Plant Physiology and Biochemistry 196 (2023) 318–327

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Monoterpene synthases of three closely related sage species (*Salvia officinalis*, *S. fruticosa* and *S. pomifera*, Lamiaceae)

Corinna Schmiderer [a,1], Ralf Steinborn [b], Johannes Novak [a,*]

[a] Institute of Animal Nutrition and Functional Plant Compounds, University of Veterinary Medicine, Veterinärplatz 1, 1210 Vienna, Austria
[b] Genomics Core Facility, VetCore, University of Veterinary Medicine, Veterinärplatz 1, 1210 Vienna, Austria

ARTICLE INFO

Keywords:
Salvia officinalis L.
Salvia fruticosa Mill.
Salvia pomifera L.
Copy number variation
Digital PCR
Gene expression
Monoterpenes
Alpha-thujone
Beta-thujone
Camphor
1,8-cineole

ABSTRACT

The diversity of plant monoterpenes is largely based on the catalytic activity of monoterpene synthases. Additionally, copy number variation of monoterpene synthase genes may contribute to the quantity of transcripts and hence to the essential oil profile.

This study used whole-genome sequencing and digital PCR for the measurement of copy number variation and quantification of gene expression in three closely related *Salvia* species, namely *Salvia officinalis*, *Salvia pomifera* and *Salvia fruticosa*. Twelve, 13 and 15 monoterpene synthase-encoding open-reading frames were predicted for *Salvia officinalis*, *Salvia pomifera* and *Salvia fruticosa*, respectively. In *Salvia officinalis*, one of the open reading frames was disrupted indicating a pseudogene. Monoterpene synthase genes were generally single copy per haploid genome, only a few were double or triple copy genes.

Expression levels of monoterpene synthases in leaves corresponded generally well with essential oil composition. In some cases, a higher expression level of a certain monoterpene synthase could be explained by its duplication or triplication. The very high content of thujones in *Salvia pomifera*, for example, was accompanied by gene duplication and increased gene expression of (+)-sabinene synthase responsible for the thujone precursor sabinene.

In *Salvia officinalis*, three individuals different in their essential oil profile showed significant differences in their monoterpene synthase expression levels corresponding roughly to the profile of the essential oils.

Transcript expression of monoterpene synthase genes were measured in leaf, calyx and corolla. The corolla differed significantly from leaves, while calyces usually showed a profile intermediary between leaf and corolla.

1. Introduction

Salvia is with around 900 species the largest genus of the Lamiaceae distributed across Europe, Asia, Africa, and America. The 30 to 40 species in the Mediterranean and the Irano-Turanic Regions are categorised within the genus as the type section *Salvia* (Hedge, 1972). This section includes three close relatives, *S. officinalis* L., *S. fruticosa* Mill (syn. *S. libanotica* Boiss. & Gaill., *S. triloba* L.). and *S. pomifera* L., essential oil storing sage species with culinary and (folk-)medicinal uses. *Salvia officinalis*, the type species of the genus, is originally native to North and Central Spain, South France and to the Western part of the Balkan Peninsula. It is widely cultivated as an herb and naturalised in parts of South and Central Europe to Asia Minor. *Salvia fruticosa* is native to the Central and Eastern Mediterranean region, and *S. pomifera* L. to Greece and Turkey. All three species are diploid with $2n = 14$ chromosomes (Löve, 1980; Maksimović et al., 2007; Esra et al., 2011). The amount of DNA per haploid genome (1C) is 0.49 pg for *S. officinalis* (Maksimović et al., 2007) and 0.84 pg for *S. fruticosa* (Bou Dagher-Kharrat et al., 2013). Hybridisation is possible between *S. officinalis* and *S. fruticosa* (Putievsky et al., 1990). The three sage species are known to accumulate bioactive secondary plant compounds from several classes such as flavonoids, phenolic acids (especially rosmarinic acid) (Pizzale et al., 2002; Lamien-Meda et al., 2010a; Cvetkovikj et al., 2013), diterpenes like carnosol and sesqui- and monoterpenes

Abbreviations: BS, bornyl diphosphate synthase; CNV, copy number variation; CS, 1,8-cineole synthase; dPCR, digital PCR; GPP, geranyl-pyrophosphate; SS, sabinene synthase; TPS, (mono)terpene synthase.

* Corresponding author.
E-mail address: Johannes.Novak@vetmeduni.ac.at (J. Novak).
[1] present address: Institute of Botany, University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria.

https://doi.org/10.1016/j.plaphy.2023.01.034
Received 28 October 2022; Received in revised form 4 January 2023; Accepted 17 January 2023
Available online 19 January 2023

(Baser et al., 1993; Länger et al., 1996; Karousou and Kokkini, 1997; Lamien-Meda et al., 2010b).

Their essential oil consists mainly of monoterpenes that accumulate in essential oil glands of the leaf epidermis. The developing leaf represents the developmental stage with the highest expression levels of (mono)terpene synthase (TPS) enzymes (Schmiderer et al., 2010).

The three species express similar monoterpene compounds that differ strikingly in quantitative composition. The essential oil of *S. officinalis* is almost colourless to pale-yellow with a typical thujone odour (Burdock, 2009). Its main compounds are 1,8-cineole (7–17%), α-thujone (1–37%), β-thujone (2–14%) and camphor (3–12%) (Lamien-Meda et al., 2010b). The essential oil composition is polymorphic and numerous chemotypes exist (Perry et al., 1999; Lamien-Meda et al., 2010b; Schmiderer et al., 2013b). The oil of *S. fruticosa* is characterised by a very high content of 1,8-cineole, accompanied by camphor, α- and β-pinene and myrcene. Both thujones are present in small quantities only (Karousou and Kokkini, 1997). The essential oil of *S. pomifera* is extremely rich in both, α- and β-thujone, while the content of 1,8-cineole is much lower than in both the other two species (Baser et al., 1993).

Therefore, the essential oil composition of *S. officinalis* is intermediate between the thujone-poor *S. fruticosa* and the thujone-rich *S. pomifera*.

Monoterpenes are competing for the same substrate, geranylpyrophosphate (GPP). 1,8-cineole synthase (CS) forms in a one-step process 1,8-cineole (Wise et al., 1998). (+)-Sabinene synthase (SS) catalyses the first step to (+)-sabinene (Wise et al., 1998), which is further hydroxylated to (+)-sabinol (Karp and Croteau, 1982). In a next step, (+)-sabinol dehydrogenase converts (+)-sabinol to (+)-sabinone (Dehal and Croteau, 1987), which is reduced by two independent double-bond reductases into (−)-α-thujone and (+)-β-thujone, respectively. Bornyl diphosphate synthase (BS) forms bornyl diphosphate, which is further transformed to borneol by bornyl diphosphate hydrolase (Croteau and Karp, 1979). Borneol dehydrogenase catalyses the conversion of borneol to camphor (Dehal and Croteau, 1987).

In some distinct casesmRNA expression is directly correlated with the end-product levels for the main monoterpenes, 1,8-cineole and camphor (transcriptionally controlled). For thujones, however, transcriptional control seems to be more complex as no direct correlation between mRNA and product levels could be detected (Schmiderer et al., 2010; Grausgruber-Gröger et al., 2012).

Here we asked whether copy number variation (CNV) and transcript expression of TPS genes contribute to the vastly different monoterpene composition of *S. officinalis*, *S. fruticosa* and *S. pomifera*. We present an inventory of TPS genes for sage that includes CNV, level of transcript expression and monoterpene profiles. The sequences of the TPS genes and their frequency per haploid genome were identified by next-generation sequencing (NGS) and verified by digital PCR (dPCR). Transcript number of TPS genes in calyx, corolla and young leaf was counted by reverse-transcription dPCR (RT-dPCR). The composition of monoterpenes in three genotypes of *S. officinalis, S. fruticosa* and *S. pomifera* was measured by gas chromatography–mass spectrometry (GC-MS). Genome, transcriptome and phenotype data facilitated correlating CNV of TPS genes to transcript expression and transcript expression to monoterpene composition.

2. Material and methods

2.1. Plant material

Clonally propagated plants of three sage species (Lamiaceae), *Salvia officinalis* ($n = 3$), *Salvia fruticosa* ($n = 1$) and *Salvia pomifera* ($n = 1$), were cultivated at the University of Veterinary Medicine Vienna. Clones were either of known geographic origin such as the samples 'So3' of *S. officinalis* descending from a wild population near Foggia (Italy) and 'Sf1' of *S. fruticosa* originating from Llogara Pass (Albania) or unknown regional provenance (samples 'So11' and 'So26' of *S. officinalis* as well as 'Sp2' of *S. pomifera*). The three genotypes of *S. officinalis* were selected after previous analysis from our own collection because of their differing chemotypes ('So3' a genotype with a relatively high content of camphor (31%), 'So11' high content of *β*-thujone (47%), 'So26' above-average content of cineole (18%)). Plant material for DNA and RNA extraction was collected repeatedly between autumn 2015 and spring 2018. Material used for RT-dPCR was collected at full bloom in spring 2018. Harvested plant material was transferred into 2-ml tubes, immediately frozen in liquid nitrogen and stored at −80 °C until RNA or DNA isolation.

2.2. DNA extraction and digestion

Extraction of genomic DNA followed a modified cetrimonium bromide (CTAB) protocol, termed "CTAB method 1" (Schmiderer et al., 2013a) that is based on Doyle and Doyle (1990). Before DNA extraction, the frozen material was ground to fine powder using a swing mill (Retsch MM301, Haan, Germany).

DNA quantity and purity were determined using a NanoDrop ND-2000c spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany). For dPCR, 1 μg of DNA was digested in a 20-μl reaction with 10 U MspI or EcoRV (Thermo Fisher Scientific, Vienna, Austria) and cleaned up with the Monarch™ PCR and DNA Cleanup Kit (5 μg) (New England Biolabs, Ipswich, MA, U.S.A.) following the protocol of the manufacturer.

2.3. NGS

Illumina's high-throughput sequencing-by-synthesis technology (HiSeq V2 sequencing, paired-end reads, and two channels) was outsourced to a commercial service provider (LGC Genomics GmbH, Berlin, Germany). Remnants of the sequencer adapter were removed from raw reads. Reads with more than one 'N' were removed and the reads trimmed at the 3′ end to obtain a minimum average Phred quality score of 25 over 10 bases. Finally, reads with a final length <20 bases were discarded. Forward and reverse reads were combined using BBMerge 34.48 (Bushnell et al., 2017).

2.4. Primer and probe design

Primers were designed using Primer Express 2.0 (Applied Biosystems, Foster City, CA, USA) (Supplementary Table 1, Supplementary Table 2) and assessed for secondary structures using the NetPrimer tool (http://www.premierbiosoft.com/netprimer/). For extendable primer dimers a Gibbs free energy change (ΔG) above −1 kcal/mol was preferred. The potential for amplicon secondary structure formation was evaluated using the DNA folding form of the UNAFold Web Server (http://www.unafold.org/mfold/applications/dna-folding-form.php) (Zuker, 2003).

Hydrolysis probes (Supplementary Table 3) were designed using the Primer Express 2.0 and the T_m Prediction tool for LNA-enhanced oligos (https://geneglobe.qiagen.com/at/tools/tm-prediction).

2.5. Copy number counting of nucleic-acid sequences by dPCR

Copies of nucleic acid sequences were counted by dPCR using the QuantStudio™ 3D Digital PCR system (Applied Biosystems, part of Thermo Fisher Scientific). The platform comprised a chip loader, chip reader, 20K Chip Kit v2 and a GeneAmp PCR System 9700 equipped with two flat blocks. The duplex dPCR assay monitored TPS targets and reference genes with 6-FAM- and Yakima Yellow™-labelled probes, respectively (Supplementary Table 3). The 18-μl reaction contained 1 × QuantStudio® 3D Digital PCR Master Mix v2, 200 nM of each primer (Sigma Aldrich, Vienna, Austria), 200 nM of each probe (Integrated DNA Technologies, Leuven, Belgium) and 2 μl of diluted DNA or cDNA (Supplementary Table 1, Supplementary Table 3). A 14.5-μl aliquot of

this mixture was applied to a dPCR chip using the chip loader. Following an initial denaturation step at 96 °C for 10 min, 45 amplification cycles (98, 55 °C for 30 s, 2 min) and a final extension (60 °C for 2 min) were performed. Reaction endpoints were measured at the Quant Studio™ 3D digital system (Thermo Fisher Scientific) and analysed in the QuantStudio® 3D Analysis Suite™ software version 3.1.4 (online at Thermo Fisher Cloud).

2.6. Assay specificity assessment

Primer and probe sequences designed for dPCR were tested for cross reactivity as follows. Genomic DNA or cDNA of each TPS gene was amplified as comprehensively as possible using two consecutive rounds of amplifications in most cases. To increase product specificity, the second reaction was performed with one or two nested primers. The 30-µl reaction included 500 nM of each primer, 1 × Phusion buffer HF and 0.3 U Phusion HS II polymerase (Thermo Fisher Scientific) (Supplementary Table 2, Supplementary Table 4). The amplification program consisted of initial denaturation at 98 °C for 60 s, 35 amplification cycles (98, 55 and 72 °C for 10, 20 and 90 s) and final elongation (72 °C for 7 min). Product composition was evaluated by electrophoresis on an 1.4% agarose gel run in 1 × sodium borate buffer (Brody and Kern, 2004). Amplicons of the expected size were cleaned with Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific). Products showing several bands were transferred to 0.5 × Rotiphorese TAE gels (Carl Roth, Graz, Austria). The separated bands were cut out and re-extracted with the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. In case of uncertainty, a product was verified by Sanger sequencing outsourced to a commercial service provider (Microsynth, Vienna, Austria). Subsequently, amplicons were analysed by quantitative PCR (qPCR). Templates diluted with Tris-EDTA buffer to correspond to *Cq* values of 22–26 were tested for cross reactivity against other TPS sequences. The extent of cross reactivity was concluded from the *Cq* difference of cross-reaction and specific signals. For qPCR conditions see below.

2.7. Sanger sequencing analysis

Sequencing of PCR amplicons was outsourced to a commercial service provider (Microsynth AG, Balgach, Switzerland).

2.8. Amplification efficiency of qPCR assays

The qPCR efficiencies for all duplex-dPCR assays (including primers and probes for one TPS and the reference gene *PEX4*) were calculated with synthetic DNA constructs and a pooled sample of genomic DNA composed of *S. officinalis* (specimen 'So3'), *S. fruticosa* (specimen 'Sf1') and *S. pomifera* (specimen 'Sp2'). The 10-µl reaction included 1 × buffer B2, 4 mM $MgCl_2$, 0.65 U HOT FIREPol® DNA Polymerase, 0.2 mM dNTPs (all Solis BioDyne, Tartu, Estonia), 250 nM of each primer, 200 nM of each probe and 1 µl template DNA. Efficiency was calculated from standard curves obtained from tenfold serial dilutions of 5 ng/µl of genomic DNA or ~2 fg/µl of synthetic DNA. The qPCR was performed using the Rotor-Gene Q (Qiagen) with an initial denaturation at 95 °C for 15 min and 45 amplification cycles (95, 57 and 72 °C for 10, 20 and 20 s). Measured efficiencies ranged from 81 to 104% (average: 94%) except of *SfTPS16* (61%) (Supplementary Table 5).

2.9. Synthetic DNA constructs

For determination of qPCR efficiency and CNV by dPCR, four double-stranded DNA constructs of 486–993 bp, termed 'GeneStrands', were used. They included target regions of the reference genes *FIM* and *PEX4* and segments of selected TPSs separated by a recognition site of a restriction endonuclease (Supplementary Data 1). GeneStrands were synthesised by Eurofins Genomics (Ebersberg, Germany).

Before use, DNA constructs were digested with MspI or EcoRV (Thermo Fisher Scientific) and cleaned up with the Monarch® PCR and DNA Cleanup Kit (5 µg) (New England BioLabs, Frankfurt, Germany). DNA was diluted in 1 × TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) to a concentration of ~2 fg/µl.

2.10. RNA extraction and cDNA synthesis

Approximately 50–90 mg of plant material flash frozen in liquid nitrogen and stored at −82 °C was subjected to extraction with the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA quantity and purity were determined using a NanoDrop ND-2000c (Peqlab Biotechnologie GmbH). Genomic DNA was removed with RNase-free DNase I (Thermo Fisher Scientific) following the manufacturer's protocol. The reverse transcription (RT) reaction was performed in a 20-µl volume using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific), 0.5 µg template RNA and 1 mM of the VN-tailed oligo$(dT)_{18}$ primer 5′-GCT GTC AAC GAT ACG CTA CGT AAC TGC ATG ACA GTG $(T)_{18}$VN, where V designates A, C or G and N any base. Contamination with DNA was monitored by a mock-RT control.

TPS transcript expression was normalised to the reference gene *PEX4* (GenBank: EU399687.1) found to be stably expressed during leaf ontogenesis in *S. officinalis* (unpublished data).

Normalised transcript expression of a particular TPS target was obtaining by dividing the cDNA copy number of the TPS gene by the cDNA copy number of the reference gene *PEX4* multiplied by 100. Finally, relative TPS transcript expression was obtained by dividing the expression of the target gene by the sum of all TPS transcripts.

2.11. Gas chromatography-mass spectrometry (GC-MS)

Plant material left-over from RNA extraction was used for GC-MS. Material remaining in the QIAshredder spin column (Qiagen) was supplemented with 0.5 ml dichloromethane, incubated at room temperature for ~ 1 min, and centrifuged for 30 s at 14,000×*g*. The flow-through was transferred to GC vials. GC-MS analysis was performed as described by Schmiderer et al. (2008). Essential oil compounds were identified using a HP 6890 gas chromatograph coupled with the quadrupole mass spectrometer HP5972 MSD (Hewlett-Packard, Palo Alto, CA, USA) fitted with a DB5-MS capillary column (30 m × 0.25 mm inner diameter, film thickness: 0.25 µm; Agilent, Palo Alto, CA, USA). Helium was used as carrier gas (average velocity: 42 cm/s), the injector temperature was set to 250 °C and the split ratio to 100:1. The temperature program started with 60 °C for 4 min, rising to 100 °C with 5 °C/min increase, and from 100 to 280 °C with 9 °C/min. The retention indices of the essential oil compounds were determined in comparison to *n*-alkane hydrocarbons (retention index standard for GC, Sigma-Aldrich, Vienna, Austria) under the same conditions. The compounds were identified comparing their mass spectra and retention indices to published data (McLafferty and Stauffer, 1989; Adams, 2007). The composition was obtained by peak-area normalisation, and the response factor for each compound was considered to equal 1.

2.12. TPS gene structure and phylogenetic reconstruction

The exon-intron structure of TPS genes was visualised using the Gene Structure Display Server (http://gsds.gao-lab.org) (Hu et al., 2015). Plastid-targeting peptides were predicted by TargetP-2.0 server (https://services.healthtech.dtu.dk/service.php?TargetP-2.0) (Almagro Armenteros et al., 2019). The evolutionary history of TPSs was reconstructed using the Maximum Likelihood method with the General Time Reversible model (Gamma distribution + invariant sites) (Nei, 2000) as implemented in MEGA11 (Tamura et al., 2021). The tree with the highest log-likelihood was displayed.

2.13. Statistical analysis

All statistical analysis were performed with R 3.6.0 (R Core Team) and RStudio (RStudio Team, 2019). PCA was performed with the R-package FactoMineR (Lê et al., 2008) and factoextra (Kassambara and Mundt, 2016).

3. Results

3.1. NGS analysis and identification of TPSs candidates

Genomic DNA isolated from a wild-derived specimen of *S. officinalis*, 'So3', was subjected to high-throughput sequencing on Illumina's platform. A total number of 77,254,984 sequencing reads with a length of 150 bp was obtained. This corresponds with a ~19-fold genome coverage, assuming a genome size of 475.30 Mb (Maksimović et al., 2007). Contigs of putative TPS genes were compiled by mapping the NGS reads against the three TPS genes reported for *S. officinalis* (AF051899 to AF051901 (Wise et al., 1998)). To improve assembly continuity, iterative extensions and merging of contigs were performed.

Orthologues were identified by carefully separating the assemblies into the individual TPS genes based on overlaps of mutations found in the subsequent assemblies of reads. Products amplified with TPS gene-specific sets of primers were verified by Sanger sequencing.

In total, twelve TPSs were determined for *S. officinalis* (Table 1). Inference of orthology performed for the twelve genes in the other two sage species, yielded two or three more members of the TPS gene family, termed *TPS15* and *TPS16* (*S. pomifera*) and *TPS15*, *TPS16* and *TPS17* (*S. fruticosa*). Of the other TPSs only *TPS13* was found to be missing in the genome of *S. pomifera*. All candidates were classified as members of the TPS subfamily b (Bohlmann et al., 1998).

3.2. Phylogenetic analysis

The phylogenetic tree resulted in three major clades (A-C, Fig. 1). **Clade A** consisted of nine TPS genes (*TPS1*, *TPS2*, *TPS5*, *TPS6*, *TPS7*, *TPS9*, *TPS15*, *TPS16* and *TPS17*) with *TPS1* and *TPS2* (almost) identical to published terpene synthases with proven activity (*Salvia officinalis* (+)-bornyl diphosphate synthase, Genbank accession no. AF051900 and *Salvia officinalis* (+)-sabinene synthase, Genbank accession no. AF051901 (Wise et al., 1998)). **Clade B** is the smallest group with *TPS11* and *TPS12* and *Rosmarinus officinalis* limonene synthase (DQ421800). **Clade C** consists of clearly two subclades, containing either *TPS3* and *TPS13* or *TPS4* and *TPS14*. *TPS3* is almost identical to *Salvia officinalis* 1, 8-cineole synthase (Wise et al., 1998). A cineole synthase from *Rosmarinus officinalis* (DQ839411) is an additional member of this clade.

3.3. Exon-intron structure, sub-cellular localisation prediction and structural elements of TPS genes

All TPS genes were structured into the common seven exons except of *TPS4* exhibiting only five exons (Fig. 2). Exon length was well conserved as indicated by low coefficients of length variation (*CV*) ranging from 0.7% (exon 4) to 3.3% (exon 2). Only the length of exon 5 of *TPS4* deviated considerably (825 bp instead of 139/140 bp) caused by the loss of introns 5 and 6. In contrast to exons, length and sequence homology of introns were highly variable (*CV*s of 87% (intron 5) to 152% (intron 4) for length variation).

Intron phasing of the six TPS introns was generally highly conserved (pattern: 0, 1, 2, 2, 0, 0) except of phasing in *TPS4* that lacks introns 5 and 6 (pattern: 0, 1, 2, 2)' and in *TPS12* – the gene with the longest introns (pattern: 0, 1, 2, 0, 1, 1).

All TPS genes encoded an N-terminal chloroplast-targeting peptide apart from *TPS14*, and contained the C-terminal motifs DDXXD (X. any amino acid), and NSE/DTE as well as the motif RR(X)8W of exon 1 (Degenhardt et al., 2009). Only in *TPS14* of *S. officinalis*, DDXXD was changed to NDXXD.

We refrained from analysing the promoters of the TPS genes since sequence information on the region upstream of the start codon was limited to only 100 to 800 bp except for two longer sequences.

Table 1
Upper part: Copy numbers of TPSs counted by dPCR in individuals of three sage species and exemplarily in artificial DNA and calibrated by the single-copy gene PEX4. TPSs deviating in copy number from one are in bold. Lower part: Relative gene expression of TPS cDNAs at full bloom in leaves, calyces and corollas of the three species. To indicate within species variation, leaf expression data of two *S. officinalis* genotypes (So11 and So26) were added.

Species (genotype)	Tissue	*TPS1*	*TPS2*	*TPS3*	*TPS4*	*TPS5*	*TPS6*	*TPS7*	*TPS9*	*TPS11*	*TPS12*[*)]	*TPS13*	*TPS14*	*TPS15*	*TPS16*	*TPS17*
TPS copy number variation																
artificial DNA		0.95	-	1.14	-	1.8	0.98	1.27	1.04	0.93	-	1.17	1.16	-	-	-
S. officinalis (So3)	leaf	1.12	1.18	**2.64**	0.78	1.69	0.97	1.03	1.04	1.11	1.03	0.97	0.76	**0**	**0**	**0**
S. fruticosa (Sf1)	leaf	1.20	0.99	1.12	1.02	1.69	1.14	1.03	1.01	0.97	**3.43**	0.83	1.04	1.13	0.90	1.46
S. pomifera (Sp2)	leaf	1.24	**2.38**	1.32	1.01	2.32	**0.51**	1.20	0.97	1.14	1.16	**0**	0.91	0.96	0.71	**0**
Relative gene expression [%]																
S. officinalis (So11)	leaf	27.2	35.7	1.5	0.6	0.2	0.2	3.0	13.4	0.2	8.9	0.0	9.1	n.d.	n.d.	n.d.
S. officinalis (So26)	leaf	39.6	21.8	1.3	0.3	0.0	0.2	1.3	19.4	0.0	6.5	0.0	9.7	n.d.	n.d.	n.d.
S. officinalis (So3)	leaf	44.1	19.4	12.0	5.0	0.0	0.1	1.0	2.9	0.0	7.5	0.1	7.9	n.d.	n.d.	n.d.
S. officinalis (So3)	calyx	20.4	9.0	4.4	6.0	52.9	0.2	0.9	0.1	0.0	3.9	0.2	2.0	n.d.	n.d.	n.d.
S. officinalis (So3)	corolla	4.3	0.4	2.6	18.7	71.1	0.1	0.1	0.2	0.5	0.9	0.1	0.9	n.d.	n.d.	n.d.
S. fruticosa (Sf1)	leaf	0.8	0.6	68.1	1.3	0.3	0.1	0.4	0.9	0.0	25.0	0.0	1.6	0.9	0.0	0.2
S. fruticosa (Sf1)	calyx	0.4	0.0	60.4	1.9	0.1	0.0	0.2	0.0	0.0	35.3	0.0	1.6	0.0	0.1	0.0
S. fruticosa (Sf1)	corolla	26.6	0.0	23.0	27.5	0.0	0.1	0.2	0.0	0.4	7.3	3.4	9.9	0.0	0.0	1.7
S. pomifera (Sp2)	leaf	0.0	53.7	0.1	1.3	n.d.	0.0	13.2	19.8	0.4	1.9	0.0	7.2	0.0	2.4	n.d.
S. pomifera (Sp2)	calyx	0.1	50.1	0.1	4.0	n.d.	0.0	17.7	17.2	1.0	0.6	0.0	5.6	0.1	3.6	n.d.
S. pomifera (Sp2)	corolla	0.0	2.5	0.0	27.4	n.d.	0.0	0.1	0.6	0.1	0.2	0.1	43.3	0.0	25.6	n.d.

*putative pseudogene in *S. officinalis* (premature stop codon).
n.d.: values not determined due to the cross-reactivity with *TPS16*.

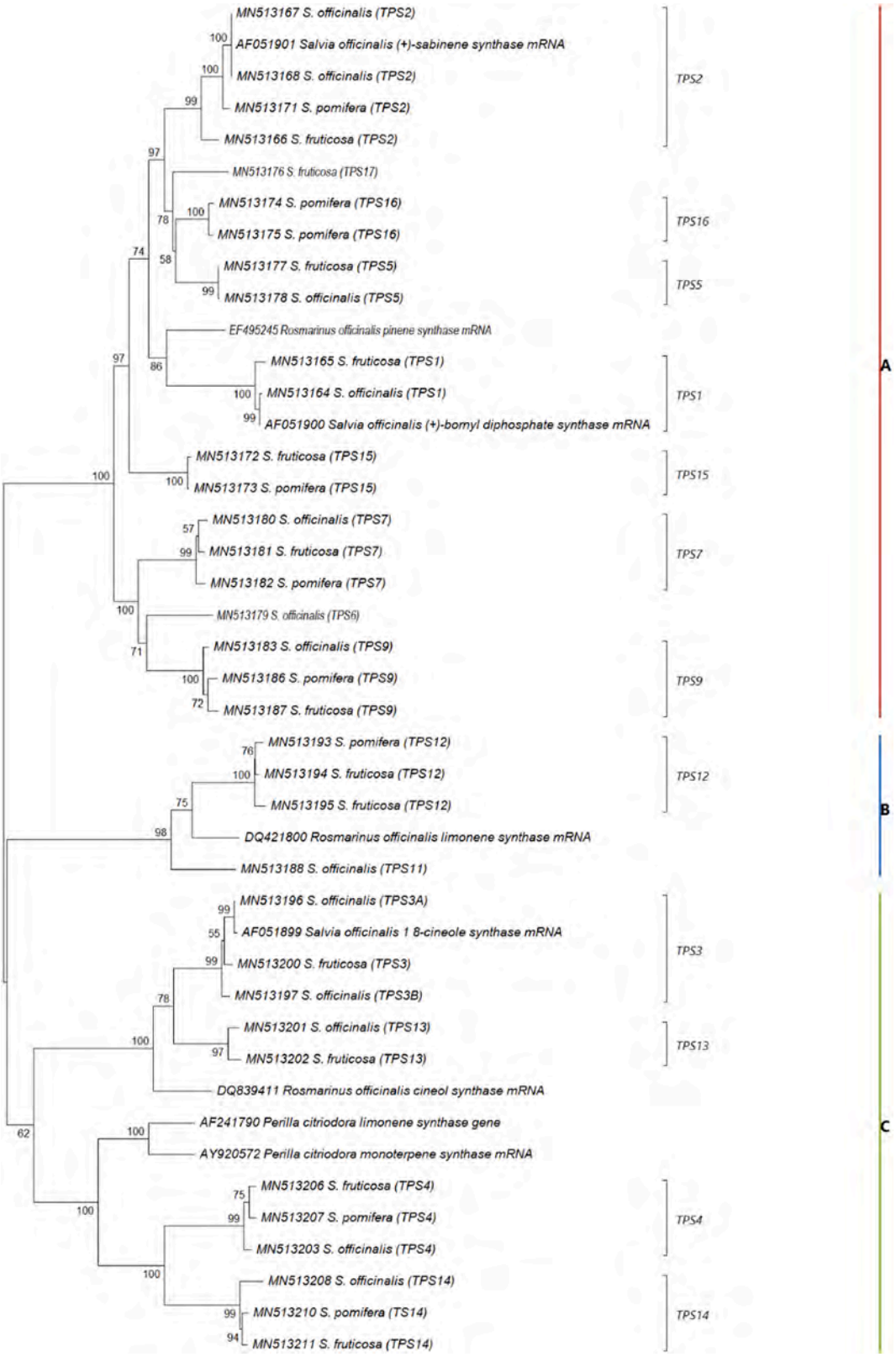

Fig. 1. Relationship of TPS mRNAs. The tree was computed using the Maximum Likelihood method under the General Time Reversible model. The tree with the highest log likelihood is shown. The percentage of trees >50%, in which the respective samples cluster together is shown. TPS1 is identical to (+)-bornyl diphosphate synthase, TPS 2 identical to (+)-sabinene synthase and TPS3 identical to 1,8-cineole synthase (Wise et al., 1998).

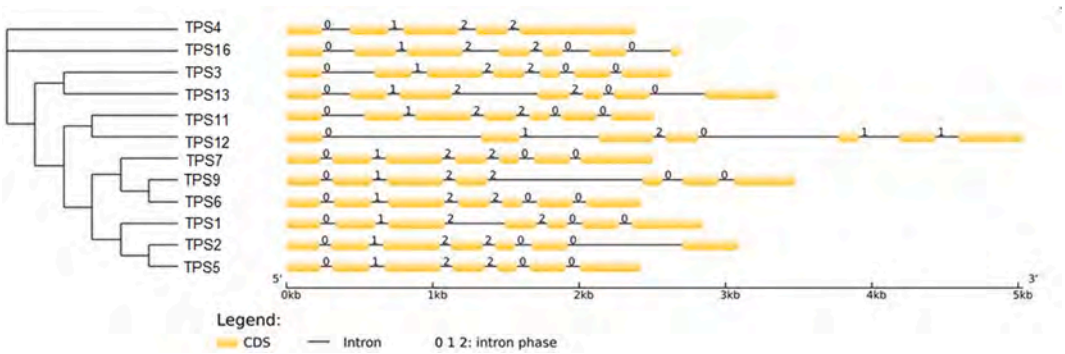

Fig. 2. Phylogenetic relationship, exon-intron structure and intron phasing of the TPSs of *S. officinalis*. Numbers: types of intron phasing (phase-0 intron does not disrupt a codon, phase-1 intron disrupts a codon between first and second base, phase-2 intron disrupts a codon between second and last base). Note that exon 7 of some TPSs is truncated.

3.4. Identification of a single-copy gene and specificity assessment of TPS assays using dPCR

dPCR, the gold-standard technique for DNA- or cDNA-sequence counting (Collier et al., 2017), was used to identify a gene present only once per haploid genome of *S. officinalis.* Four candidates were tested, namely *DEF* (Aagaard et al., 2005), *DXR* (Curto et al., 2012), *FIM* (Schultz et al., 2001) and *PEX4* (unpublished, GenBank: EU399687.1). While ~2.1 copies per haploid genome were measured for *DEF,* the single-copy gene status was confirmed for the other three genes. Out of the latter three, *PEX4* was subsequently used as calibrator to determine CNV of TPS genes.

Primers designed against the TPS genes were assessed for cross reactivity applying a ΔCq threshold of less than seven (Supplementary Data 1, Supplementary Table 6). Only some assays dropped below this line such as the *SfTPS15* assay tested against *TPS6*, *TPS7* and *TPS9* of *S. officinalis* (ΔCq values of 1.8, 4.2 and 4.1, respectively). Interestingly, cross-amplification testing of the *SoTPS5* assay against *TPS16* of *S. fruticosa* produced a ΔCq value of −6. This would indicate that the assay worked by far better in the cross-amplified gene. However, the case was not further considered since *S. officinalis* lacked *TPS16*.

To verify assay suitability and to exclude matrix effects, genomic copy number was analysed on synthetic DNA covering some TPS amplicons (Supplementary Data 1, Table 1). Only in a few cases, minor matrix effects were observed that did not affect data interpretation. In general, dPCR assays were appropriately designed as indicated by low deviation from the copy number counted for synthetic DNA constructs (variation around 1:0.93 to 1:1.26).

3.5. Copy number variation (CNV) using dPCR

CNV was higher across species than between the TPSs genes (Table 1). In all three species, *TPS1*, *TPS4*, *TPS7* to *TPS11* and *TPS14* were classified as single-copy genes. *TPS2* was single copy in *S. officinalis* and *S. fruticosa* and double copy in *S. pomifera*. *TPS3*, a single-copy gene of *S. fruticosa* and *S. pomifera*, exhibited a haploid genome copy number of 2.6 in *S. officinalis*. This might indicate a heterozygous genotype with two or three copies per haploid genome. Due to a high degree of sequence similarity of *TPS5* and *TPS8*, these two genes were analysed with a joint assay, termed *TPS5*. In this case, CNV analysis was more difficult. Therefore, the value of 1.7 measured for *S. officinalis* and *S. fruticosa* was assumed to represent a haploid-genome copy number of one for each gene. TPS5 and TPS8 of *S. pomifera* also were regarded single-copy genes, although the assay yielded a higher value (2.3 instead of 1.7) likely caused by cross reaction with TPS16 (Supplementary Table 6).

TPS6, a single-copy gene of *S. officinalis* and *S. fruticosa*, was found only in every second haploid genome set of *S. pomifera*. We considered this gene to be hemizygous in *S. pomifera* since NGS data did not indicate polymorphic primer- or probe-binding sites that might have caused allelic 'drop-out'.

TPS12, a single-copy gene in *S. officinalis* and *S. pomifera*, occurred in three copies in *S. fruticosa*. *TPS15* and *TPS16*, both single-copy genes in *S. fruticosa* and *S. pomifera*, were lacking in *S. officinalis*. *TPS17* was only present in *S. fruticosa* at the level of one or two copies.

3.6. Relative transcript expression of TPS genes

In all three species, TPS transcript expression was measured in leaf, calyx, and corolla of flowering plants. To assess intraspecific variation, leaves of two additional specimens of *S. officinalis* ('So11' and 'So26'), were analysed (Table 1). The monoterpene synthases compete for a common precursor. Therefore, for a phenotype to genotype comparison, relative gene expression was used as a better proxy for (relative) essential oil composition where each TPS or essential oil compound is reported relative to the sum of all TPSs or essential oil compounds of a sample, respectively.

Expression of *TPS1* to *TPS4* genes varied considerably in leaves of the three genotypes of *S. officinalis* (*TPS1*: 27–44%, *TPS2*: 19–36%, *TPS3*: 1.3–12%, and *TPS4*: 0.6–5%; Table 1). In contrast, the expression of the putative pseudogene *TPS12* was with ~8% quite stable. Variation of *TPS9* expression was low in 'So3' (2.9%), but moderate in the two other specimens ('So11': 13.4% and 'So26': 19.4%). *TPS5*, *TPS6*, *TPS7*, *TPS11* and *TPS13* had low or near-zero transcript levels.

The corolla of the three species was very diverse in expression. In *S. officinalis,* the calyx was intermediate between leaf and corolla. The corolla of *S. officinalis* showed an elevated expression of *TPS4* (19% instead of 5%) and much higher *TPS5* expression. In comparison, *TPS1* to *TPS3* were abundant at very low level. Calyx and leaf of *S. fruticosa* and *S. pomifera* were similar in TPS expression.

In the corolla, however, relative amounts of *TPS3* and *TPS12* were reduced to about one third, while the relative expression of *TPS1* increased to 27%.

In corolla of *S. fruticosa*, however, relative transcript abundance of *TPS3* or *TPS12* was reduced to about one third, while *TPS1* expression raised to 27%.

In *S. pomifera,* calyx and leaf showed high expression of *TPS2* (~50% overall) and a less, but still high expression of *TPS7*, *TPS9* and *TPS14*. In corolla of the species, *TPS14* was dominating (43%) followed by *TPS4* and *TPS16* (27% and 26%). *TPS2* was expressed only at a minor proportion (2.5%).

3.7. Monoterpene profiles

The composition of the main monoterpenes in the leaves of the sage species demonstrated the intermediary position of *S. officinalis* between

S. fruticosa, being very rich in 1,8-cineole, and *S. pomifera*, overly rich in thujones (Table 2). The calyx of the three species was characterised by a conspicuous increase in α- and β-pinene. While the calyx resembled the leaf in *S. officinalis* or *S. pomifera*, in *S. fruticosa* it took an intermediary position between the compositions of the leaf and the corolla. The leaf products corresponding to the three TPSs (BS, CS, SS) accounted for 94% (*S. officinalis*) and 95% (*S. pomifera*) of the total of all monoterpenes, while it composed only 65% in *S. fruticosa*, were additional monoterpenes (α- and β-pinene, myrcene) reached higher levels.

3.8. Relationship between transcript expression and monoterpene composition

According to homology of amino-acid sequence, *TPS1* encodes a (+)-bornyl diphosphate synthase, *TPS2* a (+)-sabinene synthase and *TPS3* a 1,8-cineole synthase. Calculating a principal component analysis (PCA) with the relative transcript expression and relative monoterpene composition together showed clear correlations in variables-PCA between the main monoterpenes or monoterpenes grouped according to Figs. 3 and 4. The first three axes of the PCA explained 73% of the variability at almost equal shares. The first axis distinguished the thujones from 1,8-cineole, the second axis β-pinene and the third axis the group of bornyl compounds from all other compounds. The presumed functions of TPSs were indicated by the clear co-expression of *TPS1* with the sum of bornyl compounds, of *TPS2* with the sum of sabinene compounds and of *TPS3* with 1,8-cineole. Together with *TPS2* and the thujones, three TPS genes exhibited co-expression (*TPS7* and *TPS9* and to a lesser degree also *TPS11*). Together with *TPS3* and 1,8-cineole, two TPS genes (*TPS12* and *TPS15*) and together with *TPS1* and bornyl compounds, another two TPS genes (*TPS5* and *TPS6*) were co-expressed. With β-pinene, a high correlation with *TPS4*, *TPS13*, *TPS14*, *TPS16* and *TPS17* was found.

Finally, transcript expression of *TPS1*, *TPS2* and *TPS3* was correlated to the respective end product(s) (Fig. 3). All three regressions were significant with regression coefficients of ≥0.92 indicating the transcriptional control of the mixture of monoterpenes. While the slope of the regression lines for *TPS1* and *TPS3* were quite similar (0.88 and 0.71, respectively), the slope of *TPS2* was much higher (1.6). In *S. officinalis* and *S. pomifera*, a very high proportion of monoterpene composition is explained by transcript abundance of three genes (95%). In *S. fruticosa*, however, they explain only 65% of the composition.

4. Discussion

Some *Salvia* species including the close relatives *S. officinalis*, *S. fruticosa* and *S. pomifera* produce a complex essential oil stored in specialised epidermal oil glands. The oil mixture mainly results from an accumulation process of numerous monoterpenes that are formed in parallel by different TPSs. Their relative proportions often described as 'chemotypes, are distinct across these species, but also distinct within a particular species.

In this work, we identified all TPSs including exon-intron structures, genomic copy numbers, quantified their transcript expression in leaves and flowers and the resulting composition of monoterpenes. To gain a first insight into intraspecific variation of the commercially most important species, three genotypes of *S. officinalis* were analysed.

First, TPS genes were identified in the genomic NGS data of *S. officinalis* and confirmed by the orthologues found in the genomes of *S. fruticosa* and *S. pomifera*. In addition, this search yielded three more TPS genes. We are confident to have identified most of the TPS genes of *S. officinalis*. However, this cannot be assured for the two other species lacking information on the genome sequence. Although closely related, *S. fruticosa* has a ~70% larger genome compared to *S. officinalis* (Maksimović et al., 2007; Bou Dagher-Kharrat et al., 2013). It is speculative to assume that this 1.7-fold larger genome, would harbour more TPSs in addition to the three sequences identified in this study. Currently there is no information on the genome size of *S. pomifera*.

The twelve TPS genes identified for *S. officinalis* match the range of TPS numbers reported for other genomes (six: *Arabidopsis thaliana* (Aubourg et al., 2002), eight: *Cannabis sativa* (Booth et al., 2017), 16: *Ocimum sanctum* (Kumar et al., 2018), 6: *Cinnamomum burmannii* (Ma et al., 2022), 7: *Pinus nigra* subsp. *laricio* (Alicandri et al., 2022), 19: *Vitis vinifera* (Martin et al., 2010)). A much higher number of 36 TPS genes was reported for *Eucalyptus grandis* (Külheim et al., 2015).

Transcript sequences of already characterised (+)-sabinene synthase, (+)-bornyl diphosphate synthase and 1,8-cineole synthase from *Salvia officinalis* (Wise et al., 1998) are so close to *TPS2*, *TPS1* and *TPS3*, respectively. In this case, enzymatic activity might be deduced based on sequence similarity alone. Limonene synthase of *Rosmarinus officinalis*, however, is too distantly related to predict limonene as end-product of *TPS11* or *TPS12*.

We included in the phylogenetic analysis also closely related TPSs of other genera. There is similarity between some TPSs of *Rosmarinus officinalis* and our candidates. This is not surprising because only recently, the genus *Rosmarinus* was included into the genus *Salvia* with *R. officinalis* renamed to *S. rosmarinus* Spenn. (Drew et al., 2017). Far more interesting is the grouping of some *Perilla citriodora* TPS genes in cluster C, because the genus *Perilla* is not closely related to *Salvia*. On the contrary, *Salvia* and *Perilla* are assigned to the same subfamily (Nepetoideae) but different tribes, Mentheae and Elsholtzieae, respectively. The diversification within the Nepetoideae is estimated to 57 to 52 million years (Ma) (Drew and Sytsma, 2012) or ~ 63.4 Ma (Li et al., 2017). Therefore, the origin of (at least) the clusters A and C must be older than the diversification of the Nepetoideae.

Table 2
Proportion of the most important monoterpenes at full bloom (in %).

Genotype	Tissue	α-pinene	β-pinene	myrcene	1,8-cineole	sabinene	*trans*-sabinene hydrate	α-thujone	β-thujone	s_sab	camphor	borneol	bornyl acetate	*s_born*
So11	leaf	0.3	1	0.4	**12.3**	0.6	0.5	0.9	55.3	**57.3**	19.1	1.6	2.4	***23.1***
So26	leaf	0.5	0.8	0.2	**16.7**	0.4	0.4	0.6	32.3	**33.7**	36.8	2.3	3.4	***42.5***
So3	leaf	0.2	0.5	0.1	**16.0**	0.1	0.4	30.0	2.3	**32.8**	46.7	1.8	0.4	***48.8***
So3	calyx	0.3	1.3	0.2	**31.2**	0.3	0.8	18.7	1.5	**21.2**	34.5	6.6	1.4	***42.5***
So3	corolla	2.0	14.2	1.2	**20.3**	0.0	1.7	7.8	0.5	**10.0**	34.7	10.0	0.0	***44.7***
Sf1	leaf	3.9	6.1	13.6	**52.4**	1.3	1.0	0.4	0.5	**3.2**	6.4	2.0	1.4	***9.8***
Sf1	calyx	8.0	13.5	2.0	**41.7**	1.6	1.1	0.7	0.6	**4.0**	10.1	4.5	0.5	***15.1***
Sf1	corolla	7.3	42.4	1.3	**9.3**	4.0	0.0	0.0	0.0	**4.0**	15.5	2.1	5.0	***22.6***
Sp2	leaf	0.3	0.4	0.8	**0.5**	1.2	0.1	0.3	86.3	**87.9**	2.2	0.3	4.0	***6.5***
Sp2	calyx	0.3	2.0	0.6	**0.3**	1.4	0.3	0.2	88.5	**90.4**	0.1	0.7	2.3	***3.2***
Sp2	corolla	3.5	47.1	1.5	**0.8**	3.8	0.7	0.2	34.5	**39.1**	0.0	0.6	1.1	***1.7***

Monoterpenes were grouped by pathways. So: *S. officinalis*, Sf: *S. fruticosa*, Sp: S. *pomifera*. The number after the species abbreviation indicates the genotype. s_sab … sum of components of the sabinene pathway (sabinene + trans-sabinene hydrate + alpha-thujone + beta-thujone); s_born … sum of the components of the bornyl pathway (camphor + borneol + bornyl acetate).

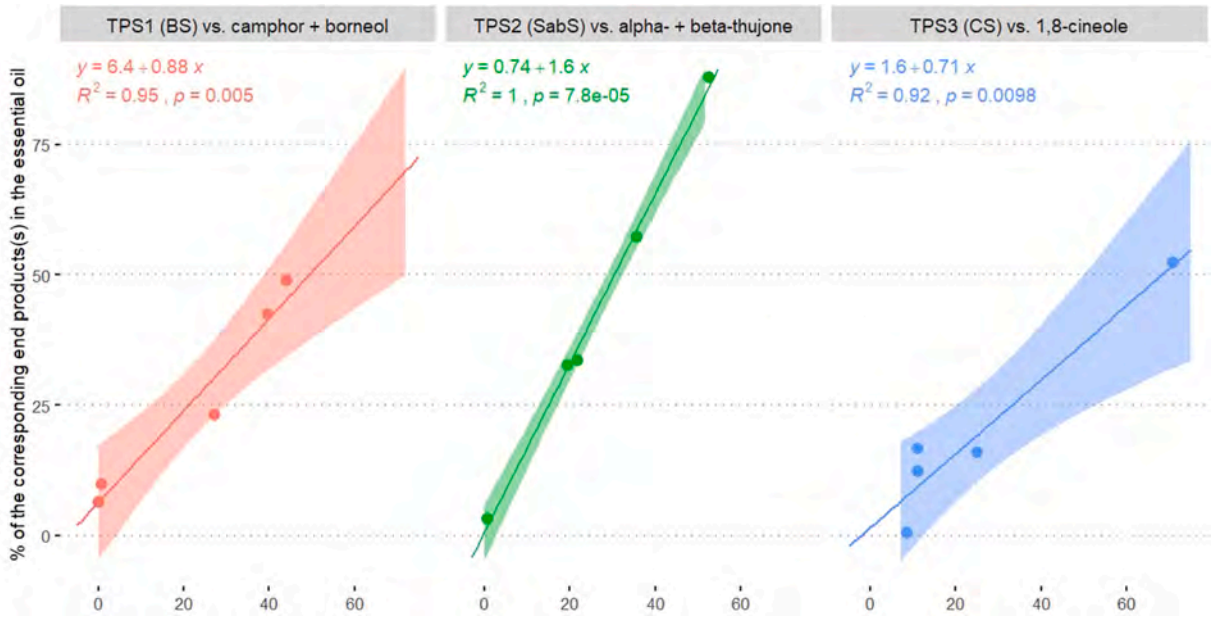

Fig. 3. Regression of gene expression relative to *PEX4* of TPSs with known function (x-axis) to their respective end product(s) (y-axis).

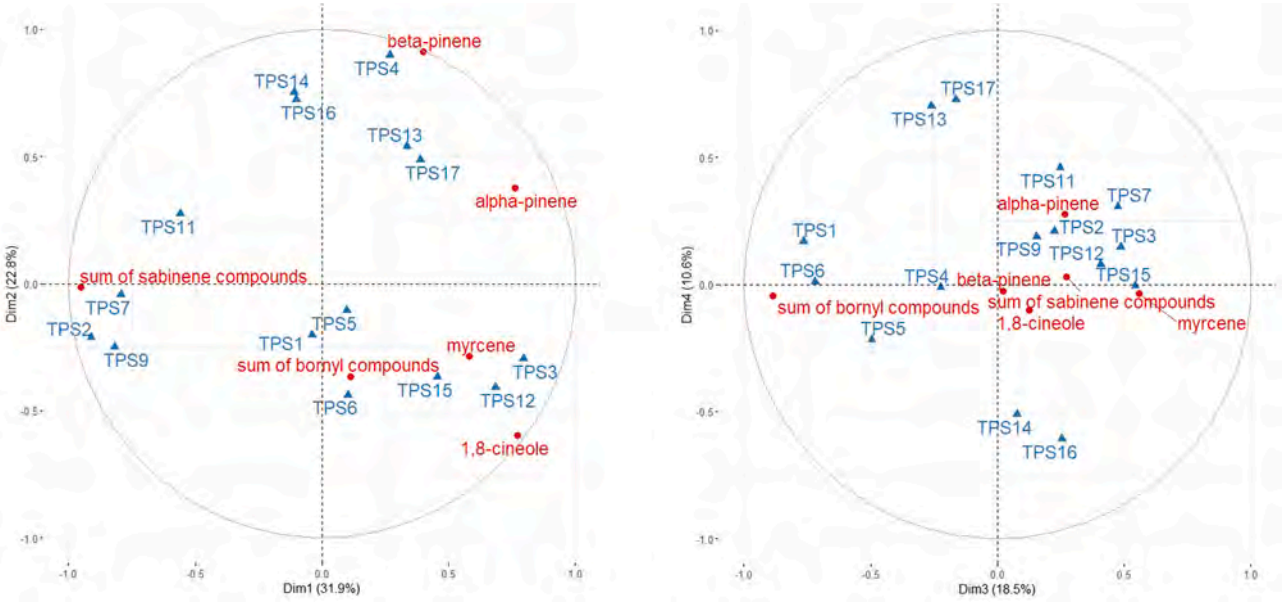

Fig. 4. Principal component analysis of the relative transcript expression of putative TPSs and the relative amounts of the main monoterpenes (left panel: dimensions 1 and 2, right panel: dimensions 3 and 4; 'sum of bornyl compounds … camphor + borneol; 'sum of sabinene compounds … sabinene + α-thujone + β-thujone).

Positions and phasings of the six introns were much conserved. For example, intron phasing was the same as in *A. thaliana* (Aubourg et al., 2002). However, intron length differed between but also within genes. One TPS gene, *TPS4*, was exceptional by losing two introns at the 3' terminus. In general, intron changes of TPS genes occur only occasionally, as an intron loss in *A. thaliana* (Aubourg et al., 2002) or an intron gain in *Vitis vinifera* (Martin et al., 2010). Intron loss is explained as a recombination event involving reverse transcribed copies of spliced mRNAs and by genomic deletions (Roy and Gilbert, 2006).

All but one of the TPS sequences determined for the three sage species exhibited the three conserved functional motifs. The only deviation was *TPS14* of *S. officinalis* where NDxxD had replaced the canonical DDxxD motif also present in *TPS14* of the related species *S. fruticosa* and *S. pomifera*. NDxxD is regarded as a neutral variant since it did not impair the catalytic activity of (+)-germacrene D synthase from *Solidago* (Prosser et al., 2004). *TPS14*, however, will not produce any monoterpene since it was the only TPS lacking a plastid targeting sequence.

CNV measurement based on dPCR was first validated on artificial DNA that contained the targets of several TPS genes and the reference sequences on one strand. This allowed to assess the reliability of each assay and subsequently the impact of the plant matrix on the performance of dPCR. Deviations from integers calculated for artificial DNA such as the 1.14 copies of *TPS3*, may be a result of unequal efficiencies or interactions of the two combined PCR reactions. Larger deviations from even multiples of the value measured for the synthetic DNA template, such as the *TPS3* copy number of *S. officinalis* (2.6 copies), are possibly due to the plant matrix.

TPS genes can be subject to CNV as reported for cultivars of rice (Yu et al., 2013). Here, we demonstrated CNV of TPS genes in the three model species of *Salvia* that ranged from one to only a few copies. While

the members of the TPS family varied little in gene-copy number, they are likely different in function. The few cases of recent CNV, gene duplication or loss, found across our sage species might also have contributed to diversify the flavour profile. The extremely thujone-rich *S. pomifera* exemplifies the influence of TPS-gene multiplication on the level of transcript expression and the monoterpene profile, namely duplication of *TPS2* encoding a TPS that forms the thujone precursor (+)-sabinene. Similarly, elevated gene dosage in combination with high transcript expression was also observed for *TPS12* of *S. fruticosa* present at three copies per haploid genome.

Earlier, very distinct monoterpene profiles were detected in different tissue types of *Salvia lavandulifolia* (Schmiderer et al., 2008). This inspired us to compare intra-individual transcript expression in leaf, calyx, and corolla. We found very diverse expression patterns such as a gradual change from leaf to corolla and completely different expression levels between corolla and leaf or calyx. It is not astonishing that the highly diverse transcription factors of leaf, calyx and corolla can also affect the expression of TPSs (Xu et al., 2019).

It is likely that the TPSs which were completely or almost identical to the functionally proven *S. officinalis* TPSs of Wise et al. (1998) produce the respective compounds in our analysed individuals. Deduction of functionality based on the correlations of our PCA analysis (Fig. 4) would be rather tempting, but not justified, but not justified considering that only similar expression is indicated, but not mandatorily similarity of function. Sometimes, despite different clade position in the phylogenetic tree, hence less overall sequence homology, some TPSs were strongly co-expressed. Therefore, analysis of sequence and function of TPS promoters should be an issue of further investigation.

In general, transcript expression alone or in combination with CNV can serve as a proxy for the monoterpene composition of the three sage species.

Contribution

Corinna Schmiderer (CS) was involved in most of the experiments (molecular biology and phytochemistry) and wrote the manuscript together with JN.

Ralf Steinborn (RS) planned and optimised digital and quantitative PCR experiments and corrected the manuscript.

Johannes Novak (JN) planned the project, performed bioinformatics and statistical analysis, and co-wrote the manuscript with CS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Alicja Czop, Marion Hacek, Tin Jurhar, Anja Panhuber and Martin Piontek for technical assistance and Martin Hofer for support of dPCR quantification. This work was funded in part by the Austrian Science Fund (FWF, grant number P 28425-B22 to JN). This study was supported by resources of the Genomics Core Facility (VetCore, University of Veterinary Medicine, Vienna).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2023.01.034.

References

Aagaard, J.E., Olmstead, R.G., Willis, J.H., Phillips, P.C., 2005. Duplication of floral regulatory genes in the Lamiales. Am. J. Bot. 92, 1284–1293.
Adams, R.P., 2007. Identification of Essential Oil Components by Gas Chromatography/mass Spectrometry, fourth ed. Allured Business Media, Carol Stream, Ill., p. 804
Alicandri, E., Covino, S., Sebastiani, B., Paolacci, A.R., Badiani, M., Sorgonà, A., Ciaffi, M., 2022. Monoterpene Synthase Genes and Monoterpene Profiles in *Pinus Nigra* Subsp, vol. 11. *laricio*. Plants, Basel, Switzerland.
Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., Heijne, G. von, Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance. 2.
Aubourg, S., Lecharny, A., Bohlmann, J., 2002. Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Mol. Genet. Genom.: Mol. Gen. Genet. 267, 730–745.
Baser, K.H., Özek, T., Kirimer, N., Tümen, G., 1993. The essential oil of *Salvia pomifera* L. J. Essent. Oil Res. 5, 347–348.
Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc. Natl. Acad. Sci. U.S.A. 95, 4126–4133.
Booth, J.K., Page, J.E., Bohlmann, J., 2017. Terpene synthases from *Cannabis sativa*. PLoS One 12, e0173911.
Bou Dagher-Kharrat, M., Abdel-Samad, N., Douaihy, B., Bourge, M., Fridlender, A., Siljak-Yakovlev, S., Brown, S.C., 2013. Nuclear DNA C-values for biodiversity screening: case of the Lebanese flora. In: Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology, 147, pp. 1228–1237.
Brody, J.R., Kern, S.E., 2004. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. Biotechniques 36, 214–216.
Burdock, G.A., 2009. Fenaroli's Handbook of Flavor Ingredients. CRC Press, Boca Raton.
Bushnell, B., Rood, J., Singer, E., 2017. BBMerge - accurate paired shotgun read merging via overlap. PLoS One 12, e0185056.
Collier, R., Dasgupta, K., Xing, Y.-P., Hernandez, B.T., Shao, M., Rohozinski, D., Kovak, E., Lin, J., Oliveira, M.L.P. de, Stover, E., McCue, K.F., Harmon, F.G., Blechl, A., Thomson, J.G., Thilmony, R., 2017. Accurate measurement of transgene copy number in crop plants using droplet digital PCR. Plant J. : Cell. Mol. Biol. (Paris, Fr., Online) 90, 1014–1025.
Croteau, R., Karp, F., 1979. Biosynthesis of monoterpenes: preliminary characterization of bornyl pyrophosphate synthetase from sage (*Salvia officinalis*) and demonstration that geranyl pyrophosphate is the preferred substrate for cyclization. Arch. Biochem. Biophys. 198, 512–522.
Curto, M.A., Puppo, P., Ferreira, D., Nogueira, M., Meimberg, H., 2012. Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Mol. Phylogenet. Evol. 63, 758–767.
Cvetkovikj, I., Stefkov, G., Acevska, J., Stanoeva, J.P., Karapandzova, M., Stefova, M., Dimitrovska, A., Kulevanova, S., 2013. Polyphenolic characterization and chromatographic methods for fast assessment of culinary *Salvia* species from South East Europe. J. Chromatogr. 1282, 38–45.
Degenhardt, J., Köllner, T.G., Gershenzon, J., 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70, 1621–1637.
Dehal, S.S., Croteau, R., 1987. Metabolism of monoterpenes: specificity of the dehydrogenases responsible for the biosynthesis of camphor, 3-thujone, and 3-lsothujone. Arch. Biochem. Biophys. 258, 287–291.
Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.
Drew, B., González-Gallegos, J.G., Xiang, C.-L., Kriebel, R., Drummond, C., Walker, J., Sytsma, K., 2017. *Salvia* united: the greatest good for the greatest number. Taxon 66, 133–145.
Drew, B.T., Sytsma, K.J., 2012. Phylogenetics, biogeography, and staminal evolution in the tribe Mentheae (Lamiaceae). Am. J. Bot. 99, 933–953.
Esra, M., Cetin, O., Kahraman, A., Celep, F., Dogan, M., 2011. A cytomorphological study in some taxa of the genus *Salvia* L. (Lamiaceae). Caryologia 64, 272–287.
Grausgruber-Gröger, S., Schmiderer, C., Steinborn, R., Novak, J., 2012. Seasonal influence on gene expression of monoterpene synthases in *Salvia officinalis* (Lamiaceae). J. Plant Physiol. 169, 353–359.
Hedge, I.C., 1972. Salvia. In: Tutin, T.G., Heywood, V.H., Burges, N.A., Valentine, D.H. (Eds.), Flora Europaea, vol. 3. Cambridge University Press, Cambridge.
Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., Gao, G., 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296–1297.
Karousou, R., Kokkini, S., 1997. Distribution and clinal variation of *Salvia fruticosa* mill. (Labiatae) on the island of crete (Greece). Willdenowia 27, 113–120.
Karp, F., Croteau, R., 1982. Evidence that sabinene is an essential precursor of C(3)-oxygenated thujane monoterpenes. Arch. Biochem. Biophys. 216, 616–624.
Kassambara, A., Mundt, F., 2016. Factoextra : Extract and Visualize the Results of Multivariate Data Analyses: R Package.
Külheim, C., Padovan, A., Hefer, C., Krause, S.T., Köllner, T.G., Myburg, A.A., Degenhardt, J., Foley, W.J., 2015. The *Eucalyptus* terpene synthase gene family. BMC Genom. 16, 450.
Kumar, Y., Khan, F., Rastogi, S., Shasany, A.K., 2018. Genome-wide detection of terpene synthase genes in holy basil (*Ocimum sanctum* L.). PLoS One 13, e0207097.
Lamien-Meda, A., Nell, M., Lohwasser, U., Börner, A., Franz, C., Novak, J., 2010a. Investigation of antioxidant and rosmarinic acid variation in the sage collection of the genebank in gatersleben. J. Agric. Food Chem. 58, 3813–3819.
Lamien-Meda, A., Schmiderer, C., Lohwasser, U., Börner, A., Franz, C., Novak, J., 2010b. Variability of the essential oil composition in the sage collection of the Genebank Gatersleben: a new viridiflorol chemotype. Flavour Fragrance J. 25, 75–82.

Länger, R., Mechtler, C., Jurenitsch, J., 1996. Composition of the essential oils of commercial samples of *Salvia officinalis* L. and *S. fruticosa* Miller: a comparison of oils obtained by extraction and steam distillation. Phytochem. Anal. 7, 289–293.
Lê, S., Josse, J., Husson, F., 2008. FactoMineR : an R package for multivariate analysis. J. Stat. Software 25.
Li, P., Qi, Z.-C., Liu, L.-X., Ohi-Toma, T., Lee, J., Hsieh, T.-H., Fu, C.-X., Cameron, K.M., Qiu, Y.-X., 2017. Molecular phylogenetics and biogeography of the mint tribe Elsholtzieae (Nepetoideae, Lamiaceae), with an emphasis on its diversification in East Asia. Sci. Rep. 7, 2057.
Löve, A., 1980. Chromosome number reports LXVII. Taxon 29, 347–367.
Ma, Q., Ma, R., Su, P., Jin, B., Guo, J., Tang, J., Chen, T., Zeng, W., Lai, C., Ling, F., Yao, Y., Cui, G., Huang, L., 2022. Elucidation of the essential oil biosynthetic pathways in *Cinnamomum burmannii* through identification of six terpene synthases. Plant Sci. : Int. J. Experiment. Plant Biol. 317, 111203.
Maksimović, M., Vidic, D., Miloš, M., Edita Šolić, M., Abadžić, S., Siljak-Yakovlev, S., 2007. Effect of the environmental conditions on essential oil profile in two Dinaric *Salvia* species: *S. brachyodon* Vandas and *S. officinalis* L. Biochem. Systemat. Ecol. 35, 473–478.
Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T., Bohlmann, J., 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol. 10, 226.
McLafferty, F.W., Stauffer, D.B., 1989. The Wiley/NBS Registry of Mass Spectral Data. Wiley, New York.
Nei, M., 2000. Molecular Evolution and Phylogenetics. Oxford University Press Incorporated, Cary, p. 348.
Perry, N.B., Anderson, R.E., Brennan, N.J., Douglas, M.H., Heaney, A.J., McGimpsey, J. A., Smallfield, B.M., 1999. Essential oils from Dalmatian sage (*Salvia officinalis* L.): variations among individuals, plant parts, seasons, and sites. J. Agric. Food Chem. 47, 2048–2054.
Pizzale, L., Bortolomeazzi, R., Vichi, S., Überegger, E., Conte, L.S., 2002. Antioxidant activity of sage (*Salvia officinalis* and *S. fruticosa*) and oregano (*Origanum onites* and *O. indercedens*) extracts related to their phenolic compound content. J. Sci. Food Agric. 82, 1645–1651.
Prosser, I., Altug, I.G., Phillips, A.L., König, W.A., Bouwmeester, H.J., Beale, M.H., 2004. Enantiospecific (+)- and (-)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif. Arch. Biochem. Biophys. 432, 136–144.
Putievsky, E., Ravid, U., Diwan-Rinzler, N., Zohary, D., 1990. Genetic affinities and essential oil composition of *Salvia officinalis* L., *S. fruticosa* Mill., *S. tomentosa* Mill. and their hybrids. Flavour Fragrance J. 5, 121–123.
R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Roy, S.W., Gilbert, W., 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat. Rev. Genet. 7, 211–221.
RStudio Team, 2019. RStudio. Integrated Development for R. RStudio, PBC, Boston, MA.
Schmiderer, C., Grassi, P., Novak, J., Franz, C., 2008. Diversity of essential oil glands of Spanish sage (*Salvia lavandulifolia* vahl, Lamiaceae). Nat. Prod. Commun. 3, 1934578X0800300.
Schmiderer, C., Grausgruber-Gröger, S., Grassi, P., Steinborn, R., Novak, J., 2010. Influence of gibberellin and daminozide on the expression of terpene synthases and on monoterpenes in common sage (*Salvia officinalis*). J. Plant Physiol. 167, 779–786.
Schmiderer, C., Lukas, B., Novak, J., 2013a. Effect of different DNA extraction methods and DNA dilutions on the amplification success in the PCR of different medicinal and aromatic plants. Zeitschrift für Arznei- und Gewürzpflanzen 18, 65–72.
Schmiderer, C., Torres-Londoño, P., Novak, J., 2013b. Proof of geographical origin of Albanian sage by essential oil analysis. Biochem. Systemat. Ecol. 51, 70–77.
Schultz, E., Carpenter, R., Doyle, S., Coen, E., 2001. The gene fimbriata interacts non-cell autonomously with floral regulatory genes. Plant J. : Cell. Mol. Biol. (Paris, Fr., Online) 25, 499–507.
Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38, 3022–3027.
Wise, M.L., Savage, T.J., Katahira, E., Croteau, R., 1998. Monoterpene synthases from common sage (*Salvia officinalis*). cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. J. Biol. Chem. 273, 14891–14899.
Xu, Y., Zhu, C., Xu, C., Sun, J., Grierson, D., Zhang, B., Chen, K., 2019. Integration of metabolite profiling and transcriptome analysis reveals genes related to volatile terpenoid metabolism in finger citron (*C. medica* var. *sarcodactylis*). Molecules 24, 2564.
Yu, P., Wang, C.H., Xu, Q., Feng, Y., Yuan, X.P., Yu, H.Y., Wang, Y.P., Tang, S.X., Wei, X. H., 2013. Genome-wide copy number variations in *Oryza sativa* L. BMC Genomics 14, 1–12.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.