# Identification, validation and cross-species transferability of novel *Lavandula* EST-SSRs

Ayelign M. Adal, Zerihun A. Demissie[1] and Soheil S. Mahmoud*

Department of Biology, University of British Columbia, 1177 Research Rd, Kelowna, BC. V1V 1V7, Canada
*Corresponding author e-mail: soheil.mahmoud@ubc.ca

[1] Current address: Department of Biological sciences, Brock University, 500 Glenridge Ave, St. Catharines, ON, L2S 3A1, Canada

## Abstract

*Main conclusion:* We identified and characterized EST-SSRs with strong discrimination power against *L. angustifolia* and *L. x intermedia* species. The markers also showed considerable cross-species transferability rate into six related *Lavandula* species.

Lavenders (*Lavandula*) are important economical crops grown around the globe for essential oil production. In an attempt to develop genetic markers for these plants, we analyzed over 13,000 unigenes developed from *L. angustifolia* and *L. x intermedia* EST databases, and identified 3,459 simple sequence repeats (SSR), which were dominated by trinucleotides (41.2%) and dinucleotides (31.45%). Approximately 19% of the unigenes contained at least one SSR marker, over 60% of which were localized in the UTRs. Only 252 EST-SSRs were 18 bp or longer from which 31 loci were validated, and 24 amplified discrete fragments with 85% polymorphism in *L. x intermedia* and *L. angustifolia*. The average number of alleles in *L. x intermedia* and *L. angustifolia* were 3.42 and 3.71 per marker with average PIC values of 0.47 and 0.52, respectively. These values suggest a moderate to strong level of informativeness for the markers, with some loci producing unique fingerprints. The cross-species transferability rate of the markers ranges in 50 to 100% across eight species. The utility of these markers was assessed in eight *Lavandula* species and 15 *L. angustifolia* and *L. x intermedia* cultivars, and the dendrogram deduced from their similarity indexes successfully delineated the species into their respective sections and the cultivars into their respective species. These markers have potential for application in fingerprinting, diversity studies and marker-assisted breeding of *Lavandula*.

**Keywords**: EST-SSR, Genetic marker, *Lavandula*, *L. angustifolia, L. x intermedia,* Polymorphism

**Abbreviations:**
CDS: protein coding regions, EST: Expressed sequence tags, EO: essential oil, *He*: genetic diversity, *Na*: number of alleles, PIC: polymorphic information content, PFR: primer flanking regions, SSR: Simple sequence repeats, UTR: Untranslated regions

## Introduction

Several members of the genus *Lavandula* (*Lamiaceae*) are cultivated worldwide for their essential oils (EOs), which are used in perfumes, cosmetic products, antiseptics, pharmaceutical preparations, alternative medicine, etc. Over 400 cultivars and their numerous hybrids are identified in the genus, which are grouped into three subgenera, eight sections and 39 species based on habit, indumentums, leaf shape and inflorescence structure. For example, the subgenus *Lavandula* comprises of three sections; *Lavandula*, *Dentatae* and *Stoechas* that are characterized by their multi-flowered cymes and woody shrubs with narrow leaves. On the other hand, the subgenus *Fabricia* – comprising the sections *Pterostoechas*, *Subnudae*, *Chaetostachy* and *Hasikenses* – and the subgenus *Sabaudia* (with *Sabaudia* as the only section) have single-flowered cymes with the former section lacking bracteoles (Upson and Andrew, 2004). These morphological markers, however, are influenced by environmental conditions and lack the power to discriminate cultivars/hybrids derived from genetically related species. Thus, developing appropriate markers that could delineate different species and their cultivars irrespective of their geographic origin and environmental conditions is of high priority as the market value and bioactivity of lavender oils largely relies on the species and/or cultivars used (Cavanagh and Wilkinson 2002). In line with this, recent *Lavandula* taxonomic classification research has focused on developing DNA-based markers. In addition to aiding genetic identification, the development of DNA-based markers would also provide an opportunity to exploit the genetic pool of *Lavandula* for targeted breeding and genetic conservation efforts.

Although different types of DNA markers have been developed and employed to characterize diverse crops, simple sequence repeats (SSRs; commonly referred to as microsatellites) are the preferred markers next to single nucleotide polymorphism (SNPs) in recent plant genetics and breeding studies. SSRs have been successfully applied in genetic variation, linkage and comparative mapping, functional diversity and evolutionary relationship analysis of several crop species including *Salvia spp* (Radosavljević et al. 2011; Radosavljević et al. 2012; Karaca et al. 2013), *Rosmarinus spp* (Segarra-Moragues and Gleiser 2008)and *Origanum vulgare* (Novak et al. 2008), *Triticum aestivum* L. (Gao et al. 2004), *Medicago truncatula* (Eujayl et al. 2004), Citrus (Liu et al. 2013) and so forth. Recent genetic studies favor SSRs over randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism due their multi-allelic nature, reproducibility, co-dominant inheritance, relative abundance, extensive genome coverage, ability to discriminate both inter- and intra-species variations, and relatively simple to analyze them (Li et al. 2004; Liu et al. 2013).

Depending on the source database, two types of SSR markers called genomic-SSRs and expressed sequence tag SSRs (EST-SSRs) are known. Genomic-SSRs are the most abundant SSRs in nature because SSRs are often located in non-coding regions of the genome. Genomic-SSRs are generally mined from SSR-enriched or non-enriched (random) genomic libraries. SSR-enriched genomic DNA libraries are constructed either through selective hybridization of genomic portions containing SSRs or by selective amplification of microsatellite containing genomic DNA fragments using SSR specific primers. On the other hand non-enriched or random libraries, as the name implies, are derived from randomly cloned genomic DNA fragments (Senan et al. 2014). However, in addition to being limited to model-organisms, genomic-SSR development is labor intensive, costly and time-consuming, and provides limited information about variations in the expressed regions of genomes (Li et al. 2004; Duran et al. 2009; Guichoux et al. 2011; Liu et al. 2013). EST-SSRs, on the other hand, are repeats located in expressed regions of a genome and mainly derived from EST databases. Although repeat numbers and total lengths of EST-SSRs are relatively small compared to genomic-SSRs (Li et al. 2004), previous reports indicated that they are: (1) uniquely informative as they represent alterations in structural and regulatory genes of an organism (Li et al. 2004; Chabane et al. 2005), (2) easily transferable among closely related species, (3) can distinguish varieties/cultivars of a given species, and (4) identified from EST databases that are cheap to develop and also not limited to model organisms. For example, Qureshi et al. (2004) reported that 26% of the EST-SSR primer pairs they tested showed intra-species polymorphism among *G. hirsutum* cultivars and 52% interspecies polymorphism between *G. hirsutum* and *G. barbadense*. EST-SSRs located in untranslated regions (UTRs) of a gene are more polymorphic than those located in coding regions (CDS) in the same gene, but the later is often associated with biochemical and/or physiological changes (Li et al. 2004).

Despite the economic importance and social value of lavenders, to our knowledge little attempt has been made to develop and apply DNA-based molecular tools to fingerprint species and cultivars, and to study the untapped genetic diversity in *Lavandula*. Upson and Andrews (2004) employed internal transcribed spacer data to discern the genetic relationship in the genus, while Hnia and Mohamed (2010) used RAPD markers to characterize the sub-genus *L.multifodia*. Recently, Karaca et al. (2013) reported the cross-genera transferability of *Salvia officinales* EST-SSRs to *Lavandula* hybrid, but only 7 of the 75 (~9%) loci were positive. To benefit *Lavandula* from recent advances in genomics, our group has recently reported three EST databases corresponding to cDNA libraries of *L. angustifolia* leaf and flower (Lane et al. 2010), and *L. x intermedia* secretory cells isolated from glandular trichome tissues (Demissie et al. 2012; Sarker et al. 2012). These libraries contain 22,290 ESTs that were assembled into 13,625 unigenes, and have already been used to isolate and functionally characterize key cDNAs involved in *Lavandula* isoprenoid biosynthesis (Demissie et al. 2011; Demissie et al. 2012; Sarker et al. 2012; Demissie et al. 2013). Given that genomic sequence of *Lavandula* is not yet available, we analyzed the above databases to identify, characterize, validate EST-SSRs and assess their cross-species transferability and suitability to analyze genetic relationship in *Lavandula*. Here, for the first time we report (1) the *in silico* identification and characterization of 3,459 novel EST-SSRs from *Lavandula,* (2) high polymorphism and cross-species transferability rate of 31 EST-SSRs, and (3) the genetic relationship assessment of eight species and fifteen cultivars belonging to *L. angustifolia* and *L. x intermedia* using the EST-SSR markers. The identified markers have shown promising polymorphism level to be developed to species and cultivar specific genetic markers. In addition, the identification of these markers would aid future *Lavandula* genetic studies including fingerprinting, diversity analysis, targeted breeding, genetic resource conservation and management practices, etc.

## Materials and methods

### EST-SSR motif identification, characterization and primer designing

The 13,625 unigenes (6,316 and 3,193 from *L. angustifolia* flower and leaf libraries, respectively, and 4,116 from *L. x intermedia* gland library) with an average length of 714 bp were screened *in silico* to identify SSR motifs using the web-based SSR mining tool 'SSR server' available at the Genome database for *Rosaceae* (GDR) (Jung et al. 2008). The following parameters were used during the screening: minimum motif length of 15 base pairs (bp) for mono- and penta-nucleotides, 12 bp for di-, tri- and tetra-nucleotide, and 18 bp for hexa-nucleotides, modified from Iorizzo et al. (2011). In addition to searching motifs from our unigenes, the SSR server was also used to calculate the number, frequency and size of repeats, location of motifs (CDS or UTR), possible primer pairs and expected amplified fragment lengths of identified motifs. The frequency of motifs residing in CDS and UTR were then manually calculated based on their location determined by the mining tool. Further, the average distance between two SSR motifs in ESTs was calculated as: EST-SSR density = (number of unigenes/number of SSRs)*average bp per unigene.

SSRs with primer-flanked regions (PFRs) identified by the SSR server tool often do not meet primer designing criterions. Thus, all unigenes, except mononucleotides, were submitted to BatchPrimer3 software to screen PFRs that meet primer designing criterions and pick their corresponding primers using the default parameters of the "SSR screening and primers" module (You et al. 2008). Depending on repeat length determined by BatchPrimer3 for PFRs meeting primer designing criterions, motifs were classified into category I and II. Category I included tri-nucleotide repeats with ≥ 18 bp and di-, tetra-, penta and hexa-nucleotides with ≥ 20 bp. Di-, tetra-, penta and hexa-nucleotide motifs with 12 - 20 bp and trinucleotides with 12 - 18 bp length were classified in category II. Since high rate of polymorphism is generally expected in longer SSRs (Temnykh et al. 2001), only category I members were advanced for further analysis. The frequency of each motif located in CDS or 5' and 3' UTR was calculated manually based on the location of the motif after determining the ORF location of the unigene using the NCBI ORF-finder software. From the 252 EST-SSRs classified in category I, 31 EST-SSRs were used for detailed validation.

### Functional annotation of SSR containing unigenes

The functional annotation of *Lavandula* unigenes containing SSR markers was performed using the Blast2GO online platform. Briefly, the sequences were blasted against the nr NCBI database using the BlastX module default parameters, gene ontology (GO) identifiers are mapped to the blast results, putative functions were assigned to the unigenes and analyzed, and finally the annotation of results were visualized using the direct acyclic graph (DAG) module (Conesa et al. 2005).

### Plant materials and genomic DNA extraction

The plant materials used in this study comprises of fifteen cultivars, eight from *L. angustifolia* and seven from *L. x intermedia* and of six cultivars from six other species, one cultivar per species, for transferability studies (Table 1). Leaf tissues of these cultivars were collected from The Okanagan Lavender and Herb Farm and The Greenery Garden Center (Kelowna, BC, Canada), freeze-dried in liquid nitrogen and genomic DNA was extracted using Geneaid Genomic DNA extraction kit (plant) (Geneaid Biotech Ltd. Taiwan) as per the manufacture's instruction. The quantity and quality of extracted DNA was determined using NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and agarose gel (1%), and the DNA was stored in – 20 °C until used.

### EST-SSR analysis and cross-species transferability study

PCR conditions (annealing temperatures) for each primer set were optimized by amplifying the corresponding DNA fragment from genomic DNAs extracted from leaf tissues of *L. angustifolia* cv. Maillette and *L. x intermedia* cv. Grosso individuals harvested from The University of British Columbia Okanagan lavender farm. The primers were then used for assessing genetic relationship by amplifying their respective loci from genomic DNAs isolated from *L. angustifolia* (8 cultivars) and *L. x intermedia* (7 cultivars). PCR amplifications were carried out at optimal conditions in a 25µL reaction volume containing 70 ng genomic DNA, 0.3 µM of each forward and reverse primers, 2.5 µl of the 10x reaction buffer, 250 µM each dNTP, 1.5 mM $MgCl_2$, 0.5 µg BSA and 1.25 unit of *Taq* DNA polymerase (NEB, Ipswich, MA). The PCR program used was: initial denaturation at 95 °C for 15 min, followed by 11 cycles of denaturation at 95 °C for 30 s, annealing temperature stepping down every cycles by 1°C from either 64 °C to 54 °C or 62 °C to 52 °C depending on the primer type, and extension at 72 °C for 2 min. This was followed by a second round amplification of 24 cycles with the following program: denaturation at 95 °C for 30

s, annealing temperature of 54 °C or 52 °C (depending on the primer type), and extension at 72 °C for 2 min, followed by a final extension at 72 °C for 10 min. Amplified fragments were visualized using SYBR safe-stained 6 % polyacrylamide gels. Clear and indisputable bands that were consistent in three separate amplifications were scored as present (1) and absent (0). Fragments with identical molecular weight across all species or cultivars were considered as monomorphic.

Cross species amplification of EST-SSRs generated from *L. angustifolia* and *L. x intermedia* were tested in six *Lavandula* species (Table 1): in PCR conditions described above. The genetic relationship among eight species, including the two EST-SSRs donor species, was analyzed using 18 cross transferred EST-SSR markers. Similarly, 24 loci were used to analyze the genetic diversities among eight *L. angustifolia* and seven *L. x intermedia* cultivars.

**Data analysis**

*Lavandula* species are characterized by complex polyploidy with significant variation in chromosome numbers, ranging from 18 – 75 chromosomes (Upson and Andrew 2004). Hence, to avoid errors associated with distinguishing alleles of homologous chromosomes, the co-dominant SSR markers were considered as dominant markers. Assuming every allele as a single locus, all alleles detected in all species were recorded as present (1) and absent (0) binary data matrix. The binary data matrix was then used to determine the total number of alleles, the number of polymorphic alleles, the number of alleles per marker (*Na*) and polymorphic alleles per marker (*Pa*). The genetic diversity ($H_e$) = 1 - (1/m)$\sum_L\sum_u P^2_{lu}$, where p*lu* is the frequency of the *u*th allele at the *l*th locus and m is the number of loci (Weir 1996) and polymorphic information content (PIC) = 1 - $\sum p^2_{i,}$ where *pi* is the frequency of the *i*th allele for individual *p* were calculated after determining the allele frequencies using PowerMarker v3.25 software (Liu and Muse 2005). The genetic similarity level among *L. angustifolia* and *L. x intermedia* cultivars, and among eight species representing four sections and one inter-section hybrid were estimated based on Jaccard's similarity coefficient and clustered with the UPGMA analysis and SAHN procedure of the NTSYS-PC v2.10t (Rolf 2000). The UPGMA dendrogram confidence limits were determined from 2000 bootstraps using WinBoot software program (Yap and Nelson 1996).

**Results**

**Identification and *in silico* characterization of EST-SSRs**

A total of 3,459 EST-SSR motifs were identified in our unigene libraries of which 1,641 were from flower database, 953 from leaves and 865 from oil glands (Figure 1). The 3,459 SSRs motifs were identified from 2,556 (18.8 %) unigenes because more than one SSR loci, up to five per unigene, were identified in some instances (Table 2). The distribution density of the EST-SSRs in *Lavandula* unigenes was one locus for every 2.81 kb distance and the number of repeat units per locus ranged from 3 for tetranucleotides - 40 for dinucleotides. The majority of the identified motifs (~60%) were located in the UTR of the unigenes. Although the EST databases of flower tissues resulted in the highest number of SSR motifs, the number of SSR motifs identified in each tissue was proportional to the number of unigenes present in the database. For example, the proportion of SSR motifs in *L. angustifolia* flower database was 26% and that of leaf was 29%.

Trinucleotide SSRs were the most abundant motifs in *Lavandula* unigenes, with 41% occurrence frequency, followed by di- (31.45%) and tetra-nucleotides (11.51%), while the frequencies of mono-, penta- and hexa-nucleotides were 6.07%, 3.79% and 5.98%, respectively (Figure 2). However, when different repeat types were considered, the dinucleotide motif type AG/GA/CT/TC showed the highest occurrence (26.2%), followed by the trinucleotide motif groups GGC/GCG/CGG/GCC/CCG/CGC and AAG/AGA/GAA/CTT/TTC/TCT with occurrence frequencies of 10 % and 8.09 %, respectively (Table 2). All other motif types had relatively low distribution, 0.35 - 5.72%, while the GC/CG motif types were completely absent in *Lavandula* unigenes. The same motif types also dominated the three tissue specific databases at comparable frequencies. For example, the dinucleotide motif group AG/GA/CT/TC is the dominant marker in flower, leaf and oil gland accounting for 26%, 27.2% and 25.6% of the total SSR motifs, respectively (Table 2).

SSRs are generally characterized by the presence of conserved flanking sequences. However, often not all EST-SSRs contain these sequences due to limitations associated with the nature of EST database or because some conserved sequences are too short to satisfy primer-designing parameters. From the 3,459 SSR motifs we identified, only 1,812 (about 52.4%) were flanked by conserved sequences, also called primer-flanking regions (PFRs), of which 1,108 were mined from flower unigene library, 300 from leaf and 404 from oil gland (Figure 1). The 1,812 motifs were grouped into two categories based on their repeat length where 252 (13.9%) EST-SSRs were clustered into category I and the rest, 1,560 (86.1%), were clustered into category II. Category II type EST-SSRs were

excluded from subsequent analysis because previous researches have established that they are inefficient in detecting polymorphism (Singh et al. 2009).

Category I member EST-SSRs were dominated by the dinucleotide repeat type AG/GA/CT/TC motif (99 of the 252 loci) followed by hexanucleotide repeats (30), the AT/TA repeat type (21), the trinucleotide repeats GGC/GCG/CGG/GCC/CCG/CGC and AAG/AGA/GAA/CTT/TTC/TCT (14 each), in their respective order (Table 3). Of the 252 category I loci, 140 (55.2%) were located in the CDS of the unigenes while 80 loci (32.1%) were found in the 5' UTR and 32 loci (12.7%) were in the 3' UTR region. This was contrary to the localization of the overall SSRs where only 40% of them were located in the CDS region. When the location of different motif types were compared, dinucleotides were predominantly found in the 5' UTR region while all the other motif types were predominantly located in the CDS region. Unlike the overall SSRs, the distribution of the dominant category I member motif types showed considerable variation among the three databases. For example, out of the 99 AG/GA/CT/TC motif types 48 (48.4%) were identified from *L. angustifolia* flower database and only 26 and 15 loci were identified from *L. x intermedia* gland and *L. angustifolia* leaf databases, respectively. Similarly, tri- and hexa-nucleotides were distributed unevenly in the three databases whereas, except for few outliers, tetra- and penta-nucleotides were generally less abundant in all databases (Table 3).

***Lavandula* EST-SSR validation and polymorphism**

Since the SSR markers were derived from *L. angustifolia* and *L. x intermedia* libraries, we used genomic DNAs extracted from the two species to study the polymorphism of 31 category I loci. Among the selected motifs 13 were dinucleotides, 12 were trinucleotides and 6 belongs to tetra-, penta- and hexa-nucleotide. In addition, 20 of the 31 randomly selected EST-SSRs were located in the CDS while seven of them were located in the 5' UTR and the remaining four in 3' UTR (Table 4). Of the total primer sets tested, 24 of them successfully amplified genomic DNA fragments in three different attempts, while seven markers failed to do so in five different attempts. All positive loci were multi-allelic and all alleles, but LAF6 in *L. angustifolia*, were polymorphic (Table 5). The LAF6 locus was polymorphic in *L. x intermedia* genomic DNA but was monomorphic in *L. angustifolia* and three loci, LAF8, LINT5 and LINT14, produced alleles greater than the expected sizes.

A total of 201 alleles (with 85% polymorphism), 103 (with 86% polymorphism) of which were from *L. angustifolia* and 98 (with 84% polymorphism) from *L. x intermedia*, were detected with an average 4.29 and 4.08 alleles per marker, respectively. The polymorphic information content (PIC) values of *L. angustifolia* plants were between 0.0 (LAF6 and LINT5) to 0.81 (LAL4) with an average value of 0.52. The PIC values for *L. x intermedia* plants were between 0.0 (LAF6 and LAF8) and 0.82 (LAF13) with an average value of 0.47. These values indicate that the selected markers have high levels of informativeness in *L. angustifolia* (0.52) but were moderate in *L. x intermedia* (0.47) according to the classification proposed by Vaiman et al. (1994). LAL4 and LAF13 loci have the highest number of alleles and PIC values in *L. angustifolia* and *L. x intermedia*, respectively, suggesting that they were the most informative markers in their respective species. In *L. angustifolia* the SSR loci LAF6 and LINT5 produced the minimum (0.0) gene diversity (*He*) level while the locus LAL4 produced the maximum *He* value (0.83). For *L. x intermedia* the minimum *He* value (0.0) was obtained at the LAF6 and LAF8 loci while the maximum value (0.84) was detected at the LAF13 locus (Table 5). This implies that these novel markers were able to detect relatively high genetic variations in both species.

**Functional annotation of SSR containing unigenes**

From the 2,556 SSR containing unigenes, BLAST hits were obtained for 2,321 (90.8%) of them while putative functions could be assigned to 1,933 (83.3%) unigenes. The majority of SSR containing unigenes have homology to nucleic acid binding and catalytic activity, 41.23 % and 38.9 %, respectively, and other functions like transporter (5.02%), transcription factors (4.24 %), structural molecule activity (3.72 %) and so forth (Figure 3). Some of the highly polymorphic loci were associated with functional genes including monoterpene synthases, implying their potential for functional adaptation genetic studies in addition to classical genetics (Table 4).

**Cross-species transferability of EST-SSR markers in *Lavandula***

As shown in Table 5, *L. angustifolia* SSRs showed 100% cross-species amplification with comparable polymorphism level in *L. x intermedia*. In fact the loci with the highest *He* and PIC value (LAF13) and highest number of alleles (LAL4) in *L. x intermedia* were identified from *L. angustifolia* (Table 5). Similarly, 100% transferability rate with comparable polymorphism level was also obtained for SSRs identified from *L. x intermedia* in *L. angustifolia*, although the number of alleles detected was considerably lower in both species. *L. angustifolia* EST-SSRs amplified a total of 75 alleles with 88% polymorphism in *L. angustifolia* and 74 alleles with 85%

polymorphism in *L. x intermedia*. On the other hand, the *L. x intermedia* EST-SSRs detected 19 alleles with 79% polymorphism in *L. x intermedia* and 14 alleles with 78% polymorphism in *L. angustifolia*. The number of alleles detected by *L. x intermedia* SSRs was low because only six of the 24 loci were from *L. x intermedia*. This considerably higher cross-species loci transferability and polymorphism level was expected since the two species are genetically related, *L. x intermedia* is a natural half progeny of *L. angustifolia*.

We also tested the cross-species transferability of these markers in six other species identified in different sections of the genus *Lavandula* (Table 1). The markers showed 100 % transferability rate in *L. latifolia* species followed by rates ranging from 50 % in *L. buchii*, *L. lusitanica* and *L. stoechas* to 83.3% in *L. x ginginsii*, a natural inter-section hybrid between *L. latifolia* and *L. dentata* (Table 6). LAF5, LAF9, LAF21 and LAL4 loci showed 100% cross-species transferability in tested species (Table 6), albeit with different level of polymorphism (data not shown). Sizes of the alleles amplified across species were within the ranges detected in the donor species for each locus, suggesting that the products were likely derived from the same loci and that the primer binding sites of the alleles were highly conserved.

**EST-SSR based genetic relationship analysis**

Eighteen cross-species amplified EST-SSR markers were used to analyze genetic similarity levels among eight different species (including donor species) identified in different section of *Lavandula* according to previous taxonomic classifications (Upson and Andrew, 2004). The genetic similarity index, as explained by the Jaccard Similarity Coefficient, among the eight species ranged between 0.11 – 0.6 (Figure 4a). The SSR loci were effective in categorizing the different species into their respective sections as described previously using both morphological and inter-spacer transcribed markers (Upson and Andrew 2004). It is, however, worth to note that although *L. x intermedia* was grouped between its parental lines in the deduced tree, it was rooted closer to *L. angustifolia* than *L. latifolia*. The most likely explanation for this is the fact that most of the polymorphic SSR loci were derived from *L. angustifolia* unigenes and also those derived from *L. x intermedia* showed considerable polymorphism in *L. angustifolia* than *L. latifolia* (Table 5 and 6). Similar trend was also observed in *L. x ginginsii* – an inter-section hybrid of *L. latifolia* and *L. dentate* – where it was closely rooted with *L. latifolia* than *L. dentate*. This was also likely because the selected markers showed 100% cross-species transferability with higher number of alleles in *L. latifolia* as opposed to the 77.8% rate seen in *L. dentate*.

Similarly, 24 loci successfully discriminated fifteen cultivars belonging to *L. angustifolia* and *L. x intermedia* according to their species with Jaccard's similarity coefficient ranging from 0.4 – 0.74. In the deduced dendrogram, the cultivars were grouped in their respective species (Figure 4b). The maximum similarity level (74% similarity) was detected between *L. angustifolia* cv Folgate and Sachet, suggesting that the markers were also effective in discriminating the cultivars from each other. Figure 5a showed LAL4 locus amplification in *L. angustifolia* and *L. x intermedia* cultivars (given in Table 1) while the amplification of LAL4 and LAL5 loci from the eight species (Table 1) was presented in Figure 5b and c, respectively.

**Discussion**

***Lavandula* EST-SSRs identification and characterization**

Of the 39 species and their numerous inter-species and inter-section hybrids, *L. angustifolia*, *L. latifolia* and their natural hybrid *L. x intermedia* are the major sources of essential oils sold worldwide. The market value and bioactivities of these oils are mainly determined by their monoterpene profiles, which in turn depend on the species and/or cultivars used and environmental conditions (Cavanagh and Wilkinson 2002). Thus, in order to ensure genetic purity of species or cultivars used, a marker that can transcend environmental effects and phenotypic overlaps among genetically related species is required. However, to our knowledge such markers have not yet been reported in *Lavandula*. SSRs are polymorphic DNA sequences with proven potential to successfully distinguish both intra-species and inter-species diversities in both model and non-model plants (Varshney et al. 2005; Sharma et al. 2007). Given that genomic sequences are not available for *Lavandula*, we identified 3,459 EST-SSR motifs from 13,625 *Lavandula* unigenes (Lane et al. 2010; Demissie et al. 2012; Sarker et al. 2012), and validated 31 of these SSRs for their polymorphism and discrimination power among cultivars and species.

Approximately 18.8% of the unigenes contained at least one SSR locus. Although this ratio was higher compared to previous reports like 7 – 10 % in cereals (Varshney et al. 2002), 7.71% in soybean (Xin et al. 2012), 9.3% in sugarcane (Singh et al. 2013) etc, it was less than the 35% ratio reported for citrus EST-SSR (Palmieri et al. 2007). The 2.81 kb per unigene SSR density obtained in this study was also higher than that reported in rice (3.4 kb), wheat (5.4 kb) and soybean (7.4 kb), but comparable with the 2.4 kb per locus density reported for *D. versipellis*

(Guo et al. 2014). These variations in the frequency, distribution and abundance of SSRs identified from different species and databases are commonly associated with the use of different search criterions, size of the database and database-mining tools employed (Sharma et al. 2009). In fact, Peng and Lapitan (2005) identified 36,520 (7.41%) EST-SSRs from wheat database that contains 492,832 ESTs compared to the 22,290 ESTs (13,625 unigenes) we had in our library.

Of the 1,812 PFR containing EST-SSRs, only category I EST-SSRs (252 of them or ~14%) were considered for further characterization. This was because previous researches had confirmed that polymorphic capacity of SSR is directly proportional to its length (Temnykh et al. 2001; Singh et al. 2009). More than 49% of the category I EST-SSRs were dinucleotide motifs, which was contrary to the overall SSR motif distribution (Table 2) where trinucleotides were dominant. This result was obtained despite using 18 bp length as the minimum requirement for trinucleotide repeat classification under this category while that of dinucleotide was set at 20 bp. Nonetheless, similar types of results have been reported in other crops (Lagercrantz et al. 1993; Grover et al. 2007). Interestingly, only ~33% of the dinucleotides were located in the CDS of the unigenes compared to ~81% of the trinucleotides and 83% of the hexanucleotides (Table 3). These results were also consistent with previous reports. For example, only 19 % of the dinucleotides in wheat were located in CDS, while 74% of their trinucleotides were located in CDS (Yu et al. 2004). This tendency of organisms to tolerate tri- and hexa-nucleotide repeats in their translated genomic region as opposed to di-, tetra- or penta-nucleotides likely stems from the fact that the later motifs are vulnerable for frame-shift mutations (Duran et al. 2009; Guichoux et al. 2011). It is also worth to note that EST-SSRs located in the 3' UTR are generally less abundant across the three *Lavandula* unigene databases, 8.73% in flower, and 1.98% in leaf and oil gland (Table 3). This was likely because our EST databases were developed by sequencing the 5' end of our cDNA library (Lane et al. 2010; Demissie et al. 2012; Sarker et al. 2012), which favored the enrichment of the 5' end sequences in the database.

**EST-SSR polymorphism**

EST-SSR polymorphism is generally lower than that of genomic-SSRs. However, previous research has shown that EST-SSRs with longer repeat size had sufficient level of polymorphism to distinguish closely related species (Yu et al. 2000; Singh et al. 2009; Dutta et al. 2011). This is particularly true for plants with complex polyploidy structure where their genome complexity favors polymorphism for a given locus. For example, the number of alleles per locus in sugarcane, a complex polyploid plant, ranged from 2 – 22 alleles with 65.5% polymorphism (Oliveira et al. 2009). In addition, Pinto et al (2006) compared 51 EST-SSRs with 50 genomic-SSRs and found that their mean discrimination power among 18 sugarcane varieties was non-significant. He also reported that dendrograms developed using the two SSR types were in agreement with documented pedigree information. Most loci in *Lavandula*, like sugarcane, were multi-allelic (with 2 – 10 alleles per locus) and 85% of them were polymorphic (Table 5). The higher allele number recorded in sugarcane could be due its superior complex polyploidy and genome size (10 Gb) (Souza et al. 2011) compared to the estimated 900 Mb size of *Lavandula* (Urwin et al. 2007; Urwin 2014). However, the loci identified from *Lavandula* were more polymorphic likely due to the heterogeneity of source material we used (Table 1) compared to the clonally propagated sugarcane varieties used by Pinto et al. (2006). Also all loci, but three, produced amplicons within the expected size ranges (Table 5), implying that most loci and their primer binding sites were highly conserved between the two species. This was not surprising considering the high genetic similarity shared between *L. angustifolia* and *L. x intermedia*. Similar high levels of loci conservation with sporadic unexpected amplicon sizes have also been reported in other crops (Pinto et al. 2004; Xin et al. 2012). The main reason for this variation is differences in number and sizes of introns among alleles (Varshney et al. 2006).

One of the major advantages of EST-SSRs is their uniqueness in revealing alterations in expressed structural and regulatory genes. Thus, the fact that putative function could be assigned to 1,933 (83.3%) of the ESTs containing SSRs (Figure 3) prompted us to study the localization of highly polymorphic loci and correlate that with their assigned function. We were particularly interested in polymorphic loci located in CDS of genes whose alteration is likely to be accompanied by phenotypic consequences. To our surprise, despite the fact that SSRs located in CDS are less polymorphic than those in UTRs (Li et al. 2004; Dutta et al. 2011) and EST-SSRs of closely related species are mostly conserved owing to their genic nature (Li et al. 2004), polymorphic loci located in CDS of functionally important lavender genes were identified. For instance, the locus LAF20 – with PIC values of 0.76 and 0.69 in *L. angustifolia* and *L. x intermedia*, respectively – was located within the coding region of a RHOMBOID-like serine protease homolog while the locus LAF18 – with PIC values of 0.59 and 0.64 in *L. angustifolia* and *L. x*

*intermedia*, respectively – was located in the coding region of an ABA inducible bHLH-type transcription factor homology. These proteins play key role in plant development and ecological interactions. Plant RHOMBOID-like serine proteases are believed to play role in correct root growth, floral development, fertility and photoprotection (Thompson et al. 2012) while bHLH-type transcription factors are involved in stress induced signal transduction pathway regulation (Nakata et al. 2013). The locus LINT12 was also located in the CDS of a gene our group previously characterized as 1,8-cineole synthase, an enzyme catalyzing the synthesis of one of the major *Lavandula* EO constituents (Demissie et al. 2012). With PIC values of 0.35 and 0.41 in *L. angustifolia* and *L. x intermedia*, respectively, the locus LINT12 showed a moderate genetic diversity level. This result was not surprising as we previously reported slight variations among the genomic sequences of 1,8-cineole synthases of *L. x intermedia*, *L. angustifolia* and *L. latifolia* (Demissie et al. 2012). In addition, the locus that produced the highest heterozygous alleles (*He* = 0.83) and genetic diversity level (PIC = 0.81) in *L. angustifolia* was located in CDS of a hypothetical protein (Table 4 and 5). This implies the potential use of EST-SSR markers in gaining insights into functional diversification, including alterations in monoterpene profile, in *Lavandula*.

**Cross-species transferability and application in genetic relationship analysis**

Prior sequence information (genomic or EST) is a prerequisite to develop SSR markers. However, since flanking sequences of SSRs are highly conserved in related species – sometimes even across genera – SSR primers obtained from well-studied species are often successfully used to cross amplify polymorphic loci from non-model species (Gupta and Varshney 2000; Pierantoni et al. 2004; Varshney et al. 2005). For example, Singh et al. (2013) reported 87 – 93 % inter-species and 80 – 87% inter-genera cross transferability for sugarcane EST-SSRs. A similar trend has been reported for other EST-SSRs as well (Sharma et al. 2009; Gong and Deng 2010; Dutta et al. 2011). In this study, *L. angustifolia* EST-SSRs showed 100% transferability and 85% polymorphism in *L. x intermedia* while those from *L. x intermedia* showed 100% transferability and 78% polymorphism in *L. angustifolia* (Table 5). Therefore, we decided to test their cross-species transferability EST-SSRs in six other *Lavandula* species that have different levels of genetic relationships with the donor species. Owing to the fact that *L. latifolia* belongs to the same section with the donor species (Upson and Andrew 2004), it was expected to find 100% transferability. The transferability rate recorded for other species (50 – 83.3%) was also consistent with the genetic distance of the species from the donor species (Table 6). For instance, 50% of the loci were amplified in *L. buchii* that shared only 11% similarity with *L. angustifolia* and *L. x intermedia* (Figure 4a). However, even this lowest transferability rate (50%) is ~490% higher than the only other SSR marker transferability study reported for *Lavandula* to date (Karaca et al. 2013).

Although the bulk of lavender oils sold worldwide are derived from *L. angustifolia* and *L. x intermedia*, their market value and application is largely determined by the species/cultivars used. Hence, obtaining a marker that has the power to discriminate these two species and their cultivars is of paramount interest both from quality maintenance point of view and genetic diversity studies aimed at improving the quality and quantity of oil production. As presented in Figure 4 a & b, the loci reported here delineated the two species and their cultivars into their respective group in separate clades. The markers were also able to discriminate all the cultivars from each other; the maximum similarity level recorded among all cultivars used in this study was the 74% similarity level between *L. angustifolia* cv Folgate and Sachet (Figure 4b). This together with the fact that some loci produced novel fingerprints with high PIC values (0.83 and 0.84 for *L. angustifolia* and *L. x intermedia*, respectively) implies the likely chance of identifying species and cultivar specific markers (Figure 5 a-c) through further detail analysis of the loci reported here and by identifying more EST-SSRs from our libraries.

In conclusion, we identified and characterized 3,459 EST-SSR markers from *Lavandula* ESTs databases. Approximately 19% of the unigenes harbor at least one SSR marker, implying the widespread distribution of microsatellites in *Lavandula* encoded genome. The selected EST-SSRs were highly polymorphic (>85%), showed considerable transferability (50 – 100%) into six other *Lavandula* species, and displayed strong discrimination power in *L. angustifolia* and *L. x intermedia* species. Thus, the identified markers could be very useful for identification of these economically important lavender species and their cultivars, and study the genetic diversity in the genus *Lavandula*. Overall, the SSR markers reported here are useful in genetic fingerprinting of economically important *Lavandula* species and cultivars. They might also have applications in quality control, targeted breeding, association mapping and assessing the genetic diversity in *Lavandula*. In addition, since the markers showed a remarkable cross-species transferability rate and polymorphism, they might be useful in fingerprinting and related genetic studies of other closely related *Lavandula* species.

**Figure and Table Legends**

**Table 1.** List of *Lavandula* species and cultivars used in this study.

**Table 2.** The occurrence of different SSR motif types in unigenes of *Lavandula* EST databases derived from flower, leaf and glandular trichome cDNA libraries.

**Table 3.** The occurrence and localization of category I SSRs in unigenes of *Lavandula* EST databases derived from flower, leaf and glandular trichome tissue cDNA libraries.

**Table 4.** List of selected *Lavandula* EST-SSRs along with their primer pairs, motif type, location and putative function.

**Table 5.** Validation of 31 EST-SSR markers in *L. angustifolia* and *L. x intermedia* species.

**Table 6.** Cross-species transferability of 18 EST-SSR markers in six *Lavandula* species.

**Figure 1.** A flow chart presenting a stepwise *in silico* analysis of EST resources to identify EST-SSRs from cDNA libraries derived flower and leaf tissues in *L. angustifolia* and glandular trichome tissues of *L. x intermedia*. *PFR*-primer flanking region, *category I*- long SSR repeats as ≥ 18 bp for trinucleotide and ≥ 20 bp for others; *category II* - short SSR repeats 12 - 18 bp for trinucleotides and between 12 and 20 bp for other motifs.

**Figure 2.** The overall frequency of SSR motifs in unigenes of *Lavandula* EST databases derived from flower, leaf and glandular trichome cDNA libraries.

**Figure 3.** Molecular level functional annotation of *Lavandula* unigenes containing SSR markers using Blast2GO software.

**Figure 4.** Phylogenic relationship among *Lavandula* species and cultivars deduced from similarity index results of the EST-SSR markers. UPGMA dendrograms showing genetic relationship among (a) eight *Lavandula* species based on 18 EST-SSRs markers; and (b) among *L. angustifolia* and *L. x intermedia* cultivars using 24 EST-SSRs markers. *L. x ginginsii* is an inter-sectional hybrid between *L. lantata* (*Lavandula* subsection) and *L. dentata* (*dentatae* section). The scale bar indicates the level of Jaccard's similarity coefficient between samples. Bootstrap values after 2000 replicates are shown if ≥ 35%.

**Figure 5**. PCR amplification of SSR loci in selected *Lavandula* species and cultivars. A) LAL4 locus amplified from genomic DNAs of *L. angustifolia* (Lane 1-8) and *L. x intermedia* (Lane 9-15) cultivars, B) LAL4 locus amplified from genomic DNAs of eight *Lavandula* species and C) LAF5 locus amplified from genomic DNAs of eight *Lavandula* species. Lane descriptions a) M: 50 bp DNA ladder (NEB, Ipswich, MA), 1) *L. a.* cv Tucker's early purple, 2) *L. a.* cv Betty's Blue, 3) *L. a.* cv Folgate, 4) *L. a.* cv Sachet, 5) *L. a.* cv Sharon Robert, 6) *L. a.* cv Royal velvet, 7) *L. a.* Subsp. angustifolia, 8) *L. a.* cv Maillette, 9) *L. x i*. cv Hidcote Giant, 10) *L. x i*. cv Grosso, 11) *L. x i*. cv Provence, 12) *L. x i*. cv Supper, 13) *L. x i*. cv Fred Boutin, 14) *L. x i*. cv Seal and 15) *L. x i*. cv Abrialli; b and c) 50 bp DNA ladder (NEB, NE), 1) *L. angustifolia,* 2) *L. latifolia,* 3) *L. x intermedia,* 4) *L. buchii,* 5) *L. dentata,* 6) *L. lusitanica*, 7) *L. x ginginsii* and 8) *L. stoechas*. PAGE (6%) stained with SYBR safe were used for resolution.

**Acknowledgements**

This work was supported through grants and/or in-kind support to SSM by UBC, Genome British Columbia, Natural Sciences and Engineering Research Council of Canada, Agriculture and Agri-Food Canada and the BC Ministry of Agriculture (through programs delivered by the Investment Agriculture Foundation of BC), Downderry Lavender (Hadlow, Tonbridge, UK), Okanagan Lavender Herb Farm (Kelowna, BC, Canada), and Sacred Mountain Lavender (Salt Spring Islands, BC, Canada). Finally, we appreciate Dr. Michael Russello for his help with manuscript preparation.

## References


Cavanagh HMA, Wilkinson JM (2002) Biological activities of lavender essential oil. Phytother Res 16:301–8. doi: 10.1002/ptr.1103

Chabane K, Ablett GA, Cordeiro GM, Valkounn J, Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. Genet Resour Crop Evol 52:903–909. doi: 10.1007/s10722-003-6112-7

Conesa A, Götz S, García-Gómez JM, Javier T, Manuel T, Montserrat R (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–6. doi: 10.1093/bioinformatics/bti610

Demissie ZA, Cella MA, Sarker LS, Thompson TJ, Rheault MR, Mahmoud SS (2012) Cloning, functional characterization and genomic organization of 1,8-cineole synthases from *Lavandula*. Plant Mol Biol 79:393–411. doi: 10.1007/s11103-012-9920-3

Demissie ZA, Erland LAE, Rheault MR, Mahmoud SS (2013) The biosynthetic origin of irregular monoterpenes in *Lavandula*: isolation and biochemical characterization of a novel cis-prenyl diphosphate synthase gene, lavandulyl diphosphate synthase. J Biol Chem 288:6333–41. doi: 10.1074/jbc.M112.431171

Demissie ZA, Sarker LS, Mahmoud SS (2011) Cloning and functional characterization of β-phellandrene synthase from *Lavandula angustifolia*. Planta 233:685–96. doi: 10.1007/s00425-010-1332-5

Duran C, Appleby N, Edwards D, Batley J (2009) Molecular genetic markers: discovery, applications, data storage and visualisation. Curr. Bioinform. 4:16–27.

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S, Singh MN, Fakrudin B, Pawan K, Wanjari KB, Varshney RK, Cook D R, Nagendra K, Singh NK (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol 11:17. doi: 10.1186/1471-2229-11-17

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–22. doi: 10.1007/s00122-003-1450-6

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–400. doi: 10.1007/s00122-003-1554-z

Gong L, Deng Z (2010) EST-SSR markers for gerbera (*Gerbera hybrida*). Mol Breed 26:125–132. doi: 10.1007/s11032-009-9380-x

Grover A, Aishwarya V, Sharma PC (2007) Biased distribution of microsatellite motifs in the rice genome. Mol Genet Genomics 277:469–80. doi: 10.1007/s00438-006-0204-y

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611. doi: 10.1111/j.1755-0998.2011.03014.x

Guo R, Mao Y-R, Cai J-R, Wang J-Y, Wu J, and Qiu Y-X (2014). Characterization and cross-species transferability of EST–SSR markers developed from the transcriptome of *Dysosma versipellis* (*Berberidaceae*) and their application to population genetic studies. Mol Breed. doi:10.1007/s11032-014-0134-z

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163-185.

Hnia C, Mohamed B (2010) Genetic diversity of *Lavandula multifida* L . (*Lamiaceae*) in Tunisia: implication for conservation. Afr J Ecol, 49: 10–20. DOI: 10.1111/j.1365-2028.2010.01223.x

Iorizzo M, Senalik DA, Grzebelus D, Bowman M, Cavagnaro PF, Matvienko M, Ashrafi H, Van Deynze A, Simon PW (2011) *De novo* assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. BMC Genomics 12:389. doi: 10.1186/1471-2164-12-389

Jung S, Staton M, Lee T, Lee T, Blenda A, Svancara R, Abbott A, Main D (2008) GDR (Genome Database for *Rosaceae*): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Res 36:D1034–40. doi: 10.1093/nar/gkm803

Karaca M, Ince AG, Aydin A, Ay ST (2013) Cross-genera transferable e-microsatellite markers for 12 genera of the *Lamiaceae* family. J Sci Food Agric 93:1869–79. doi: 10.1002/jsfa.5982

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–5.

Lane A, Boecklemann A, Woronuk GN, Sarker S, Mahmoud SS (2010) A genomics resource for investigating regulation of essential oil production in *Lavandula angustifolia*. Planta 231:835–45. doi: 10.1007/s00425-009-1090-4

Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007. doi: 10.1093/molbev/msh073

Liu K, Muse S V (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–9. doi: 10.1093/bioinformatics/bti282

Liu S-R, Li W-Y, Long D, Hu C-G, Zhang, J-Z (2013) Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis. PLoS One 8:e75149. doi: 10.1371/journal.pone.0075149

Nakata M, Mitsuda N, Herde M, Koo AJK, Moreno JE, Suzuki K, Howe G, Ohme-Takagi M (2013) A bHLH-type transcription factor, ABA-INDUCIBLE BHLH-TYPE TRANSCRIPTION FACTOR/JA-ASSOCIATED MYC2-LIKE1, acts as a repressor to negatively regulate jasmonate signaling in arabidopsis. Plant Cell 25:1641–56. doi: 10.1105/tpc.113.111112

Novak J, Lukas B, Bolzer K, Grausgruber-Gröger S, Degenhardt J (2008) Identification and characterization of simple sequence repeat markers from a glandular *Origanum vulgare* expressed sequence tag. Mol Ecol Resour 8:599–601. doi: 10.1111/j.1471-8286.2007.02059.x

Oliveira a. LM, Stoffels M, Schmid M, Reis VM, Baldani JI, Hartmann A (2009) Colonization of sugarcane plantlets by mixed inoculations with diazotrophic bacteria. Eur J Soil Biol 45:106–113. doi: 10.1016/j.ejsobi.2008.09.004

Palmieri DA, Novelli VM, Bastianel M, Cristofani-yaly M, Astúa-monge G, Carlos EF, Oliveira ACD, Machado MA (2007) Frequency and distribution of microsatellites from ESTs of citrus. Genet Mol Biol, 30:1009–1018.

Peng JH, Lapitan NL V (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics 5:80–96. doi: 10.1007/s10142-004-0128-8

Pierantoni L, Cho K-H, Shin I-S, Chiodini R, Tartarini S, Dondini L, Kang S.-J, Sansavini S (2004) Characterisation and transferability of apple SSRs to two European pear F1 populations. Theor Appl Genet 109:1519–24. doi: 10.1007/s00122-004-1775-9

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed 125:378–384. DOI: 10.1111/j.1439-0523.2006.01227.x

Pinto LR, Oliveira KM, Ulian EC, Garcia AF, Souza APD (2004) Survey in the sugarcane expressed sequence tag database ( SUCEST ) for simple sequence repeats. Genome 47:795–804. doi: 10.1139/G04-055

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: A new class of genetic markers in cotton. J. Cotton Sci. 8:112–123.

Radosavljević I, Jakse J, Javornik B, Satovic Z, Liber Z (2011) New microsatellite markers for *Salvia officinalis* (*Lamiaceae*) and cross-amplification in closely related species. Am J Bot 98:e316–8. doi: 10.3732/ajb.1000462

Radosavljević I, Satovic Z, Jakse J, Javornik B, Greguraš D, Jug-Dujaković M, Liber Z (2012) Development of new microsatellite markers for *Salvia officinalis* L. and its potential use in conservation-genetic studies of narrow endemic *Salvia brachyodon* Vandas. Int J Mol Sci 13:12082–93. doi: 10.3390/ijms130912082

Rolf JF (2000) NTSYS-PC. Numerical Taxonomy and Multivariate Analysis System, version 2.11T Exeter Software. Setauket, NY, USA.

Sarker LS, Galata M, Demissie Z a, Mahmoud SS (2012) Molecular cloning and functional characterization of borneol dehydrogenase from the glandular trichomes of *Lavandula x intermedia*. Arch Biochem Biophys 528:163–70. doi: 10.1016/j.abb.2012.09.013

Segarra-Moragues JG, Gleiser G (2008) Isolation and characterisation of di and tri nucleotide microsatellite loci in *Rosmarinus officinalis* (*Lamiaceae*), using enriched genomic libraries. Conserv Genet 10:571–575. doi: 10.1007/s10592-008-9572-7

Senan S, Kizhakayil D, Sasikumar B, Sheeja TE (2014) Methods for development of microsatellite markers : An overview. Not Sci Biol 6:1–13.

Sharma PC, Grover A, Kahl G (2007) Mining microsatellites in eukaryotic genomes. Trends Biotechnol 25:490–8. doi: 10.1016/j.tibtech.2007.07.013

Sharma RK, Bhardwaj P, Negi R, Mohapatra T, Ahuja P (2009) Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). BMC Plant Biol 9:53. doi: 10.1186/1471-2229-9-53

Singh H, Deshmukh RK, Singh A, Singh AK, Gaikwad K, Sharma TR, Mohapatra T, Singh NK (2009) Highly variable SSR markers suitable for rice genotyping using agarose gels. Mol Breed 25:359–364. doi: 10.1007/s11032-009-9328-1

Singh RK, Jena SN, Khan S, Yadav S, Banarjee N, Raghuvanshi S, Bhardwaj V, Dattamajumder SK, Kapur R, Solomon S, Swapna M, Srivastava S, Tyagi AK (2013) Development, cross-species/genera transferability of

novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. Gene 524:309–29. doi: 10.1016/j.gene.2013.03.125

Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, Henry R, Ming R, Potier B, Sluys M-A, Vincentz M, Paterson, AH (2011) The sugarcane genome challenge: strategies for sequencing a highly complex genome. Trop Plant Biol 4:145–156. doi: 10.1007/s12042-011-9079-0

Temnykh S, Declerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 1441–1452. doi: 10.1101/gr.184001.11

Thompson EP, Smith SGL, Glover BJ (2012) An *Arabidopsis* rhomboid protease has roles in the chloroplast and in flower development. J Exp Bot 63:3559–70. doi: 10.1093/jxb/ers012

Upson T, Andrews S (2004). The genus *Lavandula*. (1st ed.). Timber Press, Inc., USA.

Urwin NAR (2014) Generation and characterisation of colchicine-induced polyploid *Lavandula* × *intermedia*. Euphytica 197:331–339. doi: 10.1007/s10681-014-1069-5

Urwin NA R, Horsnell J, Moon T (2007) Generation and characterisation of colchicine-induced autotetraploid *Lavandula angustifolia*. Euphytica 156:257–266. doi: 10.1007/s10681-007-9373-y

Vaiman D, Mercier D, Moazami-Goudarzi K, Eggen A, Ciampolini R, Lépingle A, Velmala R, Kaukinen J, Varvio SL, Martin P (1994) A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. Mamm Genome 5:288–97.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55. doi: 10.1016/j.tibtech.2004.11.005

Varshney RK, Grosse I, Hähnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet 113:239–50. doi: 10.1007/s00122-006-0289-z

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546.

Weir BS (1996) Genetic data analysis. II. Methods for discrete population genetic data. 2nd edn. Sinauer Associates, Sunderland

Xin D, Sun J, Wang J, Jiang H, Hu G, Liu C, Chen Q (2012) Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. Mol Biol Rep 39:9047–57. doi: 10.1007/s11033-012-1776-8

Yap I, Nelson RJ (1996) Winboot. a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. Manila, Philippines: IRRI Discussion paper series no. 14. International Rice Research Institute.

You FM, Huo N, Gu YQ, Luo, M-C, Ma, Y, Hane, D, Lazo, GR, Dvorak, J, and Anderson, OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253. doi: 10.1186/1471-2105-9-253

Yu J, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818. doi: 10.1139/G04-057

Yu K, Park SJ, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J Hered 91:429–34.

Figure

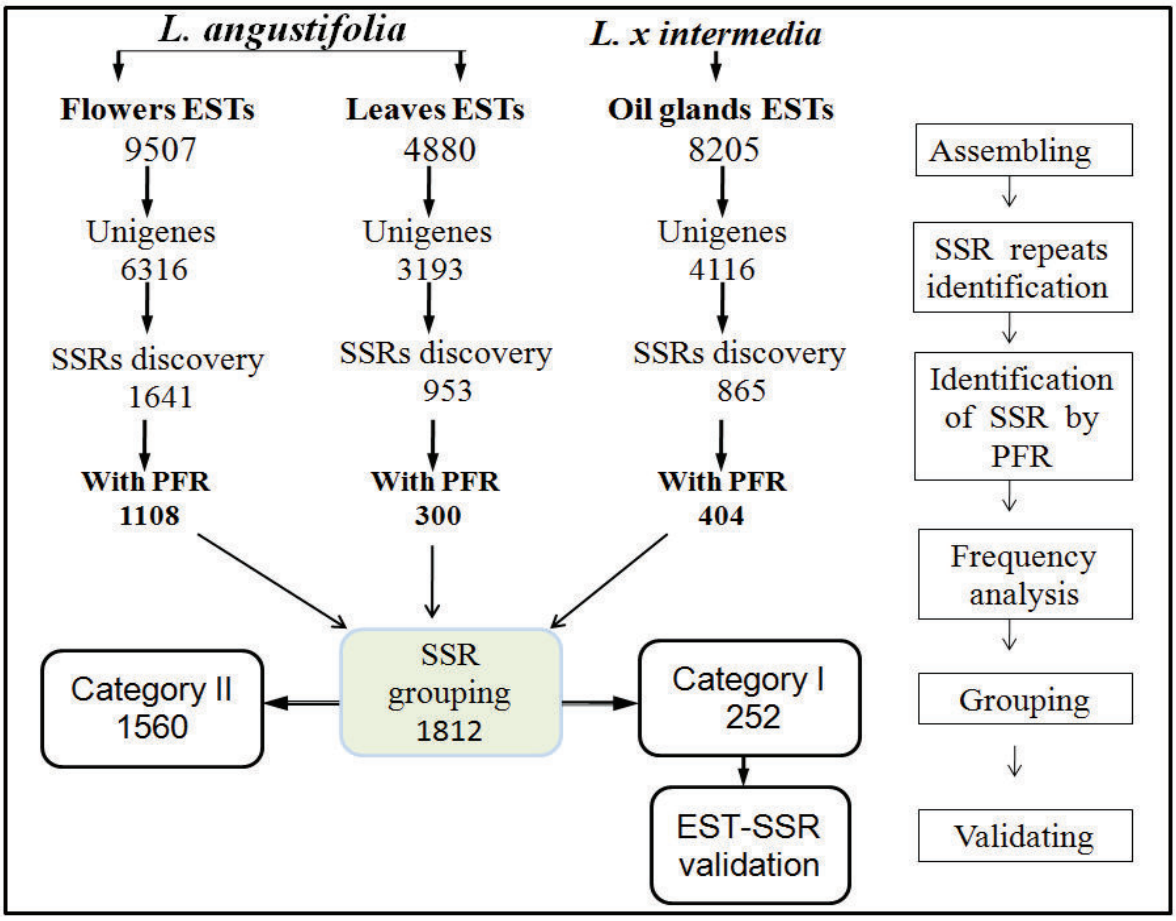


**Figure 1**

Figure

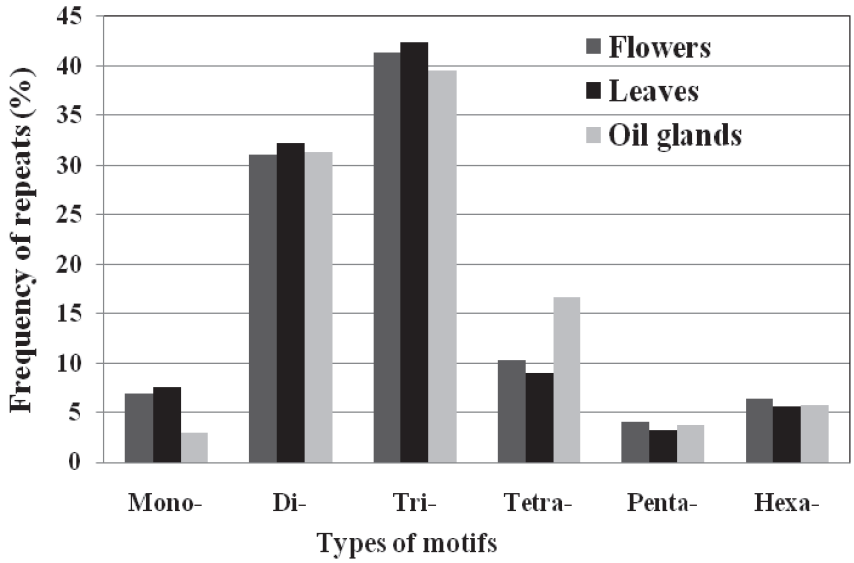


**Figure 2**

Figure

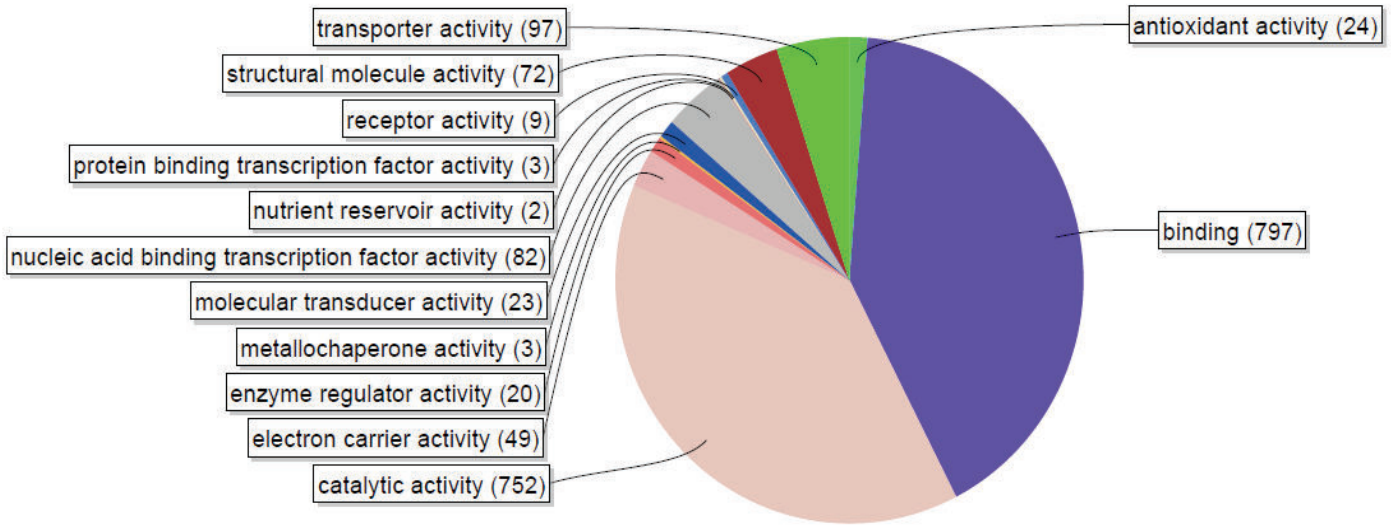


**Figure 3**

Figure

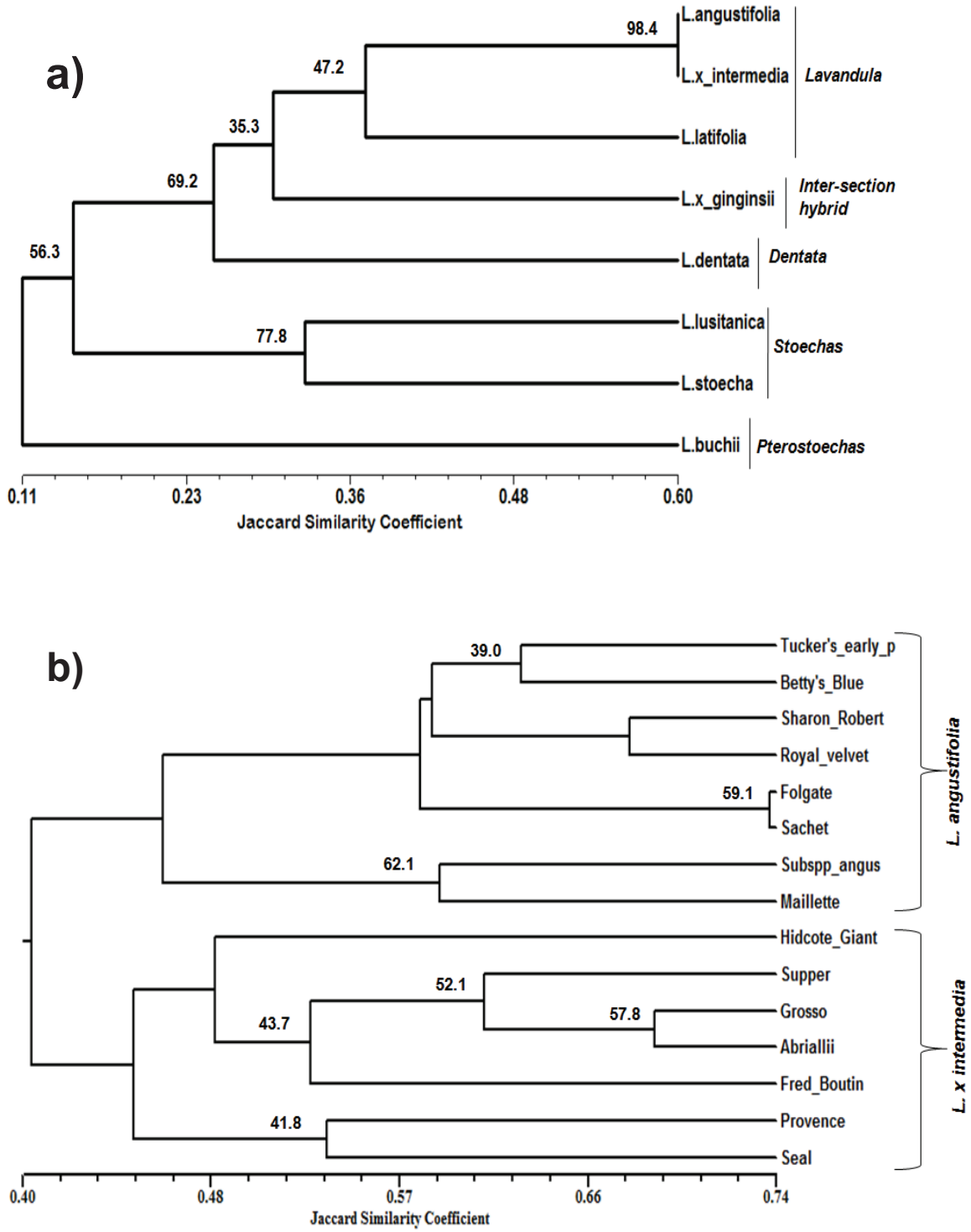


**Figure 4**

Figure

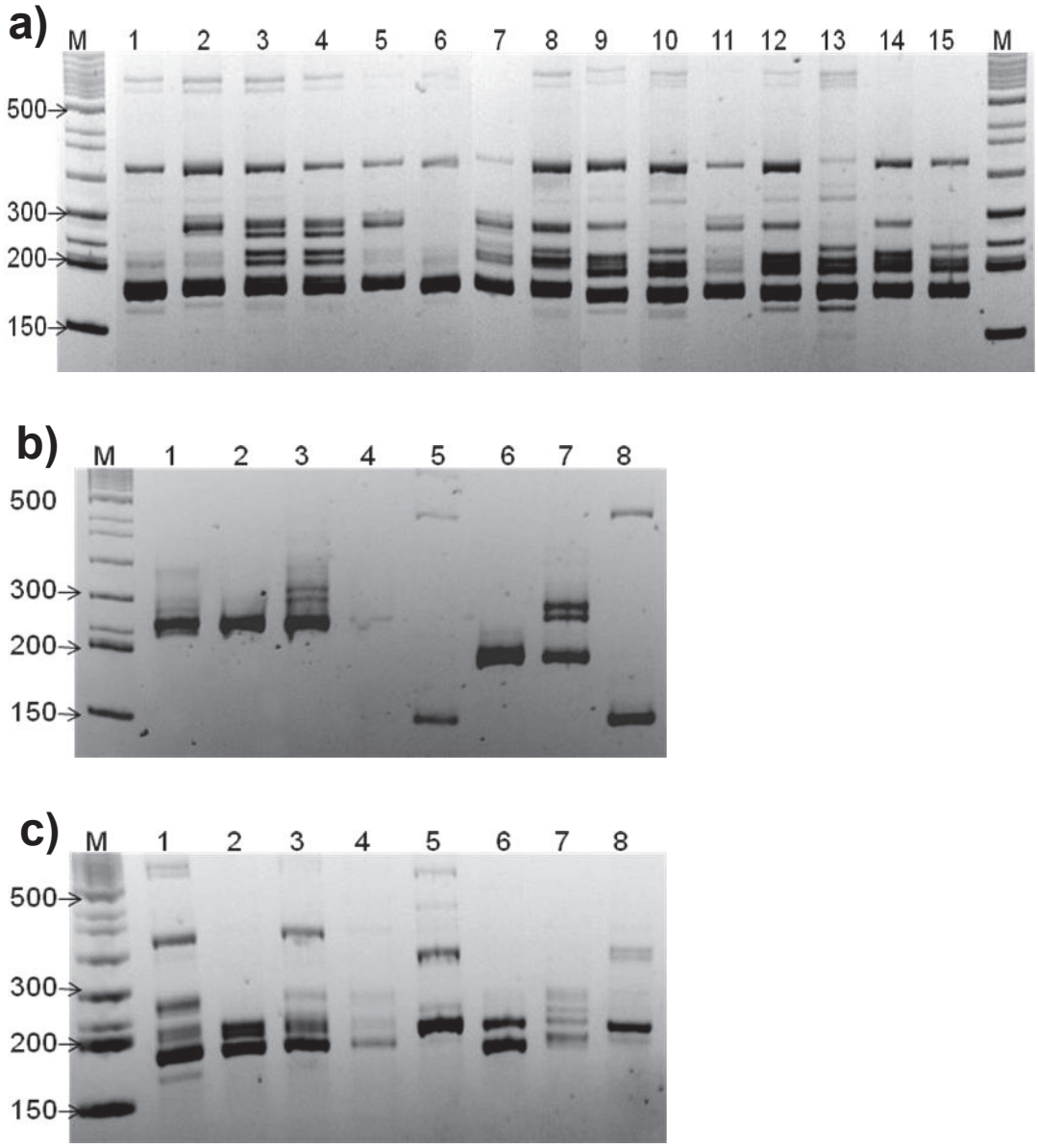


**Figure 5**

**Table 1**

| Section | Species | Chromosome numbers* | Cultivars | Origin |
|---|---|---|---|---|
| *Lavandula* | *L. angustifolia* | 2n = 36, 42, 48, 50, 54 | Tucker's early purple | OLHF |
| | | | Betty's Blue | OLHF |
| | | | Folgate | OLHF |
| | | | Sachet | OLHF |
| | | | Sharon Robert | OLHF |
| | | | Royal velvet | OLHF |
| | | | Subsp. angustifolia | OLHF |
| | | | Maillette | OLHF |
| | *L. x intermedia* | 2n= 51 | Grosso | OLHF |
| | | unknown | Provence | OLHF |
| | | | Supper | OLHF |
| | | | Hidcote Giant | OLHF |
| | | | Fred Boutin | OLHF |
| | | | Seal | OLHF |
| | | | Abrialli | OLHF |
| | *L. latifolia* | 2n=36, 48, 50, 54, 75 | Latifolia | CUBG |
| *Inter-section hybrid* | *L. x ginginsii* | unknown | Goodwin Creek Grey | GGC |
| *Dentatae* | *L. dentata* | 2n= 42, 44, 45 | Dentata | GGC |
| *Stoechas* | *L. lusitanica* | 2n = 30 | Lusi pink | GGC |
| | *L. stoechas* | | Anouk | GGC |
| *Pterostoechas* | *L. buchii* | 2n= 22 | Jagged | GGC |

*Chromosome numbers were adopted from Upson and Andrew (2004); OLHF = The Okanagan Lavender and Herb Farm, Kelowna, BC; GGC = The Greenery Garden Center, Kelowna, BC; CUBG= Cambridge University Botanic Garden, UK

Table

**Table 2**

| | Flower tissues | | Leaf tissues | | Oil glands | | Overall | |
|---|---|---|---|---|---|---|---|---|
| **Motif types** | **Occurrence** | **%** | **Occurrence** | **%** | **Occurrence** | **%** | **Occurrence** | **%** |
| **Mononucleotides** | | | | | | | | |
| A/T | 105 | 3.03 | 69 | 1.99 | 24 | 0.69 | 198 | 5.72 |
| G/C | 7 | 0.2 | 3 | 0.09 | 2 | 0.06 | 12 | 0.35 |
| ***Subtotal*** | ***112*** | ***3.23*** | ***72*** | ***2.08*** | ***26*** | ***0.75*** | ***210*** | ***6.07*** |
| **Dinucleotides** | | | | | | | | |
| AG/GA/CT/TC | 427 | 12.34 | 259 | 7.49 | 221 | 6.39 | 907 | 26.22 |
| AC/CA/TG/GT | 27 | 0.78 | 10 | 0.29 | 14 | 0.40 | 51 | 1.47 |
| AT/TA | 56 | 1.62 | 38 | 1.10 | 36 | 1.04 | 130 | 3.76 |
| GC/CG | - | - | - | - | - | - | - | - |
| ***Subtotal*** | ***510*** | ***14.74*** | ***307*** | ***8.88*** | ***271*** | ***7.83*** | ***1088*** | ***31.45*** |
| **Trinucleotides** | | | | | | | | |
| AAT/ATA/TAA/ATT/TTA/TAT | 34 | 0.98 | 14 | 0.40 | 27 | 0.78 | 75 | 2.17 |
| AAC/ACA/CAA/GTT/TTG/TGT | 24 | 0.69 | 14 | 0.40 | 13 | 0.38 | 51 | 1.47 |
| AAG/AGA/GAA/CTT/TTC/TCT | 138 | 3.99 | 76 | 2.20 | 66 | 1.91 | 280 | **8.09** |
| ACC/CCA/CAC/GGT/GTG/TGG | 80 | 2.31 | 41 | 1.19 | 31 | 0.90 | 152 | 4.39 |
| AGG/GGA/GAG/CCT/CTC/TCC | 87 | 2.52 | 49 | 1.42 | 44 | 1.27 | 180 | 5.2 |
| ATG/TGA/GAT/CAT/ATC/TCA | 49 | 1.42 | 33 | 0.95 | 34 | 0.98 | 116 | 3.35 |
| ACG/CGA/GAC/CGT/GTC/TCG | 16 | 0.46 | 9 | 0.26 | 13 | 0.38 | 38 | 1.1 |
| AGC/GCA/CAG/GCT/CTG/TGC | 77 | 2.23 | 54 | 1.56 | 37 | 1.07 | 168 | 4.86 |
| AGT/GTA/TAG/ACT/CTA/TAC | 7 | 0.20 | 7 | 0.20 | 5 | 0.14 | 19 | 0.55 |
| GGC/GCG/CGG/GCC/CCG/CGC | 167 | 4.83 | 107 | 3.09 | 72 | 2.08 | 346 | **10.00** |
| ***Subtotal*** | ***679*** | ***19.63*** | ***404*** | ***11.68*** | ***342*** | ***9.88*** | ***1425*** | ***41.19*** |
| **Tetranucleotides** | | | | | | | | |
| TAAA/TTTA | 10 | 0.29 | 3 | 0.09 | 5 | 0.14 | 18 | 0.52 |
| AAAT/ATTT | 11 | 0.32 | 4 | 0.12 | 10 | 0.29 | 25 | 0.72 |
| Others | 148 | 4.28 | 78 | 2.25 | 129 | 3.73 | 355 | 10.26 |
| ***Subtotal*** | ***169*** | ***4.89*** | ***85*** | ***2.46*** | ***144*** | ***4.16*** | ***398*** | ***11.51*** |
| **Pentanucleotides (sum)** | ***67*** | ***1.94*** | ***31*** | ***0.90*** | ***33*** | ***0.95*** | ***131*** | ***3.79*** |
| **Hexanucleotides (sum)** | ***104*** | ***3.01*** | ***54*** | ***1.56*** | ***49*** | ***1.42*** | ***207*** | ***5.98*** |
| **Overall sum** | **1641** | **47.44** | **953** | **27.55** | **865** | **25.01** | **3459** | |

Table

**Table 3**

| | *L. angustifolia* Flowers ESTs | | | *L. angustifolia* Leaves ESTs | | | *L. x intermedia* Oil glands ESTs | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **SSR repeats** | **CDS** | **5' UTR** | **3' UTR** | **CDS** | **5' UTR** | **3' UTR** | **CDS** | **5' UTR** | **3' UTR** | **Total** | **%** |
| **Dinucleotides** | | | | | | | | | | | |
| AT/TA | 5 | 3 | 8 | - | 1 | 1 | 1 | - | 2 | 21 | **8.33** |
| AG/GA/CT/TC | 15 | 32 | 11 | 4 | 9 | 2 | 15 | 11 | - | 99 | **39.29** |
| AC/CA/TG/GT | 1 | 3 | - | - | - | 1 | - | - | - | 5 | 1.98 |
| GC/CG | - | - | - | - | - | - | - | - | - | - | 0.00 |
| **Trinucleotides** | | | | | | | | | | | |
| AAT/ATA/TAA/ATT/TTA/TAT | 1 | 1 | 1 | - | 1 | - | 1 | - | - | 5 | 1.98 |
| AAG/AGA/GAA/CTT/TTC/TCT | 6 | 1 | - | 6 | - | - | - | - | 1 | 14 | **5.56** |
| AAC/ACA/CAA/GTT/TTG/TGT | 1 | - | - | 3 | - | - | - | - | - | 4 | 1.59 |
| ATG/TGA/GAT/CAT/ATC/TCA | - | 1 | - | 1 | - | - | 5 | - | - | 7 | 2.78 |
| AGT/GTA/TAG/ACT/CTA/TAC | 4 | - | - | - | - | - | 1 | - | - | 5 | 1.98 |
| AGG/GGA/GAG/CCT/CTC/TCC | 3 | - | - | 1 | - | - | 2 | - | - | 6 | 2.38 |
| AGC/GCA/CAG/GCT/CTG/TGC | 5 | 1 | - | - | 1 | - | 1 | - | 1 | 9 | 3.57 |
| ACG/CGA/GAC/CGT/GTC/TCG | 1 | - | - | 1 | - | - | - | - | 1 | 3 | *1.19* |
| ACC/CCA/CAC/GGT/GTG/TGG | 6 | - | - | 2 | - | - | 4 | - | - | 12 | 4.76 |
| GGC/GCG/CGG/GCC/CCG/CGC | 3 | 1 | - | 1 | - | 1 | 5 | 3 | - | 14 | **5.56** |
| **Tetranucleotides** | 2 | 1 | - | - | - | - | - | - | - | 3 | *1.19* |
| **Pentanucleotides** | 2 | 4 | - | 1 | 2 | - | 4 | 2 | - | 15 | 5.95 |
| **Hexanucleotides** | 15 | 3 | 2 | 3 | - | - | 7 | - | - | 30 | 11.90 |
| Total | 70 | 51 | 22 | 23 | 14 | 5 | 46 | 16 | 5 | 252 | |
| % | 27.78 | 20.24 | 8.73 | 9.13 | 5.56 | 1.98 | 18.25 | 6.35 | 1.98 | | |

CDS- coding regions and UTR- untranslated regions

Table

**Table 4**

| Genebank Accession No. | Locus | Primers (5'==>3') | Motif | SSR location | Tm * ($^{o}$C) | Expected allele (bp) | Functional annotation | |
|---|---|---|---|---|---|---|---|---|
| | *L. angustifolia* flowers | | | | | | **Protein name** | **Accession no.** |
| KM288520 | LAF1 | F: GGGATTCGACTGCCTTGATA<br>R: AATGCCATTTCCCACCACT | (AG)15 | 3' UTR | 60.04 | 246 | Uncharacterized protein | EYU29981.1 |
| KM288521 | LAF2 | F: GAAAATGATTCCGGACGAGA<br>R:TGGCCCTGGTGATTGAAATA | (TA)11 | 3' UTR | 60.01 | 245 | Cytochrome P450 | AHL46848.1 |
| KM288522 | LAF3 | F: TCTCGATGCAAACTGAATGC<br>R: ATACTCGGTGCCCAGATCAC | (AG)19 | CDS | 59.96 | 230 | ABC transporter G family member 11-like | XP_006351519.1 |
| KM288523 | LAF4 | F: AGCACGACGAGCTTTCAAGT<br>R: TTGTGCTGTTAAACCATAAGTCC | (AT)26 | CDS | 60.2 | 168 | Two-component response regulator-like APRR1-like | XP_006354770.1 |
| KM288524 | LAF5 | F: CAAATGACCCCATCAACAA<br>R: GTATGATCCCATCCCGTGAG | (CT)18 | 5' UTR | 59.75 | 225 | Root phototropism protein | XP_002525040.1 |
| KM288525 | LAF6 | F: GATGATGGGCTACTCGTGGT<br>R: GGCGATGACAGAAGAGAGA | (AGC)9 | CDS | 59.96 | 217 | SQUAMOSA promoter binding protein-like 14 | AIE89803.1 |
| KM288526 | LAF7 | F: GCTGATTCATCTTGGCCTTG<br>R: CAAATCGTTTGGAAGCACA | (ATT)15 | 3' UTR | 60.74 | 219 | Sugar phosphate/ phosphate translocator | XP_004245685.1 |
| KM288527 | LAF8 | F: TGCAATCCTCCTATGTGTTGTC<br>R: TTGGAGTTGCTGTGGATGAG | (CT)15 | CDS | 60 | 228 | No hit | |
| KM288528 | LAF9 | F: GAGCTGCGAGTGTGAGTCAG<br>R: TTTACTTGGGGGACGTTGAG | (AG)13 | CDS | 59.92 | 149 | Uncharacterized protein | EYU38816.1 |
| KM288529 | LAF10 | F: TGTCAGGATCGAAACTCGTG<br>R: ATGCACCTTTGGGATTTCAG | (AT)23 | CDS | 59.83 | 227 | Filament-like plant protein 4 | XP_008221294.1 |
| KM288530 | LAF11 | F: GCAATGTTGGAATGTGATGC<br>R: AAGCGGCAATCTTGGTAGTG | (CA)13 | 5' UTR | 59.94 | 193 | Monodehydroascorbate cytoplasmic isoform 2-like | XP_007031494.1 |
| KM288531 | LAF12 | F: AGAGCAGGCCTCTTTTGCTA<br>R: ATGAAAGGCTCAAGGGCTTC | (ATT)17 | 5' UTR | 59.36 | 194 | No hit | |
| KM288532 | LAF13 | F: TCCTTCTCCCTCTCCCTCTC<br>R: AGCTATCCGAGCACGACAAT | (TCTCCC)5 | 5' UTR | 59.88 | 188 | Cyclin-d1-1-like | CAB61221.1 |
| KM288533 | LAF14 | F: GTCCCACTCCCACAACTCAT<br>R: TCCTTCTTGGTCAACAATTTCA | (CT)35 | 3' UTR | 59.82 | 237 | Uncharacterized protein | XP_002299474.1 |
| KM288534 | LAF15 | F: AAGCGGAAGTGGATTCATGT<br>R: GCAAGATTGCATTAGCACGA | (TGC)7 | CDS | 59.56 | 154 | 39S ribosomal protein L27, mitochondrial-like | XP_006356646.1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| KM288535 | LAF16 | F: CGCTCTATCCCTTTCCCTCT<br>R: AGCCGCTTGTGACTTTCTCT | (TCCCTC)5 | CDS | 59.81 | 139 | E3 ubiquitin-protein ligase | XP_008236747.1 |
| KM288536 | LAF18 | F: TTCACCCGGAATCTTTACCA<br>R: CAAATTCCCTGCAACCAATC | (CGTT)6 | CDS | 60.3 | 256 | Transcription factor ABA-Inducible bHLH-type | XP_003596884.1 |
| KM288537 | LAF19 | F: TCATGGAGCAGCAGAACATC<br>R: CGAAGTCCTGGTCCAATAGC | (GCA)10 | CDS | 59.95 | 153 | Uncharacterized protein | EYU24058.1 |
| KM288538 | LAF20 | F: TATCGGACACCGCTCCAAT<br>R: AAGAGGGGAATCAACCAAGC | (CCG)9 | CDS | 61.43 | 187 | RHOMBOID-like 1 | XP_007027004.1 |
| KM288539 | LAF21 | F: GAGGAGGAGGAGGAAGTGCT<br>R: ATACTCGGTGCCCAGATCAC | (GAC)10 | CDS | 59.95 | 222 | Nucleolin-like | XP_006366238.1 |
| | *L. angustifolia* leaves | | | | | | | |
| KM288540 | LAL2 | F: TTGGCAGCATCAGAGACAAC<br>R: CTTGG AGGTTGGGAAGTTT | (ACA)12 | CDS | 59.99 | 201 | Nucleic acid binding | XP_002531176.1 |
| KM288541 | LAL3 | F: TCTGTTGTTGCCTCCTCTCC<br>R: AAGTGGGAAAACGGAAAAGG | (TCTGC)4 | 5'UTR | 60.39 | 190 | No hit | |
| KM288542 | LAL4 | F: AAGTTTCCCTCTGCCTCCTC<br>R: AGAGGCCGTAGCTGTCTTCA | (AAC)8 | CDS | 59.82 | 186 | Uncharacterized protein | EPS61380.1 |
| | *L. x intermedia* oil glands | | | | | | | |
| KM288543 | LINT1 | F: GGCACATTGGGGTACAAGAT<br>R: GCTGATCCAGGCTTCTCATC | (CAC)10 | CDS | 59.68 | 191 | N-alpha-acetyltransferase 10-like | XP_004514569.1 |
| KM288544 | LINT4 | F: CATCTCCTCCCCTCCCTACT<br>R: CGACGGAGAAGAGTGACGAC | (CT)15 | CDS | 59.51 | 185 | No hit | |
| KM288545 | LINT5 | F: TGCGAGAAGACCGTTTATCC<br>R: CCAAGCATCAAAACAGCAAA | (TC)18 | CDS | 60.21 | 214 | Pollen-specific protein SF3-like | XP_004138434.1 |
| KM288546 | LINT6 | F: TTTTCTCTTTGGTTTCATGAGC<br>R: AATCCTGGTTCCTCGGAGTT | (GA)11 | 5'UTR | 58.49 | 170 | Uncharacterized protein | EYU26600.1 |
| KM288547 | LINT10 | F: TTAGCCGAGGCACTAGAGGA<br>R: CTTGGCCTCCTTCTTCTTCA | (ATG)7 | CDS | 60.11 | 142 | Uncharacterized protein | CDSP13488.1 |
| KM288548 | LINT11 | F: AAGGCGAGAAAAGCATTTGA<br>R: CTGAGCCTTCACGGTTCTTC | (GATTT)4 | 5' UTR | 59.96 | 263 | Ras-related protein raba2a-like | XP_004230439.1 |
| KM288549 | LINT12 | F: GAAACCCTCCAATCCATCCT<br>R: GCCGAAAGCAATTCAATGTT | (ATC)7 | CDS | 60.13 | 226 | 1,8-cineole synthase | AFL03421.1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| KM288550 | LINT14 | F: CCTCACGAACCACACCATC<br>R: TTAGGAAGGAGGAGCAACCA | (CCACA T)6 | CDS | 59.95 | 232 | No hit | |

*Tm = the melting temperature for every primer pair. However, best PCR amplification done using touchdown PCR (TD-PCR) programs: (**A**) Tm= 64-54 $^{o}$C (-1 in every cycle) for 11 cycles and 54 $^{o}$C for 24 cycles worked well for majority, (**B**) 62-52 $^{o}$C (-1 in every cycle) for 11 cycles and 54 $^{o}$C for 24 cycles for LAF2, LAF10, LAF11, LAF15, LAF18, LAF20, LAF21 and LINT12, and (**C**). Both programs were employed for LAF3, LAF7, LAF12, LAF14, LAL3, LINT1 and LINT11.

Table

**Table 5**

| | *L. angustifolia* (n=8) | | | | *L. x intermedia* (n=7) | | | | Allele size ranges (bp) |
|---|---|---|---|---|---|---|---|---|---|
| **Loci name** | $N_a$ | $P_a$ | $H_E$ | *PIC* | $N_a$ | $P_a$ | $H_E$ | *PIC* | |
| LAF1 | 6 | 6 | 0.8 | 0.77 | 6 | 6 | 0.78 | 0.75 | 240-400 |
| LAF2 | 4 | 4 | 0.68 | 0.62 | 2 | 2 | 0.47 | 0.36 | 240-350 |
| LAF3 | - | - | - | - | - | - | - | - | No amplicon |
| LAF4 | 4 | 4 | 0.66 | 0.61 | 3 | 2 | 0.5 | 0.36 | 135-170 |
| LAF5 | 4 | 4 | 0.61 | 0.54 | 4 | 4 | 0.5 | 0.38 | 225-320 |
| LAF6 | 3 | 0 | 0 | 0 | 5 | 1 | 0 | 0 | 150-325 |
| LAF7 | - | - | - | - | - | - | - | - | No amplicon |
| LAF8 | 2 | 2 | 0.38 | 0.3 | 1 | 1 | 0 | 0 | 275-350 |
| LAF9 | 4 | 4 | 0.72 | 0.67 | 3 | 3 | 0.64 | 0.54 | 140-200 |
| LAF10 | 2 | 2 | 0.35 | 0.29 | 2 | 2 | 0.41 | 0.32 | 175-450 |
| LAF11 | 4 | 4 | 0.73 | 0.68 | 4 | 4 | 0.69 | 0.63 | 180-380 |
| LAF12 | - | - | - | - | - | - | - | - | No amplicon |
| LAF13 | 7 | 4 | 0.55 | 0.5 | 7 | 7 | 0.84 | 0.82 | 150-500 |
| LAF14 | - | - | - | - | - | - | - | - | No amplicon |
| LAF15 | 5 | 4 | 0.72 | 0.67 | 5 | 5 | 0.79 | 0.75 | 145-200 |
| LAF16 | 6 | 6 | 0.75 | 0.72 | 1 | 1 | 0 | 0 | 130-160 |
| LAF18 | 3 | 3 | 0.66 | 0.59 | 4 | 4 | 0.7 | 0.64 | 200-300 |
| LAF19 | 5 | 5 | 0.76 | 0.72 | 5 | 5 | 0.75 | 0.71 | 145-200 |
| LAF20 | 5 | 5 | 0.79 | 0.76 | 4 | 4 | 0.73 | 0.69 | 165-260 |
| LAF21 | 4 | 3 | 0.64 | 0.57 | 2 | 1 | 0 | 0 | 140-400 |
| LAL2 | 7 | 7 | 0.8 | 0.77 | 7 | 6 | 0.76 | 0.72 | 170-420 |
| LAL3 | - | - | - | - | - | - | - | - | No amplicon |
| LAL4 | 10 | 8 | 0.83 | 0.81 | 9 | 5 | 0.78 | 0.74 | 160-365 |
| ***Subtotal*** | ***85*** | ***75 (88%)*** | | | ***74*** | ***63 (85%)*** | | | |
| ***Mean*** | ***4.72*** | ***4.17*** | ***0.64*** | ***0.59*** | ***4.11*** | ***3.5*** | ***0.52*** | ***0.47*** | |
| LINT1 | - | - | - | - | - | - | - | - | No amplicon |
| LINT4 | 6 | 5 | 0.78 | 0.74 | 7 | 5 | 0.74 | 0.69 | 130-200 |
| LINT5 | 1 | 1 | 0 | 0 | 2 | 2 | 0.5 | 0.38 | 250-275 |
| LINT6 | 2 | 2 | 0.22 | 0.19 | 3 | 2 | 0.32 | 0.27 | 165-350 |
| LINT10 | 3 | 2 | 0.41 | 0.37 | 3 | 2 | 0.5 | 0.38 | 135-150 |
| LINT11 | - | - | - | - | - | - | - | - | No amplicon |
| LINT12 | 2 | 2 | 0.46 | 0.35 | 2 | 2 | 0.41 | 0.32 | 175-265 |
| LINT14 | 4 | 2 | 0.3 | 0.25 | 7 | 6 | 0.79 | 0.75 | 330-550 |
| ***Subtotal*** | ***18*** | ***14 (78%)*** | ***2.17*** | ***1.9*** | ***24*** | ***19 (79%)*** | ***3.26*** | ***2.79*** | |
| ***Mean*** | ***3*** | ***2.33*** | ***0.36*** | ***0.32*** | ***4*** | ***3.17*** | ***0.54*** | ***0.47*** | |
| **Grand total** | **103** | **89 (86%)** | - | - | **98** | **82 (84%)** | - | - | |
| **Overall mean** | **4.29** | **3.71** | **0.57** | **0.52** | **4.08** | **3.42** | **0.53** | **0.47** | |
| **Max** | **10** | **8** | **0.83** | **0.81** | **9** | **7** | **0.84** | **0.82** | |
| **Min** | **1** | **0** | **0** | **0** | **1** | **1** | **0** | **0** | |

n= number of cultivars; Na= Number of alleles; Pa= Polymorphic alleles; *He*= genetic diversity; PIC= Polymorphic information content

Table

**Table 6**

| **Loci name** | ***L. latifolia*** | ***L. buchii*** | ***L. lusitanica*** | ***L. dentata*** | ***L. x ginginsii*** | ***L stoechas*** | **Allele size (bp)** |
|---|---|---|---|---|---|---|---|
| **LAF2** | + | + | - | + | + | + | 240-350 |
| **LAF4** | + | - | - | - | + | + | 135-170 |
| **LAF5** | + | + | + | + | + | + | 225-320 |
| **LAF6** | + | + | + | + | + | - | 150-325 |
| **LAF8** | + | - | - | - | + | - | 275-350 |
| **LAF9** | + | + | + | + | + | + | 140-200 |
| **LAF11** | + | - | + | + | + | + | 180-380 |
| **LAF13** | + | - | - | + | + | - | 150-500 |
| **LAF15** | + | + | - | - | + | - | 145-200 |
| **LAF18** | + | - | + | + | + | + | 200-300 |
| **LAF19** | + | - | + | + | - | - | 145-200 |
| **LAF20** | + | - | - | + | - | - | 165-260 |
| **LAF21** | + | + | + | + | + | + | 140-400 |
| **LAL2** | + | - | - | - | + | - | 170-420 |
| **LAL4** | + | + | + | + | + | + | 160-365 |
| **LINT4** | + | + | - | + | + | - | 130-200 |
| **LINT6** | + | - | - | + | + | + | 165-350 |
| **LINT14** | + | + | + | + | - | - | 330-550 |
| **TR (%)** | **100** | **50** | **50** | **77.7** | **83.3** | **50** | |

TR= cross-species transferability rate