Pak. J. Bot., 47(4): 1347-1357, 2015.

THE FIRST INSIGHT INTO THE *SALVIA* (LAMIACEAE) GENOME VIA BAC LIBRARY CONSTRUCTION AND HIGH-THROUGHPUT SEQUENCING OF TARGET BAC CLONES

DA CHENG HAO[1*], SONIA VAUTRIN[2], CHI SONG[3], YING JIE ZHU[3], HELENE BERGES[2], CHAO SUN[3] AND SHI LIN CHEN[3,4*]

[1]*Biotechnology Institute, School of Environment, Dalian Jiaotong University, Dalian 116028, China;*
[2]*INRA-CNRGV, UR1258, 31326 Castanet Tolosan Cedex, France;*
[3]*Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China;*
[4]*Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[*]*Corresponding author's e-mail: hao@djtu.edu.cn, slchen@implad.ac.cn*

Abstract

Salvia is a representative genus of Lamiaceae, a eudicot family with significant species diversity and population adaptibility. One of the key goals of *Salvia* genomics research is to identify genes of adaptive significance. This information may help to improve the conservation of adaptive genetic variation and the management of medicinal plants to increase their health and productivity. Large-insert genomic libraries are a fundamental tool for achieving this purpose. We report herein the construction, characterization and screening of a gridded BAC library for *Salvia officinalis* (sage). The *S. officinalis* BAC library consists of 17,764 clones and the average insert size is 107 Kb, corresponding to ~3 haploid genome equivalents. Seventeen positive clones (average insert size 115 Kb) containing five terpene synthase (TPS) genes were screened out by PCR and 12 of them were subject to Illumina HiSeq 2000 sequencing, which yielded 28,097,480 90-bp raw reads (2.53 Gb). Scaffolds containing sabinene synthase (Sab), a Sab homolog, TPS3 (kaurene synthase-like 2), copalyl diphosphate synthase 2 and one cytochrome P450 gene were retrieved via *de novo* assembly and annotation, which also have flanking noncoding sequences, including predicted promoters and repeat sequences. Among 2,638 repeat sequences, there are 330 amplifiable microsatellites. This BAC library provides a new resource for Lamiaceae genomic studies, including microsatellite marker development, physical mapping, comparative genomics and genome sequencing. Characterization of positive clones provided insights into the structure of the *Salvia* genome. These sequences will be used in the assembly of a future genome sequence for *S. officinalis*.

Key words: *Salvia officinalis*, Gridded BAC library, High-throughput sequencing, Terpene synthase, Genome characterization, Microsatellite.

Introduction

The family Lamiaceae (mint family) has a cosmopolitan distribution. The enlarged Lamiaceae contains about 236 genera and 6,900 to 7,200 species, many of which contribute significantly to the endemic flora of their distribution area (Theodoridis *et al*., 2012). Plants of Lamiaceae are becoming ideal model species for ecological and evolutionary studies. For instance, phylogeny and historical biogeography of *Isodon* (Lamiaceae) suggest rapid species radiation in Southwest China and Miocene overland dispersal into Africa (Yu *et al*., 2014); chemical ecology of *Ocimum basilicum* indicates that derivatives of essential oil of this plant are used as antifeedants against larvae of the gypsy moth (Popović *et al*., 2013); *Scutellaria baicalensis* showed synergistic physiological ecology responses to drought and rewatering (Zhang *et al*., 2013); The regional demographic expansions of *Lagochilus* Bunge ex Bentham (Lamiaceae) in arid Northwest China reflects a major influence of geologic and climatic events on the evolution of species of *Lagochilus* (Meng & Zhang 2013). Lamiaceae plant populations have been used for both short-term and long-term ecological investigations. For instance, *Prunella vulgaris* is used in both standard greenhouse and field trial to test the ecotoxicological effects of herbicides on plant communities (Pfleeger *et al*., 2012); *Melissa officinalis* and other Lamiaceae species are used to explore biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria (Vokou *et al*., 2012); *Lavandula stoechas* is used to assess the intraspecific variability in fire-related regeneration traits (Moreira *et al*., 2012); *Salvia cyclostegia* is used to explore possible ecological functions of the lever-like stamens and the floral design (Zhang *et al*., 2011).

The largest Lamiaceae genera is *Salvia* (more than 900 species), followed by *Scutellaria* (360) and *Stachys* (300). *Salvia* consists of shrubs, herbaceous perennials, and annuals, which are adapted to various environments throughout the Old World and the New World, with three distinct regions of diversity: Central and South America (around 500 species), Central Asia and Mediterranean (250), and Eastern Asia (100; Li *et al*., 2013). *Salvia* species and population have unique value in ecological and evolutionary studies. For example, *Salvia lavandulifolia* was used to reveal that contrasting seasonal overlaps between primary and secondary growth are linked to wood anatomy in Mediterranean sub-shrubs (Camarero *et al*., 2013); *S. farinacae* was introduced in two populations of *Taraxacum officinale* to examine its effect on pollinators' foraging behavior on *Taraxacum* (Lázaro & Totland 2010); microsatellite markers of *S. officinalis* provide essential genomic tools for future studies of Lamiaceae populations (MER Primer Development Consortium 2010; Radosavljević *et al*., 2012). Since more molecular data are available for *Salvia* than for other Lamiaceae genera (Ma *et al*., 2012), it is argued that *Salvia* plants are prime candidates for further research on the complexities of plant-environment interaction with regard to molecular ecology and ecological genomics.

Terpenoids are the main components of Lamiaceae

essential oil (Crocoll *et al*., 2010), which showed antimicrobial and antifeedant activities and play a vital role in plant-environment interaction (Popović *et al*., 2013; Tabti *et al*., 2014). Volatile monoterpenoids produced by *Salvia leucophylla* exert allelopathic effects by inhibiting cell proliferation and DNA synthesis in the root apical meristem of *Brassica campestris* seedlings (Nishida *et al*., 2005). Variation of terpenoid composition of essential oil might have a significant potential to change the ecology, and provide the tactic for plants to survive in new niches in the distinct environment. The enormous diterpenoid chemodiversity of Lamiaceae also contributes to population fitness in the ever-changing terrestrial environments (Jassbi *et al*., 2006). However, the complete genome sequence of Lamiaceae species has not been determined and the gene content of the Lamiaceae genome is unknown. Additionally, although the terpene synthase genes of *Salvia* have been studied (Wise *et al*., 1998; Caniard *et al*., 2012), very few analyses have been conducted at the genomic level (Ma *et al*., 2012). Bacterial artificial chromosome (BAC) libraries are critical for identifying full-length genomic sequences, correlating genetic and physical maps, and comparative genomics. Sequencing large insert clones to completion is useful for characterizing specific genomic regions and closing gaps in whole genome sequencing projects. To obtain a more precise picture of the actual genes and their products, especially defence-related terpene synthases, encoded by *Salvia* genome, we constructed the first *Salvia* BAC library and evaluated the positive BAC clones by high-throughput next generation sequencing (NGS) and sequence annotation.

Materials and Methods

***Salvia officinalis* BAC library construction:** Seeds of *S. officinalis* (No. 51219) were obtained from Kew botanic garden, UK. Seeds were germinated in the glass house and eight-week old plantlets were used for high molecular weight (HMW) genomic DNA (gDNA) extraction. Genomic DNA was prepared from 20g of sampled leaf material, as described by Peterson *et al*. (2000) with the following modifications: (1) Sucrose based Extraction Buffer was 0.01 M Tris, 0.1 M KCl, 0.01 M EDTA pH 9.4, 500 mM sucrose, 4 mM spermidine, 1 mM spermine tetrahydrochloride, 0.1% w/v ascorbic acid, 0.25% w/v PVP 40,000, and 0.13% w/v sodium diethyldithiocarbamate, (2) lysis buffer was 1% w/v sodium lauryl sarcosine, 0.3 mg/ml proteinase K, 0.13% w/v sodium diethyldithiocarbamate, (3) after lysis of the nuclei, agarose plugs were pre-washed 1h in 0.5M EDTA pH9.1 at 50°C, 1h in 0.05M EDTA pH8 at 4°C, and then stored at 4°C.

Embedded DNA was partially digested with *Hind*III, size selected, eluted and ligated into pIndigoBAC-5 *Hind*III-Cloning Ready vector (Epicentre Biotechnologies, Madison, Wisconsin). After electroporation in *Escherichia coli* T1 resistant DH10B electrocompetent cells (Invitrogen, Carlsbad, California), transformants were incubated in SOC medium for one hour under agitation, and 50 µL of transformants were plated on LB agarose medium containing chloramphenicol (at 12.5 µg /mL), 80 mg/mL X-gal and 100 mg/mL IPTG for cell counting and insert size assessment. BAC DNA from randomly selected colonies was isolated using NucleoSpin96 fast purification kit (Macherey-Nagel, Düren, Germany) and digested with *Not*I. Digested BAC DNA was separated on 0.8% agarose gel by pulsed-field electrophoresis in 0.25× TBE buffer at 6 V/cm with a linear ramp pulse time of 5–15s. Insert sizes were estimated with reference to the MidRangeI PFG marker (New England Biolabs). Colony picking was carried out on the validated transformants using a robotic workstation QPix2 XT (Genetix, Molecular Devices). The resulting BAC library was named Sof-B-I and represents ~3-fold coverage of *S. officinalis* genome (46 plates, 17,764 BAC clones, and mean insert size of 107 Kb).

Pool production and BAC library screening: The BAC library was screened by PCR for five target genes, terpene synthase (TPS) genes encoding sabinene synthase (Sab), bornyl diphosphate synthase (Bor), copalyl diphosphate synthase (CPS) 2, kaurene synthase (KS), and diterpene synthase (TPS) 3. We used all available *Salvia* DNA sequence information to design PCR primers that are, to the best of current knowledge, specific for the target genes. Degenerate PCR primers (Table 1) were designed to amplify CPS2, KS and TPS3 fragments from *S. officinalis* gDNA. The obtained PCR products were transferred into pGEMT-easy TA cloning vector (Promega) and sequenced. BAC library screen primers of CPS2, KS and TPS3 were designed based on the sequencing results to obtain PCR amplicons of 120-144bp. BAC screen primers of Sab and Bor were designed from NCBI GenBank sequences AF051901 and AF051900 (Wise *et al*., 1998) respectively. All screen primers were tested by regular PCR using *S. officinalis* gDNA. To substantially decrease the number of PCR reactions requested for library screening, the 384 BAC clones of each plate of the library were pooled together. DNA amplification was then performed on the resulting 46 plate-pools using Genomiphi.v2 kit (GE Healthcare) (Dean *et al*., 2001), under supplier specifications. Amplification yield was checked on agarose gel, and samples were diluted 1:200 in ultra-pure sterile water.

BAC library screening was performed by real-time PCR on a LightCycler LC480 (Roche Diagnostics, Indianapolis, Indiana) with the following conditions: 5 min activation at 95°C followed by 45 cycles consisting of 2s at 95°C and 5s at 55°C. A single fluorescent read was taken immediately following the 5-s extension time. A melting cycle was performed following the acquisition cycles by constant temperature ramping from 65 to 95°C with data acquisition every 0.2°C. Ssofast EvaGreen Supermix (Bio-Rad Laboratories, Hercules, California) was used at a final 1× concentration and a low primer final concentration was chosen to prevent non-specific amplification (200 nM). After validation of optimal cycling conditions on *Salvia* gDNA control (52-62°C annealing temperature gradient tested for the five markers), 1µl of each diluted plate-pools DNA was used as template in a 10µl reaction volume. Positive samples were selected based on their melting temperature (+/- 0.5°C compared to gDNA control), and their sizes were validated on 2% agarose gel. For each highlighted positive plate, BAC clones from all columns and lines were pooled together (24 samples pooled for each line, 16 samples pooled for each column). A second screening step was carried out on 0.5µl of these 40 line/column-pools raw cultures in order to identify the coordinates of positive BAC clone. Finally all positive clones were validated individually on 1µl of diluted clone cultures (1:20).

Table 1 PCR primers used in this study.

Oligo name	Tm (ºC)	Sequence (5'-3')	Use
Offsab61F	61	GAA TTC CCT CCA CAA CTT TG	BAC library screen
Offsab180R	63.1	CGG ATT TGG TGA GGT TTT TC	BAC library screen
Offbor61F	64.3	CCT CCA CAA CTT GGA GAG GA	BAC library screen
Offbor180R	64.5	GAT CGT CGG ATT TGA TGA GC	BAC library screen
offTPS3F	63.3	GTTTGCATTTTTTGGAGWTG	Amplify TPS3-like sequence from *S. officinalis* gDNA
offTPS3F1	63.3	ATC GGA TTT GAG ATT GCG TT	BAC library screen
offTPS3R	59.7	TTT TAA GCT CCA AAT CCC TC	BAC library screen
offCPS2F	55.7	GCATCTCCAATRTACAATCT	Amplify CPS2-like sequence from *S. officinalis* gDNA
offCPS2R	56.9	GGA TAT CAC CCA TYT TTC TT	Amplify CPS2-like sequence from *S. officinalis* gDNA
offCPS2F2	56.5	GCA TCT CCA ATG TAC AAT CT	BAC library screen
offCPS2R1	56.9	GGA TAT CAC CCA TTT TTC TT	BAC library screen
offKSF	59.5	GACACCCTYCAAAGGTTGGG	Amplify KS-like sequence from *S. officinalis* gDNA
offKSR	59.9	ACTCGCAAAAGYCTAAATGCC	Amplify KS-like sequence from *S. officinalis* gDNA
offKSF1	63.8	GAATC AGAAG CCGAT GCTTC	BAC library screen
offKSR1	62.1	ACTCGCAAAAGTCTAAATGCC	BAC library screen

High-throughput sequencing of positive BAC clones, scaffold assembly and analysis: BAC DNA was extracted from 12 isolated positive clones and then fragmented randomly. The overhangs resulting from fragmentation were converted into blunt ends by using T4 DNA polymerase, Klenow fragment and T4 polynucleotide kinase. After adding an 'A' base to the 3' end of the blunt phosphorylated DNA fragments, adapters were ligated to the ends of the DNA fragments. After electrophoresis, DNA fragments of desired length were gel purified, selectively enriched and amplified by PCR to construct sequencing library. DNA cluster was subject to HiSeq 2000 (Illumina) paired end sequencing. After sequencing, the raw reads were filtered, including removing adapter sequences, reads with ≥5% unknown bases and low-quality reads from raw reads. The redundant reads from PCR duplication were also removed. The starting five bp were deleted from the raw reads. Contamination was removed by mapping the sequences of *E. coli* DH10B and pIndigoBAC-5 vector to the filtered reads with SOAP2 software (http://soap.genomics.org.cn/ soapaligner.html). Parameters are: -m 200 -x 800, with other default ones. Unpaired reads were removed and only read pairs were used for the following analysis.

HiTEC software (http://www.csd.uwo.ca/~ilie/HiTEC/) was used in the error correction of the non-contaminated reads. Parameters are: genomeLength 100,000, perBaseErrorRate 1. Reads after error correction were used for de novo sequence assembly. SOAPdenovo software (http://soap.genomics.org.cn/soapdenovo.html) parameters: -K 43 -R, max_rd_len=85, avg_ins=500, reverse_seq=0, rd_len_cutoff=85, rank=1, pair_num_cutoff=3, map_len=32.

Two methods were combined to annotate the repeat sequences: one is the homologous prediction based on RepBase18.06 (http://www.girinst.org/repbase), while MAKER software (Cantarel *et al*., 2008) and default parameters were used; the other is the *de novo* prediction based on sequence self-comparison in RepeatModeler1.0.7 (http://repeatmasker.org/RepeatModeler.html). BatchPrimer3 (You *et al*., 2008) was employed to design PCR primers in the flanking regions of the detected SSRs, setting a minimum product size of 100 bp, a minimum primer length of 18 bp, a minimum GC content of 30%, a melting temperature between 50 and 70 °C and a maximum melting temperature difference between primers of 8°C.

MAKER software was used for gene annotation. Parameters are: organism_type = eukaryotic, protein = uniprot_swissprot.fasta, augustus_species = arabidopsis. Protein homologous prediction with UniProt/Swiss-Prot database, as well as ab initio gene prediction with AUGUSTUS model and repeat sequence masking, was used in annotation pipeline. FancyGene (Rambaldi & Ciccarelli, 2009) and PlantCARE (Lescot *et al*., 2002) were used to visualize gene structure and predict *cis*-acting regulatory elements respectively. InterProScan (v5, http://www.ebi.ac.uk/interpro/download.html), KEGG online annotation tool (http://www.genome.jp/tools/kaas/) and blast2go (Conesa & Götz 2008) v2.5 were used in gene functional annotation. Evolutionary analyses and phylogenetic tree reconstruction were conducted in MEGA6 (Tamura *et al*., 2013).

Results

BAC library construction: Three different batches of ligations were validated based on their transformation efficiency. BAC clones were obtained for three DNA fractions, i.e., A, B and C, which were from the same nuclei extraction and DNA sizing steps. Fraction A has ~7100 unique clones, and the mean insert size was estimated to be 92 kb (1.09× coverage based on haploid genome size; Fig. 1); fraction B has ~10170 unique clones, and the mean insert size was estimated to be 118 kb (2× coverage); fraction C has high percentage of empty clones and low transformation efficiency, therefore it was not exploited further.

Screening and sequencing of the BAC clones: Clones from fractions A and B were screened with five probes (Table 1). Seven and 11 positive clones were found from fractions A and B respectively. Nine, three, one, two and three positive clones were obtained for Sab, Bor, TPS3, KS and CPS2 respectively. Insert sizes vary between 70 and 165 Kb (excluding clone 17K11, a putative cross contamination by 17K13), with an estimated average of 115 Kb. In BES, one end of clone 2O24, 17K13 and 34K8 was found to represent *S. officinalis* Sab, while one end of clone 22C6 is 84% identical to *S. miltiorrhiza* CPS2. Retrotransposon and simple sequence repeat (SSR, microsatellite) were also found at some BAC ends (data not shown).

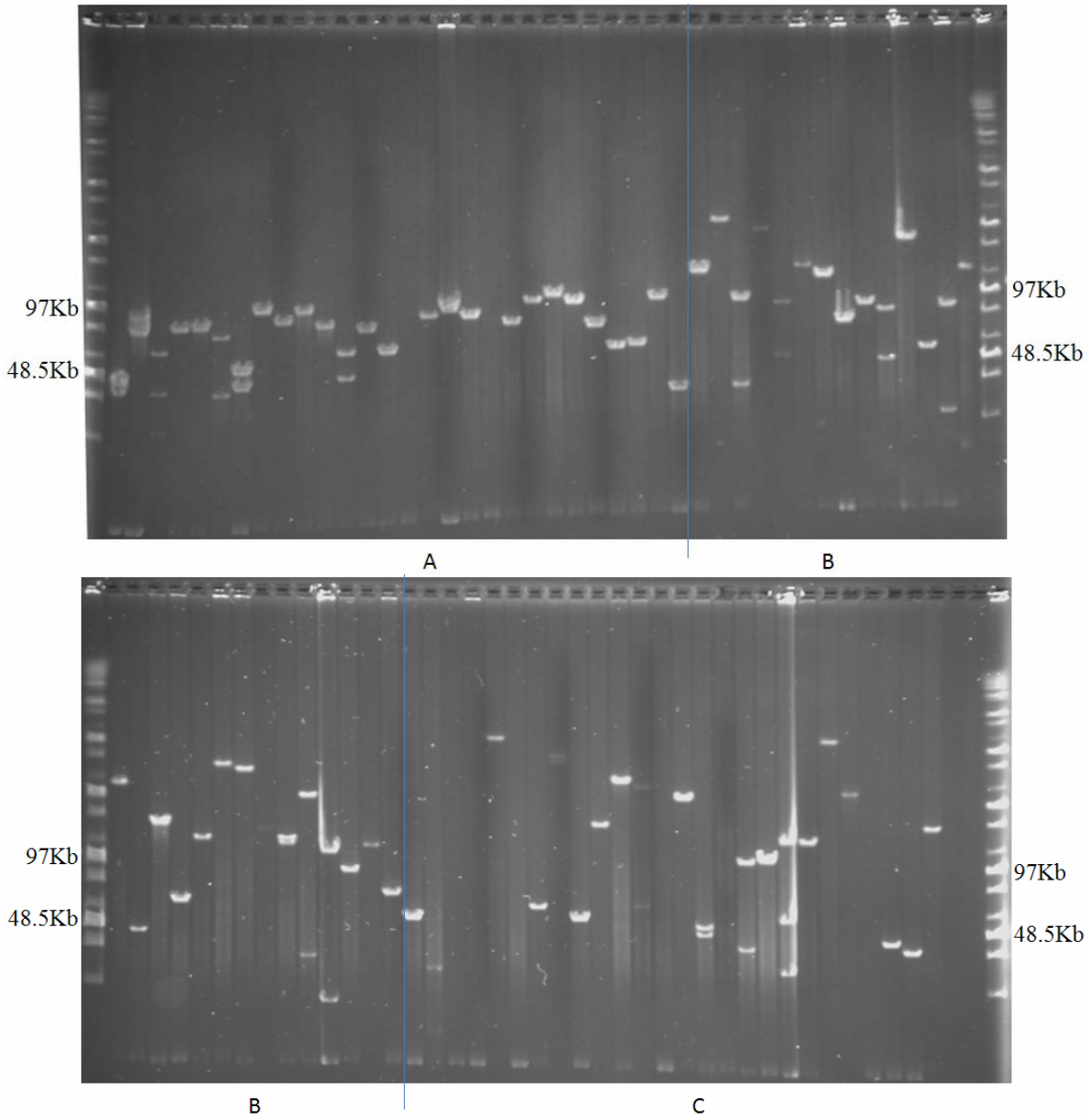

Fig. 1 Insert size estimation of BAC clones obtained from *S. officinalis* gDNA fractions A, B and C.

HiSeq 2000 sequencing of 12 BAC clones with longer inserts yielded 28,097,480 raw reads (2.53 Gb) with an average GC content of 41.3%. BAC sequencing reads can be retrieved with NCBI SRA accession no. SRX361905 and SRP030644. After filtering, the base number was reduced to 2.39 Gb but the quality metrics Q20 and Q30 did not alter significantly. After de-contamination and further filtering, 17,749,752 reads (1.51 Gb) were obtained for error correction and de novo sequence assembly (Table 2). On the average, 22 scaffolds (5-42) per BAC clone were obtained. BAC clone 24G14 has the least scaffold (5), followed by 32N10 (7) and 14G1 (9). BAC 14G1 has the largest N50 size (31,917 bp), followed by 32N10 (30,038 bp) and 24G14 (22,156 bp). The average GC content (36.5%) is lower than that of raw reads.

BAC sequence annotation: BAC sequences were annotated for repeat sequences, protein-coding genes and their functions. Repeat sequences account for 30.76% of BAC sequences (Table 2) and include all major classes (Fig. 2A). Most LTR (long terminal repeat) retrotransposons are Gypsy (261) and Copia (85), while LINE1 (L1, 14) is the most common non-LTR retrotransposon. The most common DNA transposon is hAT (42), followed by EnSpm (15) and MuDR (12). Among 1648 SSRs (Fig. 2), di-nucleotide SSR (298) is most prevalent, followed by tri-nucleotide (288) and hexa-nucleotide (262). SSRs are most abundant in BAC clone 39M8 (258), followed by 31P3 (250) and 32N10 (131). High-quality primers were obtained for 330 SSRs (67% of the total; Table 3) with an expected product size ranging from 100 bp to 300 bp (data available upon request).

Table 2. BAC *de novo* sequence assembly and annotation.

BAC clone	Scaffold no.	Total scaffold (>500bp) size(bp)[1]	N50 size(bp)[2]	GC%	Repeat no.	Repeat footprint (%)[3]	Gene no.	Gene footprint (%)	Exon no.	Exon footprint (%)
37P12	42	111339	5875	37.09	147	22.62	1	2.33	3	1.44
5P21	13	104146	8626	36.03	179	15.58	1	0.72	1	0.72
17K13	13	98845	6042	36.66	154	22.52	2	6.09	11	3.35
17L2	32	93710	3164	36.01	224	33.92	1	2.75	6	2.12
27F5	31	98952	4244	35.7	294	37.5	1	3.25	7	1.91
4J9	37	99901	2137	37.67	186	23.21	1	8.92	13	3.87
14G1	9	109431	31917	35.86	168	4.9	2	5.12	11	3.85
39M8	42	225771	4429	37.59	392	13.89	7	8.98	39	4.38
32N10	7	124347	30038	37.98	184	9.03	7	16.67	56	8.66
31P3	19	116022	6557	33.36	353	13.86	3	2.29	4	2.1
22C6	14	110317	6739	39.5	193	19.72	2	1.73	4	1.54
24G14	5	95095	22156	34.74	164	15.6	3	11.16	18	5.32

[1]Only scaffolds of longer than 500 bp are included. [2]All scaffolds are summed sequentially from large value to small value. N50 size is the length of the last added scaffold once the sum reaches 50% of total scaffold size. [3]Percentage of the full length of the annotated BAC sequences (not including unknown bases)

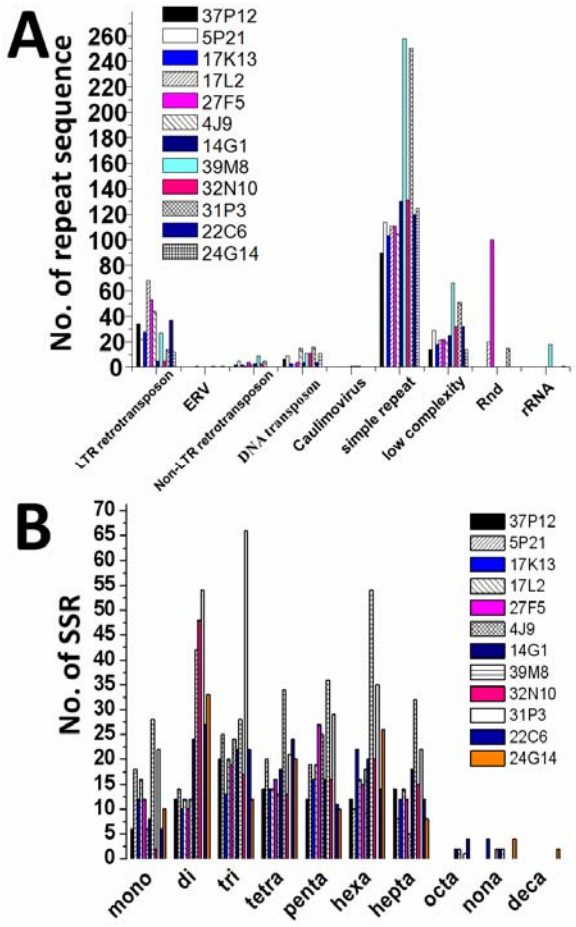

Fig. 2. Summary of repeat sequences of 12 BAC clones. A. Repeat elements; B. SSRs.

Thirty one genes were predicted, accounting for 10.98% of BAC sequences (Table 2). Full length Sab gene (Wise *et al*., 1998) was found in scaffold two of BAC clone 27F5 (Table 4), which is flanked by upstream hexa-nucleotide (ATTTAT) and downstream mono-nucleotide (A) SSRs. Bor gene was not found, instead a Sab homologous gene (Fig. 3) was predicted in scaffold three of BAC 14G1, which is closer to sabinene synthase of *Salvia* species than to bornyl diphosphate synthase and pinene synthase. Both Sab and Sab homolog belong to TPS-b clade. Similar to other Lamiaceae monoterpene synthases (Lima *et al*., 2013), they have seven exons (Fig. 4). The predicted Sab homolog protein has 586 amino acids, which is 77% (465/602) identical to pinene synthase of *Rosmarinus officinalis* and 79% (465/588) identical to sabinene synthase of *Salvia pomifera*. Full length TPS3 (632 aa) was found in scaffold one of BAC 32N10, which is 65% (431/660) identical to TPS3 of *S. sclarea* (Caniard *et al*., 2012) and 56% (385/689) identical to KSL2 of *S. miltiorrhiza* (Ma *et al*., 2012; Fig. 3), all of which belong to TPS-e clade and might be involved in diterpene synthesis. Full length CPS2 (712 aa) was found in scaffold two of BAC 24G14, which is 89% identical to *R. officinalis* CPS (AHL67261; Brückner *et al*., 2014) and 87% identical to *S. miltiorrhiza* CPS2 (Ma *et al*., 2012). These CPSs belong to TPS-c clade (Fig. 3) and are involved in diterpene synthesis. All above TPSs have common ancestor and are derived from multiple gene duplication events. Intriguingly, a cytochrome P450 (CYP) gene was found in scaffold one of 24G14, which is 62% (249/403) identical to CYP11 of *S. miltiorrhiza* (Guo *et al*., 2013; Fig. 5). A CYP gene fragment lacking C-terminal portion was found in scaffold four of 24G14, which is 58% (193/335) identical to CYP11 of *S. miltiorrhiza*. Bor and KS genes were not found in sequenced BAC clones. In BAC 5P21, a CYP gene fragment similar to CYP76AH1 of *S. miltiorrhiza* (Guo *et al*., 2013) was found in scaffold one, which lacks N-terminal portion. A putative KS pseudogene was predicted in the same scaffold, as there are multiple stop codons in the sequence. In BAC 31P3, three CYP gene fragments similar to CYP76AH1 of *S. miltiorrhiza* were predicted: one lacks C-terminal portion and the other two lack N-terminal portion.

Analysis of upstream sequences for *cis*-regulatory elements covered 1500 bp upstream of the ATG start codon for Sab, Sab homolog, CPS2 and CYP11. Only 395 bp upstream of TPS3 are available for promoter prediction as this gene is at the end of scaffold one of BAC 32N10. The most common *cis*-regulatory elements TATA-box and CAAT-box are present in the region immediately upstream of the start codon of all genes. Many regulatory elements involved in defense response were also found. For instance, TC-rich repeats, which are known to be involved in defense and stress responsiveness (Hamberger *et al*., 2009), are present in all genes except TPS3; TCA-element is involved in salicylic acid responsiveness (Hamberger *et al*., 2009) and is found in all genes except TPS3; ARE (anaerobic responsive element; Olive *et al*., 1991) is essential for the anaerobic induction and is present in Sab, Sab homolog and TPS3; CGTCA-motif is involved in methyl jasmonate (MeJA) responsiveness (Bhat *et al*., 2014) and is found in CPS2 and CYP11.

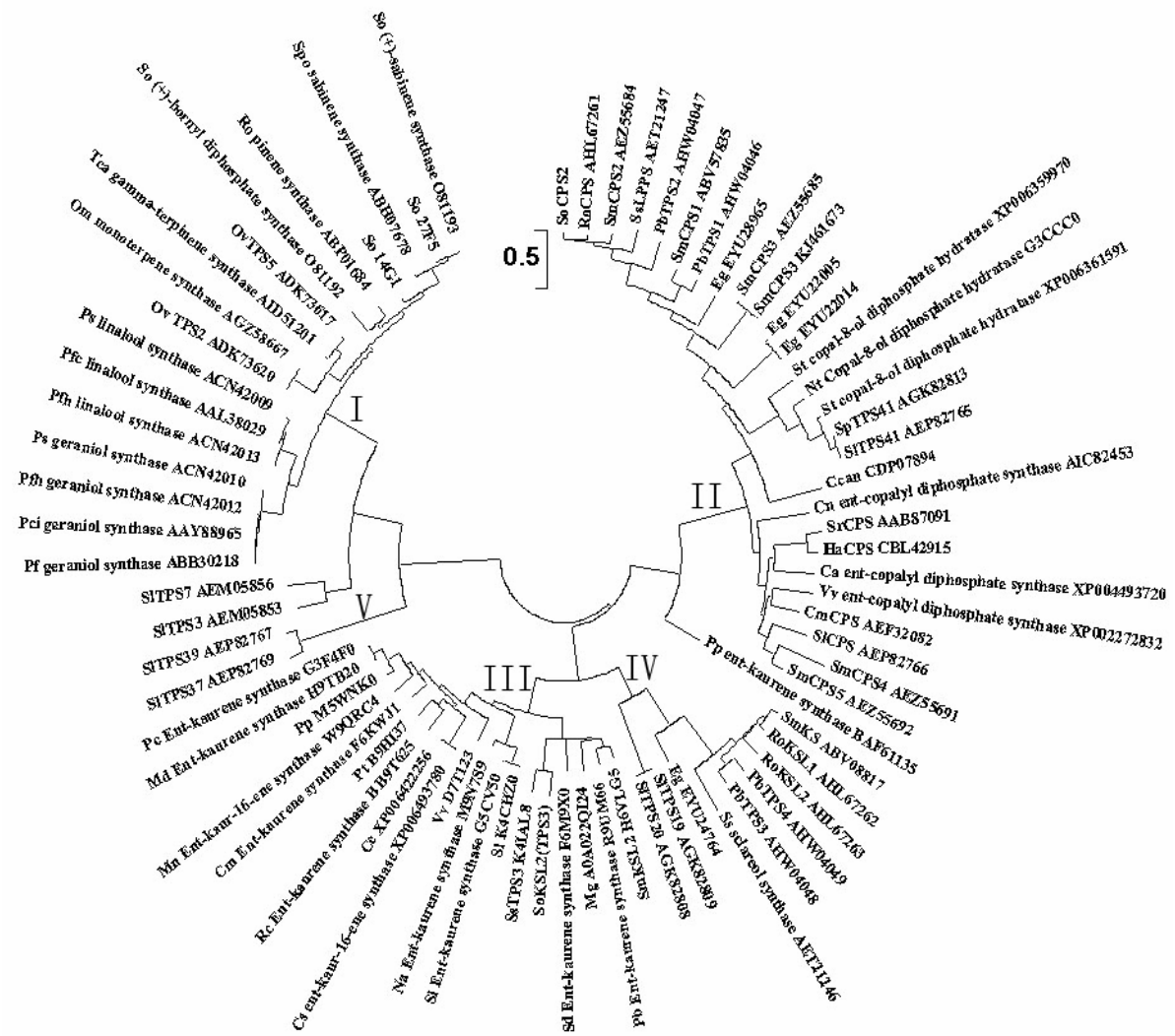

Fig. 3 Phylogenetic analysis of 77 Lamiaceae terpene synthases. The evolutionary history was inferred by using the Maximum Likelihood (ML) method based on the JTT+I+G model. The tree with the highest log likelihood (-57571.4726) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, parameter = 2.2283). The rate variation model allowed for some sites to be evolutionarily invariable (+I, 0.3325% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 77 amino acid sequences. There were a total of 1088 positions in the final dataset. I, TPS-b clade; II, TPS-c; III, TPS-e; IV, TPS-e/f; V, TPS-g. The moss ent-kaurene synthase belongs to TPS-h/c clade. NCBI GenBank or Uniprot accession numbers are shown for all sequences. Ca, *Cicer arietinum*; Cc, *Citrus clementina*; Ccan, *Coffea canephora*; Cm, *Castanea mollissima*; Cn, *Cocos nucifera*; Cs, *Citrus sinensis*; Eg, *Erythranthe guttata*; Ha, *Helianthus annuus*; Md, *Malus domestica*; Mg, *Mimulus guttatus*; Mn, *Morus notabilis*; Na, *Nicotiana attenuata*; Nt, *Nicotiana tabacum*; Om, *Origanum majorana*; Ov, *Origanum vulgare*; Pb, *Plectranthus barbatus*; Pc, *Pyrus communis*; Pci, *Perilla citriodora*; Pf, *Perilla frutescens*; Pfc, *Perilla frutescens var. crispa*; Pfh, *Perilla frutescens var. hirtella*; Pp, *Physcomitrella patens*; Ps, *Perilla setoyensis*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Ro, *Rosmarinus officinalis*; Sd, *Scoparia dulcis*; Sl, *Solanum lycopersicum*; Sm, *Savia miltiorrhiza*; So, *Salvia officinalis*; Sp, *Solanum pimpinellifolium*; Spo, *Salvia pomifera*; Sr, *Stevia rebaudiana*; Ss, *Salvia sclarea*; St, *Solanum tuberosum*; Tca, *Thymus caespititius*; Vv, *Vitis vinifera.*

Table 3. SSRs in 12 positive BAC clones that have PCR primers.

SSR class	No. of SSR obtained by MAKER + *de novo* prediction	Density (SSR no./Mb)	No. of SSR predicted by BatchPrimer3	Minimum repeat no.	Average length (bp)	No. of PCR primers	PCR product size (bp)
Dinucleotide	298	212.8	219	6	17.3	140	100-293
Trinucleotide	288	205.7	209	4	15.8	134	108-300
Tetranucleotide	221	157.8	34	4	17.0	27	113-220
Pentanucleotide	236	168.5	22	4	23.1	18	128-296
Hexanucleotide	262	187.1	11	4	28.9	11	138-263
Total	**1305**	**932.1**	**495**		**17.4**	**330**	**100-300**

Numbers of di-, tri-, tetra-, penta-, and hexa- nucleotide SSRs occurring in the *S. officinalis* genome. For each SSR type, the average length, the density (calculated as the number of SSRs per Mb) and the number of potentially amplifiable loci for which high-quality PCR primers were designed is reported

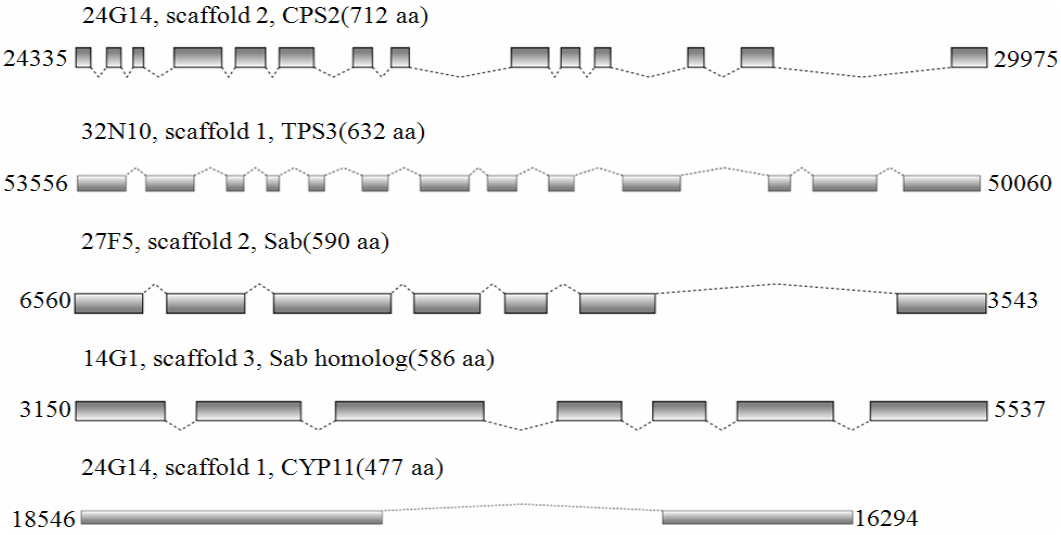

Fig. 4. Exon-intron structure of five terpenoid synthesis-related genes. Total length of the predicted protein is shown in parenthesis next to gene name. Box: exons; Line: introns. The genomic locations of these genes on the scaffolds are shown.

Table 4. Genes in the 12 BAC clones of *S. officinalis* predicted by MAKER.

Gene	Scaffold no.[1]	Gene position (strand), exon no.	Annotation
37P12			
1	6	3537-5224(-),3	XP_006338722,PREDICTED: probable protein S-acyltransferase 7-like,*Solanum tuberosum*
5P21			
1	1	7491-8123(+),1	AGN04215,CYP76AH1,*Salvia miltiorrhiza*
17K13			
1	6	1401-4366(+),6	XP_008221227,PREDICTED: GDSL esterase/lipase EXL3-like,*Prunus mume*
2	2	4660-6300(+),5	O81193,(+)-sabinene synthase,*S. officinalis*
17L2			
1	30	4691-6624(+),6	O81193,(+)-sabinene synthase,*S. officinalis*
27F5			
1	2	3543-6560(-),7	O81193,(+)-sabinene synthase,*S. officinalis*
4J9			
1	Contig1071[2]	385-6325(+),13	XP_006355174,PREDICTED: prolyl endopeptidase-like,*Solanum tuberosum*
14G1			
1	3	3150-5537(+),7	ABP01684,pinene synthase,*Rosmarinus officinalis*
2	4	25888-28612(-),4	XP_006346695,PREDICTED: pentatricopeptide repeat-containing protein,*Solanum tuberosum*
39M8			
1	Contig2591	148-1200(+),4	ADK73617,terpene synthase 5,*Origanum vulgare*
2	29	1583-4139(+),4	XP_004244963,PREDICTED: pentatricopeptide repeat-containing protein,*Solanum lycopersicum*
3	Contig2507	80-1146(-),3	ABP01684,pinene synthase,*Rosmarinus officinalis*
4	3	22855-25068(+),4	XP_002282196,PREDICTED: glutaredoxin-C1,*Vitis vinifera*
5	3	34338-37670(-),9	XP_004241509,PREDICTED: anthranilate phosphoribosyltransferase,chloroplastic-like isoform 2,*Solanum lycopersicum*
6	3	6184-10058(-),14	XP_007036268,Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B beta isoform 2,*Theobroma cacao*
7	Contig2505	408-812(+),1	1211235DD,NADH dehydrogenase 2-like ORF 180,*Nicotiana tabacum*
32N10			
1	1	16360-17370(-),3	AAF32410,α-expansin 2,*Triphysaria versicolor*
2	1	19165-26400(-),16	XP_008387770,PREDICTED: ribonuclease II,chloroplastic/mitochondrial,*Malus domestica*
3	1	50060-53556(-),13	AFU61899,diterpene synthase 3, partial,*Salvia sclarea*
4	Contig689	69-1960(-),6	XP_004249535,PREDICTED: probable serine/threonine-protein kinase NAK-like,*Solanum lycopersicum*
5	3	12874-13280(+),3	ABR92329,putative metallothionin 2a,*Salvia miltiorrhiza*
6	3	19296-22929(-),11	XP_002270444,PREDICTED: kinesin-like protein KIF22-B,*Vitis vinifera*
7	7	869-2852(+),4	XP_004246409,PREDICTED: probable protein phosphatase 2C 23-like isoform 2,*Solanum lycopersicum*
31P3			
1	1	4934-6022(-),2	AGN04215,CYP76AH1,*Salvia miltiorrhiza*
2	1	3877-4440(-),1	AGN04215,CYP76AH1,*Salvia miltiorrhiza*
3	7	4549-5202(+),1	AGN04215,CYP76AH1,*Salvia miltiorrhiza*
22C6			
1	4	2550-3636(-),2	AGN04218,cytochrome P450(P11),*Salvia miltiorrhiza*
2	4	189-718(-),2	no plant homology
24G14			
1	1	16294-18546(-),2	AGN04218,cytochrome P450(P11),*Salvia miltiorrhiza*
2	2	24335-29975(+),14	AHL67261,copalyl diphosphate synthase,*Rosmarinus officinalis*
3	4	1621-3467(+),2	AGN04218,cytochrome P450(P11),*Salvia miltiorrhiza*

[1]All genes are found in the respective single scaffold, some of which are on the same scaffold (Datasets S1-S12). [2]There is no unknown base (N) in a contig, while a scaffold is composed of a few contigs separated by N

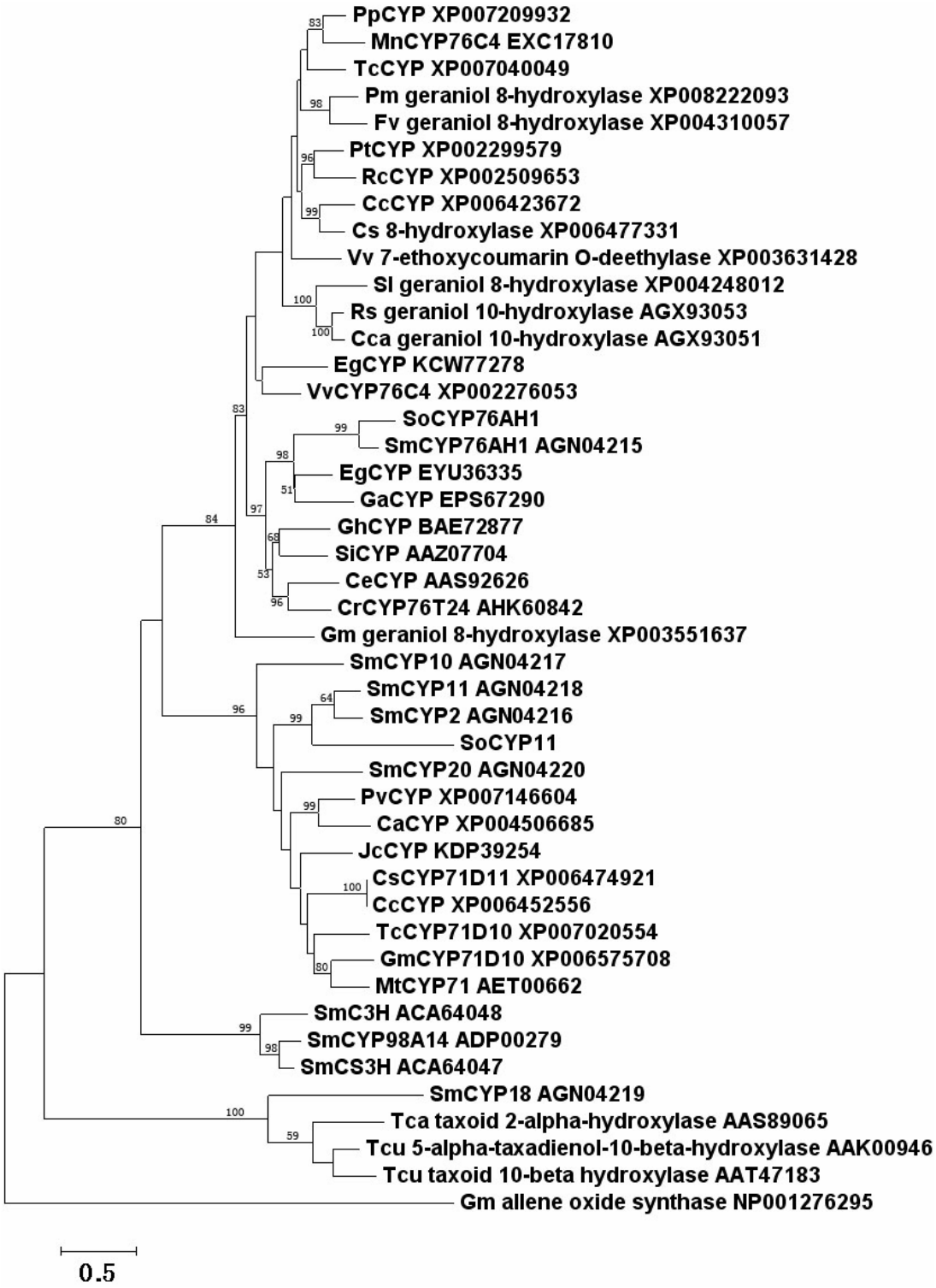

Fig. 5. Phylogenetic position of *S. officinalis* CYPs found in the sequenced BAC clones. The evolutionary history was inferred by using the ML method based on the JTT + I + G model (G, parameter = 2.0598; I, 1.3353% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 45 amino acid sequences. There were a total of 660 positions in the final dataset. Bootstrap values of >50% are shown on the branch. Ca, *Cicer arietinum*; Cc, *Citrus clementina*; Cca, *Cinchona calisaya*; Ce, *Centaurium erythraea*; Cr, *Catharanthus roseus*; Cs, *Citrus sinensis*; Eg, *Erythranthe guttata*; Fv, *Fragaria vesca subsp. vesca*; Ga, *Genlisea aurea*; Gh, *Glandularia × hybrida*; Gm, *Glycine max*; Jc, *Jatropha curcas*; Mn, *Morus notabilis*; Mt, *Medicago truncatula*; Pm, *Prunus mume*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Pv, *Phaseolus vulgaris*; Rc, *Ricinus communis*; Rs, *Rauvolfia serpentina*; Si, *Sesamum indicum*; Sl, *Solanum lycopersicum*; Sm, *Savia miltiorrhiza*; So, *Salvia officinalis*; Tc, *Theobroma cacao*; Tca, *Taxus canadensis*; Tcu, *Taxus cuspidata*; Vv, *Vitis vinifera*.

Discussion

High density filter and hybridization are commonly used in plant BAC library screening (Gonthier *et al*., 2010; Kovach *et al*., 2010; Paiva *et al*., 2011), which is labor intensive and the radiolabelled probes are formidable. Quantitative real-time PCR (qPCR) was used in BAC library screening of lake whitefish (Jeukens *et al*., 2011) but has not been used for plant. Herein we present a universal and rapid qPCR-based method to select positive BAC clones in a short time and at reasonable cost, using the *S. officinalis* BAC library as proof of concept. qPCR approach proved effective and convenient in obtaining at least one positive clone for each marker. Traditional PCR was also used in plant BAC library screening (Hamberger *et al*., 2009). However, agarose gels are not needed in qPCR, and since qPCR tubes are never opened after a run, cross-contamination is minimized. In the future, this approach tends to predominate over the laborious hybridization method and traditional PCR.

Traditional Sanger sequencing is still useful in BAC clone sequencing and assembly (Hamberger *et al*., 2009; Kovach *et al*., 2010; Wegrzyn *et al*., 2013). Sanger sequencing, when combined with high-throughput NGS, is useful in generating full-length draft assembly (Jeukens *et al*., 2011). However, the efficiency of 454 pyrosequencing alone in achieving high-quality genomic assembly of a non-model organism has been shown (Jeukens *et al*., 2011). The feasibility of using 454 sequencing in the assembly of plant BAC clones has been confirmed (Paiva *et al*., 2011; Wegrzyn *et al*., 2013). A more cost-effective NGS platform, Illumina GAII, has been used in BAC clone sequencing of Zhikong scallop (Zhao *et al*., 2012). In the context of the rapid development of NGS technology and based on our previous experiences of fosmid library (Hao *et al*., 2011a) and plant transcriptome sequencing (Hao *et al*., 2011b; 2012), we choose to use Illumina HiSeq 2000 in BAC cloning sequencing as it is more powerful and more efficient than GAII. Although the single sequence assembly was not obtained for BAC clone, the full length of three target genes (Sab, TPS3 and CPS2) and the flanking sequences were recovered for further analysis. Assembly gaps might be caused by highly abundant repeat sequences. Although the putative *Salvia* genome size is just ~600 Mb, the predicted repeat sequences account for 30.76% of the sequenced BAC clones (Table 2). The 454 sequencing results of lake whitefish BAC clones showed that four out of five target genes were split into multiple contigs (Jeukens *et al*., 2011), which had to be manually assembled. Our assembly results are better, as all terpene synthesis-related genes were assembled onto the single scaffold. Although target genes KS and Bor were not present in the assembly, two other terpene synthesis-related genes, Sab homolog and CYP11, were serendipitously found (Table 4 and Fig. 4). It is not surprising that Bor probe fished out a Sab homolog, as these monoterpene synthase genes have common ancestor and are evolutionarily young (Fig. 3). Many of these enzymes from sage generate overlapping multiple products from geranyl diphosphate (Wise *et al*., 1998). It will be interesting to characterize the recombinant Sab homolog protein to compare the product profile of these highly similar enzymes.

Multiple CYP11 sequences, either full length or fragment, coexists with CPS2 in BAC 24G14, raising the possibility that these terpene synthesis-related genes might comprise a genomic cluster. Both *S. officinalis* CYP11 and *S. miltiorrhiza* counterpart are closely related to CYP71D subfamily members (Fig. 5), which catalyze regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates (Takahashi *et al*., 2007). Metabolic gene cluster has been found in *Solanum lycopersicum* (Matsuba *et al*., 2013), which contains both CPS and sesquiterpene synthase genes. In promoter analysis, some *cis*-regulatory elements have been found to be present in the upstream sequences of both CPS2 and CYP11 (data not shown), implying that the expression of both genes could be co-regulated.

High-throughput genomic sequences can be readily mined for molecular genetic markers such as SSRs. We have utilized the *S. officinalis* genomic BAC clones for this application. The novel microsatellite markers developed here add to the previously identified species-specific markers (Anon., 2010; Radosavljević *et al*., 2012), as well as recently developed microsatellite markers with general utility across the Lamiaceae (Karaca *et al*., 2013). Microsatellites developed from genomes are more possible to be selectively neutral, because unlike those from transcriptomes, they are not linked to expressed genes, and could thus be used with more confidence in population genetic analyses. BAC-end sequences have been commonly used to find SSRs in various plants (Faivre Rampant *et al*., 2011; Sharma *et al*., 2012), the sequencing of which is costly and time-consuming. Our study highlights the usefulness of high-throughput full-length sequencing of BAC clones for rapid and cost-effective SSR discovery. Lamiaceae plants have common ancestor, and many of their genomic regions are well-conserved, therefore a reasonable transferability rate of the SSR markers across Lamiaceae species has been shown (Radosavljević *et al*., 2012; Karaca *et al*., 2013), making them particularly useful for comparative studies within the family.

Conclusion

In conclusion, a *Salvia* BAC library has been developed for the first time to facilitate map-based cloning of genes that play central roles in plant defense, species divergence and adaptation. Sequencing of the *Salvia* genome is paving the way for comparative and functional analysis of adaptive evolution. The *S. officinalis* BAC library can be screened for more candidate genes of interest. The large number of potentially amplifiable SSRs detected in our study represents an important resource for population genetics, comparative genomics and association studies. The complete *S. officinalis* BAC library and isolated clones are available upon request (contact the corresponding author).

Acknowledgements

This study is supported by China Scholarship Council (201208210122) and the Ministry of Science and Technology national support program (No. 2012BAI29B01). We thank Prof. Anne Osbourn (John Innes Centre, UK) for screen primer preparation and BAC DNA extraction in her lab.

References

Anonymous. 2010. MER Primer Development Consortium. Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2009-30 November 2009. *Mol. Ecol. Resour.*, 10(2): 404-408.

Bhat, W.W., S. Razdan, S. Rana, N. Dhar, T.A. Wani, P. Qazi, R. Vishwakarma and S.K. Lattoo. 2014. A phenylalanine ammonia-lyase ortholog (PkPAL1) from Picrorhiza kurrooa Royle ex. Benth: Molecular cloning, promoter analysis and response to biotic and abiotic elicitors. *Gene*, 547(2): 245-256.

Brückner, K., D. Božić, D. Manzano, D. Papaefthimiou, I. Pateraki, U. Scheler, A. Ferrer, R.C. de Vos, A.K. Kanellis and A. Tissier. 2014. Characterization of two genes for the biosynthesis of abietane-type diterpenes in rosemary (*Rosmarinus officinalis*) glandular trichomes. *Phytochemistry*, 101: 52-64.

Camarero, J.J., S. Palacio and G. Montserrat-Martí. 2013. Contrasting seasonal overlaps between primary and secondary growth are linked to wood anatomy in Mediterranean sub-shrubs. *Plant Biol.*, 15(5): 798-807.

Caniard, A., P. Zerbe, S. Legrand, A. Cohade, N. Valot, J.L. Magnard, J. Bohlmann and L. Legendre. 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.

Cantarel, B.L., I. Korf, S.M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado and M. Yandell. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.*, 18(1): 188-196.

Conesa, A. and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Inter. J. Plant Genomics*, 2008: 619832.

Crocoll, C., J. Asbach, J. Novak, J. Gershenzon and J. Degenhardt. 2010. Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Mol. Biol.*, 73(6): 587-603.

Dean, F.B., J.R. Nelson, T.L. Giesler and R.S. Lasken. 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.*, 11(6): 1095-1109.

Faivre Rampant, P., I. Lesur, C. Boussardon, F. Bitton, M.L. Martin-Magniette, C. Bodénès, G. Le Provost, H. Bergès, S. Fluch, A. Kremer and C. Plomion. 2011. Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. *BMC Genomics*, 12: 292.

Gonthier, L., A. Bellec, C. Blassiau, E. Prat, N. Helmstetter, C. Rambaud, B. Huss, T. Hendriks, H. Bergès and M.C. Quillet. 2010. Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Res. Notes.*, 3: 225.

Guo, J., Y.J. Zhou, M.L. Hillwig, Y. Shen, L. Yang, Y. Wang, X. Zhang, W. Liu, R.J. Peters, X. Chen, Z.K. Zhao and L. Huang. 2013. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc. Nati. Acad. Sci. U.S.A.*, 110(29): 12108-12113.

Hamberger, B., D. Hall, M. Yuen, C. Oddy, B. Hamberger, C.I. Keeling, C. Ritland, K. Ritland and J. Bohlmann. 2009. Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. *BMC Plant Biol.*, 9: 106.

Hao, D.C., G. Ge, P.G. Xiao, Y. Zhang and L. Yang. 2011b. The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PLoS One*, 6(6): e21220.

Hao, D.C., L. Yang and P.G. Xiao. 2011a. The first insight into the *Taxus* genome via fosmid library construction and end sequencing. *Mol. Genet. Genomics*, 285(3): 197-205.

Hao, D.C., P. Ma, J. Mu, S. Chen, P. Xiao, Y. Peng, L. Huo, L. Xu and C. Sun. 2012. *De novo* characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. *Sci. China Life Sci.*, 55(5): 452-466.

Jassbi, A.R., M. Mehrdad, F. Eghtesadi, S.N. Ebrahimi and I.T. Baldwin. 2006. Novel rearranged abietane diterpenoids from the roots of *Salvia sahendica*. *Chem. & Biodiversity*, 3(8): 916-922.

Jeukens, J., B. Boyle, I. Kukavica-Ibrulj, J. St-Cyr, R.C. Lévesque and L. Bernatchez. 2011. BAC library construction, screening and clone sequencing of lake whitefish (*Coregonus clupeaformis*, Salmonidae) towards the elucidation of adaptive species divergence. *Mol. Ecol. Resour.*, 11(3): 541-549.

Karaca, M., A.G. Ince, A. Aydin and S.T. Ay. 2013. Cross-genera transferable e-microsatellite markers for 12 genera of the Lamiaceae family. *J. Sci. Food Agri.*, 93(8): 1869-1879.

Kovach, A., J.L. Wegrzyn, G. Parra, C. Holt, G.E. Bruening, C.A. Loopstra, J. Hartigan, M. Yandell, C.H. Langley, I. Korf and D.B. Neale. 2010. The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics*, 11: 420.

Lázaro, A. and O. Totland. 2010. Population dependence in the interactions with neighbors for pollination: A field experiment with *Taraxacum officinale*. *Am J Bot.*, 97(5): 760-769.

Lescot, M., P. Déhais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouzé and S. Rombauts. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, 30(1): 325-327.

Li, M.H., Q. Li, Y. Liu, L. Huang and P.G. Xiao. 2013. Pharmacophylogenetic study on plants of genus *Salvia* L. from China. *Chin. Herb. Med.*, 5(3): 164-181.

Lima, A.S., J. Schimmel, B. Lukas, J. Novak, J.G. Barroso, A.C. Figueiredo, L.G. Pedro, J. Degenhardt and H. Trindade. 2013. Genomic characterization, molecular cloning and expression analysis of two terpene synthases from *Thymus caespititius* (Lamiaceae). *Planta*, 238(1): 191-204.

Ma, Y., L. Yuan, B. Wu, X. Li, S. Chen and S. Lu. 2012. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J Exp Bot.*, 63(7): 2809-2823.

Matsuba, Y., T.T. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schäfer, D. Kudrna, R.A. Wing, A.M. Bolger, B. Usadel, A. Tissier, A.R. Fernie, C.S. Barry and E. Pichersky. 2013. Evolution of a complex locus for terpene biosynthesis in solanum. *Plant Cell*, 25(6): 2022-2036.

Meng, H.H. and M.L. Zhang. 2013. Diversification of plant species in arid Northwest China: species-level phylogeographical history of *Lagochilus Bunge ex Bentham* (Lamiaceae). *Mol Phylogenet Evol.*, 68(3): 398-409.

Moreira, B., C. Tavsanoglu and J.G. Pausas. 2012. Local versus regional intraspecific variability in regeneration traits. *Oecologia*, 168(3): 671-677.

Nishida, N., S. Tamotsu, N. Nagata, C. Saito and A. Sakai. 2005. Allelopathic effects of volatile monoterpenoids produced by *Salvia leucophylla*: Inhibition of cell proliferation and DNA synthesis in the root apical meristem of *Brassica campestris* seedlings. *J. Chem. Ecol.*, 31(5): 1187-1203.

Olive, M.R., W.J. Peacock and E.S. Dennis. 1991. The anaerobic responsive element contains two GC-rich sequences essential for binding a nuclear protein and hypoxic activation of the maize Adh1 promoter. *Nucleic Acids Res.*, 19(25): 7053-7060.

Paiva, J.A., E. Prat, S. Vautrin, M.D. Santos, H. San-Clemente, S. Brommonschenkel, P.G. Fonseca, D. Grattapaglia, X. Song, J.S. Ammiraju, D. Kudrna, R.A. Wing, A.T. Freitas, H. Bergès and J. Grima-Pettenati. 2011. Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genomics*, 12: 137.

Peterson, D.G., J.P. Tomkins, D.A. Frisch, R.A. Wing and A.H. Paterson. 2000. Construction of plant Bacterial Artificial Chromosome (BAC) libraries: an illustrated guide. *J Agri Genomics*, 2000: 5.

Pfleeger, T., M. Blakeley-Smith, G. King, E. Henry Lee, M. Plocher and D. Olszyk. 2012. The effects of glyphosate and aminopyralid on a multi-species plant field trial. *Ecotoxicology*, 21(7): 1771-1787.

Popović, Z., M. Kostić, S. Stanković, S. Milanović, I. Sivčev, I. Kostić and P. Kljajić. 2013. Ecologically acceptable usage of derivatives of essential oil of sweet basil, *Ocimum basilicum*, as antifeedants against larvae of the gypsy moth, *Lymantria dispar*. *J. Insect Sci.*, 13: 161.

Radosavljević, I., Z. Satovic, J. Jakse, B. Javornik, D. Greguraš, M. Jug-Dujaković and Z. Liber. 2012. Development of new microsatellite markers. for *Salvia officinalis* L. and its potential use in conservation-genetic studies of narrow endemic *Salvia brachyodon* Vandas. *Inter. J. Mol. Sci.*, 13(9): 12082-12093.

Rambaldi, D. and F.D. Ciccarelli. 2009. FancyGene: dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics*, 25(17): 2281-2282.

Sharma, M.K., R. Sharma, P. Cao, J. Jenkins, L.E. Bartley, M. Qualls, J. Grimwood, J. Schmutz, D. Rokhsar and P.C. Ronald. 2012. A genome-wide survey of switchgrass genome structure and organization. *PLoS One*, 7(4): e33892.

Tabti, L., A. Dib Mel, N. Gaouar, B. Samira and B. Tabti. 2014. Antioxidant and antifungal activity of extracts of the aerial parts of *Thymus capitatus* (L.) Hoffmanns against four phytopathogenic fungi of *Citrus sinensis*. *Jundishapur J. Nat. Pharm. Prod.*, 9(1): 49-54.

Takahashi, S., Y.S. Yeo, Y. Zhao, P.E. O'Maille, B.T. Greenhagen, J.P. Noel, R.M. Coates and J. Chappell. 2007. Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates. *J Biol Chem.*, 282(43): 31744-31754.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*, 30(12): 2725-2729.

Theodoridis, S., A. Stefanaki, M. Tezcan, C. Aki, S. Kokkini and K.E. Vlachonasios. 2012. DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey). *Mol Ecol Resour.*, 12(4): 620-633.

Vokou, D., K. Vareli, E. Zarali, K. Karamanoli, H.I. Constantinidou, N. Monokrousos, J.M. Halley and I. Sainis. 2012. Exploring biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria. *Microb. Ecol.*, 64(3): 714-724.

Wegrzyn, J.L., B.Y. Lin, J.J. Zieve, W.M. Dougherty, P.J. Martínez-García, M. Koriabine, A. Holtz-Morris, P. deJong, M. Crepeau, C.H. Langley, D. Puiu, S.L. Salzberg, D.B. Neale and K.A. Stevens. 2013. Insights into the loblolly pine genome: characterization of BAC and fosmid sequences. *PLoS One*, 8(9): e72439.

Wise, M.L., T.J. Savage, E. Katahira and R. Croteau. 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(24): 14891-14899.

You, F.M., N. Huo, Y.Q. Gu, M.C. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak and O.D. Anderson. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, 9: 253.

Yu, X.Q., M. Maki, B.T. Drew, A.J. Paton, H.W. Li, J.L. Zhao, J.G. Conran and J. Li. 2014. Phylogeny and historical biogeography of *Isodon* (Lamiaceae): rapid radiation in south-west China and Miocene overland dispersal into Africa. *Mol. Phylogenet Evol.*, 77: 183-194.

Zhang, B., R. Classen-Bockhoff, Z.Q. Zhang, S. Sun, Y.J. Luo and Q.J. Li. 2011. Functional implications of the staminal lever mechanism in *Salvia cyclostegia* (Lamiaceae). *Ann. Bot.*, 107(4): 621-628.

Zhang, Y.G., M. Han, X. Jiang, S.N. Zhao and L.M. Yang. 2013. Physiological ecology responses of *Scutellaria baicalensis* to drought and rewatering. *Zhongguo Zhong Yao Za Zhi*, 38(22): 3845-3850.

Zhao, C., T. Zhang, X. Zhang, S. Hu and J. Xiang. 2012. Sequencing and analysis of four BAC clones containing innate immune genes from the Zhikong scallop (*Chlamys farreri*). *Gene*, 502(1): 9-15.

(Received for publication 12 May 2014)