Fig. 6B).

SrKSL encoded a 782 amino acid long protein that shares 61.2% identity with SrKS1-1 (hereafter referred to as SrKS1). It contains the conserved DDxxD and NSE/DTE motifs of class I TPS for metal-dependent ionization of the prenyl diphosphate substrate (Supplemental Fig. S11). Similar to the relation between *SrCPS* and *SrCPS2*, the expression patterns of *SrKS1* and *KSL* were clearly distinct from each other. Fig. 6B shows that *SrKS1* was predominantly expressed more than 5 times in L-T than T, but *SrKSL* expression level was more than 8 times higher in T compared to L-T. These results suggest that *SrCPS* and *SrKS1* are involved in SG biosynthesis in L-T as reported (Richman et al., 1999) whereas *SrCPS2* and/or *SrKSL* are likely to function in the biosynthesis of other diterpenoids in trichomes.

Plant diTPSs, CPS and KS have a transit peptide sequence at their N-termini for plastidic targeting (Toyomasu and Sassa, 2010). Similarly, a putative N-terminal plastidic transit peptide sequence of 52 amino acids for SrCPS2 and 62 amino acids for SrKSL were also predicted (Supplemental Fig. S10 and Supplemental Fig. S11). In confirmation of this prediction, both SrCPS2-YFP and SrKSL-YFP were found to be localized in chloroplasts of *N. benthamiana* leaves (Fig. 6C).

Functional characterization of SrCPS2 and SrKSL

The different spatial expression patterns between *SrCPS* and *SrCPS2* suggests that *SrCPS2* plays little role in the synthesis of copalyl diphosphate (CPP) from GGPP. To determine the catalytic activity of this diTPS *in vitro*, N-terminally truncated recombinant proteins that lack the putative transit peptide was expressed in *E. coli* BL21(DE3). 6His-tagged purified recombinant proteins were used for *in vitro* assays using GGPP as the common substrate for diTPS activity. Reaction products were then dephosphorylated by alkaline phosphatase for analysis by GC-MS. Using SrCPS as a control, a single peak for copal-15-ol, which is a dephosphorylated form of CPP (Supplemental Fig. S12A; Richman et al., 1999), was detected in the reaction mix. Moreover, SrCPS produced *ent*-kaurene in a sequential reaction with SrKS1 (Supplemental Fig. S12B and Fig. S12C; Richman et al., 1999). However, assay with SrCPS2 produced labd-13-en-8,15-diol (labdenediol) as a major component instead, suggesting that SrCPS2 is a labda-13-en-8-ol (or copal-8-ol) diphosphate synthase (LPPS) that catalyzes the formation of labda-13-en-8-ol diphosphate (LPP) from GGPP (Fig. 7A). We also detected additional minor components, a racemic mixture of manoyl oxide and *epi*-manoyl oxide from the SrCPS2 product profile, but these were previously suggested to be the byproducts of a nonenzymatic reaction (Caniard et al., 2012; Zerbe et al., 2013; Pateraki et al., 2014). SrCPS2-catalyzed reaction products were identified by comparison to authentic standards of labdane-type diterpenes based on both retention time and mass spectra (Fig. 7A and Fig. 7B; Falara et al., 2010). Moreover, mass spectra of the three compounds were nearly identical to the published fragmentation patterns of labdenediol, manoyl oxide and *epi*-manoyl oxide (Falara et al., 2010; Caniard et al., 2012; Zerbe et al., 2013; Pateraki et al., 2014).

With the newly identified *SrKSL* being highly expressed in trichomes, it seemed possible that SrKSL functions downstream of SrCPS2 to cyclize LPP. Therefore, we carried out further *in*

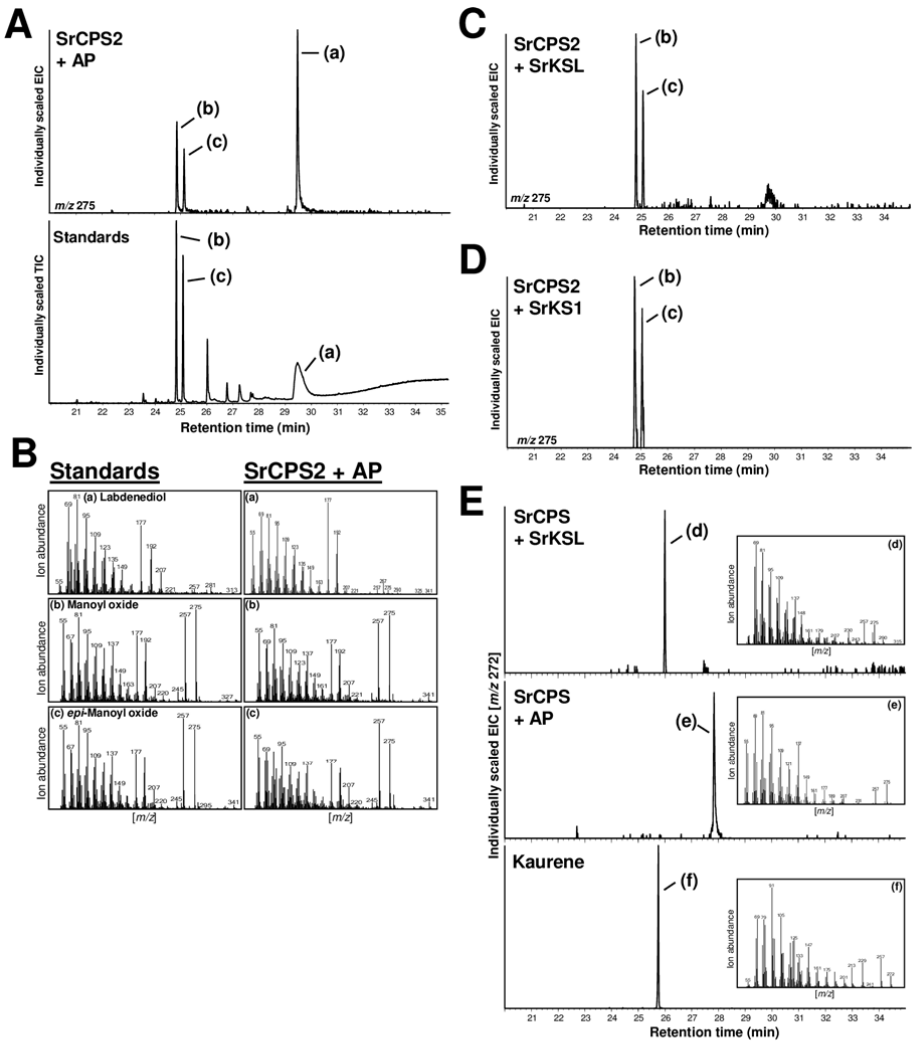


Figure 7. *In vitro* characterization of diTPSs from *S. rebaudiana*. A, GC trace of reaction product from *in vitro* assay of recombinant SrCPS2 with GGPP as a substrate. Reaction product was dephosphorylated by alkaline phosphatase (AP) for GC-MS analysis. Peaks were compared with authentic standards. B, Mass spectra of peak (a), (b) and (c) obtained from authentic standards and reaction product in (A). C to E, GC traces of reaction products from coupled *in vitro* assays of recombinant SrCPS2 + SrKSL (C), SrCPS2 + SrKS1 (D) or SrCPS + SrKSL (E) with GGPP as substrate. Peak (a), labdenediol; peak (b), manoyl oxide; peak (c), *epi*-manoyl oxide; peak (d), unidentified compound; peak (e), *ent*-copalol; peak (f), *ent*-kaurene. *m/z*, mass-to-charge ratio.TIC, total ion chromatogram; EIC, extracted-ion chromatogram.

372 *vitro* enzyme assays containing both SrCPS2 and SrKSL as enzyme sources and GGPP as

373 substrate. Fig. 7C and Supplemental Fig. S13A show that the coupled enzyme assay with

374 SrCPS2 and SrKSL produced both manoyl oxide and *epi*-manoyl oxide, but labdenediol was

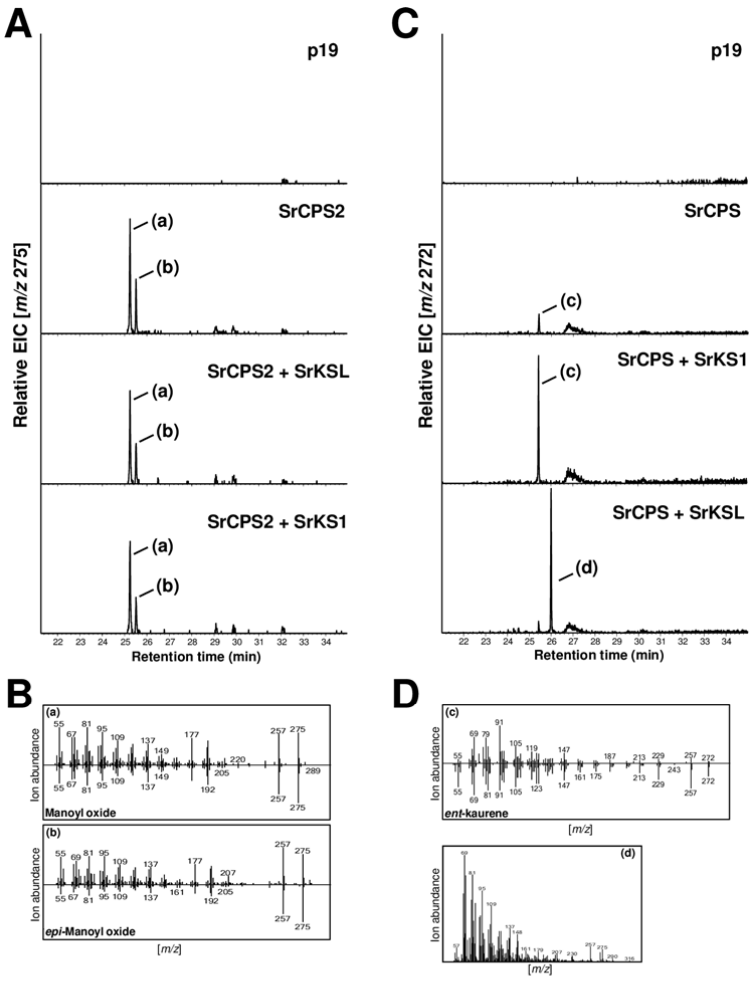


Figure 8. *In vivo* characterization of diTPSs from *S. rebaudiana*. SrCPS, SrCPS2, SrKS1, and SrKSL were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi by GC-MS. The resulting peaks were identified through comparison with reference mass spectra of the NIST 2014 library. A and C, GC traces of hexane extracts from *N. benthamiana* leaves expressing SrCPS2, SrCPS2+SrKSL, and SrCPS2+SrKS1 (A) or SrCPS, SrCPS+SrKS1, and SrCPS+SrKSL (C). *N. benthamiana* leaves expressing p19 was used as a control. B and D, Mass spectra obtained from peak (a), (b), (c) and (d) and reference spectra from the NIST 2014 library. Peak (a), manoyl oxide; peak (b), *epi*-manoyl oxide; peak (c), *ent*-kaurene; peak (d), unidentified compound. *m/z*, mass-to-charge ratio. EIC, extracted-ion chromatogram.

no longer detected. Similarly, transient expression of SrCPS2 and SrKSL in *N. benthamiana* using Agrobacterium infiltration also yielded manoyl oxide and *epi*-manoyl oxide in the infiltrated leaves (Fig. 8A and Fig. 8B). These results suggest that SrKSL catalyzes the conversion of LPP to manoyl oxide and *epi*-manoyl oxide. On the other hand, the expression

379 of SrCPS and SrKS1 in *N. benthamiana* produced *ent*-kaurene in the infiltrated leaves, as
380 expected (Fig. 8C and Fig. 8D).

381 However, in our *in vitro* enzyme assay combining SrCPS2 and SrKS1, the peaks for manoyl
382 oxide and *epi*-manoyl oxide were also detected (Fig. 7D and Supplemental Fig. S13B). This
383 result was unexpected, since SrKS1 has been identified as a typical kaurene synthase
384 (Richman et al., 1999). Furthermore, *in vivo* transient assays in *N. benthamiana* where
385 *SrCPS2* was infiltrated alone or in combination with *SrKS1* also resulted in the detection of
386 these two peaks (Fig. 8A). These results suggest that SrKS1 and other endogenous diTPSs of
387 *N. benthamiana* may carry out this catalytic reaction as well. Nevertheless, since *SrCPS2* and
388 *SrKSL* have similar expression patterns in Stevia leaves, it is perhaps likely that they catalyze
389 consecutive steps in the conversion of GGPP to manoyl oxide and *epi*-manoyl oxide
390 specifically in trichomes. Then, we also tested the SrCPS and SrKSL combination.
391 Interestingly, a novel peak similar to (-)-copalol (or *ent*-copalol) accordingly to the NIST
392 library was detected in this combination (Fig. 7E). This compound was also observed by
393 coexpression of SrCPS and SrKSL in *N. benthamiana* (Fig. 8C), confirming a specialized
394 function of these two enzymes. However, it was neither *ent*-copalol nor *ent*-kaurene, since its
395 retention time did not correspond to the peaks for *ent*-copalol obtained from *in vitro* assay
396 with SrCPS hydrolysed by alkaline phosphatase (SrCPS+AP) or for *ent*-kaurene standard
397 (Fig. 7E). Thus, the identification of this novel compound remains to be further investigated.
398 Note that we only detected this peak when full-length SrKSL including the putative transit
399 peptide was used in the reaction mix.

400

401

Discussion

Differential distribution of diterpenoids in different cell types of Stevia leaves

Stevia accumulates SGs largely in the leaves rather than other tissues such as stems and roots (Brandle and Rosa, 1992; Kumar et al., 2012). Here we found that most SGs are detected in L-T where genes encoding CPS, KS and KO related to steviol biosynthesis are expressed (Fig. 2A and Fig. 6B; Humphrey et al., 2006). This implies that SGs are stored in the place of their synthesis, and are not transported to other tissues or cells. On the other hand, Stevia leaf trichomes stored a variety of mono-, sesqui- and other labdane-type di-terpenes (Fig. 2B and Table I). In general, compounds that can be detected by GC-MS are relatively volatile, indicating that volatile organic compounds (VOCs) of Stevia mostly accumulate in the trichomes. Our result is consistent with other reports that VOCs such as terpenoids or phenylpropanoids are stored in glandular trichomes with a storage cavity, e.g. peltate trichomes of mint and capitate trichomes of basil (Turner et al., 2000; Gang et al., 2002). Tobacco glandular trichomes are capable of synthesizing diterpenoids such as *cis*-abienol and duvatrienediol and the later compound could be found in Stevia trichomes as well (Keene and Wagner, 1985; Kandra and Wagner, 1988; Guo et al., 1994; Fig. 2B and Table I). These observations suggest that most VOCs detected by GC-MS might be accumulated in glandular trichomes rather than the other two non-glandular trichome types. Our results demonstrated that there is differential accumulation of diterpenoids in different cell types of the Stevia leaf tissues through a biochemical specialization.

RNA-seq data of Stevia leaf tissues distinguished gene members involved in SG or other labdane-type diterpenoid biosynthesis

425 The diterpene steviol, the aglycone of SGs is known to be synthesized through the MEP
426 pathway from two isoprene units, IPP and DMAPP (Totte et al., 2000; 2003; Supplemental
427 Fig. S4). In addition to eight genes identified previously (Totte et al., 2003; Kumar et al.,
428 2012), our RNA-seq found additionally six genes (*SrDXS1*, *SrDXS2*, *SrDXS3*, *SrDXR2*,
429 *SrHDR2* and *SrIDI2*) encoding enzymes involved in the MEP pathway and IDI. Among four
430 *SrDXS* genes, only *SrDXS1* transcripts expressed abundantly in L-T sample (Fig. 4B). It
431 implies that SrDXS1 rather than SrDXS4 which was previously implicated in SG
432 biosynthesis (Totte et al., 2003) may be involved in SG biosynthesis in the L-T sample. On
433 the other hand, two genes, *SrDXS2* and *SrDXS4* encoding proteins of the DXS2 clade were
434 predominantly expressed in T (Fig. 4B). These results suggest that SrDXS2 and SrDXS4
435 might be related to the biosynthesis of other volatile terpenoids including labdane-type
436 diterpenoids in trichome cells.

437 As mediators of the MEP pathway and an isomerase of IPP or DMAPP, genes encoding DXR,
438 MCT, CMK, MDS, HDS, HDR and IDI of Stevia have been previously reported (Totte et al.,
439 2003; Kumar et al., 2012; Supplemental Fig. S4). Most genes except *DXR*, *HDR* and *IDI*
440 have been found in our RNA-seq data set as a single copy and the difference of their
441 expression levels was less than 2-4 fold in both L-T and T samples (Fig. 4C and
442 Supplemental Fig. S4) implying these gene functions both in primary metabolism as well as
443 terpenoid biosynthesis. SrDXR1, SrHDR1 and SrIDI1 have been suggested to be involved in
444 SG biosynthesis (Kumar et al., 2012). However, our RNA-seq and qRT-PCR analyses
445 showed that *SrHDR1* and *SrIDI1* transcripts were more abundant in T sample, suggesting the
446 important role in the biosynthesis of other volatile terpenoids rather than SGs. Based on the
447 expression levels shown in Fig. 4C, it is more likely that SrDXR1, SrHDR2 and SrIDI2 are
448 associated with SGs production.

In plant genomes, GGPPS is usually encoded by a *GGPPS* gene family comprising of 2-12 members (Beck et al., 2013). Among the three *SrGGPPSs* identified here, *SrGGPPS1* has been previously implicated in SG biosynthesis (Kumar et al., 2012). However, *SrGGPPS1* was in fact predominantly expressed in trichomes. Moreover it was almost undetectable in the L-T sample where SGs are synthesized (Fig. 5B), which suggests that it is likely to function in the biosynthesis of other diterpenoids rather than SGs. Both *SrGGPPS2* and *SrGGPPS3* identified here showed similar expression pattern expressing more abundantly in the L-T sample, but SrGGPPS2 was located in cytosol (Fig. 5B and Fig. 5C). These results propose that the newly identified *SrGGPPS3* is solely involved in SG biosynthesis.

It has been reported that *SrCPS* and *SrKS1* are highly expressed in mature leaf tissue rather than stems and roots (Richman et al., 1999; Kumar et al., 2012). We confirmed that both transcripts were more than 5 times abundant in L-T where SGs are stored than in T (Fig. 6B). However, their homologs, *SrCPS2* and *SrKSL* showed opposite expression patterns in L-T and T (Fig. 6B). For this reason, we speculated that they may function differently for the biosynthesis of other diterpenes in trichomes. Phylogenetic analysis further supported a possible functional difference as it placed SrCPS2 on the same branch as GrTPS1 (Fig. 6A), which was identified as a LPPS from *Grindelia robusta* (Zerbe et al., 2013). Although SrKSL is closest to SrKS1, it also falls into same subgroup containing GrTPS6 (Fig. 6A), which was characterized as a manoyl oxide synthase working in conjunction with GrTPS1 (Zerbe et al., 2013). Our *in vitro* and *in vivo* enzymatic assays demonstrated this speculation confirming that SrCPS2 is a LPPS (Fig. 7A). Genes encoding LPPS that catalyzes the conversion of GGPP to LPP has also been reported in a few other plants. *SsLPPS* sequence were detected by deep 454-sequencing of cDNAs from clary sage (*Salvia sclarea*) calices where sclareol, a labdane-type diterpene alcohol is mainly accumulated (Caniard et al., 2012). On the other hand, *CfTPS1* was highly expressed in the root cork of *Coleus forskohlii* where another

474 labdane-type diterpenoid forskolin is synthesized (Pateraki et al., 2014). In particular, *CcCLS*
475 from *Cistus creticus* was highly expressed in leaf trichomes as was the case with *SrCPS2*
476 shown here (Falara et al., 2010). Although the *Grindelia robusta* GrTPS1 is placed closest to
477 SrCPS2 on the phylogenetic tree its expression pattern has not yet been investigated (Fig. 6A;
478 Zerbe et al., 2013). Our *in vitro* and *in planta* coupled enzyme assays with SrCPS2 and
479 SrKSL newly identified SrKSL as a manoyl oxide synthase. Interestingly, SrKS1 was also
480 capable of producing two manoyl oxide stereoisomers in combination with SrCPS2 *in vitro*
481 as well as *in planta* (Fig. 7D and Fig. 8A). We postulate that the conversion of LPP to these
482 two manoyl oxide stereoisomers may not require a highly specific enzyme or SrKS1 might
483 have a dual function to use both CPP and LPP as substrates. The latter hypothesis is made
484 more likely by the recent finding that SmKSL2 from *Salvia miltiorrhiza* reacted with both
485 CPP and LPP to catalyze the formation of *ent*-kaurene and *epi*-manoyl oxide in a sequential
486 reaction with SmCPS5 and SmCPS4, respectively (Cui et al., 2015).

487 Another coupled enzyme assays with SrCPS and SrKSL produced an unidentified compound
488 *in vitro* as well as *in planta*, which was identified as *ent*-copalol by the NIST 2014 library
489 (Fig. 7E and Fig. 8C). GC-MS has proven to be effective in resolving stereoisomers of many
490 terpenoid compounds (Armstrong et al., 1996, Huang et al., 2009). Since the peak observed
491 in assays combining SrCPS and SrKSL has a different retention time with a similar mass
492 spectrum as *ent*-copalol, we postulate that the unidentified compound could be a stereoisomer
493 of *ent*-copalol.

494

495 **Distinct functions of diTPSs in Stevia leaves**

496 The extensive differences in the function of Stevia diTPSs were also clearly reflected in the
497 differences in metabolites derived from L-T and T samples (Fig. 2 and Fig. 9). We were able

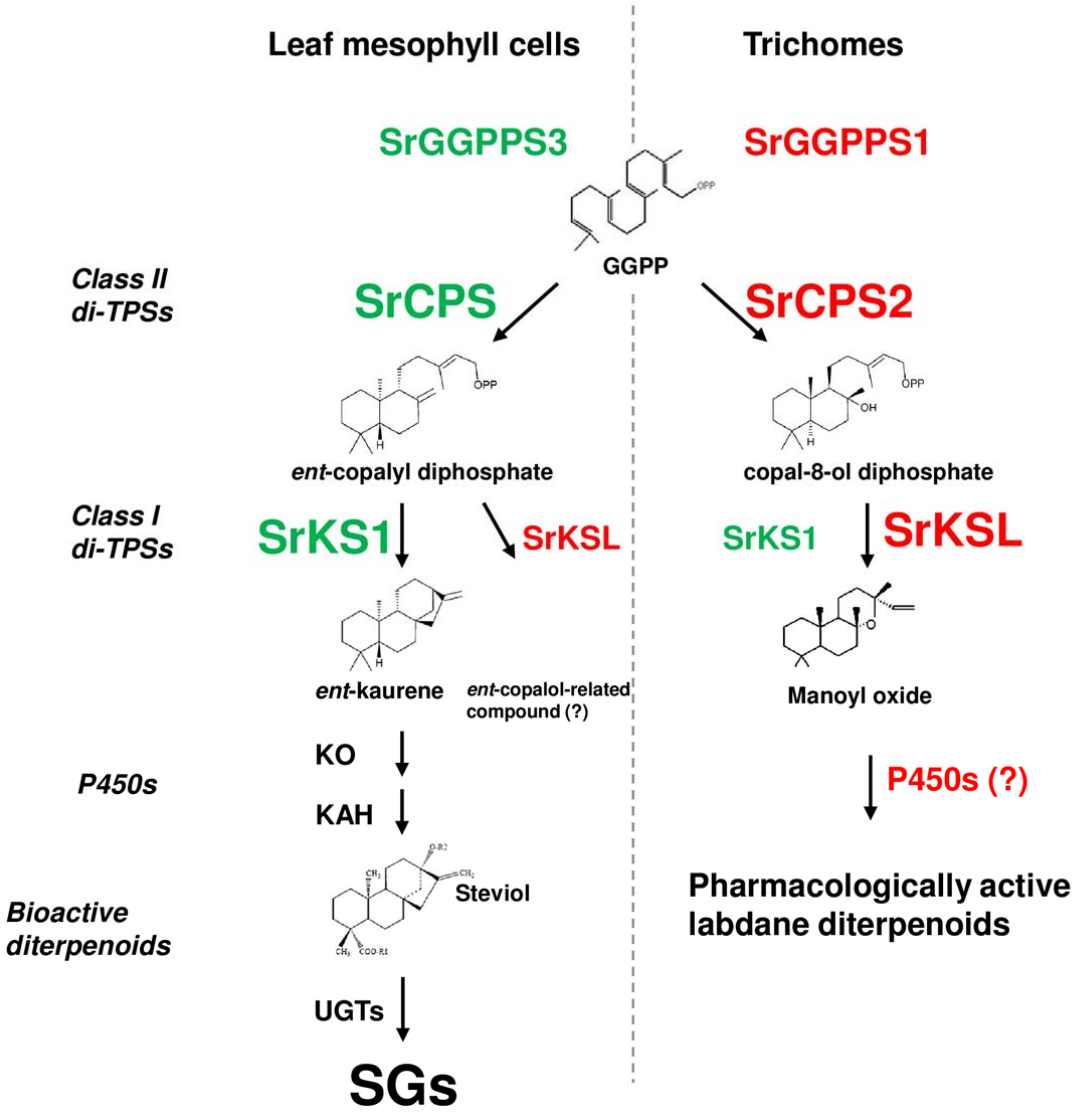


Figure 9. The biosynthetic routes of SGs and other labdane-type diterpenoids from GGPP in different parts of Stevia leaves. For these distinct pathways, SrGGPPS1 and SrGGPPS3 may function separately in trichomes and leaf mesophyll cells, respectively.

to detect a range of SGs in the L-T samples. This result is in agreement with our transcriptome and qRT-PCR data showing that *SrCPS* and *SrKS1* which are involved in SG biosynthesis are highly expressed in the L-T sample (Fig. 6B and Fig. 9). It also suggests that SGs are mostly stored in the same tissue where they are produced. Instead, Stevia trichomes produced a variety of terpenoids rather than SGs, most of which were volatiles. Manoyl oxide was detected among trichome metabolites (Fig. 2B; Fig. 9 and Table I) correlating with the high transcript levels of *SrCPS2* and *SrKSL* in trichomes. This further suggests that

505 enzymes encoded by these two genes are relevant in the production of manoyl oxide (Fig. 6B;
506 Fig. 7C; Fig. 8A and Fig. 9). It is also possible that SrKS1 may have a redundant role of
507 SrKSL for manoyl oxide biosynthesis in trichomes, since it catalyzed the formation of
508 manoyl oxide and *epi*-manoyl oxide in a sequential reaction with SrCPS2 and its transcripts
509 could be detected in T sample as well (Fig. 6B; Fig. 7D; Fig. 8A and Fig. 9). In *C. forskohlii*,
510 manoyl oxide can be found in oil bodies of the root cork cells and is known to be a precursor
511 to forskolin (Pateraki et al., 2014). We suggest that manoyl oxide in Stevia may act as a
512 precursor to other as yet unidentified diterpenoids, since forskolin is not found in Stevia
513 trichomes (Fig. 9).

514 Many diterpenoids are further modified by varying degrees of oxidation. These modifications
515 give rise to a diversity of structural classes which provide further functional groups for
516 additional modifications, such as glycosylation, acetylation and methylation. In plants, the
517 cytochrome P450s are mainly responsible for the oxidation of terpenoids (Boutanaev et al.,
518 2015). In Stevia, kaurene is oxidized and hydroxylated to form steviol by KO and KAH,
519 respectively (Humphrey et al., 2006). The latter enzyme performs the first committed step in
520 SG biosynthesis providing an aglycone, steviol, which is sequentially glycosylated by a series
521 of UGTs in Stevia leaves. We found that both P450 genes, *SrKO1* and *SrKAH* were
522 abundantly expressed in L-T sample compared to the T sample (Fig. 6B). The results of
523 transcript analysis correlated with those of chemical analysis. However, we did not find new
524 Stevia *KAH* homologue in our RNA-seq dataset suggesting a non-redundant role of SrKAH
525 in SG biosynthesis.

526 Recently, functional diterpenoids with biological activities for treatment of a wide range of
527 medical conditions have been reported from diverse medicinal plants (Guo et al., 2013; Zerbe
528 et al., 2013; King et al., 2014; Pateraki et al., 2014). Although manoyl oxide itself is known
529 to be an anticancer compound, it is considered as the molecular core for a large series of

bioactive derivatives (Pateraki et al., 2014). In addition to manoyl oxide, we found several labdane-type and oxidized bicyclic diterpenoids such as oxomanoyl oxide and labda-8(20),13-dien-15-oic acid (agatholic acid) from Stevia trichomes (Fig. 2 and Table I); the functions of all these terpenoids are unknown. Querying the transcriptome of Stevia trichomes, we found more than 70 candidate P450s that were preferentially expressed in the T-samples with a minimum of 4-fold increase in expression level compared to the L-T sample. Future work should be directed toward the identification of cytochrome P450s involved in the enzymatic conversion of manoyl oxide to bioactive labdane diterpenoids in the Stevia trichomes (Fig. 9).

MATERIALS AND METHODS

Plant materials

Stevia (*Stevia rebaudiana* Bertoni) seeds were germinated on soil and grown in a greenhouse under natural light condition in Singapore. Four-week old *Nicotiana benthamiana* plants grown in a greenhouse were used for subcellular localization and *in vivo* characterization of Stevia genes.

Trichome isolation

Trichomes were isolated using a glass bead abrasion method (Lange et al., 2000; Jin et al., 2014). One-month old Stevia plants grown in a greenhouse were used. Young leaves (20-30 g; 1-2 cm in length) were placed in 50-mL test tube containing ice-cold imbibition buffer (1 mM aurintricarboxylic acid, 5 mM thiourea, and 2 mM DTT, pH 6.6) and soaked for 1 h. After removing the imbibition buffer, glass beads (425-600 μm, Sigma-Aldrich) were added

553 to the tube with an extraction buffer (25 mM MOPSO, pH 6.6, 200 mM D-sorbitol, 10 mM
554 sucrose, 0.5 mM sodium phosphate, 1% [wt/vol] polyvinylpyrrolidone-40, 0.6% [wt/vol]
555 methyl cellulose, 1 mM aurintricarboxylic acid, 5 mM thiourea, and 2 mM DTT) and the
556 content vigorously vortexed for 1 min. The mixture was first passed through a cell strainer
557 (mesh size 100 μm) to remove debris and trichomes were collected using a 20 μm strainer.
558 The collected trichomes were washed several times with the same buffer before being frozen
559 in liquid nitrogen for further use.

560 To obtain leaf tissues from which trichomes were removed (leaf-trichomes), a cold-brushing
561 method was used to completely remove trichomes (Wang et al., 2001). The quality of
562 trichomes or leaf-trichomes was monitored under a dissecting microscope.

563

564 **RNA isolation for RNA sequencing**

565 Total RNA was extracted from the isolated trichomes and leaf-trichomes using the
566 Spectrum^{TM} Plant total RNA kit (Sigma-Aldrich), which included RNase-free DNaseI
567 treatment to remove genomic DNA. The RNA quantity was determined with Nanodrop
568 spectrophotometer, ND-1000 (Thermo Fisher Scientific). The RNA integrity number (RIN)
569 was measured to evaluate the RNA quality using Agilent 2100 bioanalyzer and RNA 6000
570 Nano Labchip Kit (Agilent Technologies). RNA samples with a RIN value of $7<x<10$ were
571 processed for RNA-seq by the Rockefeller University Genomics Resource Center (New York,
572 USA) using a HiSeq 2000 (Illumina).

573

574 **RNA-seq *de novo* assembly**

575 The RNA libraries prepared using the TruSeq RNA Sample Preparation Kits v2, set A (RS-
576 122-2001, Illumina Inc.) were qualified by the Agilent 2200 TapeStation system (Agilent
577 Inc.). The qualified libraries were run on single lanes for 100 cycles (strand specific single-
578 end) on HiseqTM 2000 (illumine Inc.). Raw reads were analysed by FastQC
579 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) for the quality control. The
580 Trinity method was used for *de novo* assembly of short sequence reads due to the absence of
581 a reference Stevia genome (Grabherr et al., 2011). The assembled unigenes were annotated
582 based on sequence similarities to sequences in the NCBI nr database and also the protein
583 sequence databases from *Arabidopsis thaliana*, *Vitis vinifera, Glycine max* and *Oryza sativa*.

584

585 **Quantitative real time PCR (qRT-PCR)**

586 qRT-PCR was performed to investigate gene expression pattern in different Stevia tissues.
587 One µg of total RNA from each sample was used for cDNA synthesis with M-MLV
588 Superscript II (Promega). cDNA samples were quantified by Applied Biosystems (AB)
589 7900HT fast real-time PCR system and AB power SYBR green PCR master mix (Life
590 Technologies). Fold change values of target gene transcripts were subsequently normalized
591 by dividing the ΔCt values by the ΔCt values of control gene transcripts. Gene-specific
592 primers for qRT-PCR were designed using the Primer3 program (http://bioinfo.ut.ee/primer3-
593 0.4.0/) and are listed in Supplemental Table S1. Each PCR product obtained from regular
594 PCR was verified by sequencing after cloning into pGEM-T Easy vector (Promega). A
595 melting curve analysis in Applied Biosystems (AB) 7900HT fast real-time PCR system was
596 performed to assess the specificity of the amplified PCR product. All reactions were carried
597 out in technical triplicates and biological replicates. A mock reaction containing all the RT-

PCR reagents except the reverse transcriptase was used as a negative control. Stevia *Actin* gene was used as an internal control for normalization.

Isolation of full-length ORF of Stevia genes and vector construction for *Agrobacterium*-mediated gene expression

Full-length open reading frames (ORFs) of four *SrDXSs* (*SrDXS1-4*), three *SrGPPSs* (*SrGPPS1-3*), *SrCPS*, and *SrKS1* were amplified by PCR from Stevia leaf-derived cDNA. Primer sets listed in Supplemental Table S1 were used to amplify gene products that were flanked by attB sites for Gateway cloning (Invitrogen). Purified PCR products were cloned into pDONR221 (Invitrogen) and verified by sequencing. Partial sequences for *SrCPS2* and *SrKSL* were obtained from Stevia RNA-seq data. Using these partial sequences, we designed Rapid Amplification of cDNA Ends (RACE) primers (Supplemental Table S1) and performed 5' RACE experiment using SMARTer RACE cDNA Amplification Kit (Clontech). The resulting sequences of *SrCPS2* and *SrKSL* obtained by 5' RACE were confirmed by sequencing after cloning into pDONR221.

For yellow fluorescent protein (YFP)-fusion construct, the pDONR221 clone harbouring each gene was integrated into the destination vector, pBA-DC-YFP expression vector, which contained a CaMV35S promoter and a C-terminal in frame YFP by LR Clonase (Invitrogen). The final constructs were transformed into *Agrobacterium tumefaciens* GV3101by electroporation (Bio-Rad).

Sequence identification, multiple sequence alignments and phylogenetic analysis

620 DNA sequences were edited and assembled using DNASTAR Lasergene 8 (DNASTAR,
621 Inc.). Sequence alignments were conducted using Biology WorkBench 3.2
622 (http://seqtool.sdsc.edu/CGI/BW.cgi). Phylogenetic analysis was performed using the
623 Maximum Likelihood method in Molecular Evolutionary Genetics Analysis (MEGA) version
624 6 program (Tamura et al., 2013). Abbreviations and GenBank accession numbers of proteins
625 in phylograms are listed in Supplemental Table S2.

626

627 **Subcellular localization**

628 To determine the subcellular localization of C-terminal in-frame YFP-tagged proteins, three
629 *SrGPPSs* (*SrGPPS1-3*), *SrCPS, SrCPS2*, *SrKS1* and *SrKSL* were transiently expressed in
630 leaves of 4-week old *N. benthamiana* plant as described in Jin et al. (2014). Infiltrated plants
631 were incubated in a growth chamber at 24°C, under long day condition (16 h light/8 h dark).
632 After 3 days, the leaves were excised, mounted onto slides and imaged by confocal laser
633 scanning microscopy (Carl Zeiss LSM5 Exciter). The 488- and 514-nm lines of an argon
634 laser were used to excite GFP and YFP, respectively. Band pass was set to 500 to 550 nm,
635 and long pass was set to 560 nm. Images were recorded and processed using the Zeiss LSM
636 Image Browser (Carl Zeiss).

637

638 ***In vitro* functional characterization of Stevia diTPSs**

639 To determine the function of Stevia diTPSs, we performed single or coupled enzyme assays
640 with recombinant proteins. PCR-amplified *SrCPS*, *SrCPS2*, *SrKS1*, and *SrKSL* cDNAs were
641 inserted into pET28b plasmid (Novagen) to construct the vectors for the production of
642 recombinant N-terminal poly-histidine (His)-tagged proteins. The final constructs were

643 transformed into *E. coli* BL21(DE3)pLysS (Invitrogen). *E.coli* culture was treated with 0.2
644 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16°C for 12 h to induce His-tagged
645 protein expression, and induced cells were then harvested by centrifugation. After adding
646 binding buffer (20 mM HEPES, pH 7.5, 0.5 M NaCl, 25 mM Imidazole, 5% (v/v) glycerol),
647 one protease inhibitor cocktail tablet per 50 mL (Roche), and 0.1 mg/L lysozyme to the cell
648 pellet, cells were subsequently lysed by sonication. The cell lysate was centrifuged for 30 min
649 at 14,000g and the supernatant was used for purification of the recombinant proteins using
650 Ni-NTA affinity chromatography. Proteins bound to Ni-NTA agarose (Qiagen) were eluted
651 from the column with elution buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 250 mM
652 Imidazole) and dialyzed against a dialysis buffer (50 mM HEPES, pH 7.5, 0.1 M KCl, 1 mM
653 DTT and 5% [v/v] glycerol).

654 *In vitro* enzyme assay for CPS activity was performed in a final volume of 500 µL of reaction
655 buffer (50 mM HEPES pH 7.5, 100 mM KCl, 7.5 mM $MgCl_2$, 5 mM DTT, 5% [v/v] glycerol,
656 20 µM GGPP) with about 5 µg of the purified protein. The reaction mixture was incubated at
657 30°C for 2 h. After incubation, 10 units of FastAP Thermosensitive Alkaline Phosphatase
658 (Thermo Scientific) was added to hydrolyse copalyl diphosphate. The reaction mixture was
659 overlaid with 500 µL hexane to trap hexane-soluble copalol and then incubated at 37°C for 4
660 h. The dephosphorylated compounds were then extracted two times with 500 µL hexane. The
661 hexane fractions were completely dried under a stream of N_2 gas at room temperature and
662 redissolved in 100 µL of hexane for GC-MS analysis.

663 Coupled enzyme assay was performed using two recombinant proteins to determine if they
664 catalyze consecutive steps in a catalytic pathway. The enzymatic reaction and compound
665 extraction were as described above except for the alkaline phosphatase treatment. Enzyme
666 reaction with SrCPS and SrKS1 were used as reference (Richman et al., 1999). *In vitro*
667 enzyme assays were done at least in biological triplicates.

668

669 ***In vivo* functional characterization of Stevia diTPSs**

670 *In vivo* characterization of diTPSs was carried out by *Agrobacterium*-mediated transient gene
671 expression in *N. benthamiana* leaves. Overnight cultures of *Agrobacterium* strains harbouring
672 *SrCPS*, *SrCPS2*, *SrKS1*, and *SrKSL* constructs as described above were used alone or in
673 combinations. Harvested cultures were suspended in a solution containing 10 mM $MgCl_2$, 10
674 mM MES pH 5.6 and 100 µM acetosyringone. After 1 h incubation at room temperature, the
675 *Agrobacterium* mixture was infiltrated into the underside of *N. benthamiana* leaves using a
676 needleless syringe. The infiltrated plants were incubated in the growth chamber at 24°C for 3
677 days. After incubation, four to five infiltrated leaves were frozen in liquid nitrogen and
678 homogenized with a pre-chilled mortar and pestle. About 500 mg of leaf powder were
679 dissolved in 500 µL ethyl acetate containing 1 µL (10 mg/mL) of camphor (Sigma) as an
680 internal standard and incubated on a horizontal shaker at 200 rpm for 2 h. After centrifugation
681 of the slush at 13,000 *g* for 10 min, the ethyl acetate upper layer was transferred into a new
682 Eppendorf tube and mixed with 300 mg anhydrous Na_2SO_4 to remove water. Following an
683 additional centrifugation for 1 min to separate the Na_2SO_4, the extract was transferred into a 2
684 mL amber glass GC sample vial for GC-MS analysis. *In vivo* characterization of diTPSs was
685 done at least in biological triplicates.

686

687 **GC-MS analysis**

688 GC-MS analysis was performed on Agilent 7890A GC (Agilent Technologies) system and an
689 Agilent-Technologies 5975C inert XL Mass selective detector, equipped with a HP-5MS UI
690 column (30 m × 0.25 mm × 0.25 µm; Agilent Technologies). Helium was used as the carrier

691 gas at a flow rate of 1mL min^{-1}. Samples (5 µL) were injected onto the column at 250°C in
692 the splitless mode, and a temperature gradient of 8°C min^{-1} from 50°C (1 min hold) to 300 °C
693 (5 min hold) was applied. Data were processed by MSD ChemStation Data Analysis (Agilent
694 Technologies). The chemical components of Stevia trichomes were identified by comparison
695 of their mass spectra with those in NIST 2014 library data of GC-MS and with literature data
696 (Adams, 2007) along with the retention indices (RI) associated with a series of n-alkanes
697 (C7-C30) mix standard (Sigma-Aldrich, 49451-U). Camphor (10 mg/mL) was used as an
698 internal standard. The amount of each compound was calculated by measuring its peak area
699 related to that of a known amount of camphor. The chemical analysis of Stevia trichomes
700 were done at least in biological triplicates. The identified components along with their RI
701 and relative % values are listed in Table I. Compound identification by *in vitro* and *in planta*
702 assays was done by comparison to authentic standards, reference spectra from literature and
703 databases (NIST 2014 library).

704

705 **HPLC analysis of steviol glycosides from Stevia leaves**

706 Fresh Stevia leaves, leaf-trichomes or isolated trichomes were frozen in liquid nitrogen and
707 ground with a pre-chilled mortar and pestle. About 100 mg of the samples was extracted in 1
708 mL of water at 70˚C for 3 h. The extract was centrifuged at 1,500 x *g* for 15 min and the
709 supernatant was filtered through a 0.45 µm filter. The filtrate was further extracted using
710 Finisterre SPE column C2 (Teknokroma) which was pre-treated with methanol followed by
711 water. 2 mL of the filtrate was loaded on the cartridge and allowed to flow through. The
712 cartridge was washed with water followed by acetonitrile:water (20:80, v/v) and air dried for
713 3 min. Samples were then eluted in 1 mL methanol:acetonitrile (50:50, v/v) and filtered using
714 a 0.45 µm nylon centrifuge tube (Corning).

715 HPLC analysis of the samples were carried out on Shidmadzu Nexera X2 UHPLC system
716 using a Shim-pack VP-ODS column (250 mm × 4.6 mm, i.d. 5 μm, particle size 4.6 μm) and
717 detected by a photodiode array detector (SPD-M30A with high sensitivity cell). 5 μL of
718 sample was injected and the elution was performed over 24 min with a 30-80% acetonitrile
719 gradient at a flow rate of 1.0 mL/min according to protocol by Shimadzu. Column oven was
720 maintained at 40˚C. Peak assignment for the absorbance spectrum was based on comparison
721 with elution profile of known standards (complete stevia standards kit, KIT-00019565-005,
722 ChromaDex) at a wavelength of 210 nm. HPLC analysis of Stevia samples were done at least
723 in biological triplicates.

724

725 **Accession numbers**

726 The RNA-seq data supporting the result of this article is available in the DNA Data Bank of
727 Japan (DDBJ: http://www.ddbj.nig.ac.jp/) with accession number, DRA003752. Nucleotide
728 sequences from this study were submitted to the National Center for Biotechnology
729 Information (NCBI)/GenBank using BankIt with accession numbers: SrDXS1 (KT276229),
730 SrDXS2 (KT276230), SrDXS3 (KT276231), SrDXS4 (KT276232), SrDXR2 (KT276233),
731 SrHDR2 (KT276234), SrIDI1 (KT276235), SrIDI2 (KT276236), SrGGPPS2 (KT276237),
732 SrGGPPS3 (KT276238), SrCPS2 (KT276239), and SrKSL (KT276240).

733

734 **ACKNOWLEDGEMENTS**

735 The purified chromatographic fraction from yeast culture transformed with *C. creticus* copal-
736 8-ol diphosphate synthase (CcCLS) used as a standard were kindly gifted from Dr Angelos K.
737 Kanellis (Aristotle University of Thessaloniki, Greece). Dr Hiroshi Kawaide (Tokyo

738 University of Agriculture and Technology, Japan) kindly provided *ent*-kaurene. We thank Dr
739 Rajani Sarojam for help in trichomes isolation and Lay-Peng Tan (Agilent Technologies
740 Singapore Pte Ltd) for technical assistance in GC-MS analysis. We also thank the Temasek
741 Life Sciences Laboratory central facilities for support on scanning electron microscopy and
742 confocal microscopy.

743

744 **LIST OF AUTHOR CONTRIBUTIONS**

745 M.J.K and J.Z performed the molecular and biochemical experiments; J.J and L.W performed
746 the RNA-seq and transcriptome analyses; N.H.C coordinated RNA-seq and transcriptome
747 analyses, and revised the article; I.C.J conceived the research, designed the experiments, and
748 wrote the article.

749

750 **FIGURE LEGENDS**

751 **Figure 1.** Glandular and filamentous trichomes of *S. rebaudiana*. A, Scanning electron
752 microscopy (SEM) images of a Stevia leaves showing three types of trichomes. (a) glandular
753 trichome; (b) long non-glandular filamentous trichome; (c) short non-glandular filamentous
754 trichome. B, A detailed view of the three types of trichomes of Stevia. Scale bars are included
755 on all images.

756

757 **Figure 2.** Distribution of diterpenoids in Stevia leaf tissues. Tissue-specific abundance of
758 metabolites was evaluated by UHPLC (A) and GC-MS (B). A, Comparison of SG content
759 between leaf-trichomes and trichomes. Standards contain a mixture of nine known SGs. Peak
760 (a) Reb D; peak (b) Reb A; peak (c) Stevioside; peak (d) Reb F; peak (e) Reb C; peak (f)
761 Dulcoside A; peak (g) Rubusoside; peak (h) Reb B; peak (I) Steviobioside. B, GC traces
762 showing the difference between leaf-trichomes and trichomes. The arrow indicates the peak

763 of camphor (10 μg/μL), the internal standard used in the assay. The peaks numbered in GC
764 traces were identical to those listed in Table I.
765
766 **Figure 3.** Gene Ontology (GO) analysis of more abundantly expressed unigenes in leaf-
767 trichomes (A) or trichomes (B). X-axis, log(1/P-value). P-value is the hypergeometic test
768 result for each GO terms.
769
770 **Figure 4.** Analysis of genes involved in MEP pathway from *S. rebaudiana*. A, Phylogenetic
771 analysis of DXSs from Stevia. The maximum likelihood tree was constructed by MEGA 6
772 program from an alignment of full-length SrDXSs with other plant DXSs. Abbreviations and
773 accession numbers of proteins are listed in Supplemental Table S2. B, Relative expression
774 levels of *SrDXS* genes in leaf-trichomes (L-T) and trichomes (T). C, Relative expression
775 levels of *SrDXRs*, *SrMCT*, *SrCMK*, *SrMDS*, *SrHDS*, *SrHDRs* and *SrIDIs* in L-T and T.
776 Transcript levels of genes expressed in L-T and T were measured by qRT-PCR.
777 Amplification of *Actin* mRNA was used for normalization.
778
779 **Figure 5.** Analysis of GGPPSs from *S. rebaudiana*. A, Phylogenetic analysis of GGPPSs
780 from Stevia. The maximum likelihood tree was constructed by MEGA 6 program from an
781 alignment of full-length SrGGPPSs with other plant GGPPSs, GPPSs, GPPS.SSUIs,
782 GPPS.SSUIIs and FPPSs. Abbreviations and accession numbers of proteins are listed in
783 Supplemental Table S2. B, Relative expression levels of *SrGGPPS* genes in leaf-trichomes
784 (L-T) and trichomes (T). Transcript levels of genes expressed in L-T and T were measured by
785 qRT-PCR. Amplification of *Actin* mRNA was used for normalization. C, Subcellular
786 localization of SrGGPPs. YFP-fused SrGGPPSs (SrGGPPS1-YFP, SrGGPPS2-YFP, and
787 SrGGPPS3-YFP) were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-

mediated infiltration and visualized 3 dpi (days post-infiltration) using YFP channel of a confocal microscope. Auto, chlorophyll autofluorescence; YFP, YFP channel image; Light, light microscope image; Merged, merged image between Auto and YFP. Scale bars, 10 μm.

Figure 6. Analysis of diterpene synthases (diTPSs) from *S. rebaudiana*. A, Phylogenetic analysis of SrCPS2 and SrKSL from Stevia. The maximum likelihood tree was constructed by MEGA 6 program from an alignment of full-length SrCPS2 and SrKSL with other plant TPSs. Abbreviations and accession numbers of proteins are listed in Supplemental Table S2. B, Relative expression levels of Stevia diTPSs (*SrCPS*, *SrCPS2*, *SrKS1*, and *SrKSL*) and P450s (*SrKO1* and *SrKAH*) genes in leaf-trichomes (L-T) and trichomes (T). Transcript levels of genes expressed in L-T and T were measured by qRT-PCR. Amplification of *Actin* mRNA was used for normalization. C, Subcellular localization of Stevia diTPSs. YFP-fused SrCPS, SrCPS2 and SrKSL (SrCPS-YFP, SrCPS2-YFP, and SrKSL-YFP) were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration and visualized 3 dpi (days post-infiltration) using YFP channel of a confocal microscope. Auto, chlorophyll autofluorescence; YFP, YFP channel image; Light, light microscope image; Merged, merged image between Auto and YFP. Scale bars, 10 μm for SrCPS-YFP or 20 μm for SrCPS2-YFP and SrKSL-YFP.

Figure 7. *In vitro* characterization of diTPSs from *S. rebaudiana*. A, GC trace of reaction product from *in vitro* assay of recombinant SrCPS2 with GGPP as a substrate. Reaction product was dephosphorylated by alkaline phosphatase (AP) for GC-MS analysis. Peaks were compared with authentic standards. B, Mass spectra of peak (a), (b) and (c) obtained from authentic standards and reaction product in (A). C to E, GC traces of reaction products from coupled *in vitro* assays of recombinant SrCPS2 + SrKSL (C), SrCPS2 + SrKS1 (D) or SrCPS

+ SrKSL (E) with GGPP as substrate. Peak (a), labdenediol; peak (b), manoyl oxide; peak (c), *epi*-manoyl oxide; peak (d), unidentified compound; peak (e), *ent*-copalol; peak (f), *ent*-kaurene. *m/z*, mass-to-charge ratio.TIC, total ion chromatogram; EIC, extracted-ion chromatogram.

Figure 8. *In vivo* characterization of diTPSs from *S. rebaudiana*. SrCPS, SrCPS2, SrKS1, and SrKSL were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi by GC-MS. The resulting peaks were identified through comparison with reference mass spectra of the NIST 2014 library. A and C, GC traces of hexane extracts from *N. benthamiana* leaves expressing SrCPS2, SrCPS2+SrKSL, and SrCPS2+SrKS1 (A) or SrCPS, SrCPS+SrKS1, and SrCPS+SrKSL (C). *N. benthamiana* leaves expressing p19 was used as a control. B and D, Mass spectra obtained from peak (a), (b), (c) and (d) and reference spectra from the NIST 2014 library. Peak (a), manoyl oxide; peak (b), *epi*-manoyl oxide; peak (c), *ent*-kaurene; peak (d), unidentified compound. *m/z*, mass-to-charge ratio. EIC, extracted-ion chromatogram.

Figure 9. The biosynthetic routes of SGs and other labdane-type diterpenoids from GGPP in different parts of Stevia leaves. For these distinct pathways, SrGGPPS1 and SrGGPPS3 may function separately in trichomes and leaf mesophyll cells, respectively.

833 **Table I.** Chemical composition of trichomes from *S. rebaudiana*.

No[a]	Compounds	RT (min)[b]	RI[c]	Formula	RA (%)[d]
1	α-Pinene	9.07	922	$C_{10}H_{16}$	0.13
2	(+)-Sabinene	9.74	972	$C_{10}H_{16}$	0.15
3	β-Pinene	9.84	979	$C_{10}H_{16}$	1.10
4	β-Linalool	11.82	1,126	$C_{10}H_{18}O$	0.17
5	L-α-Terpineol	13.49	1,186	$C_{10}H_{18}O$	0.06
6	β-Elemene	16.82	1,389	$C_{15}H_{24}$	1.84
7	Caryophyllene	17.37	1,424	$C_{15}H_{24}$	2.15
8	α-Bergamotene	17.45	1,432	$C_{15}H_{24}$	0.94
9	α-Humulene	17.89	1,452	$C_{15}H_{24}$	1.18
10	Germacrene D	18.29	1,484	$C_{15}H_{24}$	1.10
11	β-Selinene	18.40	1,489	$C_{15}H_{24}$	0.29
12	Elixene	18.52	1,492	$C_{15}H_{24}$	1.27
13	γ-Cadinene	18.74	1,513	$C_{15}H_{24}$	0.24
14	δ-Cadinene	18.82	1,524	$C_{15}H_{24}$	0.17
15	(±)-trans-Nerolidol	19.27	1,561	$C_{15}H_{26}O$	0.93
16	Farnesol	19.47	1,620	$C_{15}H_{26}O$	0.17
17	Germacrene D-4-ol	19.69	1,660	$C_{15}H_{26}O$	0.56
18	α-epi-Cadinol	20.58	1,638	$C_{15}H_{26}O$	0.27
19	α-Cadinol	20.79	1,652	$C_{15}H_{26}O$	0.28
20	Isoaromadendrene epoxide	21.08	1,579	$C_{15}H_{24}O$	0.40
21	Ledene oxide-(II)	22.18	1,683	$C_{15}H_{24}O$	0.09
22	[2-(4-Methoxy-phenyl)-[1,3]dioxolan-2-yl]-acetic acid	23.46	1,930	$C_{12}H_{14}O_5$	0.15
23	Manoyl oxide	25.32	1,988	$C_{20}H_{34}O$	1.07
24	n-Octadecanol	25.95	2,053	$C_{18}H_{38}O$	2.39
25	Ethanol, 2-[2-[4-(1,1,3,3-tetramethylbutyl)phenoxy]ethoxy]-	26.93	2,126	$C_{18}H_{30}O_3$	0.28
26	Stearyl acetate	27.30	2,160	$C_{20}H_{40}O_2$	0.33
27	3-Oxomanoyl oxide	27.90	2,219	$C_{20}H_{32}O_2$	2.73
28	(+)-Agathadiol	28.80	2,357	$C_{19}H_{30}O_2$	4.80
29	(+)-Copaiferic acid	28.96	2,369	$C_{20}H_{32}O_2$	4.62
30	4,8,13-Duvatriene-1,3-Diol	29.28	2,400	$C_{21}H_{28}O_2$	0.72
31	4-(3-Hydroxy-3-methylpentyl)-3,4a,8,8-tetramethyldecahydro-1-naphthalenol	29.57	2,415	$C_{20}H_{34}O_2$	1.30
32	Androstan-17-one, 3-methoxy-16,16-dimethyl-, (3β,5α)-	29.86	2,418	$C_{22}H_{36}O_2$	1.90
33	Labda-8(20),13-dien-15-oic acid, 19-hydroxy-, methyl ester, (E)-	30.45	2,429	$C_{21}H_{32}O_3$	29.64
34	Pregnane-18,20-diol, (5α)-	31.09	2,449	$C_{21}H_{36}O_2$	5.63
35	Allopregnane-3α,20α-diol	31.47	2,475	$C_{21}H_{36}O_2$	12.20
36	17-α-hydroxypregnenolone	31.57	2,480	$C_{20}H_{32}O_2$	9.97
37	2-[5-(2,2-Dimethyl-6-methylene-cyclohexyl)-3-methyl-pent-2-enyl]-[1,4]benzoquinone	31.84	2,489	$C_{21}H_{28}O_2$	2.91
38	Isophthalic acid, 3,7-dimethyloct-6-enyl propyl ester	31.96	2,492	$C_{21}H_{30}O_4$	3.30
39	3-Ethyl-5-(2'-ethylbutyl)octadecane	32.13	2,599	$C_{26}H_{54}$	0.76
40	Allopregnane-7α,11α-diol-3,20-dione	32.36	2,613	$C_{21}H_{32}O_4$	0.39
41	Ethanol, 2-[2-[2-[2-[p-(1,1,3,3-tetramethylbutyl)phenoxy]ethoxy]ethoxy]ethoxy]-	32.64	2,676	$C_{22}H_{38}O$	0.97
42	n-Heptacosane	33.06	2,700	$C_{27}H_{56}$	0.45
					100.00

834 [a]Compound listed in order of elution in a HP-5MS UI column; [b]RT, retention time (min); [c]RI,
835 retention indices calculated against C_7-C_{30} n-alkanes on the HP-5MS UI column; [d]RA,
836 relative amount (%); ratio expressed against the sum of all peaks.

837

838

839 **Table II.** Overview of the assembly results of RNA-seq.

Tissue	Total reads	# Unigenes	# Annotation	% Annotation
Leaf-trichomes	75,097,166	32,336	24,041	74.3
Trichomes	85,902,623	45,918	32,692	71.2

840

841

842 **Supplemental Data**

843 **Supplemental Figure S1**. Stevia tissues for RNA-seq.

844 **Supplemental Figure S2.** Quality of strand specific RNA-seq derived from two Stevia
845 tissues.

846 **Supplemental Figure S3.** Heat map of transcript expression in leaf-trichomes and trichomes.
847 **Supplemental Figure S4.** Biosynthesis of steviol glycosides via the MEP pathway.

848 **Supplemental Figure S5.** Comparison of deduced amino acid sequences of four SrDXS
849 small gene family.

850 **Supplemental Figure S6.** Comparison of deduced amino acid sequences of SrDXRs.

851 **Supplemental Figure S7.** Comparison of deduced amino acid sequences of SrHDRs.

852 **Supplemental Figure S8.** Comparison of deduced amino acid sequences of SrIDIs.

853 **Supplemental Figure S9.** Comparison of deduced amino acid sequences of SrGGPPSs.

854 **Supplemental Figure S10.** Comparison of deduced amino acid sequences of SrCPS and
855 SrCPS2.

856 **Supplemental Figure S11.** Comparison of deduced amino acid sequences of SrKS1-1 and
857 SrKSL.

858 **Supplemental Figure S12.** GC-MS analysis of *in vitro* assay with diTPSs from *S.*
859 *rebaudiana*.

860 **Supplemental Figure S13.** Mass spectra obtained from coupled *in vitro* assays of
861 recombinant SrCPS2 and SrKSL or SrCPS2 and SrKS1.

862 **Supplementary Table S1.** Oligonucleotide primers used in this study.

863 **Supplementary Table S2.** GenBank accession numbers of proteins used in phylograms.

864 **Supplementary File S1.** Differential gene expression between leaf-trichomes and trichomes.

865

Parsed Citations

Adams RP (2007) Identification of essential oil components by gas chromatography/mass spectrometry, Ed 4th. Allured Publishing Corporation, Carol Stream, Illinois.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Adam P, Hecht S, Eisenreich W, Kaiser J, Grawert T, Arigoni D, Bacher A, Rohdich F (2002) Biosynthesis of terpenes: studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. Proc Natl Acad Sci USA 99: 12108-12113.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Alvarez MA (2014) Plant Biotechnology for Health: From Secondary Metabolites to Molecular Farming. Springer. International Publishing Switzerland pp33-59.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Armstrong DW, Zhou EY, Zukowski J, Kosmowska-Ceranowicz B (1996) Enantiomeric composition and prevalence of some bicyclic monoterpenoids in amber. Chirality 8: 39-48.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodríguez-Concepción M, Gruissem W, Vranová E (2013) Characterization of the GGPP synthase gene family in Arabidopsis thaliana. Plant Mol Biol 82: 393-416.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Bleeker PM, Spyropoulou EA, Diergaarde PJ, Volpin H, De Both MT, Zerbe P, Bohlmann J, Falara V, Matsuba Y, Pichersky E, Haring MA, Schuurink RC (2011) RNA-seq discovery, functional characterization, and comparison of sesquiterpene synthases from Solanum lycopersicum and Solanum habrochaites trichomes. Plant Mol Biol 77: 323-336.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A (2015) Investigation of terpene diversification across multiple sequenced plant genomes. Proc Natl Acad Sci USA 112: E81-88.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brandle JE, Richman A, Swanson AK, Chapman BP (2002) Leaf ESTs from Stevia rebaudiana: a resource for gene discovery in diterpene synthesis. Plant Mol Biol 50: 613-622.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brandle JE, Rosa N (1992) Heritability for yield, leaf:stem ratio and stevioside content estimated from a landrace cultivar of Stevia rebaudiana. Can J Plant Sci 72: 1263-1266.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Brandle JE, Telmer PG (2007) Steviol glycoside biosynthesis. Phytochemistry 68: 1855-1863.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Caniard A, Zerbe P, Legrand S, Cohade A, Valot N, Magnard JL, Bohlmann J, Legendre L (2012) Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in Salvia sclarea (L.) and their relevance for perfume manufacture. BMC Plant Biol. 12: 119.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66: 212-229.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Chen J, Hou K, Qin P, Liu H, Yi B, Yang W, Wu W (2014) RNA-Seq for gene identification and transcript profiling of three Stevia rebaudiana genotypes. BMC Genomics 15: 571.
Pubmed: Author and Title
CrossRef: Author and Title

Google Scholar: Author Only Title Only Author and Title

Cho BO, Ryu HW, So Y, Cho JK, Woo HS, Jin CH, Seo KI, Park JC, Jeong IY (2013) Anti-inflammatory effect of austroinulin and 6-O-acetyl-austroinulin from Stevia rebaudiana in lipopolysaccharide-stimulated RAW264.7 macrophages. Food Chem Toxicol 62:638-644.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cordoba E, Porta H, Arroyo A, San Román C, Medina L, Rodríguez-Concepción M, León P (2011) Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. J Exp Bot 62: 2023-2038.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cordoba E, Salmi M, León P (2009) Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. J Exp Bot 60: 2933-2943.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cui G, Duan L, Jin B, Qian J, Xue Z, Shen G, Snyder JH, Song J, Chen S, Huang L, Peters RJ, Qi X (2015) Functional divergence of diterpene syntheses in the medicinal plant Salvia miltiorrhiza Bunge.

Plant Physiol pp.00695.2015. [Epub ahead of print]
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Cui H, Zhang ST, Yang HJ, Ji H, Wang XJ (2011) Gene expression profile analysis of tobacco leaf trichomes. BMC Plant Biol 11: 76-86.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

De S, Mondal S, Banerjee S (2013) Introduction to stevioside, in Stevioside: Technology, Applications and Health, John Wiley & Sons, Oxford. pp. 1-18.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Falara V, Pichersky E, Kanellis AK (2010) A copal-8-ol diphosphate synthase from the angiosperm Cistus creticus subsp. creticus is a putative key enzyme for the formation of pharmacologically active, oxygen-containing labdane-type diterpenes. Plant Physiol 154: 301-310.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Gang DR, Beuerle T, Ullmann P, Werck-Reichhart D, Pichersky E (2002) Differential production of meta hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. Plant Physiol 130: 1536-1544.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Geuns JM (2003) Stevioside. Phytochemistry 64: 913-921.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Glas JJ, Schimmel BC, Alba JM, Escobar-Bravo R, Schuurink RC, Kant MR (2012) Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. Int J Mol Sci 13: 17077-17103.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644-652.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, Zhang X, Liu W, Peters RJ, Chen X, Zhao ZK, Huang L (2013) CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. Proc Natl Acad Sci USA 110: 12108-12113.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Guo Z, Severson RF, Wagner GJ (1994) Biosynthesis of the diterpene cis-abienol in cell-free extracts of tobacco trichomes. Arch

Biochem Biophys 308: 103-108.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Harada E, Kim JA, Meyer AJ, Hell R, Clemens S, Choi YE (2010) Expression profiling of tobacco leaf trichomes identifies genes for biotic and abiotic stresses. Plant Cell Physiol 51: 1627-1637.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hedden P, Phillips AL (2000) Gibberellin metabolism: new insights revealed by the genes. Trends Plant Sci 5: 523-530.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Huang K, Armstrong DW (2009) GC-MS analysis of crocetane, phytane and some of their stereoisomers using cyclodextrin-based stationary phases. Organic Geochemistry 40: 283-286.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Humphrey TV, Richman AS, Menassa R, Brandle JE (2006) Spatial organisation of four enzymes from Stevia rebaudiana that are involved in steviol glycoside synthesis. Plant Mol Biol 61: 47-62.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Hunter WN (2007) The non-mevalonate pathway of isoprenoid precursor biosynthesis. J Biol Chem 282: 21573-21577.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Ibrahim NA, El-Gengaihi S, Motawe H, Riad SA (2007) Phytochemical and biological investigation of Stevia rebaudiana Bertoni; 1-labdane-type diterpene. Eur Food Res Technol 224: 483-488.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jin J, Kim MJ, Dhandapani S, Tjhang JG, Yin JL, Wong L, Sarojam R, Chua NH, Jang IC (2015) Floral transcriptome of Ylang Ylang (Cananga odorata var. fruticosa) uncovers biosynthetic pathways for volatile organic compounds and a multifunctional and novel sesquiterpene synthase. J Exp Bot 66: 3959-3975.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Jin J, Panicker D, Wang Q, Kim MJ, Liu J, Yin J-L, Wong L, Jang IC, Chua NH, Sarojam R (2014) Next generation sequencing unravels the biosynthetic ability of Spearmint (Mentha spicata) peltate glandular trichomes through comparative transcriptomics. BMC Plant Biology 14: 292.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kandra L, Wagner GJ (1988) Studies of the site and mode of biosynthesis of tobacco trichome exudate components. Arch Biochem Biophys 265: 425-432.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Keene CK, Wagner GJ (1985) Direct demonstration of duvatrienediol biosynthesis in glandular heads of tobacco trichomes. Plant Physiol 79: 1026-1032.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

King AJ, Brown GD, Gilday AD, Larson TR, Graham IA (2014) Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. Plant Cell 26: 3286-3298.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Kumar H, Kaul K, Bajpai-Gupta S, Kaul VK, Kumar S (2012) A comprehensive analysis of fifteen genes of steviol glycosides biosynthesis pathway in Stevia rebaudiana (Bertoni). Gene 492: 276-284.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R (2000) Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. Proc Natl Acad Sci USA 97: 2934-2939.
Pubmed: Author and Title

CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Liang, PH, Ko TP, Wang AHJ (2002) Structure, mechanism and function of prenyltransferases. Eur J Biochem 269: 3339-3354.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Markovic IS, Đarmati ZA, Abramovic BF (2008) Chemical composition of leaf extracts of Stevia rebaudiana Bertoni grown experimentally in Vojvodina. J Serb Chem Soc 73: 283-297.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

McGarvey DJ, Croteau R (1995) Terpenoid metabolism. Plant Cell 7: 1015-1026.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Obtani K, Yamasaki K (2002) Methods to improve the taste of the sweet principles of Stevia rebaudiana. In: Kinghorn, A.D. (Ed.), Stevia, The genus Stevia. Taylor & Francis, London and New York, pp. 138-159.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, Zerbe P, Bach SS, Møller BL, Bohlmann J, Hamberger B (2014) Manoyl oxide (13R), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in Coleus forskohlii. Plant Physiol 164: 1222-1236.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Peters RJ (2010) Two rings in them all: the labdane-related diterpenoids. Nat Prod Rep 27: 1521-1530.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Philippe RN, De Mey M, Anderson J, Ajikumar PK (2014) Biotechnological production of natural zero-calorie sweeteners. Curr Opin Biotechnol 26: 155-161.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Phillips MA, Walter MH, Ralph SG, Dabrowska P, Luck K, Urós EM, Boland W, Strack D, Rodríguez-Concepción M, Bohlmann J, Gershenzon J (2007) Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (Picea abies). Plant Molecular Biology 65: 243-257.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. Nucleic Acids Research 33: W116-W120.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Richman AS, Gijzen M, Starratt AN, Yang Z, Brandle JE (1999) Diterpene synthesis in Stevia rebaudiana: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway. Plant J 19: 411-421.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle JE (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana. Plant J 41: 56-67.
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Sallaud C, Giacalone C, Töpfer R, Goepfert S, Bakaher N, Rösti S, Tissier A (2012) Characterization of two genes for the biosynthesis of the labdane diterpene Z-abienol in tobacco (Nicotiana tabacum) glandular trichomes. Plant J 72: 1-17.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Starratt AN, Kirby CW, Pocs R, Brandle JE. (2002) Rebaudioside F, a diterpene glycoside from Stevia rebaudiana. Phytochemistry 59: 367-370.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725-2729.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Toyomasu T, Sassa T (2010) Diterpenes. In: Mander L, Lui H-W, eds. Comprehensive natural products II. Chemistry and biology, Vol. 1. Oxford: Elsevier, pp. 643-672.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Totte´ N, Charon L, Rohmer M, Compernolle F, Baboeuf I, Geuns JMC (2000) Biosynthesis of the diterpenoid steviol, an ent-kaurene derivative from Stevia rebaudiana Bertoni, via the methylerythritol phosphate pathway. Tetrahedron Lett 41: 6407-6410.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Totté N, Van den Ende W, Damme EJMV, Compernolle F, Baboeuf I, Geuns JMC (2003) Cloning and heterologous expression of early genes in gibberellin and steviol biosynthesis via the methylerythritol phosphate pathway in Stevia rebaudiana. Can J Bot 81: 517-522.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Turner GW, Gershenzon J, Croteau RB (2000) Distribution of peltate glandular trichomes on developing leaves of peppermint. Plant Physiol 124: 655-664.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Walter M, Hans J, Strack D (2002) Two distantly-related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. Plant J 31: 243-254.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Wang E,Wang R, Deparasis J, Loughrin JH, Gan S, Wagner GJ (2001) Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. Nature Biotechnology 19: 371-374.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zerbe P, Bohlmann J (2015) Plant diterpene synthases: exploring modularity and metabolic diversity for bioengineering. Trends Biotechnol pii: S0167-7799(15)00097-9.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zerbe P, Hamberger B, Yuen MM, Chiang A, Sandhu HK, Madilao LL, Nguyen A, Hamberger B, Bach SS, Bohlmann J (2013) Gene discovery of modular diterpene metabolism in nonmodel systems. Plant Physiol 162: 1073-1091.

Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Zi J, Mafu S, Peters RJ (2014) To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. Annu Rev Plant Biol 65: 259-86.

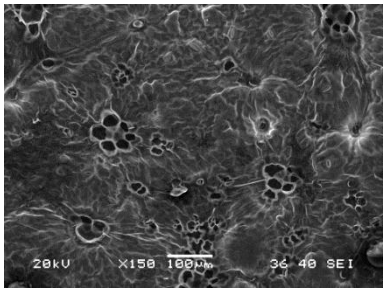
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

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

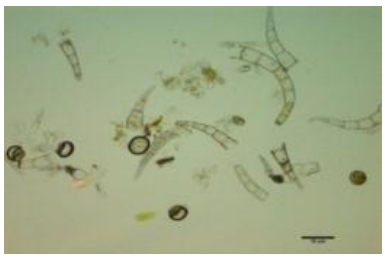
Pubmed: Author and Title
CrossRef: Author and Title
Google Scholar: Author Only Title Only Author and Title

Supplemental Figure S1

A



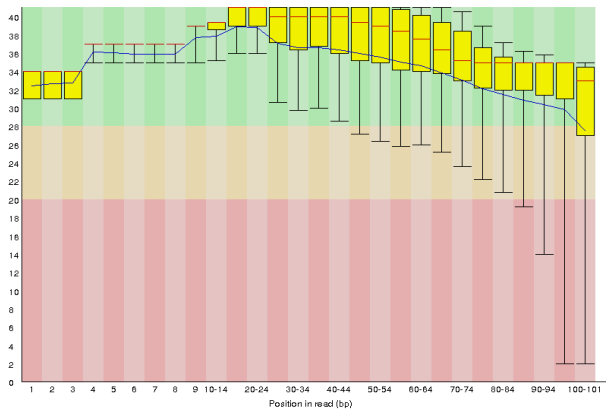
B



Supplemental Figure S1. Stevia tissues for RNA-seq.
A, leaf-trichomes; B, isolated trichomes.

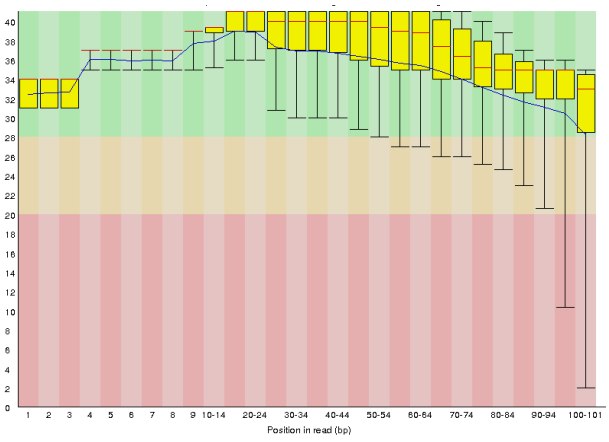
Supplemental Figure S2

A



leaf-trichomes

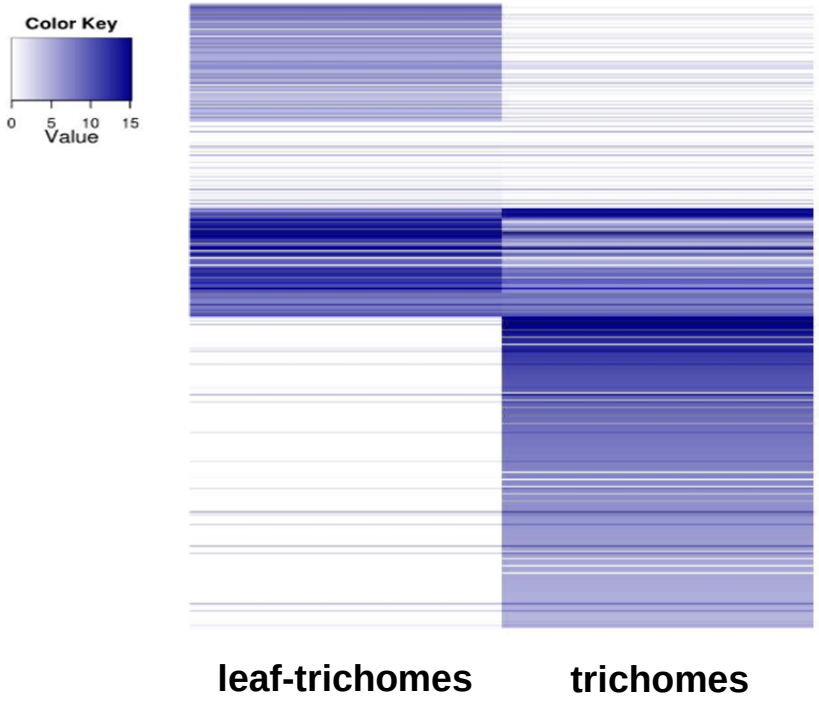
B



trichomes

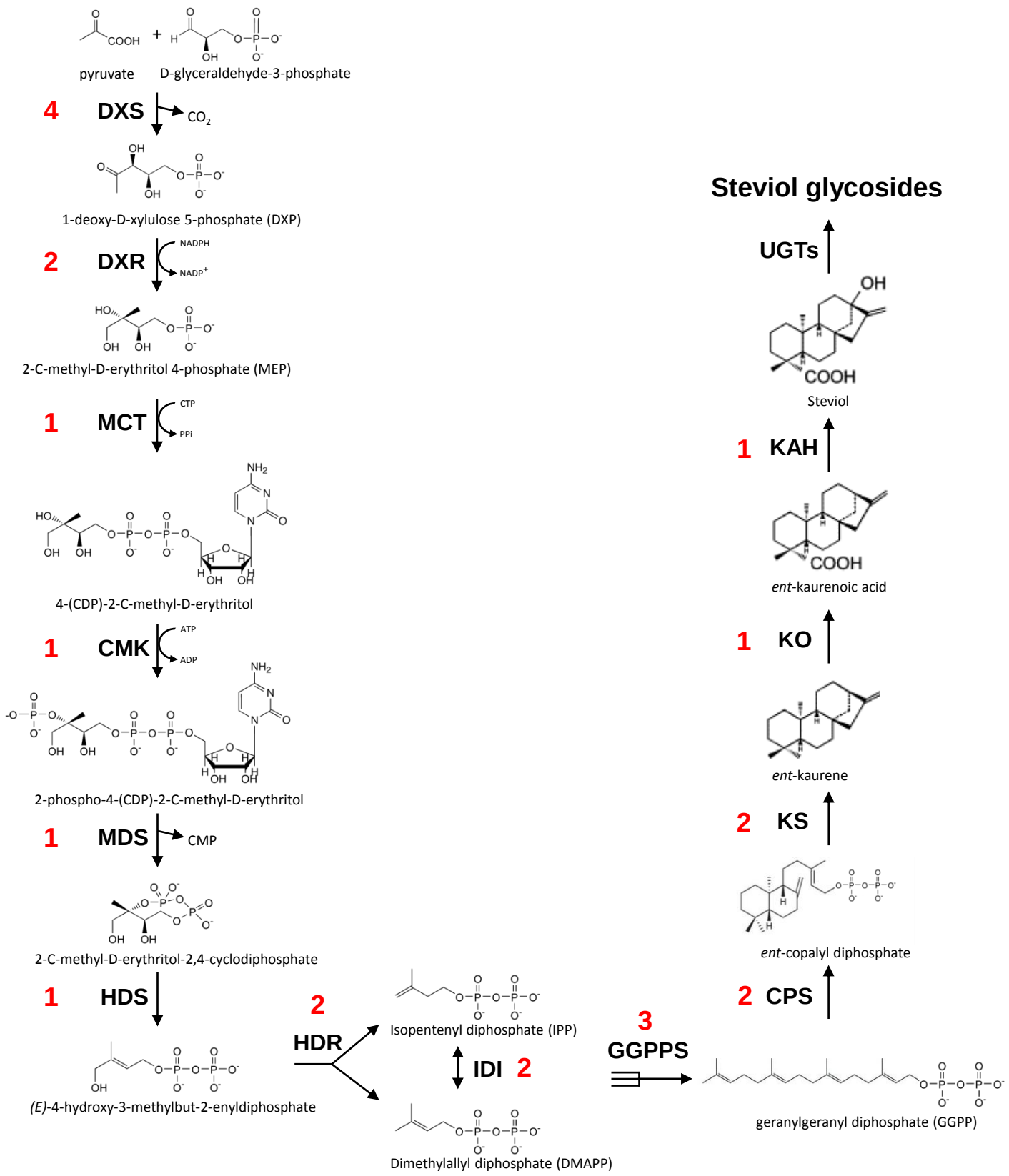
Supplemental Figure S2. Quality of strand specific RNA-seq derived from two Stevia tissues. The sequence quality was evaluated by FastQC. X-axis, position of 101 bp reads; Y-axis, score for each position.

Supplemental Figure S3



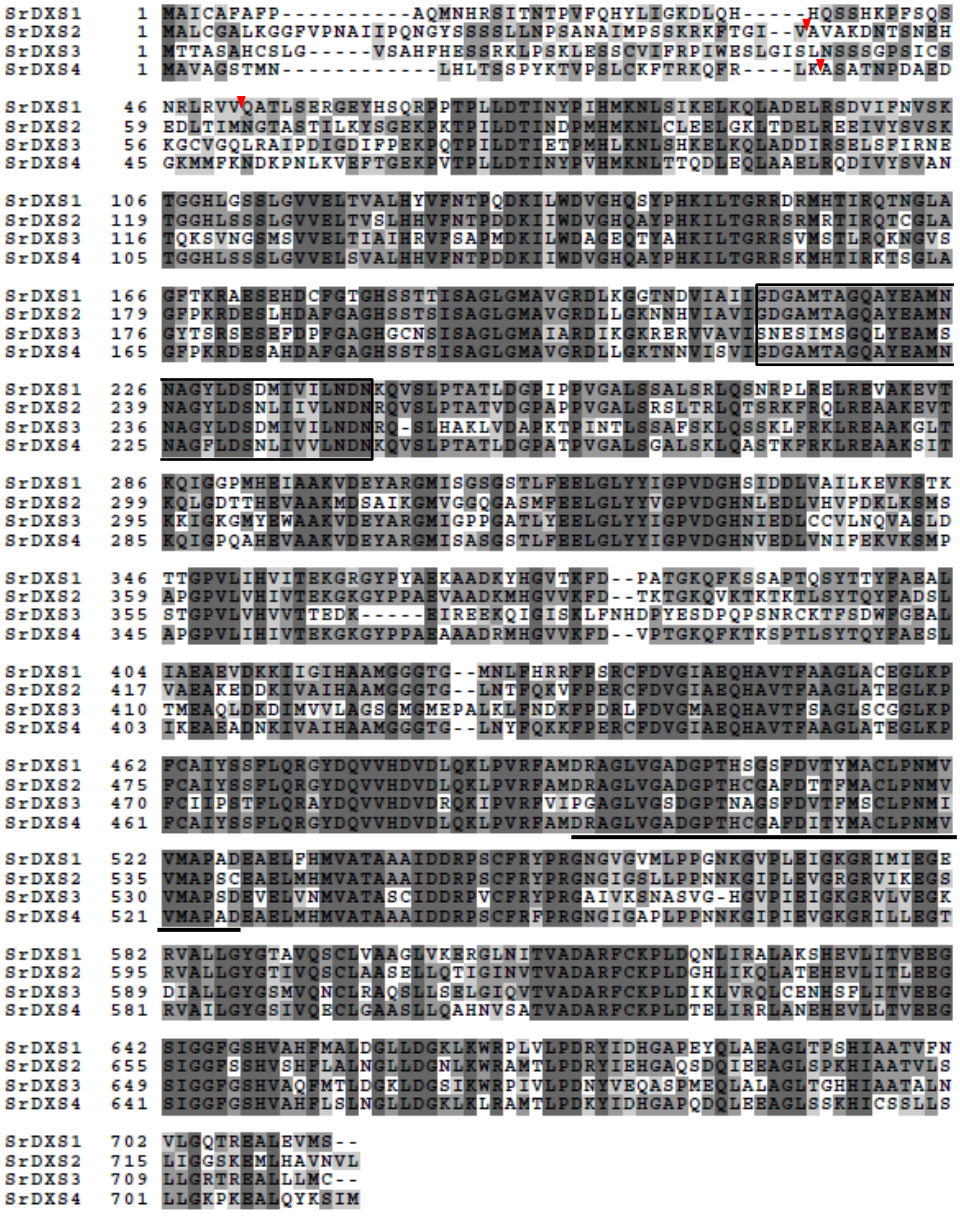
Supplemental Figure S3. Heat map of transcript expression in leaf-trichomes and trichomes.

Supplemental Figure S4



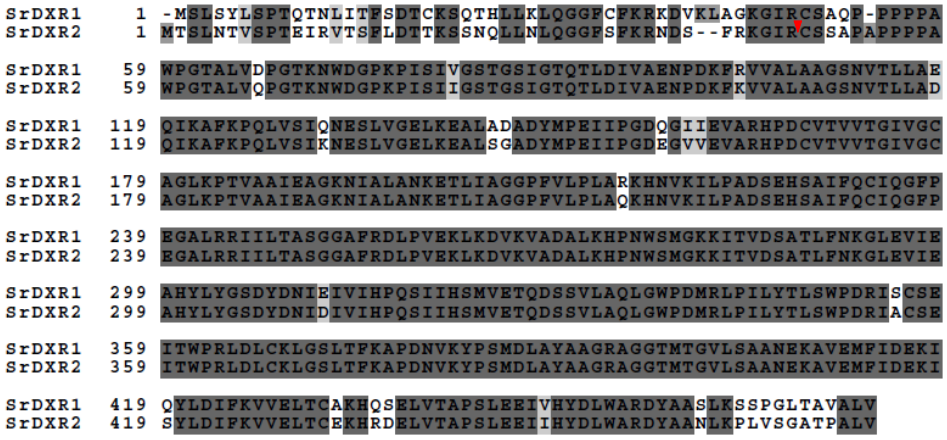
Supplemental Figure S4. The biosynthesis of steviol glycosides via the MEP pathway. The number in red represents full-length ORFs found in Stevia RNAseq.

Supplemental Figure S5



Supplemental Figure S5. Comparison of deduced amino acid sequences of four SrDXS small gene family. The consensus thiamine pyrophosphatase-binding domain and the pyridine-binding DRAG domain are indicated by the open box and the horizontal line, respectively. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program.

Supplemental Figure S6



Supplemental Figure S6. Comparison of deduced amino acid sequences of SrDXRs. The arrowhead denotes the predicted cleavage site of plastidial transit peptide by the ChloroP program.

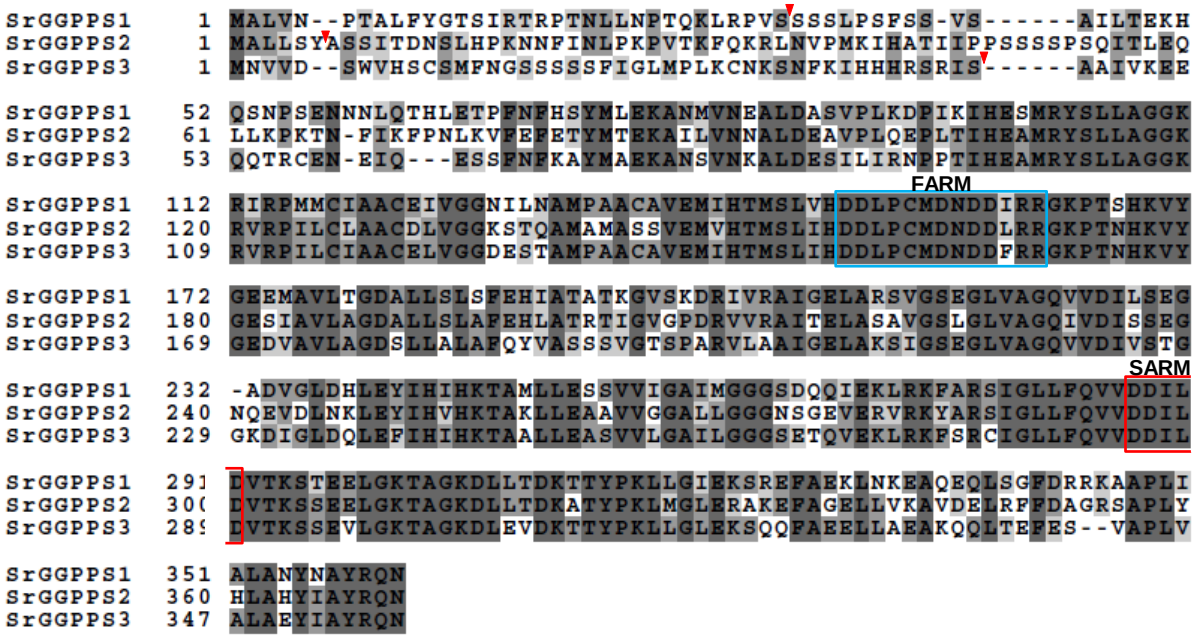
Supplemental Figure S7

Supplemental Figure S7. Comparison of deduced amino acid sequences of SrHDRs. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program. Dots mark the position of conserved cysteine residues.

Supplemental Figure S8

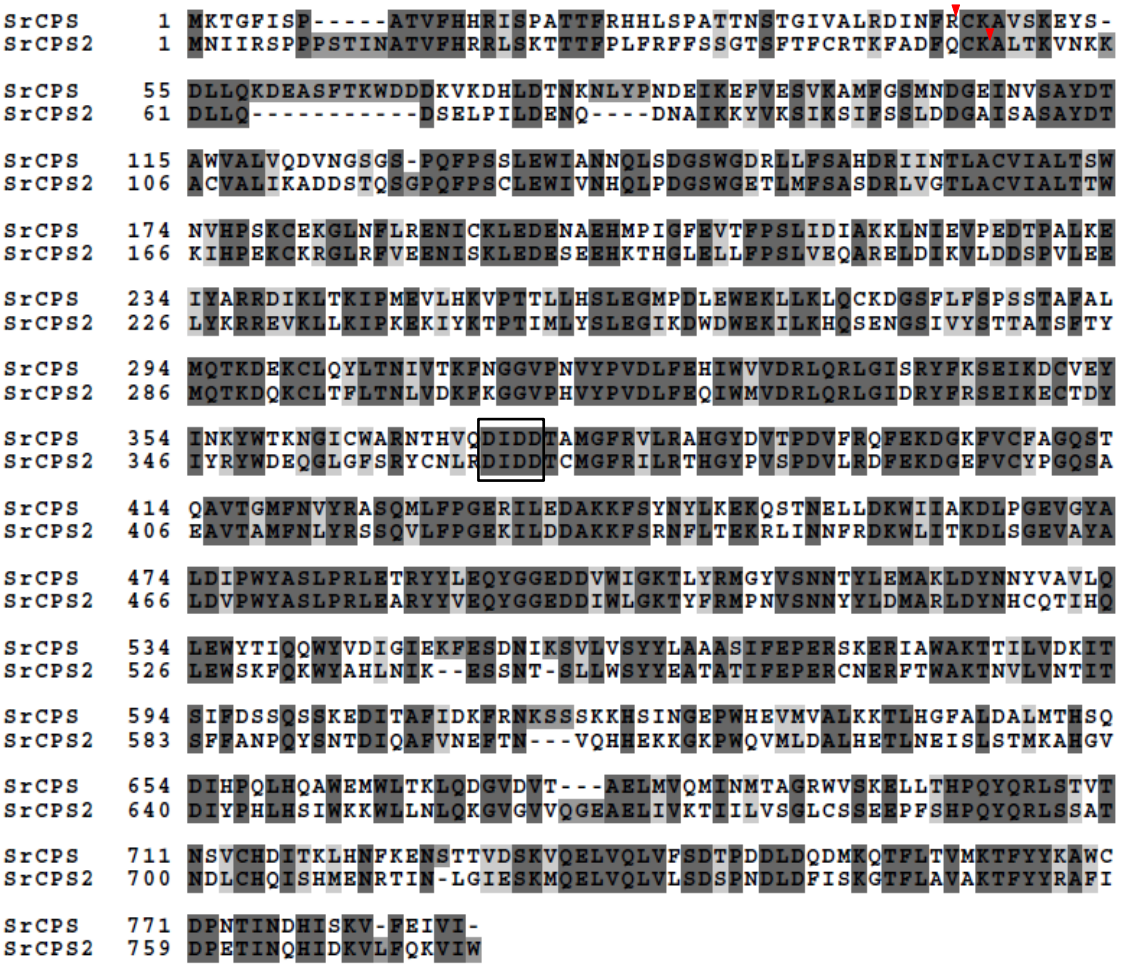
Supplemental Figure S8. Comparison of deduced amino acid sequences of SrIDIs. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program. The open boxes represent the NUDIX (Nucleoside Diphosphate linked to X) motif which is the characteristic of IDIs. SrIDI (Accession number, DQ989585) was previously reported (Kumar et al., 2012).

Supplemental Figure S9



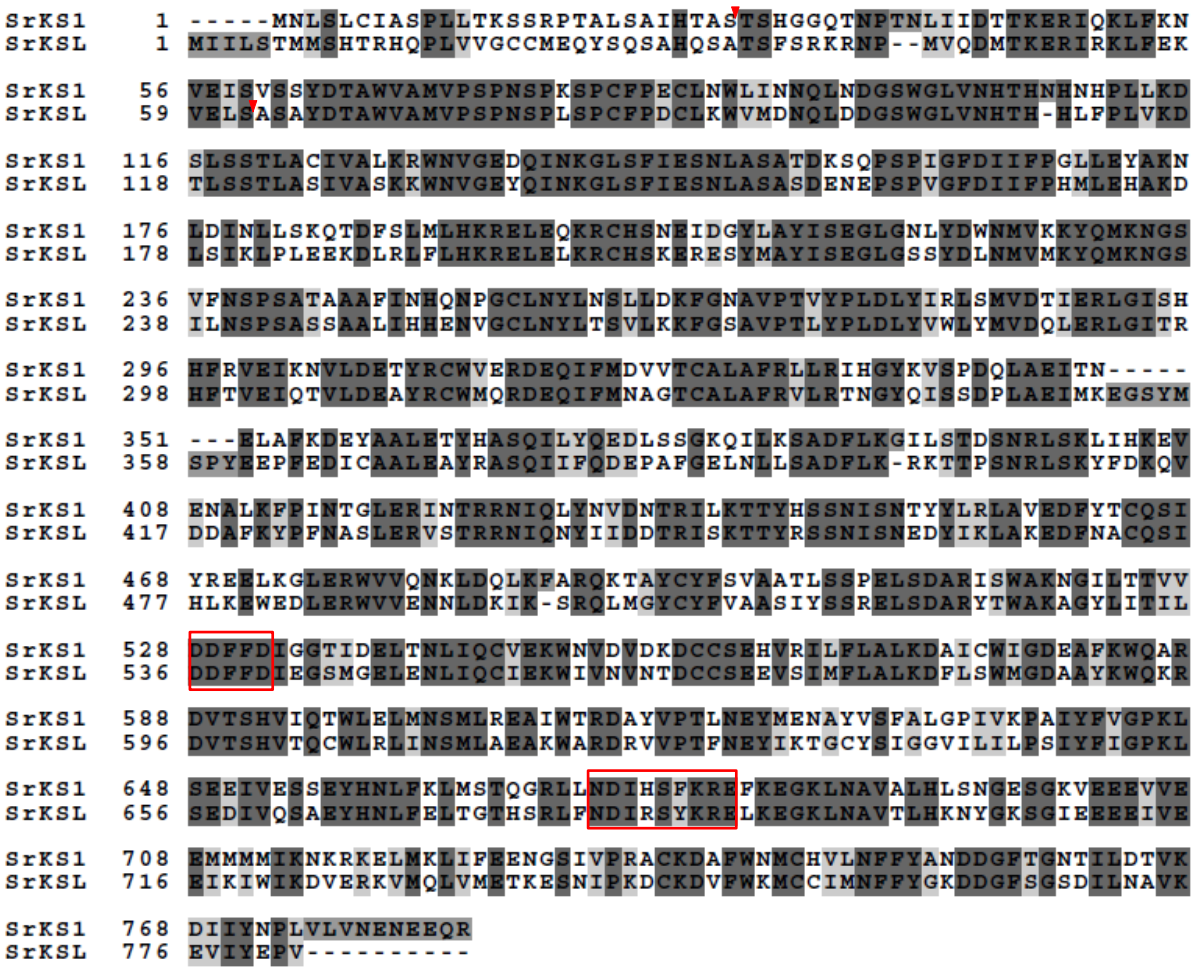
Supplemental Figure S9. Comparison of deduced amino acid sequences of SrGGPPSs. Two conserved aspartate-rich domains, the first aspartate-rich motif (FARM; DDx9RR) and the second aspartate-rich motif (SARM; DDxxD) are indicated by the blue and red open boxes, respectively. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program.

Supplemental Figure S10



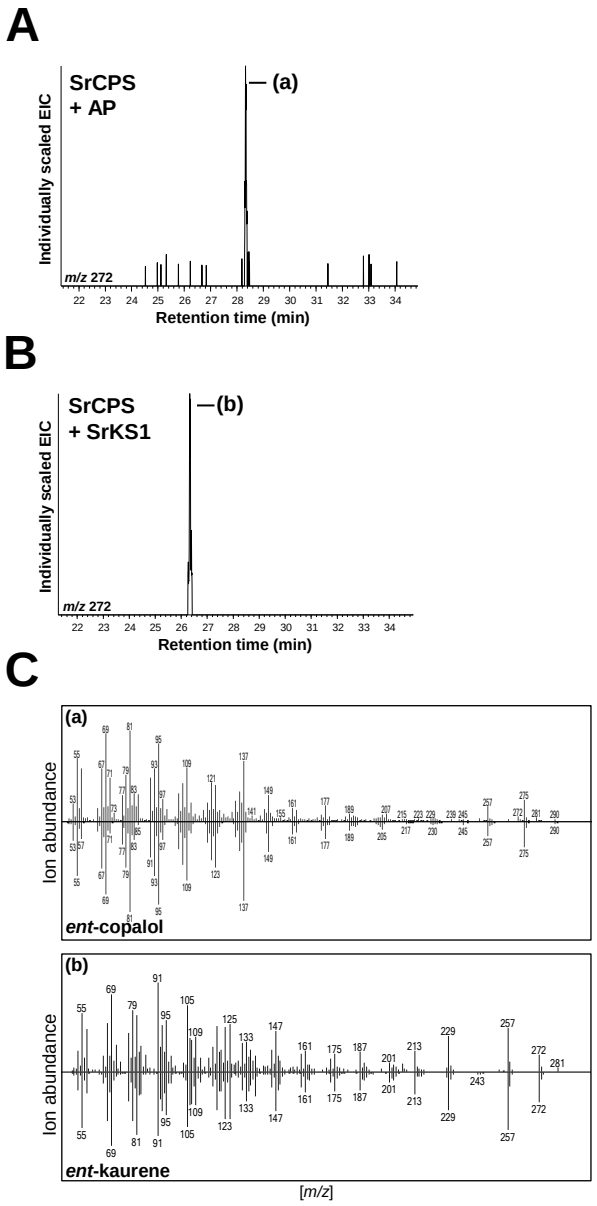
Supplemental Figure S10. Comparison of deduced amino acid sequences of SrCPS and SrCPS2. The open box indicates a class II TPS characteristic DxDD motif for a protonation-initiated cyclization of GGPP. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program.

Supplemental Figure S11



Supplemental Figure S11. Comparison of deduced amino acid sequences of SrKS1 and SrKSL. The open boxes indicate the conserved DDxxD and NSE/DTE motifs of class I TPS for metal-dependent ionization of the prenyl diphosphate substrate. The arrowheads denote the predicted cleavage sites of plastidial transit peptides by the ChloroP program.

Supplemental Figure S12

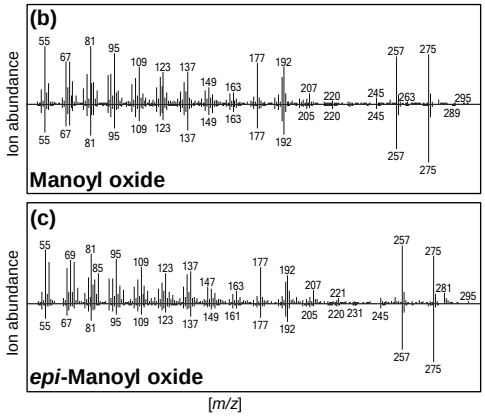


Supplemental Figure S12. GC-MS analysis of *in vitro* assay with diTPSs from *S. rebaudiana*. A. GC trace of reaction product from *in vitro* assay of recombinant SrCPS with GGPP as substrate. Reaction products were dephosphorylated by alkaline phosphatase (AP) for GC-MS analysis. B, GC trace of reaction product from coupled *in vitro* assays of recombinant SrCPS and SrKS1 with GGPP as substrate. C. Mass spectra obtained from peak (a) and (b) and reference spectra from the NIST 2014 library. peak (a), *ent*-copalol; peak (b), *ent*-kaurene. *m/z*, mass-to-charge ratio.

Supplemental Figure S13

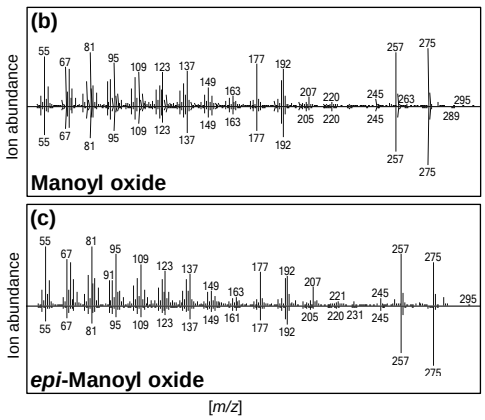
A

SrCPS2 + SrKSL



B

SrCPS2 + SrKS1



Supplemental Figure S13. Mass spectra obtained from coupled *in vitro* assays of recombinant SrCPS2 and SrKSL or SrCPS2 and SrKS1. peak (b), manoyl oxide; peak (c), *epi*-manoyl oxide. *m/z*, mass-to-charge ratio.

Supplemental Table S1. Oligonucleotide primers used in this study.

Name	Forward	Reverse
for qRT-PCR		
SrDXS1	CCGTTGCAGACGCTCGGTTTTG	GCACCATGGTCAATGTAACGGTC
SrDXS2	GGAAGAGTGATCAAAGAGGGAAGTA	TAGGACCTCATGCTCAGTTGCTAG
SrDXS3	CACATGTTGCACAATTTATGACTCTTG	ATAGAACCGCGATAAATTGTTGTCTTG
SrDXS4	TCACCGACACTTTCGTATACTCAG	CGAAAGTAACTGCGTGTTGTTCC
SrGGPPS1	GTTGCGAAAGTTTGCGAGATCGAT	TGCGGCGATCAAACCCCGACAATT
SrGGPPS2	AGCATCCGCGGTGGGTTCTTTAG	TGCATACTTTCTAACCCTTTCCACCT
SrGGPPS3	GACATCGGACTGGATCAACTGGAG	CCCAACAGTTTCGGATACGTCGTT
SrCPS	TTCCGGTGTAAAGCGGTATC	CATTGCTTTCACGCTCTCAA
SrCPS2	TGCAATGAGCGATTTACTTGGGCCA	AAGGGTTTCATGCAGTGCATCCAAC
SrKS1	TCCGGCTTTCTATGGTTGAC	AACCGAAAGGCTAAAGCACA
SrKSL	AGAGATCGTGTGGTGCCAACGTTTA	AGGCGAGAATGAGTGCCCGTTA
SrKO1	TCGATTAAAACCGGAGCAAC	CCCAAAACAGCGGTCAGTAT
SrKAH	CCGACTCCATTCACCTTTGT	ACTTGATGGGGATGAAGACG
SrACTIN	TCTTGATCTTGCTGGTCGTG	GAGCAAGAACTTGAAACCGC
for BP cloning		
SrDXS1	*AAAAAGCAGGCTTC***ATG**GCGATTTGTGCCTTTGCATTCCCG	*AGAAAGCTGGGTG*TGACATAACCTCCAGAGCCTCTCGG
SrDXS2	*AAAAAGCAGGCTTC***ATG**GCTTTATGTGGTGCTTTGAAGGGTG	*AGAAAGCTGGGTG*TAATACATTGACAGCATGTAGCATCTCCTTGC
SrDXS3	*AAAAAGCAGGCTTC***ATG**ACTACTGCTTCTGCACATTGTTCTTTGG	*AGAAAGCTGGGTG*ACACATCAAAAGAAGAGCTTCACGGGTTC
SrDXS4	*AAAAAGCAGGCTTC***ATG**GCGGTTGCAGGATCGACCATGAA	*AGAAAGCTGGGTG*CATTATTGATTTGTATTGAAGTGCTTCTTTAGGT
SrGGPPS1	*AAAAAGCAGGCTTC***ATG**GCTCTTGTAAATCCCACAGCTTT	*AGAAAGCTGGGTG*TCAGTTTTGCCTATAAGCATTGTAAT
SrGGPPS2	*AAAAAGCAGGCTTC***ATG**GCTTTGCTTTCTTATGCATCTTCT	*AGAAAGCTGGGTG*ATTTTGTCGGTAAGCAATATAATGTGC
SrGGPPS3	*AAAAAGCAGGCTTC***ATG**AATGTGGTTGATTCATGGGTTCAC	*AGAAAGCTGGGTG*GTTTTGCCGGTACGCTATATACTC
SrCPS	*AAAAAGCAGGCTTC***ATG**AAGACCGGCTTCATCTCTCCC	*AGAAAGCTGGGTG*TCATATTACAATCTCGAACACCTTGG
SrCPS2	*AAAAAGCAGGCTTC***ATG**GCTTTGCTTTCTTATGCATCTTC	*AGAAAGCTGGGTG*ATTTTGTCGGTAAGCAATATAATGTGC
SrKS1	*AAAAAGCAGGCTTC***ATG**AATCTTTCACTATGCATTGCGTC	*AGAAAGCTGGGTG*TTACCTTTGTTCTTCATTTTCATTCACA
SrKSL	*AAAAAGCAGGCTTC***ATG**ATCATATTATCTACCATGATGTC	*AGAAAGCTGGGTG*GACTGGTTCATATATGACTTCCTTC
for 5'-RACE		
	R1	**R2**
SrKSL	GATCTGTGATGCCCGATAGGCTTCGAGAG	CACCCGAAAAGCTAGAGCACAAGTCC
SrCPS2	GAGGACCACTTTGTGTACTATCATCGGC	GCGACAAAACGTGAAGGAGGTGCCAGAAC
for *E.coli* expression vectors		
SrCPS-F	aaGGATCCg**ATG**AAGACCGGCTTCATCTC	aaGCGGCCGCTCATATTACAATCTCGAACACC
SrCPS-M	aaGGATCCgATGTGTAAAGCGGTATCCAAAGAG	aaGCGGCCGCTCATATTACAATCTCGAACACC
SrCPS2-F	aaGCGGCCGCa**ATG**AATATCATCAGATCACCACCGCC	aaCTCGAGCTACCAAATTACCTTTTGAAACAACACT
SrCPS2-M	aaGCGGCCGCaATGGCGCTAACCAAAGTCAACAA	aaCTCGAGCTACCAAATTACCTTTTGAAACAACACT
SrKS1-F	aaGCGGCCGCa**ATG**AATCTTTCACTATGCATTG	aaCTCGAGTTACCTTTGTTCTTCATTTTCATTC
SrKS1-M	aaGCGGCCGCaATGACTTCACATGGTGGACAAACT	aaCTCGAGTTACCTTTGTTCTTCATTTTCATTC
SrKSL-F	aaGCGGCCGCa**ATG**ATCATATTATCTACCATGATGTC	aaCTCGAGTTAGACTGGTTCATATATGACTTCC
SrKSL-M	aaGCGGCCGCa**ATG**GCATCTGCATATGACACAGCATGG	aaCTCGAGTTAGACTGGTTCATATATGACTTCC

GGATTC, *Bam* HI; GCGGCCGC, *Not* I; CTCGAG, *Xho* I

Supplemental Table S2. Annotation and Genbank accession numbers of proteins used in the phylogenetic trees.

Enzyme name	Accession No.	Species
DXS (1-deoxy-D-xylulose 5-phosphate synthase)		
AtDXS1	At4g15560	*Arabidopsis thaliana*
AtDXS2	At3g21500	*Arabidopsis thaliana*
AtDXS3	At5g11380	*Arabidopsis thaliana*
CrDXS	CAA07554	*Chlamydomonas reinhardtii*
OsDXS1	NP_001055524	*Oryza sativa*
OsDXS2	NP_001059086	*Oryza sativa*
OsDXS3	BAA83576	*Oryza sativa*
PtDXS1	XP_002312717	*Populus trichocarpa*
PtDXS2A	XP_002303416	*Populus trichocarpa*
PtDXS2B	XP_002331678	*Populus trichocarpa*
PtDXS3	XP_002308644	*Populus trichocarpa*
PaDXS1	ABS50518	*Picea abies*
PaDXS2A	ABS50519	*Picea abies*
PaDXS2B	ABS50520	*Picea abies*
SmDXS1	ACF21004	*Salvia miltiorrhiza*
SmDXS2	ACQ66107	*Salvia miltiorrhiza*
SmDXS3	AEZ55686	*Salvia miltiorrhiza*
SmDXS4	AEZ55687	*Salvia miltiorrhiza*
SmDXS5	AEZ55688	*Salvia miltiorrhiza*
ZmDXS1	NP_001157805	*Zea mays*
ZmDXS2	ABP88135	*Zea mays*
ZmDXS3	ADN22972	*Zea mays*
GGPPS.LSU (Geranylgeranyl diphosphate synthase large subunit)		
AmGPPS.LSU	AAS82860	*Antirrhinum majus*
AtGGPPS1	NP_195399	*Arabidopsis thaliana*
AtGGPPS2	NP_179960	*Arabidopsis thaliana*
AtGGPPS3	NP_188073	*Arabidopsis thaliana*
MpGPPS.LSU	AF182828	*Mentha×piperita*
SmGGPPS1	ACJ66778	*Salvia miltiorrhiza*
SmGGPPS2	AEZ55682	*Salvia miltiorrhiza*
SmGGPPS3	AEZ55683	*Salvia miltiorrhiza*
SmGPPS.LSU	AEZ55681	*Salvia miltiorrhiza*
GGPPS.SSUI (Geranylgeranyl diphosphate synthase small subunit I)		
AmGPPS.SSUI	AAS82859	*Antirrhinum majus*
MpGPPS.SSUI	AF182827	*Mentha×piperita*
SmGPPS.SSUI	AEZ55678	*Salvia miltiorrhiza*
GGPPS.SSUII (Geranylgeranyl diphosphate synthase small subunit II)		
AtGPPS.SSUII	AAG40013	*Arabidopsis thaliana*
OsGPPS.SSUII	EAY87007	*Oryza sativa*
SmGPPS.SSUII.1	AEZ55679	*Salvia miltiorrhiza*
SmGPPS.SSUII.2	AEZ55680	*Salvia miltiorrhiza*
GPPS (Geranyl diphosphate synthase)		
AtGPPS1	NP_001031483	*Arabidopsis thaliana*
CrGPPS	ACC77966	*Catharanthus roseus*
SlGPPS	ABB88703	*Solanum lycopersicum*
SmGPPS	AEZ55677	*Salvia miltiorrhiza*
FPPS (Farnesyl diphosphate synthase)		
AtFPPS1	NP_199588	*Arabidopsis thaliana*
HbFPPS	AAM98379	*Hevea brasiliensis*
MpFPPS	AF384040	*Mentha×piperita*
SmFPPS	ABV08819	*Salvia miltiorrhiza*
VvFPPS	AAX76910	*Vitis vinifera*

(Continued)

Enzyme name	Accession No.	Species	Enzyme function
TPS-c			
AtCPS	NP_192187	*Arabidopsis thaliana*	copalyl diphosphate synthase
CcCLS	E2IHE0	*Cistus creticus*	copal-8-ol diphosphate hydratase
CfTPS1	AHW04046	*Coleus forskohlii*	TPS1
CfTPS2	AHW04047	*Coleus forskohlii*	TPS2
CmCPS1	AAD04292	*Cucurbita maxima*	copalyl diphosphate synthase1
CmCPS2	AAD04293	*Cucurbita maxima*	copalyl diphosphate synthase2
GrTPS1	AGN70887	*Grindelia robusta*	labd-13-en-8-ol diphosphate synthase
IeCPS1	AEP03177	*Isodon eriocalyx*	copalyl diphosphate synthase1
IeCPS2	AEP03175	*Isodon eriocalyx*	copalyl diphosphate synthase2
LsCPS1	BAB12440	*Lactuca sativa*	copalyl diphosphate synthase1
NtLPPS	G3CCC0	*Nicotiana tabacum*	copal-8-ol diphosphate hydratase
OsCPS1	NP_001046550	*Oryza sativa*	copalyl diphosphate synthase
SdCPS	BAB03594	*Scoparia dulcis*	copalyl diphosphate synthase
SmCPS1	ABV57835	*Salvia miltiorrhiza*	copalyl diphosphate synthase1
SmCPS2	AEZ55684	*Salvia miltiorrhiza*	copalyl diphosphate synthase2
SrCPS	AAB87091	*Stevia rebaudiana*	copalyl diphosphate synthase
SsLPPS	AFU61897	*Salvia sclarea*	labd-13-en-8-ol diphosphate synthase
ZmCPS	NP_001105257	*Zea mays*	copalyl diphosphate synthase
TPS-d			
AbCAS	H8ZM73	*Abies balsamea*	cis-abienol synthase
AgAS	Q38710	*Abies grandis*	abietadiene synthase
GbLS	Q947C4	*Ginkgo biloba*	levopimaradiene synthase
PaLAS	Q675L4	*Picea abies*	levopimaradiene/abietadiene synthase
PaIso	Q675L5	*Picea abies*	isopimaradiene synthase
TbTs	Q41594	*Taxus brevifolia*	taxadiene synthase
TPS-e			
AtKS	AAC39443	*Arabidopsis thaliana*	kaurene synthase
CmKS	Q39548	*Cucurbita maxima*	kaurene synthase
GrTPS6	AGN70886	*Grindelia robusta*	manoyl oxide synthase
LsKS1	BAB12441	*Lactuca sativa*	kaurene synthase1
OsKS1	AAQ72559	*Oryza sativa*	kaurene synthase1
PtKS	XP_002311286	*Populus trichocarpa*	kaurene synthase
SdKS	AEF33360	*Scoparia dulcis*	kaurene synthase
SsdiTPS3	AFU61899	*Salvia sclarea*	TPS3
SrKS1	AF097310	*Stevia rebaudiana*	kaurene synthase1
TPS-e/f			
CfTPS3	AHW04048	*Coleus forskohlii*	TPS3
CfTPS4	AHW04049	*Coleus forskohlii*	TPS4
NtABS	G3CCC1	*Nicotiana tabacum*	cis-abienol synthase
ShSBS	B8XA41	*Solanum habrochaites*	santalene and bergamotene synthase
SlPHS1	NP_001234629	*Solanum lycopersicum*	β-phellandrene synthase
SmKSL	ABV08817	*Salvia miltiorrhiza*	kaurene synthase like
SsSS	AFU61898	*Salvia sclarea*	sclareol synthase
Bifunctional CPS/KS			
PpCPS/KS	BAF61135	*Physcomitrella patens*	*ent*-kaurene synthases