

# Eugenol synthase genes in floral scent variation in *Gymnadenia* species

Alok K. Gupta · Ines Schauvinhold · Eran Pichersky · Florian P. Schiestl

Received: 9 July 2013 / Revised: 7 August 2014 / Accepted: 31 August 2014 / Published online: 20 September 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Floral signaling, especially through floral scent, is often highly complex, and little is known about the molecular mechanisms and evolutionary causes of this complexity. In this study, we focused on the evolution of “floral scent genes” and the associated changes in their functions in three closely related orchid species of the genus *Gymnadenia*. We developed a benchmark repertoire of 2,571 expressed sequence tags (ESTs) in *Gymnadenia odoratissima*. For the functional characterization and evolutionary analysis, we focused on eugenol synthase, as eugenol is a widespread and important scent compound. We obtained complete coding complementary DNAs (cDNAs) of two copies of putative *eugenol synthase* genes in each of the three species. The proteins encoded by these cDNAs were characterized by expression and testing for

activity in *Escherichia coli*. While *G. odoratissima* and *Gymnadenia conopsea* enzymes were found to catalyze the formation of eugenol only, the *Gymnadenia densiflora* proteins synthesize eugenol, as well as a smaller amount of isoeugenol. Finally, we showed that the eugenol and isoeugenol producing gene copies of *G. densiflora* are evolutionarily derived from the ancestral genes of the other species producing only eugenol. The evolutionary switch from production of one to two compounds evolved under relaxed purifying selection. In conclusion, our study shows the molecular bases of eugenol and isoeugenol production and suggests that an evolutionary transition in a single gene can lead to an increased complexity in floral scent emitted by plants.

**Keywords** ESTs · Eugenol · Floral scents · *Gymnadenia* · Isoeugenol · Pollination · VOC

**Electronic supplementary material** The online version of this article (doi:10.1007/s10142-014-0397-9) contains supplementary material, which is available to authorized users.

A. K. Gupta (✉) · F. P. Schiestl  
Institute of Systematic Botany, University of Zurich, 8008 Zurich, Switzerland  
e-mail: ethalok@yahoo.com

A. K. Gupta  
Institute of Integrative Biology, Plant Ecological Genetics, ETH Zurich, 8092 Zurich, Switzerland

I. Schauvinhold · E. Pichersky  
Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI, USA

I. Schauvinhold  
Laboratorio de Cinética y Fotoquímica, Centro de Investigaciones y Transferencia de Sgo. del Estero, Universidad Nacional de Santiago del Estero, Santiago del Estero, Argentina

*Present Address:*

A. K. Gupta · I. Schauvinhold · E. Pichersky  
Department of Ecology and Evolution, University of Lausanne, Le Biophore, 1015 Lausanne, Switzerland

## Introduction

Many insect-pollinated plants communicate with their pollinators through a number of floral signals, such as fragrance, unique flower shape, and color (Delle-Vedove et al. 2011; Huber et al. 2005; Salzmann et al. 2007; Salzmann and Schiestl 2007; Schiestl and Glaser 2012; Streinzer et al. 2010; van der Niet et al. 2011a, b; Vereecken and Schiestl 2009). Floral scents are among the key signals in many plant-pollinator systems for attracting pollinators from both short and long distances (Brodmann et al. 2008; Huber et al. 2005; Majetic et al. 2007; Miyake and Yafuso 2003; Negre et al. 2003; Plepys et al. 2002) and play a vital role in the reproductive success of night-flowering plants, wherein visual cues are usually of minor importance for mediating interactions with pollinators (Balao et al. 2011; Dötterl and Jürgens 2005; Gupta et al. 2012; Kondo et al. 2006; Wälti et al. 2009). In addition, several studies have established that floral volatiles can serve multiple

functions, such as attracting pollinators and deterring the nectar thieves simultaneously (Junker and Blüthgen 2010; Shuttleworth and Johnson 2009; Theis 2006; Theis et al. 2009). In plant signaling, volatile blends derived from vegetative tissues are generally assumed to enhance predator-mediated defenses, whereas floral-based chemicals are strictly considered insect attractants (Heil 2008; Pichersky and Gershenzon 2002; Raguso 2008; van Dam 2009). Increasing evidence, however, suggests that floral-specific chemical substances can also play a central role in guarding the reproductive organs from the antagonists (Kessler and Baldwin 2007; Kessler et al. 2013; Schiestl et al. 2011).

*Gymnadenia odoratissima*, *Gymnadenia conopsea*, and *Gymnadenia densiflora* rarely hybridize despite being sympatrically distributed (Lönn et al. 2006; Schiestl and Schlüter 2009; Soliva and Widmer 1999). Therefore, besides postzygotic isolation barriers, floral scent has been argued to be one of the most crucial factors to maintain reproductive isolation among these species (Huber et al. 2005; Schiestl and Schlüter 2009). Although three *Gymnadenia* species emit quantitatively and qualitatively different blends of approximately 50 volatile compounds to attract different suits of pollinators, eugenol and benzyl acetate are the only two shared active compounds among the scent bouquets of these species (Huber et al. 2005; Schiestl et al. 2011). However, in addition to these active compounds, other compounds found in blend can also influence plant performance. For example, Schiestl et al. (2011) uncovered that the key active compounds (e.g., eugenol and phenylacetaldehyde) emitted by *G. odoratissima* are positively associated with fitness. However, “non-active” compounds, such as (*Z*)-isoeugenol, limonene, and  $\alpha$ -pinene are negatively correlated with fitness, possibly by interaction with blends of active volatile compounds that altogether can reduce pollinator visitations or increase florivores attractiveness.

In orchids, most of the previous molecular studies have predominantly made progress either in identifying MADS-box genes for flower development (Aceto and Gaudio 2011; Mondragón-Palomino and Theißen 2009; Mondragón-Palomino and Theißen 2011; Mondragón-Palomino and Theissen 2008; Xu et al. 2006) or in molecular phylogeny (Bellusci et al. 2008; Douzery et al. 1999; Inda et al. 2012; Mant et al. 2002; Soliva et al. 2001; van der Niet et al. 2011b). Our current knowledge of biosynthetic pathways that produce scent compounds in orchids, however, is still in its infancy. The food rewarding *Gymnadenia* species emits a large amount of different scent compounds, and therefore, it provides an excellent model system to understand the evolutionary processes underlying striking scent variability and to study the genes responsible for scent biosynthesis. Surprisingly, to date, nothing is known about the scent biosynthesis in this species.

Despite the increasing number of expressed sequence tags (ESTs) and genomic resources for several orchid species in public databases, sequence information for *G. odoratissima*

had remained elusive until the present study (Chan et al. 2011; Fu et al. 2012; Hsiao et al. 2006, 2011; Monteiro et al. 2012; Pan et al. 2012; Sedeek et al. 2013; Teh et al. 2011). Therefore, to characterize genes involved in floral- and scent-related pathways, we have constructed one standard and one subtracted complementary DNA (cDNA) library and developed a benchmark EST resource of 2,571 sequences from *G. odoratissima*. The floral scent compounds, eugenol and isoeugenol, which constitute a group of phenylpropene compounds, are structurally distinguished from each other only by the double-bond position of the propene side chain (Koeduka et al. 2006, 2008). These compounds have also received historical attention due to their flavoring and antimicrobial properties (Ahmed et al. 2000; Koeduka et al. 2006; Moleyar and Narasimham 1992; Sangwan et al. 1990). Enzymes responsible for the synthesis of eugenol, isoeugenol, and related compounds have been functionally characterized in model plants, such as *Petunia* (Koeduka et al. 2008), *Clarkia* (Koeduka et al. 2006, 2008), *Ocimum* (Koeduka et al. 2006), and *Anise* (Koeduka et al. 2009). All of these studies confer that eugenol and isoeugenol are catalyzed by two different types of enzyme (Koeduka et al. 2006, 2008). However, there is still a lack of comprehensive knowledge about evolution of such scent producing genes in plants. In this study, with the use of functional genomics and evolutionary approaches, we investigate a *G. odoratissima* EST resource as a starting point and further address about eugenol and isoeugenol emission in naturally occurring three *Gymnadenia* species, and also about the identification, expression, and evolution of the genes underlying their synthesis.

## Methods

### Plant materials

*G. odoratissima*, *G. conopsea*, and *G. densiflora* flowers and leaves were harvested from natural populations of Ofenpass (*G. odoratissima*), Albulapass (*G. conopsea*), and Münstertal (*G. densiflora*) in Switzerland. These samples were immediately frozen in liquid nitrogen and stored in a  $-80^{\circ}\text{C}$  freezer until subsequent analyses. Total RNA was isolated from these floral and leaf materials using RNeasy mini kit (Qiagen AG, Hombrechtikon, Switzerland). The necessary permits for all collection locations were obtained from the Amt für Natur und Umwelt, Chur, Switzerland, for the described field studies.

### Standard cDNA library construction

Approximately 1–2 mg of poly(A)<sup>+</sup> RNA from *G. odoratissima* flowers was obtained using Oligotex mRNA isolation kit following the manufacturer’s instructions (Qiagen AG,

Hombrechtikon, Switzerland). The Creator smart cDNA library construction kit (Clontech, Palo Alto, CA, USA) was employed for the construction of one standard cDNA library, and larger than 1 kb inserts were selected using Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Orange, CA, USA) instead of using the size selected columns provided with cDNA library kit (Clontech, Palo Alto, CA, USA). Prepared oligo(dT)-primed inserts were spliced into pDNR-LIB vector (Clontech, Palo Alto, CA, USA). Subsequently, the ligation mixture was transformed in to XL-10 gold Kan<sup>+</sup> ultra-competent cells of *Escherichia coli* (Stratagene, La Jolla, CA, USA). White colonies were screened to ensure the efficiency of prepared library and insert size range through colony PCR reactions. The library was preserved for long-term storage at –80 °C with the addition of 20 % (v/v) of glycerol in a final concentration and delivered to Purdue University sequencing platform, West Lafayette, IN, USA, for sequencing.

#### Suppressive subtractive hybridization cDNA library construction

*G. odoratissima* first-strand cDNA was prepared from 2 µg of poly(A)<sup>+</sup> floral RNA (tester population) and leaf (driver population) following Smart PCR cDNA synthesis kit specifications (Clontech, Palo Alto, CA, USA). The subtraction was based on Clontech kit (Clontech, Palo Alto, CA, USA) and was performed as described in the *PCR-Select cDNA subtraction kit user manual*. Subsequently to this step, resulting cDNA pool was used for size selection (>1 kb) as described above. The gel purified cDNA was end-repaired using DNA terminator kit (Lucigen Corporation, Middleton, WI, USA), and blunt end-polished cDNA fragments were cloned into pSMART HC kan vector (Lucigen Corporation, Middleton, WI, USA). The resulting ligation mixture was transformed into XL-10 gold ultracompetent cells (Stratagene, La Jolla, CA, USA). One suppressive subtractive hybridization (SSH) cDNA library (flower vs leaf) was constructed and delivered for sequencing as described above.

#### Bioinformatic tools

In this study, one standard (flower specific) and one SSH (flower vs leaf) cDNA libraries were constructed from *G. odoratissima*. Altogether, 2,571 ESTs from the total pool of 3,456 cDNA clones were achieved with a high confidence base call. Among these high-quality ESTs, 1,376 ESTs were generated from SSH library, whereas 1,195 ESTs were obtained from standard library. The chromatograms retrieved from sequencing platform were analyzed using the Phred base-calling program with default parameters (Ewing et al. 1998). Sequences were subjected to remove vector sequences with the use of Cross-match program. Masked sequences were

filtered out from the pool of total sequences. We analyzed these sequences with their corresponding quality files produced from Phred analysis for generating a tentatively unique gene (TUG) set from each library. These sequences were used for Phrap assembly program using default parameters. Clustering step estimated sequence redundancy in libraries. BlastN searches against “other EST database” and protein similarities (BlastX) against non-redundant (nr) protein database (Altschul et al. 1990) were analyzed with an *E*-value cutoff of  $\leq e^{-10}$  and  $\leq e^{-6}$ , respectively.

#### Assignment of gene ontology terms

Unigenes from both the libraries were annotated with gene ontology terms using Blast2GO software (Conesa et al. 2005). These unigenes were annotated using BlastX search against GenBank nr database. The best 20 hits were evaluated for assigning the gene ontology (GO) terms with an *E*-value cutoff of  $\leq e^{-6}$ .

#### Volatile collection for quantitative eugenol and isoeugenol GC analysis

A total of 14 to 20 *G. odoratissima* (Ofenpass), *G. conopsea* (Albulapass), and *G. densiflora* (Münstertal) individuals were chosen for scent collections. All volatile collections were sampled during the day time (between 11 AM and 5 PM) between July and August 2008 in the field. Each undamaged whole inflorescence was separately covered with an oven roasting bag (Toppits®, Cafresco frischhalteprodukte GmbH & Co., Minden, Germany), and these bags were subsequently closed with a plastic wire at both ends. A battery-operated pump (PAS-500 Personal Air Sampler, Spectrex, Redwood city, California, USA) was employed for each setup to absorb the volatile compounds into glass tubes prepared with 100 mg of Tenax TA (Gerstel GmbH & Co., Mülheim an der Ruhr, Germany). Each volatile collection was carried out for 30 min. In addition to *Gymnadenia* volatile collections, control samples were also collected in order to determine the possible air contaminants in the surrounding environment. After sealing both ends of each glass tube with Teflon tape, each sample was kept in a Gerstel TDS storage tube (Gerstel GmbH & Co., Mülheim an der Ruhr, Germany) and stored at –20 °C.

Collected scent samples were processed within 5 days using a gas chromatograph (GC, Agilent 6890 N) fitted with both a HP5 column (30-m×0.32-mm internal diameter×0.25-µm film thickness) and an Agilent 5975 mass selective detector connected to a thermal desorption system (TDS2, Gerstel, Mülheim, Germany) for scent analysis. The TDS temperature was programmed to rise from 30 °C (0.5-min hold) to 240 °C (1-min hold) at 60 °C/min. The cold injection system (CIS) temperature was programmed to rise from –50 °C (0.5-min hold) to 150 °C (0.5-min hold) at 16 °C/s

and to 250 °C (0.5-min hold) at 12 °C/s. The oven temperature of the GC6890 was adjusted to rise from 50 °C (3-min hold) to 230 °C at 8 °C/min. Agilent ChemStation and MSD ChemStation E.02.00.493 (Agilent Technologies, Palo Alto, CA, USA) software were not only used for identification of peak but also for retention time. Synthetic eugenol (Sigma-Aldrich, Switzerland) and isoeugenol (Sigma-Aldrich, Switzerland) standards were analyzed to compare the retention time and for calibration of the peak area. Quantification of eugenol and isoeugenol amounts were performed in terms of nanograms per liter of air sampled. To statistically compare amounts of eugenol and isoeugenol emitted by plants of the three species, we used a one-way ANOVA with LSD post hoc tests. Correlations between eugenol and isoeugenol were assessed by Pearson product moment correlations (SPSS Inc., Chicago).

#### Isolation of full-length *Gymnadenia* eugenol cDNAs and preparation of constructs

Several partial sequences showing high similarities with *eugenol synthase* genes from standard and subtracted EST libraries were initially screened for the isolation of full-length *G. odoratissima eugenol synthase* genes. The largest sequence among these partial sequences was further used to design gene-specific forward and reverse primers. Two full-length cDNA sequences encoding *G. odoratissima eugenol synthase1* (*GoEGS1*) and *G. odoratissima eugenol synthase2* (*GoEGS2*) were isolated using SMART-RACE cDNA amplification kit (Clontech). Amplified cDNA fragments were ligated into the pSMART HC kan vector (Lucigen Corporation, Middleton, WI, USA). Cloned sequences were verified using the ABI Prism Big Dye terminator cycle sequencing kit on ABI automated sequencers (Applied Biosystems) following the manufacturer's instructions. The complete coding regions were obtained for *G. conopsea eugenol synthase1* (*GcEGS1*), *G. conopsea eugenol synthase2* (*GcEGS2*), *G. densiflora (iso)eugenol synthase1* (*GdIEGS1*), and *G. densiflora (iso)eugenol synthase2* (*GdIEGS2*) based on the *GoEGS1* and *GoEGS2* sequences. RT-PCR reactions were set up for amplifying coding regions (ORFs) of *GoEGS1*, *GoEGS2*, *GcEGS1*, *GcEGS2*, *GdIEGS1*, and *GdIEGS2* from flower cDNA using forward and reverse primers but without stop codons in the end of sequences (Supplementary Table 1). The PCR conditions were as follows: 2 min initial denaturation at 94 °C, 30 cycles of 1 min at 94 °C, 1 min at 55 °C and 2 min at 72 °C. The final primer extension cycle was extended to 9 min. The amplified *Gymnadenia* coding regions were spliced into the expression vector using pEXP-CT/Topo TA expression kit according to the manufacturer's instructions. The orientation of all prepared constructs was confirmed by complete sequencing.

#### Biochemical test for activity in *E. coli*

The expression vectors with the cDNAs were transferred to *E. coli* cells, and the cells were grown as previously described (Koeduka et al. 2006). After induction and the addition of coniferyl alcohol, cells were grown overnight. Thereafter, the medium was collected and assayed for the presence of eugenol and isoeugenol (Koeduka et al. 2006).

#### Semi-quantitative RT-PCR analyses

To test gene expression patterns between flower and leaf tissue, two individuals each from natural populations of *G. odoratissima* (Ofenpass), *G. conopsea* (Davos), and *G. densiflora* (Münstertal) were selected. Total RNA followed by mRNA isolation from these individuals was performed as described above. About 1 µg of poly(A)<sup>+</sup> RNA was reverse-transcribed using Omniscript RT kit (Qiagen AG, Hombrechtikon, Switzerland). The cDNA amplification was performed in a Biometra thermocycler (Biometra GmbH, Germany) following these PCR conditions: 2 min initial denaturation at 94 °C, 38 cycles of 1 min at 94 °C, 1 min at 55 °C, and 2 min at 72 °C, followed by final primer extension for 5 min at 72 °C. Amplification of the cDNA coding for the putative *G. odoratissima glyceraldehyde-3-phosphate dehydrogenase* (*GADPH*), a ubiquitously expressed gene, served as a control for cDNA synthesis and PCR efficiency in all samples (Supplementary Table 1). Seven microliters of the amplified PCR products from each sample was electrophoresed on 1.5 % TAE agarose gels and stained with ethidium bromide.

#### Phylogeny reconstruction and selection analysis

*GoEGS2* gene sequence was exploited for searching its homologous sequences from NCBI using BlastN algorithm (nr/nt database) with the “somewhat similar sequences” option (Max target sequences 100, expect threshold 10, world size 11, and match/mismatch scores 2–3). These sequences were added to the pool of six *Gymnadenia eugenol* gene sequences and other functionally characterized *eugenol* or *isoeugenol synthase* genes. The sequence alignments were manually performed by employing BioEdit v7.0.5, and poorly aligned sequences were not taken into account for subsequent analyses. Mr. Modeltest 2.3 (Nylander 2004) and MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) were applied for determining the best substitution model and approximating the posterior probabilities of the phylogenetic tree by employing Markov chain Monte Carlo (MCMC) method, respectively. BaseML, a part of PAML 4.4 package (Yang 2007), was implemented for optimizing the branch lengths of resultant consensus tree. In addition, CodeML was further applied for a

codon-based model for calculating the synonymous and non-synonymous substitution rates.

## Results

### *G. odoratissima* EST library overview

Of the 2,571 high-quality *G. odoratissima* EST sequences, 1,195 and 1,376 EST sequences were retained after vector clipping, adaptor removal, and trimming of poor-quality sequence from the standard and subtraction cDNA libraries, respectively. Table 1 illustrates the summary statistics for the produced EST sequences from each library. Altogether, the success rate for obtaining sequences of high-quality sequence data was 76.4 %. The average sequence length for both libraries was 541 bp with the proportion of observed redundancy of 33 % (standard library) and 66 % (subtraction library).

### Unigene blast searches with the publically available database

Our BLASTX results suggest that 12.09 % (standard) and 17.37 % (subtraction) of the unigenes represent no significant similarities with other plant sequences, indicating that these sequences may have a specific role in *Gymnadenia* and other closely related plant species. Furthermore, BlastN similarity searches against EST others db yielded 61.97 and 59.11 % unigenes similarities from standard and subtracted EST libraries, respectively. Supplementary Tables S2 and S3 summarize the most abundant ESTs from standard and subtraction EST collections, and these transcripts possess at least eleven or more overlapping sequences in each contig with its putative identification determined by BLASTX analysis.

**Table 1** *Gymnadenia odoratissima* EST sequence summary

Total EST sequenced from both libraries	3,456
Number of cleaned EST sequences	2,571
Total EST sequences from standard library	1,920
Number of cleaned EST sequences from standard library	1,195
Numbers of contigs from standard library	306
Numbers of singlets from standard library	496
Total EST sequences from subtraction library	1,536
Number of cleaned EST sequences from subtraction library	1,376
Numbers of contigs from subtraction library	275
Numbers of singlets from subtraction library	197

Gene ontology annotations for standard and subtraction libraries

Of the total 802 *Gymnadenia* standard library unigenes, 556 unigenes (69.33 %) were defined by biological process terms, 518 (64.59 %) were defined by cellular component terms, and 572 (71.32 %) were defined by molecular function terms using Blast2GO. Relative frequencies from both libraries were determined and compared using GO terms (Supplementary Fig. S1). In total, 802 unigene sequences of standard library were annotated in 4,500 GO terms. We also estimated that a total of 266 unigenes from standard library were having enzyme code, whereas for subtraction library, only 144 unigenes were having enzyme code. To examine relative differential expression of transcripts between both libraries, we also compared the gene ontology annotations between standard and subtraction libraries (Supplementary Fig. S1). A total of 291 (61.65 %), 255 (54.03 %), and 239 (50.64 %) unigenes from subtraction EST libraries were defined by molecular function, biological process, and cellular component, respectively. Altogether, 472 unigenes from subtraction library were annotated in 2,113 GO terms.

### Emission of eugenol and isoeugenol in *Gymnadenia* species

In the three investigated species, eugenol was produced by *G. odoratissima* and *G. densiflora*, but only in trace amounts in a few individuals of *G. conopsea* ( $F_{2,46}=23.02$ ,  $P<0.001$ ; Table 2). Isoeugenol was found predominantly in *G. densiflora* and only in trace amounts in some individuals of the other two species ( $F_{2,46}=11.14$ ,  $P<0.001$ ; Table 2). There was a strong positive correlation between eugenol and isoeugenol in *G. densiflora* ( $r=0.75$ ,  $P=0.001$ ).

### Identification of *Gymnadenia* cDNAs encoding enzyme capable of synthesizing eugenol and isoeugenol

Two complete coding cDNA sequences of 984-bp long were identified and named as GoEGS1 and GoEGS2. Both nucleotide sequences encoded a peptide sequence of 328 aa with a calculated molecular mass of 36.2 kDa and 6.13 pI (isoelectric point). The deduced proteins of GoEGS1 and GoEGS2 showed 55 and 54 % amino acid identity with *Clarkia* EGS2 and *Petunia* EGS1, respectively (Supplementary Fig. S2). A total of four complete coding regions each with 984-bp long were isolated from *G. conopsea* and *G. densiflora* species and designated as GcEGS1, GcEGS2, GdIEGS1, and GdIEGS2 with calculated molecular masses of 36.25, 36.21, 36.33, and 36.33 kDa, respectively. All these sequences were nearly identical, varying by 1 to 12 amino acids between same species. Since it has been established that *E. coli* cells growing on coniferyl alcohol will take up this compound and convert it to coniferyl acetate, the substrate of both eugenol and

**Table 2** Quantification of eugenol and isoeugenol emission from intact inflorescences of three *Gymnadenia* species

Species	Population	Total number of individuals sampled	Eugenol	Isoeugenol
<i>G. odoratissima</i>	Ofenpass	20	60.5±13.96 (n=19)	1.27±0.34 (n=15)
<i>G. conopsea</i>	Albulapass	14	0.53±0.36 (n=4)	0.28±0.28 (n=1)
<i>G. densiflora</i>	Münstertal	15	839.47±178.25 (n=15)	196.66±62.63 (n=15)

Eugenol and isoeugenol compounds were collected from each sample for 30 min and identified by dynamic headspace method using GC-MS. The mean volatile emission are shown with standard errors, and volatile were collected in nanograms per liter of air sampled. The number of individuals (*n*) produced volatile are indicated in brackets

isoeugenol synthases (Koeduka et al. 2006), we tested the activity of the *Gymnadenia* enzymes by expressing them in *E. coli* cells. These cells expressing GoEGS1, GoEGS2, GcEGS1, and GcEGS2 and growing in the presence of coniferyl alcohol produced only eugenol, whereas the *E. coli* cells expressing GdIEGS1 and GdIEGS2 produced eugenol, as well as a trace amount of isoeugenol (Figs. 1 and 2; Supplementary Fig. S3 and Supplementary Table S4).

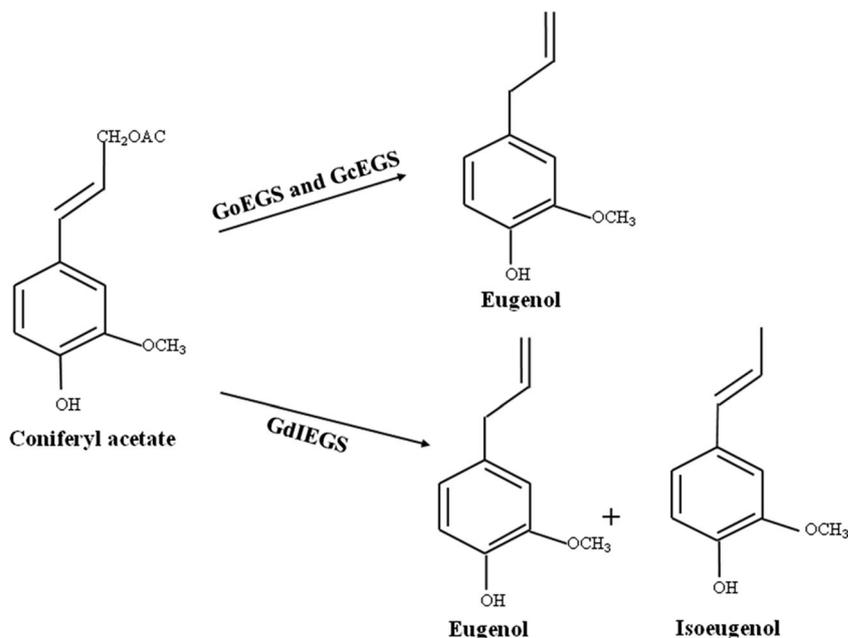
#### Gene expression in *Gymnadenia* floral and leaf tissue

All the *EGS* and *IEGS* genes are expressed in floral tissues that are the key source for fragrance emission in *Gymnadenia* species (Stpiczynska 2001). None of the *Gymnadenia EGS* showed any expression in leaf tissue, indicating that all functionally characterized *EGS* and *IEGS* in this study are florally expressed (Supplementary Fig. S4).

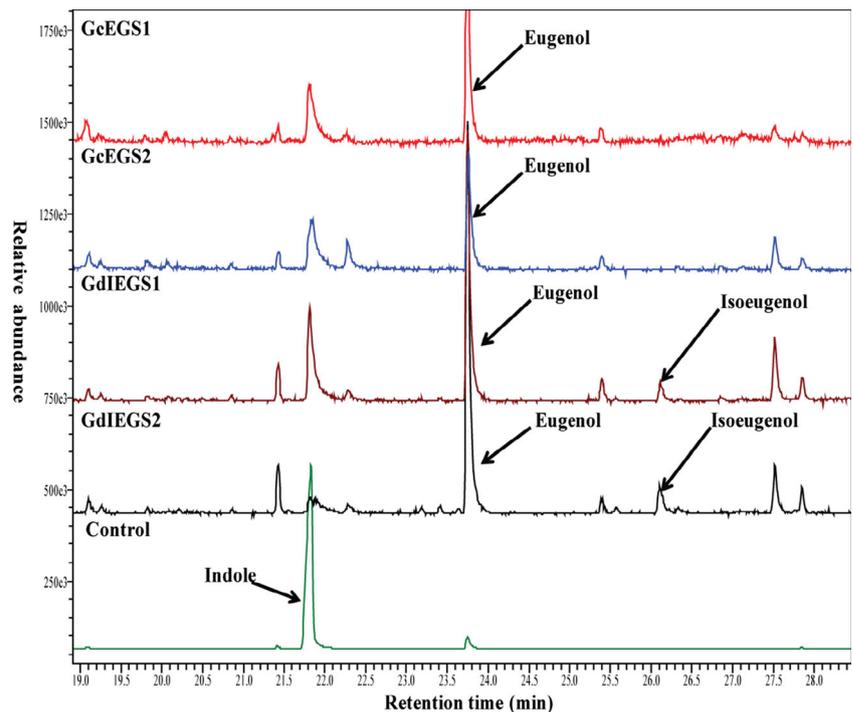
#### Evolutionary analysis

Altogether, 56 sequences were used for evolutionary analysis. All the functionally characterized *Gymnadenia eugenol* gene sequences form one distinct clade together with two monocot sequences (Supplementary Fig. S5). Furthermore, our phylogenetic analysis is also consistent with the previous study showing that *CbIGS1*, *CbEGS1*, *PhIGS1*, *PaAIS1*, and *ObEGS1* are closely related and clustered into a single clade (Koeduka et al. 2009). Our results further determine that *CbEGS2* and *PhEGS1* are only distantly related to all other functionally characterized eugenol and isoeugenol producing genes and belong to completely separate clade. Our selection analysis revealed significant relaxed purifying selection ( $\omega=0.48$ ;  $P=0.024$ ) acting only on the *GdIEGS1* and *GdIEGS2* clade, while no other functionally characterized *GoEGS1*, *GoEGS2*, *GcEGS1*, and *GcEGS2* genes showed evidence for selection (Supplementary Fig. S5, Supplementary Tables S5 and S6). Furthermore, our evolutionary analyses also indicate that *Petunia EGS1* is under purifying selection.

**Fig. 1** Pathway for biosynthesis of eugenol and isoeugenol in *Gymnadenia* species



**Fig. 2** A representative GC-MS profile of volatiles extracted from *Gymnadenia conopsea* and *Gymnadenia densiflora* genes. Volatiles were collected from the spent medium of cells expressing *GcEGS1*, *GcEGS2*, *GdIEGS1*, and *GdIEGS2*, and these cells were grown in the presence of coniferyl alcohol



## Discussion

In this study, we identified the genes encoding eugenol and (iso)eugenol synthase in the orchid genus *Gymnadenia*. We show that a single enzyme, (iso)eugenol synthase in *G. densiflora* catalyses eugenol together with trace amounts of isoeugenol and this double-function is evolutionarily derived from the single-product enzyme. Although it was previously known that single scent enzymes can produce several different scent compounds (Falara et al. 2011; Garms et al. 2010; Martin et al. 2004), we demonstrate here that the increased functional complexity evolved under relaxed purifying selection. The production of more than one compound by a single enzyme may contribute to the reproductive biology and/or defence system of the plant.

Besides other sequenced ESTs from orchid genera, such as *Phalaenopsis*, *Oncidium*, *Ophrys*, and *Vanda* using Sanger or next-generation sequencing approaches (Chan et al. 2011; Hsiao et al. 2006, 2011; Sedeek et al. 2013; Tan et al. 2005; Teh et al. 2011), our study reports a significant resource of 2,571 ESTs derived from floral tissue of *Gymnadenia odoratissima* (Orchidoideae, Orchidinae). Floral tissues were preferred for libraries construction to cover the large range of genes involved in flower senescence, flower pigmentation, and fragrance-related attributes. The primary objective of these EST generations was to identify the floral- and fragrance-related genes in our focal species. In the standard *Gymnadenia* library, transcripts that had third highest expression were found to be homologous to *phenylacetaldehyde synthase* gene. Fragrance-related transcripts further served as

the basis for gene characterization in *Gymnadenia* species. Identified putative genes are the first representation of “scent genes” in *Gymnadenia*. On the other hand, a highly expressed transcript found in our subtraction library was cytochrome P450, which is not only entailed in primary/secondary metabolism, but probably also in promoting longer survival ability (Hsiao et al. 2011). Despite the small number of sequences generated for *G. odoratissima*, it was striking that EST sequences corresponded to all important categories of gene ontology terms. Altogether, it verifies that our small number of ESTs build up an important resource for future molecular research in orchids.

To date, NADP-dependent reductases that produce eugenol, isoeugenol, and t-anol have only been identified in eudicot species (Koeduka et al. 2006, 2008, 2009). Although some orchids produce copious amount of these compounds, no monocot NADP-dependent phenylpropane enzyme has been reported up to now (Huber et al. 2005; Johnson et al. 2007; Nishida et al. 2004; Schiestl et al. 2011). The four *GoEGS1*, *GoEGS2*, *GcEGS1*, and *GcEGS2* are responsible for formation of eugenol, as evidenced by our functional characterization. Considering the high levels of sequence similarities among all the characterized *Gymnadenia* enzymes, it was expected that *GdIEGS1* and *GdIEGS2* will also catalyze the formation of eugenol only. However, these enzymes catalyze a mixture of eugenol and isoeugenol. It is common for terpene synthases to be able to form multiple products from a single prenyl diphosphate substrate (Chen et al. 2011). Whereas terpene synthases often show multiple products,

as also recently shown in orchids, *Vanda Mimi Palmer* (Song et al. 2012) and *Phalaenopsis bellina* (Hsiao et al. 2008), previously characterized eugenol synthase and isoeugenol synthase made only one product. Koeduka et al. (2008) has shown that the amino acids at positions 84 and 87 in *CbIGS1* and similar positions in other functionally characterized proteins, are crucial for specifying either eugenol or isoeugenol production.

We also show that the *Gymnadenia* enzymes share a high degree of protein identity with *Clarkia* and *Petunia* EGSs (Koeduka et al. 2008). Moreover, the two sequences from *G. odoratissima*, *G. conopsea*, and *G. densiflora* were nearly identical, and it is as yet not clear whether they represent two different alleles of the same locus or two duplicated gene copies. Interestingly, we found an evolutionary switch from the enzymatic production of one scent compound in *G. odoratissima* and *G. conopsea*, to the production of two compounds in *G. densiflora*. Since we also found significant values of selection for this evolutionary switch, we interpret this transition, leading to the production of an additional scent compound, to be mediated by relaxed purifying selection ( $\omega=0.48$ ). At present, it is, however, unknown what could be the possible fitness impact of the additional volatile production. We suggest two hypotheses that may account for the evolution of isoeugenol emission in *G. densiflora*:

1. Better discrimination by pollinators: Although *G. conopsea* and *G. densiflora* are very similar in their floral color and morphology, they may be incompatible due to different genome size and ploidy levels at least in some populations (Trávníček et al. 2011). Enhanced differences in floral scent may increase prezygotic isolation by pollinators and thus increase reproductive success in mixed populations. However, there is evidence for little prezygotic, pollinator-mediated isolation among the two *Gymnadenia* species (Jersáková et al. 2010).
2. Better defense: Isoeugenol has been shown to be under negative directional selection mediated by pollinators in *G. odoratissima*, suggesting it can deter insects from flowers (Schiestl et al. 2011). Similarly, in *Petunia*, isoeugenol emission has recently been shown in reducing the floral damage from its florivores by 50 % (Kessler et al. 2013). Isoeugenol may thus play a role in reducing insect-mediated damage to flowers, and eventually benefit plant fitness.

**Acknowledgments** The authors appreciate Prof. Natalia Dudareva and Dr. Phillip SanMiguel for their help in sequencing cDNA libraries. Thanks are also due to Edward Connor for providing help in volatile collections and Philipp Schlüter for assisting in PAML program. We also thank the “Amt für Natur und Umwelt, Chur” for issuing collection permits. This work was financed by the Swiss National Science Foundation grant to FPS (SNF; project no. 31003A-112342) and ETH Zürich doctoral scholarship to AKG.

## References

- Aceto S, Gaudio L (2011) The MADS and the beauty: genes involved in the development of orchid flowers. *Curr Genomics* 12:342–356
- Ahmed M, Amin S, Islam M, Takahashi M, Okuyama E, Hossain CF (2000) Analgesic principle from *Abutilon indicum*. *Pharmazie* 55: 314–316
- Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Balao F, Herrera J, Talavera S, Dötterl S (2011) Spatial and temporal patterns of floral scent emission in *Dianthus innoxianus* and electroantennographic responses of its hawkmoth pollinator. *Phytochemistry* 72:601–609
- Bellusci F, Pellegrino G, Palermo A, Musacchio A (2008) Phylogenetic relationships in the orchid genus *Serapias* L based on noncoding regions of the chloroplast genome. *Mol Phylogenet Evol* 47:986–991
- Brodmann J, Twele R, Francke W, Holzler G, Zhang Q, Ayasse M (2008) Orchids mimic green-leaf volatiles to attract prey-hunting wasps for pollination. *Curr Biol* 18:740–744
- Chan W-S, Abdullah JO, Namasivayam P (2011) Isolation, cloning, and characterization of fragrance-related transcripts from *Vanda Mimi Palmer*. *Sci Hortic* 127:388–397
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 66:212–229
- Conesa A, Gotz S, Garcia-Gomez J, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Delle-Vedove R, Juillet N, Bessièrre J-M, Grison C, Barthes N, Paillet T, Dormont L, Schatz B (2011) Colour-scent associations in a tropical orchid: three colours but two odours. *Phytochemistry* 72:735–742
- Dötterl S, Jürgens A (2005) Spatial fragrance patterns in flowers of *Silene latifolia*: lilac compounds as olfactory nectar guides? *Plant Syst Evol* 255:99–109
- Douzery EJP, Pridgeon AM, Kores P, Linder HP, Kurzweil H, Chase MW (1999) Molecular phylogenetics of disease (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. *Am J Bot* 86:887–899
- Ewing B, Hillier L, Wendl M, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schillmiller AL, Last RL, Schuurink RC, Pichersky E (2011) The tomato terpene synthase gene family. *Plant Physiol* 157:770–789
- Fu SF, Tsai T, Chen YR, Liu CP, Haiso LJ, Syue LH, Yeh HH, Huang HJ (2012) Characterization of the early response of the orchid, *Phalaenopsis amabilis*, to *Erwinia chrysanthemi* infection using expression profiling. *Physiol Plant* 145:406–425
- Garms S, Köllner TG, Boland W (2010) A multiproduct terpene synthase from *Medicago truncatula* generates cadalane sesquiterpenes via two different mechanisms. *J Org Chem* 75:5590–5600
- Gupta AK, Akhtar TA, Widmer A, Pichersky E, Schiestl FP (2012) Identification of white campion (*Silene latifolia*) guaicol *O*-methyltransferase involved in the biosynthesis of veratrole, a key volatile for pollinator attraction. *BMC Plant Biol* 12:158
- Heil M (2008) Indirect defence via tritrophic interactions. *New Phytol* 178:41–61
- Hsiao Y, Tsai W, Kuoh C, Huang T, Wang H, Wu T, Leu Y, Chen W, Chen H (2006) Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce the monoterpene biosynthesis pathway. *BMC Plant Biol* 6:14
- Hsiao Y, Jeng M, Tsai W, Chung Y, Li C, Wu T, Kuoh C, Chen W, Chen H (2008) A novel homodimeric geranyl diphosphate synthase from

- the orchid *Phalaenopsis bellina* lacking a DD(X)2-4D motif. *Plant J* 55:719–733
- Hsiao Y-Y, Chen Y-W, Huang S-C, Pan Z-J, Fu C-H, Chen W-H, Tsai W-C, Chen H-H (2011) Gene discovery using next-generation pyrosequencing to develop ESTs for *Phalaenopsis* orchids. *BMC Genomics* 12:360
- Huber F, Kaiser R, Sauter W, Schiestl F (2005) Floral scent emission and pollinator attraction in two species of *Gymnadenia* (Orchidaceae). *Oecologia* 142:564–575
- Inda LA, Pimentel M, Chase MW (2012) Phylogenetics of tribe orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Ann Bot* 110:71–90
- Jersáková J, Castro S, Sonk N, Milchreit K, Schödelbauerová I, Tolasch T, Dötterl S (2010) Absence of pollinator-mediated pre-mating barriers in mixed-ploidy populations of *Gymnadenia conopsea* s.l. (Orchidaceae). *Evol Ecol* 24:1199–1218
- Johnson SD, Ellis A, Dötterl S (2007) Specialization for pollination by beetles and wasps: the role of lollipop hairs and fragrance in *Satyrium microrrhynchum* (Orchidaceae). *Am J Bot* 94:47–55
- Junker RR, Blüthgen N (2010) Floral scents repel facultative flower visitors, but attract obligate ones. *Ann Bot* 105:777–782
- Kessler D, Baldwin IT (2007) Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*. *Plant J* 49:840–854
- Kessler D, Diezel C, Clark DG, Colquhoun TA, Baldwin IT (2013) *Petunia* flowers solve the defence/apparency dilemma of pollinator attraction by deploying complex floral blends. *Ecol Lett* 16:299–306
- Koeduka T, Fridman E, Gang DR, Vassão DG, Jackson BL, Kish CM, Orlova I, Spassova SM, Lewis NG, Noel JP, Baiga TJ, Dudareva N, Pichersky E (2006) Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proc Natl Acad Sci* 103:10128–10133
- Koeduka T, Louie GV, Orlova I, Kish CM, Ibdah M, Wilkerson CG, Bowman ME, Baiga TJ, Noel JP, Dudareva N, Pichersky E (2008) The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. *Plant J* 54:362–374
- Koeduka T, Baiga TJ, Noel JP, Pichersky E (2009) Biosynthesis of t-anethole in Anise: characterization of t-anol/isoeugenol synthase and an O-methyltransferase specific for a C7-C8 propenyl side chain. *Plant Physiol* 149:384–394
- Kondo M, Oyama-Okubo N, Ando T, Marchesi E, Nakayama M (2006) Floral scent diversity is differently expressed in emitted and endogenous components in *Petunia axillaris* lines. *Ann Bot* 98:1253–1259
- Lönn M, Alexandersson R, Gustafsson S (2006) Hybrids and fruit set in a mixed flowering-time population of *Gymnadenia conopsea* (Orchidaceae). *Hereditas* 143:222–228
- Majetic C, Raguso R, Tonsor S, Ashman T-L (2007) Flower color-flower scent associations in polymorphic *Hesperis matronalis* (Brassicaceae). *Phytochemistry* 68:865–874
- Mant J, Schiestl F, Peakall R, Weston P (2002) A phylogenetic study of pollinator conservatism among sexually deceptive orchids. *Evolution* 56:888–898
- Martin DM, Fäldt J, Bohlmann J (2004) Functional characterization of nine norway spruce *TPS* genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol* 135:1908–1927
- Miyake T, Yafuso M (2003) Floral scents affect reproductive success in fly-pollinated *Alocasia odora* (Araceae). *Am J Bot* 90:370–376
- Moleyar V, Narasimham P (1992) Antibacterial activity of essential oil components. *Int J Food Microbiol* 16:337–342
- Mondragón-Palomino M, Theissen G (2008) MADS about the evolution of orchid flowers. *TRENDS Plant Sci* 13:51–59
- Mondragón-Palomino M, Theißen G (2009) Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. *Ann Bot* 104:583–594
- Mondragón-Palomino M, Theißen G (2011) Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the ‘orchid code’. *Plant J* 66:1008–1019
- Monteiro F, Sebastiana M, Figueiredo A, Sousa L, Cotrim H, Pais M (2012) Labellum transcriptome reveals alkene biosynthetic genes involved in orchid sexual deception and pollination-induced senescence. *Funct Integr Genom* 12:693–703
- Negre F, Kish CM, Boatright J, Underwood B, Shibuya K, Wagner C, Clark DG, Dudareva N (2003) Regulation of methylbenzoate emission after pollination in Snapdragon and *Petunia* flowers. *Plant Cell Online* 15:2992–3006
- Nishida R, Tan K-H, Wee S-L, Hee AK-W, Toong Y-C (2004) Phenylpropanoids in the fragrance of the fruit fly orchid, *Bulbophyllum cheiri*, and their relationship to the pollinator, *Bactrocera papayae*. *Biochem Syst Ecol* 32:245–252
- Nylander J (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University
- Pan IC, Liao D-C, Wu F-H, Daniell H, Singh ND, Chang C, Shih M-C, Chan M-T, Lin C-S (2012) Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical *Oncidium* breeding. *PLoS ONE* 7:e34738
- Pichersky E, Gershenzon J (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr Opin Plant Biol* 5:237–243
- Plepyš D, Ibarra F, Löfstedt C (2002) Volatiles from flowers of *Platanthera bifolia* (Orchidaceae) attractive to the silver Y moth, *Autographa gamma* (Lepidoptera: Noctuidae). *Oikos* 99:69–74
- Raguso RA (2008) Wake up and smell the roses: the ecology and evolution of floral scent. *Annu Rev Ecol Evol Syst* 39:549–569
- Ronquist F, Huelsenbeck J (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Salzmänn CC, Schiestl FP (2007) Odour and colour polymorphism in the food-deceptive orchid *Dactylorhiza romana*. *Plant Syst Evol* 267:37–45
- Salzmänn C, Nardella A, Cozzolino S, Schiestl FP (2007) Variability in floral scent in rewarding and deceptive orchids: the signature of pollinator-imposed selection? *Ann Bot* 100:757–765
- Sangwan NK, Verma BS, Verma KK, Dhindsa KS (1990) Nematicidal activity of some essential plant oils. *Pestic Sci* 28:331–335
- Schiestl FP, Glaser F (2012) Specific ant-pollination in an alpine orchid and the role of floral scent in attracting pollinating ants. *Alp Bot* 122:1–9
- Schiestl FP, Schlüter PM (2009) Floral isolation, specialized pollination, and pollinator behavior in orchids. *Annu Rev Entomol* 54:425–446
- Schiestl FP, Huber F, Gomez J (2011) Phenotypic selection on floral scent: trade-off between attraction and deterrence? *Evol Ecol* 25:237–248
- Sedeek KE, Qi W, Schauer MA, Gupta AK, Poveda L, Xu S, Liu ZJ, Grossniklaus U, Schiestl FP, Schlüter PM (2013) Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS ONE* 8:e64621
- Shuttleworth A, Johnson S (2009) Specialized pollination in the African milkweed *Xysmalobium orbiculare*: a key role for floral scent in the attraction of spider-hunting wasps. *Plant Syst Evol* 280:37–44
- Soliva M, Widmer A (1999) Genetic and floral divergence among sympatric populations of *Gymnadenia conopsea* s.l. (Orchidaceae) with different flowering phenology. *Int J Plant Sci* 160:897–905
- Soliva M, Koecyan A, Widmer A (2001) Molecular phylogenetics of the sexually deceptive orchid genus *Ophrys* (Orchidaceae) based on nuclear and chloroplast DNA sequences. *Mol Phylogenet Evol* 20:78–88
- Song AAL, Abdullah JO, Abdullah MP, Shafee N, Rahim RA (2012) Functional expression of an orchid fragrance gene in *Lactococcus lactis*. *Int J Mol Sci* 13:1582–1597
- Stpiczynska M (2001) Osmophores of the fragrant orchid *Gymnadenia conopsea* L. (Orchidaceae). *Acta Soc Bot Pol* 70:91–96

- Streinzer M, Ellis T, Paulus H, Spaethe J (2010) Visual discrimination between two sexually deceptive *Ophrys* species by a bee pollinator. *Arthropod Plant Interact* 4:141–148
- Tan J, Wang H, Yeh K (2005) Analysis of organ-specific, expressed genes in *Oncidium* orchid by subtractive expressed sequence tags library. *Biotechnol Lett* 27:1517–1528
- Teh S, Chan W, Abdullah J, Namasivayam P (2011) Development of expressed sequence tag resources for *Vanda Mimi* Palmer and data mining for EST-SSR. *Mol Biol Rep* 38:3903–3909
- Theis N (2006) Fragrance of Canada thistle (*Cirsium arvense*) Attracts both floral herbivores and pollinators. *J Chem Ecol* 32:917–927
- Theis N, Kesler K, Adler LS (2009) Leaf herbivory increases floral fragrance in male but not female *Cucurbita pepo* subsp. *texana* (Cucurbitaceae) flowers. *Am J Bot* 96:897–903
- Trávníček P, Kubátová B, Čum V, Rauchová J, Krajníková E, Jersáková J, Suda J (2011) Remarkable coexistence of multiple cytotypes of the *Gymnadenia conopsea* aggregate (the fragrant orchid): evidence from flow cytometry. *Ann Bot* 107:77–87
- van Dam NM (2009) Belowground herbivory and plant defenses. *Annu Rev Ecol Evol Syst* 40:373–391
- van der Niet T, Hansen DM, Johnson SD (2011a) Carrion mimicry in a South African orchid: flowers attract a narrow subset of the fly assemblage on animal carcasses. *Ann Bot* 107:981–992
- van der Niet T, Liltved W, Johnson S (2011b) More than meets the eye: a morphological and phylogenetic comparison of long-spurred, white-flowered *Satyrium* species (Orchidaceae) in South Africa. *Bot J Linn Soc* 166:417–430
- Vereecken NJ, Schiestl FP (2009) On the roles of colour and scent in a specialized floral mimicry system. *Ann Bot* 104:1077–1084
- Wälti M, Page P, Widmer A, Schiestl F (2009) How to be an attractive male: floral dimorphism and attractiveness to pollinators in a dioecious plant. *BMC Evol Biol* 9:190
- Xu YF, Teo LL, Zhou J, Kumar PP, Yu H (2006) Floral organ identity genes in the orchid *Dendrobium crumenatum*. *Plant J* 46:54–68
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591