Chapter from the book *Data Mining Applications in Engineering and Medicine*
Downloaded from: http://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine
Discovering Fragrance Biosynthesis Genes from Vanda Mimi Palmer Using the Expressed Sequence Tag (EST) Approach

Seow-Ling Teh, Janna Ong Abdullah, Parameswari Namasivayam and Rusea Go

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/46833

1. Introduction
1.1. Orchidaceae

Orchidaceae is the largest family of angiosperms with an estimation of 17000 to 35000 species in 880 genera (Chai & Yu, 2007). In Malaysia, more than 230 orchid genera and 4000 species had been discovered (Go et al., 2012). In Peninsular Malaysia, a total of 898 species in 143 genera are currently recognised (Go et al., 2010). The amazing vast diversity of types and forms enable the Orchidaceae to be successfully distributed and colonised almost every habitats worldwide (Arditti, 1992). As a result of selective forces from evolution, orchids are found to be evolved from its ancestral forms and adapted well to their present habitats (Aceto & Gaudio, 2011). Associated with their diverse floral morphology and physiology properties, they have drawn the attention of botanists and scientists for centuries. There are orchids which resemble moths (Phalaenopsis), butterflies (Oncidium papillo), the slippers of Aphrodite or moccasins (Paphiopedilum or Cypripedium), dancing ladies (Oncidium), spiders (Brassia), scorpions (Arachnis) and bees (Ophrys) (Teoh, 1980).

Similar to other angiosperms, two whorls of perianth segment can also be found in orchids. The outer and inner whorls of the orchid flowers consist of three petals and three sepals. The labellum or lip (one of the petals), is distinctly evolved from the other two morphologically and physically (Arditti, 1992). The lifespan of opened-orchid flowers can range from as short as one day to as long as 270 days (Micheneau et al., 2008).

Most orchids are epiphytic that obtain their support from trees but not for nutrition while the rest are terrestrial plants (Rada & Jaimez, 1992). Orchids’ cultivation was famous since 5000 years ago in China where Cymbidium was grown as potted plants, and Vanda and
Aerides were suspended from baskets (Teoh, 1980). Orchid hybrids can be divided into three groups which are hybrid species (produced by selfing a species or by crossing two plants belonging to the same species), inter-specific hybrid (produced from the crossing between two different species belonging to the same genus or by second crossing with other inter-specific hybrids) and inter-generic hybrid (produced from the crossings of orchids belonging to different genera).

Orchid hybrids cultivation started since 1856 by John Dominy (http://www/ionopsis.com/hybridization.htm). Calanthe dominii was created from the crossing between Calanthe masuca and Calanthe furcated. Orchid hybrids made up majority of the commercial orchids in Malaysia and Singapore. There are lots of reasons that led to the wide area coverage of orchid hybrids cultivation, such as the ease of cultivation, free blooming habit, and compactness and fantastic arrays of shapes, colours and flowers (Kishor et al., 2006).

The export of orchids from Malaysia, Thailand and Singapore contributed RM200 million annually in the world floral (Ooi, 2005). According to the Japan Florists’ Telegraph Delivery Association, cut flower in Japan constituted around 13.3% of the imported market in the year 2010 with Malaysia having the largest market share of imports, which accounted for 23.4% (7,648 million yen), followed by Columbia at 19.2% and China at 10.4%. Malaysian orchids consisted of 8.1% of the imported cut flower orchid to Japan.

1.2. Vandaceous orchids

Tropical Asia is the native home for approximately 50 vandaceous orchids species. They are distributed in Sri Lanka and southern India to New Guinea, northern Australia and Solomon Islands, and north to China, Taiwan and the Philippines. Thailand was found to be predominated with 11 vanda species. In Thailand, vanda is a vital commercial orchid. Most of them exhibit monopodial growth where their leaves are varies according to habitat. Vandas have many different colours, and majority are yellowish-brown with dark brown spots.

The vanda has been designated as the ‘Queen Orchid of the East’ due to its robust and large rounded flowers (Teo, 1981). Most of the vandaceous orchids prefer sunlight but some are well adapted to shady areas. Like any other tropical orchids, they require warm temperature with good aeration. The vandaceous orchids take around three and half to 10 years to become mature flowering plants (Kishor et al., 2008). Once matured, this orchid genus blooms every two to three months with the flowers lasting two to three weeks. As a result of land development, 28 orchid species have been listed as endangered species on Appendix II of the Convention on International Trade in Endangered Species (CITES) and prohibited from worldwide export. Among those orchids, two belong to vandas (Vanda coerulea and Vanda sanderiana).

It is impossible to differentiate or identify an orchid species based on the vegetative parts of the plant alone. Hence, a convenient and flower-independent method to allow quick assessment of a given orchid vegetative specimen for species confirmation can be achieved with the help of molecular markers. To date, approximately 50 species are registered under vandas in the Royal Horticultural Society (RHS) database due to their commercial importance.
1.3. *Vanda Mimi Palmer*

*Vanda* Mimi Palmer (VMP) (Figure 1) is one of those hybrids known to be highly sought after mainly for its fragrance rather than its flower shape or colour. It is a hybrid of *Vanda Tan Chay Yan* and *Vanda tessellate* (Roxb.) Hk.f. ex G. Don, and was registered on 1st December 1963 by the Gem Nursery, Singapore. It inherited its sweet fragrance and tri-colour flower from *V. tessellate*, while its terete-shaped is inherited from *V. Tan Chay Yan*. It has purplish green, but brown-speckled, petals and sepals with a dark blue lip. This hybrid is a frequent bloomer throughout the year with maximal floral scent emission when the flower is fully bloomed (an average bloom diameter of 5.0-7.0 cm). Fragrance emission is detected only when the flower bud starts to open as none is detected from the closed bud [average size of 0.8-1.4 cm](Chan *et al*., 2011; Janna *et al*., 2005). It has been recognised for its extraordinary fragrance as evident with awards won from the competitions organised by the Royal Horticultural Society of Thailand in 1993 and the 17th World Orchid Conference in 2002.

*Figure 1. Vanda Mimi Palmer, a hybrid crossing between Vanda Tan Chay Yan and Vanda tessellate (Roxb.) Hk.f. ex G. Don.*

Orchids cultivation entails hard work as the orchids can be easily infected by orchid-infecting viruses. More than 50 orchid-infecting viruses have been detected worldwide, with cymbidium mosaic virus (CymMV) and odontoglossum ringspot tobamovirus (ORSV) infections being the most prevalent (Sherpa *et al*., 2007). Infected orchid cultivars usually exhibit blossoms with brown necrotic streaks and other necrosis symptoms. Infected flowers are smaller in size and poorer in quality. This has caused severe economic damages in the cut flower and potted plant industry (Ajjiukittira *et al*., 2002; Sherpa *et al*., 2007; Vejaratpimol *et al*., 1999). CymMV infection is dominant and extremely stable in Orchidaceae, and it was found to be prevalent in VMP. A screen of our VMP cDNA library revealed a 1.6% contamination rate with CymMV genes (Teh *et al*., 2011). Markers might be useful in facilitating the screening of virus-infected stock plants to minimise losses incurred in the floral industry. So far, the identification of CymMV infection is done through serological,
bioassay or electron microscopy. Those techniques include enzyme-linked immunosorbent assays (ELISA), dot-blot immunoassay (DBIA), rapid immunofilter paper assay (RIPA), immunosorbent electron microscope (ISEM), DIG-labelled cRNA probes, reverse transcription polymerase chain reaction (RT-PCR), quartz crystal microbalance-based DNA biosensors and TaqMan real-time quantitative RT-PCR (Eun et al., 2002; Eun & Wong, 2000; Hsu et al., 1992; Hu et al., 1998; Khentry et al., 2006; Rani et al., 2010).

2. Importance of floral fragrance

Floral fragrance plays various functions in both the floral and vegetative organs. Fragrance is defined as a highly complex component of floral phenotype for its dynamic patterns of emission and chemical composition (Raguso, 2008). Due to their restriction to specific lineages and interactions in species-specific ecology, these have led to their designation as specialized or secondary metabolites (Pichersky et al., 2006).

Floral fragrance has a significant impact in plant reproduction as it is a selective attractant in a variety of animal pollinators especially insects. Pollinators such as bees, butterflies and moths are able to discriminate visitation on plants based on the compositions of the floral scent. This plant-insect interaction has led to many successful pollinations and development of fruits in many crop species (Majetic et al., 2009; Shuttleworth & Johnson, 2009).

Anti-microbial or anti-herbivore properties of floral volatiles could be used by plants to protect their vital floral reproductive parts from potential predators. Two types of plant defences can be characterised based on floral volatiles property, that are direct and indirect defences such as herbivore-induced volatiles signals, and visual and olfactory floral signals to attract pollinators (Schiestl, 2010). Indirect plant defences protect the plants by minimizing damage to plant tissues through attracting arthropods that prey or parasitize the herbivores. This general property has been reported in cabbage (Park et al., 2005), Lotus japonicas (Arimura et al., 2004) and cucumber (Agrawal et al., 2002). On the other hand, direct defences take place when herbivore-induced volatiles repel or intoxicate the herbivores and pathogens. For example, some herbivore-induced monoterpenes and sesquiterpenes tend to react with various reactive oxygen species and thus protect the plants from internal oxidative damage (Dudareva et al., 2004). Terp et al. (2006) reported that lipoxygenases are produced in Brassica napus seeds upon wounding and pathogen attack. Complexity in floral scent chemicals was found to be useful in protecting the plants’ reproductive structures from herbivores and ants (Schiestl, 2010).

Through the discovery of pollinator-attracting floral scents as the source of olfactory pleasure since ancient times, humans had figured out unique values in certain types of floral scents and exploited them to cultivate and propagate specific plant species. For centuries, flowers with vibrant colours and scents have been used by people to enhance their beauty and this was seen in almost all major civilizations. Large numbers of aromatic plants have been used as flavourings, preservatives and herbal remedies (Pichersky et al., 2006). Their economic importance also relies on their petals which are found to be the main site of natural fragrances and flavourings in most of the plants (Baudino et al., 2007).
2.1. Fragrance biosynthesis pathways

Plants are known with the capability of synthesizing many volatile metabolites, either primary or secondary metabolites with variety of functions (Pichersky et al., 2006). However, volatile esters formation is not restricted to plant kingdom but also in yeast and fungi especially in the fermentation industry (Beekwilder et al., 2004). Floral scent is made up of a complex mixture of low-molecular-weight lipophilic compounds which are typically liquids with high vapour pressures (Vainstein et al., 2001). With the discovery of novel techniques including gas chromatography-nuclear magnetic resonance (GC-NMR), gas chromatography-mass spectrometry (GC-MS), headspace based techniques in volatiles detection and analyses, the number of identified volatile compounds has increased tremendously (Gonzalez-Mas et al., 2008; Mohd-Hairul et al., 2010; Nojima et al., 2011).

Most of the volatile compounds are derived from three major biosynthesis pathways: phenylpropanoids, fatty acid derivatives and terpenoids. They are thus classified into three major categories: terpenes, lipid derivatives and aromatic compounds. The terpenes are the largest class of plant volatiles, which consist of monoterpenene alcohols and sesquiterpenes. There are also other terpene derivatives like ketones that are present in low quantities but have significant contributions to the floral fragrance. The second largest family of plant volatiles is the aromatic compounds. Most of them are derived from the intermediates in the benzenoid biosynthesis pathway that resulted in the synthesis of phenylalanine from the shikimate pathway, followed by a wide range of primary metabolites (eugenol, a lignin precursor, is one of them) and secondary non-volatile compounds (this was well reviewed in Bick & Lange, 2003; Pichersky & Dudareva, 2007).

The third category of plant volatiles is the lipid derivatives from the oxidative cleavage and decarboxylation of a variety of fatty acids which lead to shorter-chain volatiles with aldehyde and ketone moieties formation (reviewed in Baysal & Demirdoven, 2007). There are also other plant volatiles especially those with nitrogen or sulfur, which are produced through the cleavage of modified amino acids or their precursors (Cherri et al., 2007; Pichersky et al., 2006).

2.2. Discovery of fragrance-related genes

Due to the invisibility of floral scents and its dynamic nature, the study on flower scent is limited. There is no convenient plant model system that allows chemical and biochemical studies on floral scents. The well-established Arabidopsis as a plant model system failed to serve this purpose as the detection of volatiles production by its flower was barely detectable (Vainstein et al., 2001). To date, the characterization and elucidation of enzymes and genes involved in flower scent production are still not as advanced as the biochemical study on the scent constituents. Petunia hybrid is one of the plant model systems that has been used to study the biological importance of floral scent (Spitcer et al., 2007). However, limited information is available for the floral fragrance synthesized in petunia flowers as well as its cell specificity in fragrance production (van Moerkercke et al., 2012).
‘Scent’ enzymes can be identified through in vitro characterization of their enzymatic activities with substrates predicted to be the precursors of known products in the tissues from which they are derived (Dudareva et al., 2004). Besides, they can also be determined by identifying the homologue sequences in genomic or EST databases, expressing the protein in heterologous system like bacterial system followed by biochemical testing of the enzymatic activities with various substrates (Adelene et al., 2012; Bradbury, 2007).

For more than 400 orchids (including both species and hybrids) that were discovered to emit fragrance (Frowine, 2005), in-depth scientific studies on orchid fragrance barely covered 2 percents of the fragrant orchids. Sadly, fragrance study in orchid is not as well established as in other flowers such as rose (Guterman et al., 2002), petunia (Verdonk et al., 2003) and Clarkia breweri (Dudareva et al., 1996; Dudareva et al., 1998). The advent of GC-MS technique enabled the evaluation of floral scent components quantitatively and qualitatively (Vainstein et al., 2001). Mohd-Hairul et al. (2010) showed that the scent of VMP was predominated by terpenoid, benzenoid and phenylpropanoid compounds through GC-MS analysis (Table 1). It was interesting to note that a comparison of the volatiles captured from VMP with both of its parents revealed that the scent was dissimilar to its fragrance parent, V. tessellate. Such biochemical data is useful for subsequent work in the identification of fragrance-related genes in VMP.

3. Expressed sequence tags (ESTs)

Expressed Sequence Tags (ESTs) are partial sequences generated from single-pass sequencing (5’- or 3’- end) from a reverse transcription of mRNA representing tissue of interest or a particular developmental stage of an organism (Adam et al., 1991). In plant, EST approach was first used in the model plant A. thaliana for genome and in vivo analysis of their gene products (Hofte et al., 1994). Generally, the generated ESTs are highly informative at the middle DNA sequences but its redundancy rate at both ends can reach up to 3% (Nagaraj et al., 2006; Rudd, 2003). Thus, pre-processing of ESTs is required prior to further analyses.

EST libraries for many plant species such as P. equestris (Schauer) Rchb.f. (Tsai et al., 2006; Tsai et al., 2011), cycads (Brenner et al., 2003), mints (Lindqvist et al., 2006), apple (Newcomb et al., 2006), Jatropha curcas (Chen et al., 2011) and Prunus mume (Li et al., 2010) have been developed and deposited into the GenBank database. One of the databases for the submission and deposition of ESTs sequences is the database of expressed sequence tags (dbEST). dbEST is a collection of cDNA sequences and ESTs information from different variety of organisms. From the dbEST released on April 1, 2012 (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), the publicly available ESTs data were 72,316,247.

3.1. Importance of ESTs

EST is a fast, efficient and valuable tool for gene expression, genome annotation and evolitional studies. Analysis of ESTs provides a platform for functional genomics study as well as uncovering the potentially novel genes, and poses an avenue for genome sequencing
projects (Ayeh, 2008; Kisiel & Podkowinski, 2005; Li et al., 2010; Lindqvist et al., 2006). Early EST projects were focused mostly on economically important plants and crop plants. In subsequent years, more EST projects on plants yet-to-achieve high economical impact started to materialise. The development of ESTs for some of the plants species from the early years until now is summarised in Table 2.

EST also proves to be a beneficial resource for comparative genomic studies in plant. Hsiao et al. (2006) deduced monoterpene biosynthesis pathway and identified a few fragrance-related genes in Phalaenopsis bellina by making a comparison with the floral EST library of P. equestris (scentless species). EST is also useful in the development and mining of microsatellite markers such as simple sequence repeats (SSRs) (Alba et al., 2004). Lindqvist et al. (2006) reported the development of mint’s EST for the identification of genetic markers for Hawaiian endemic mints. Moreover, EST provides the basis for the understanding of metabolic regulation mechanisms (Chen et al., 2011; Li et al., 2010).

<table>
<thead>
<tr>
<th>Peak</th>
<th>Relative retention time (min)</th>
<th>Main spectrum fragments (m/z)</th>
<th>Compound name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monoterpene</td>
<td>1</td>
<td>8.636</td>
<td>36,41,53,67,79,93,105,121</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.147</td>
<td>41,43,59,81,93,112</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.592</td>
<td>41,43,69,71,93,107,121,136</td>
</tr>
<tr>
<td>Sesquiterpene</td>
<td>10</td>
<td>16.492</td>
<td>41,43,69,71,93,107,123,136, 162</td>
</tr>
<tr>
<td>Benzenoid</td>
<td>4</td>
<td>9.702</td>
<td>51,77,105,136</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10.030</td>
<td>39,51,65,78,91,105,122</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>11.926</td>
<td>39,43,65,78,91,104</td>
</tr>
<tr>
<td>Indole</td>
<td>9</td>
<td>13.084</td>
<td>39,50,63,74,90,117</td>
</tr>
<tr>
<td>Formanilide</td>
<td>7</td>
<td>12.260</td>
<td>39,52,65,76,93,161</td>
</tr>
</tbody>
</table>

Table 1. Volatile compounds emitted by fully open flower of Vanda Mimi Palmer with their relative retention times and spectral fragments. This table is adapted from Mohd-Hairul et al. (2010).
<table>
<thead>
<tr>
<th>Organism</th>
<th>ESTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zea mays (maize)</td>
<td>2,019,114</td>
</tr>
<tr>
<td>Arabidopsis thaliana (thale cress)</td>
<td>1,529,700</td>
</tr>
<tr>
<td>Glycine max (soybean)</td>
<td>1,461,624</td>
</tr>
<tr>
<td>Oryza sativa (rice)</td>
<td>1,252,989</td>
</tr>
<tr>
<td>Triticum aestivum (wheat)</td>
<td>1,073,877</td>
</tr>
<tr>
<td>Panicum virgatum (switchgrass)</td>
<td>720,590</td>
</tr>
<tr>
<td>Brassica napus (oilseed rape)</td>
<td>643,874</td>
</tr>
<tr>
<td>Hordeum vulgare + subsp. vulgare (barley)</td>
<td>501,838</td>
</tr>
<tr>
<td>Vitis vinifera (wine grape)</td>
<td>446,639</td>
</tr>
<tr>
<td>Nicotiana tabacum (tobacco)</td>
<td>334,384</td>
</tr>
<tr>
<td>Pinus taeda (loblolly pine)</td>
<td>328,662</td>
</tr>
<tr>
<td>Malus x domestica (apple tree)</td>
<td>324,742</td>
</tr>
<tr>
<td>Picea glauca (white spruce)</td>
<td>313,110</td>
</tr>
<tr>
<td>Gossypium hirsutum (upland cotton)</td>
<td>297,239</td>
</tr>
<tr>
<td>Solanum lycopersicum (tomato)</td>
<td>297,142</td>
</tr>
<tr>
<td>Medicago truncatula (barrel medic)</td>
<td>269,238</td>
</tr>
<tr>
<td>Solanum tuberosum (potato)</td>
<td>249,761</td>
</tr>
<tr>
<td>Lotus japonicas</td>
<td>242,432</td>
</tr>
<tr>
<td>Mimulus guttatus</td>
<td>231,095</td>
</tr>
<tr>
<td>Raphanus sativus (radish)</td>
<td>110,006</td>
</tr>
<tr>
<td>Petunia x hybrid</td>
<td>50,705</td>
</tr>
<tr>
<td>Zingiber sp.</td>
<td>38,190</td>
</tr>
<tr>
<td>Elaeis guineensis</td>
<td>40,737</td>
</tr>
<tr>
<td>Petunia axillaris subsp. axillaris</td>
<td>11,078</td>
</tr>
<tr>
<td>Phalaenopsis equestris</td>
<td>5,604</td>
</tr>
<tr>
<td>Rosa hybrid cultivar</td>
<td>5,565</td>
</tr>
<tr>
<td>Phalaenopsis violacea</td>
<td>2,359</td>
</tr>
<tr>
<td>Dendrobium officinale</td>
<td>800</td>
</tr>
<tr>
<td>Oncidium hybrid cultivar</td>
<td>280</td>
</tr>
<tr>
<td>Phalaenopsis amabilis</td>
<td>103</td>
</tr>
</tbody>
</table>

*Table 2.* Summary of some of the plants species available in dbEST from the early years until 1 April 2012.
3.2. Discovery of fragrance-related genes from VMPESTs

Expressed sequence tags (ESTs) have been developed from many monocot plant species including *P. equestris* (Tsai *et al.*, 2006), ginger (Chandasekar *et al.*, 2009), oil palm (Low *et al.*, 2008), wheat (Zhang *et al.*, 2004) and maize (Fernandes *et al.*, 2002). To date, more than 1700 floral fragrance compounds have been identified through biochemical analysis with the majority identified from higher plants whilst others are from animals, insects, marine organisms, algae and bacteria. Unfortunately, the developments of the biochemical aspects of fragrance compounds are not in par with the molecular information on fragrance-related genes. More so, reported molecular information from the vandaceous orchids is still extremely scarce. Thus far, the only available fragrance-related molecular works on orchid are from the genera *Phalaenopsis* and *Vanda* (from our research group).

A VMP floral cDNA library was previously constructed from opened flowers at different developmental stages and time-points (Chan *et al.*, 2009). All the cDNA clones with the inserts sizes of 0.5 kb to 1.6 kb were mass excised and single-pass 5’-sequenced. From our attempt, a total of 2,132 ESTs was generated. This VMP dbEST (designated as VMPEST) was clustered, annotated and further classified with Gene Ontology (GO) identifier into three categories: Molecular Functions (51.2%), Cellular Components (16.4%) and Biological Processes (24.6%). Around 3.1% of the VMPEST had hits with other orchid species such as dendrobium, phalaenopsis, oncidium, *Aerides japonia*, and *Aranda Deborah*. A number of fragrance-related transcripts were identified (Table 3; Teh, 2011) by comparing Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with the three major volatiles biosynthesis pathways (terpenoid, benzenoid and phenylpropanoid) of well-studied scented flowers such as *Antirrhinum majus*, *Clarkia breweri*, *Petunia* hybrid and *Rosa* hybrid (Boatright *et al.*, 2004; Lavid *et al.*, 2002).

From the VMPEST, several fragrance-related transcripts were selected for full-length isolation and expression analysis using real-time quantitative RT-PCR. They were *Vanda Mimi Palmer* acetyl-CoA acetyltransferase (designated as VMPACA, in press), *Vanda Mimi Palmer* 3-hydroxy-3-methylglutaryl-coenzyme A reductase (VMPHMGR), *Vanda Mimi Palmer* 1-deoxy-D-xylulose 5-phosphate synthase (VMPDXPS), *Vanda Mimi Palmer* linalool synthase (VMPLis) and *Vanda Mimi Palmer* lipoxigenase (VMPLox) (Teh, 2011). Gene specific primers were designed and synthesised for 5’-RACE targeting at the incomplete 5’-ends of each of the aforementioned gene transcript. The full length cDNA sequences of those transcripts were deduced, amplified, followed by sequencing and sequence analysis. Real-time quantitative RT-PCR were performed using cDNA templates extracted from different types of VMP tissues (sepals, petals, lips, stems, stalks, columns, roots and leaves), from full-bloom flowers taken at different time points (2-hour intervals within 24-hour duration) and from different flower developmental stages (buds, blooming and full-bloom flower). Overall, the expression profiles revealed that the transcripts exhibited a developmentally regulated pattern in the different flower developmental stages. Majority of the transcripts were highly expressed in the full-bloom stage, with the sepals and petals having the highest expression levels.
<table>
<thead>
<tr>
<th>GenBank Accession</th>
<th>dbEST id</th>
<th>Putative identity</th>
<th>E-value</th>
<th>Score (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GW392501</td>
<td>68671343</td>
<td>3-hydroxy-3-methylglutaryl-coenzyme A reductase 2 [Gossypium hirsutum]</td>
<td>2.00E-108</td>
<td>358</td>
</tr>
<tr>
<td>GW392566</td>
<td>68671408</td>
<td>LOX1 (Lipoxygenase 1); lipoxygenase [Arabidopsis thaliana]</td>
<td>1.00E-50</td>
<td>203</td>
</tr>
<tr>
<td>GW392695</td>
<td>68671537</td>
<td>linalool synthase 2 [Clarkia breweri]</td>
<td>9.00E-19</td>
<td>97.8</td>
</tr>
<tr>
<td>GW392657</td>
<td>68671499</td>
<td>trans-caffeoyl-CoA 3-O-methyltransferase [Populus trichocarpa]</td>
<td>2.00E-52</td>
<td>209</td>
</tr>
<tr>
<td>GW392731</td>
<td>68671573</td>
<td>lipoxygenase 1 [Brassica napus]</td>
<td>2.00E-15</td>
<td>85.9</td>
</tr>
<tr>
<td>GW392740</td>
<td>68671582</td>
<td>acyltransferase [Vanda hybrid cultivar]</td>
<td>3.00E-70</td>
<td>268</td>
</tr>
<tr>
<td>GW392813</td>
<td>68671655</td>
<td>cinnamyl alcohol dehydrogenase [Populus trichocarpa]</td>
<td>2.00E-73</td>
<td>279</td>
</tr>
<tr>
<td>GW393688</td>
<td>68672530</td>
<td>carboxyl methyltransferase [Crocus sativus]</td>
<td>5.00E-58</td>
<td>228</td>
</tr>
<tr>
<td>GW392895</td>
<td>68671737</td>
<td>cinnamyl-CoA reductase [Saccharum officinarum]</td>
<td>2.00E-46</td>
<td>189</td>
</tr>
<tr>
<td>GW392922</td>
<td>68671764</td>
<td>putative 1-deoxy-D-xylulose 5-phosphate synthase [Hevea brasiliensis]</td>
<td>6.00E-62</td>
<td>241</td>
</tr>
<tr>
<td>GW393960</td>
<td>68672802</td>
<td>putative acetyl-CoA C-acyltransferase [Oryza sativa Japonica Group]</td>
<td>4.00E-103</td>
<td>378</td>
</tr>
<tr>
<td>GW393331</td>
<td>68672173</td>
<td>lipoxygenase 1 [Brassica napus]</td>
<td>4.00E-27</td>
<td>125</td>
</tr>
<tr>
<td>GW393619</td>
<td>68672461</td>
<td>resveratrol O-methyltransferase [Vitis vinifera]</td>
<td>8.00E-56</td>
<td>221</td>
</tr>
<tr>
<td>GW393628</td>
<td>68672470</td>
<td>s-adenosylmethionine synthetase, putative [Ricinus communis]</td>
<td>2.00E-59</td>
<td>233</td>
</tr>
<tr>
<td>GW394168</td>
<td>68673010</td>
<td>linalool synthase-like protein [Oenothera arizonica]</td>
<td>6.00E-41</td>
<td>171</td>
</tr>
<tr>
<td>GW393499</td>
<td>68672341</td>
<td>acetyl-CoA acetyltransferase, cytosolic 1 [Zea mays]</td>
<td>1.00E-136</td>
<td>488</td>
</tr>
</tbody>
</table>

**Table 3.** Selected putative fragrance-related ESTs generated from *Vanda* Mimi Palmer. This table is adapted from Teh et al. (2011).

Among all the transcripts analysed, VMPHMGR was selected for functional analysis. HMGR is a very well-studied enzyme in cholesterol synthesis especially in animal system. A functional enzymatic assay performed showed VMPHMGR was functionally active in *Escherichia coli*, catalysing the conversion of HMG-CoA to mevalonate derivatives, which are commonly used in the metabolic biosynthesis of steroids, terpenoids and carotenoids (Teh, 2011). This VMPHMGR shows high sequence identity to the HMGR found in other plant species, with 76% sequence similarity with *Oryza sativa* Indica group. Although its N-terminal end differs distinctly in length and amino acids compositions, its C-terminal catalytic domain shows high sequence similarity with other plant species (unpublished data). This highlights the importance of information and evolution of gene of similar function in different organisms.
4. Simple sequence repeat and its importance

Simple Sequence Repeat (SSR) or microsatellite is a short tandem repeats of a unique DNA sequence with one to six nucleotides motif (Jacob et al., 1991). SSR is a famous molecular marker because of its hyper variability, relative abundance, highly reproducible, multiallelic diversity, co-dominantly inherited and extensive coverage of the genome (Mohan et al., 1997).

Owing to its desirable genetic attributes, SSRs have been utilized in genetic and genomic analyses including genetic mapping, marker assisted plant breeding, development of linkage map, and ecology studies (Kalai et al., 2011; Sonah et al., 2011; Yue et al., 2006). Yue et al. (2006) reported the usage of SSRs in the protection of new Dendrobium varieties.

To date, the used of SSRs markers have been reported in several monocot and dicot species including raspberry and blackberry (Stafne et al., 2005), rice (Chakravarthi & Naravaneni, 2006), common bean (Yu et al., 2000), Brachypodium (Sonah et al., 2011) and Dendrobium (Yue et al., 2006). However, the effort to develop SSR markers for orchids is limited to several species: Chinese orchid [Cymbidium spp.] (Huang et al., 2010), Phalaenopsis (Bory et al., 2008; Hsu et al., 2011), Brazilian orchid (Epipendrum fulgens) (Pinheiro et al., 2008), Dendrobium (Yue et al., 2006) and Vanda (Phuekvilai et al., 2009).

So far, the reported SSRs generated from vandaceous orchids were used as selective marker only. Phuekvilai et al. (2009) generated SSRs from 33 vandas species for the sole purpose of identifying and evaluating the purity of cultivar in commercial samples. However, the identification and development of SSRs for VMP will be channelled towards facilitating the screening of any potential fragrance-related transcripts from closely related species. Besides, it will be used to determine the extent of inter-species transferability of genes, which had been reported in many plant species (Chapman et al., 2009; Stafne et al., 2005; Wang et al., 2004).

4.1. Data mining of VMPEST-SSR

In recent years, genic microsatellite or EST-SSRs which is less time consuming and relatively easy to develop has replaced the genomics SSRs (Sharma et al., 2007). The publicly available ESTs sequences facilitate the development of SSRs by using the SSR identification tools. These search tools include MISA (MircroSatellite), SSRIT (SSR Identification Tool), SciRoKo, TRF (Tandem Repeat Finder), Sputnik, SSRfinder, SAT (SSR Analysis Tool), Poly and SSR Primer. It is deemed important to choose a search tool which is user-friendly and has unlimited access to a non-redundant database (this was well reviewed by Kalai et al., 2011). The first attempt to develop such EST-SSR marker was in rice by Miyao et al. (1996).

SSRIT which is accessible at URL (http://www.gramene.org) was used to identify the SSR motifs in our VMPEST. The script assessed the sequences uploaded in FASTA-formatted files and detected the SSR motifs, the number of repeats as well as identified the sequence corresponded to the SSRs. A total of 98 (9.4%) unigenes containing 112 SSRs with motifs length ranging from two to six nucleotides were detected from VMPEST.
The VMPEST-SSRs were classified into 2 groups with 88.4% belonging to Class I \( (n \geq 20 \text{ nucleotides}) \) and 11.6% belonging to Class II \( (12 \leq n \leq 20 \text{ nucleotides}) \) according to their lengths and genetic marker potential (Teh et al., 2011). Such groupings had been reported in rice by Temnykh et al. (2001). In their study, they revealed that higher rates of polymorphism occurred in transcripts with longer SSR sequences. Likewise, Song et al. (2010) reported that SSR which covered longer portion of sequence (repeat number less than 35) were deemed better for development of genetic markers because they associated with expressed portion of the genome, thus facilitate better understanding of associated protein(s). From our study, majority of the EST-SSR sequences were categorised as ‘ideal’ repeat \( (n \geq 20 \text{ nucleotides}) \), which correlated well to the SSRs mined from tea \([Camellia sinensis L.]\) (Sharma et al., 2009).

Sharma et al. (2009) stressed that the length of repeats and the tools used in the EST-SSRs mining play a significant role in EST-SSR occurrences. The di-nucleotide motif (AT/TA) was present with the most abundance (33.9%) in our VMPEST-SSR. Such observation of A- and T-rich signatures being the most common repeat motifs in the VMPEST-SSRs is also reflected in the early findings of Chagne et al. (2004), Lagercrantz et al. (1993), and Morgante & Olivieri (1993). Interestingly, Blair et al. (2009) found that this motif occurred mainly in the 3’-end of the common bean cDNA clones.

Nevertheless, whatever mined results we obtained from our study, each SSR needs to be validated with further analyses such as selection of SSR primers and screening for utility in vandaceous orchids.

Besides EST-SSR, there are other alternative data mining techniques such as expressed sequence tag-single nucleotide polymorphism (EST-SNP) in maize (Batley et al., 2003), barley (Varshney et al., 2007), melon (Deleu et al., 2009), citrus (Jiang et al., 2010), and cocoa (Allegre et al., 2011), expressed sequence tag-rapid amplified polymorphic DNA (EST-RAPD) in oil palm (Balakrishnan & Vadivel, 2012), and expressed sequence tag-sequence tagged site (EST-STS) in wheat (Leonard et al., 2008; Naji et al., 2008). All the aforementioned techniques are useful for functional genetic diversity estimation of GenBank collections and valuable for use in marker-assisted programmes. However, each technique has its good and down sides that might affect its eventual development.

5. Conclusion

The VMPEST dataset is a potential asset in facilitating the molecular biology and cloning of more genes involved in the fragrance biosynthesis pathway(s). Several fragrance-related transcripts were identified from our VMPEST including \( VMPACA, VMPHMG, VMPDXP, VMPLox \) and \( VMPLis \). The functional enzymatic assay that was performed on the selected transcript \( (VMPHMG) \) proved to be functionally active in its catalysis reaction in a heterologous system. The detected SSRs loci and microsatellite motifs that had hits with fragrance-related genes in the GenBank are believed to be a valuable resource especially to researchers involved in studying diversity to access the functional diversity of fragrant vandaceous orchids and their linkages to other orchids.
Discovering Fragrance Biosynthesis Genes from Vanda Mimi Palmer Using the Expressed Sequence Tag (EST) Approach

Author details

Seow-Ling Teh and Janna Ong Abdullah
Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Malaysia

Parameswari Namasivayam
Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Malaysia

Rusea Go
Biology Department, Faculty of Science, Universiti Putra Malaysia, Malaysia

Acknowledgement

Authors would like to thank the Ministry of Higher Education Malaysia and Universiti Putra Malaysia for financial support through the Fundamental Research Grant Scheme (02-12-10-1002FR) and Research University Grant Scheme (05-04-0551RU), respectively.

6. References


Discovering Fragrance Biosynthesis Genes from Vanda Mimi Palmer Using the Expressed Sequence Tag (EST) Approach


