The ecological and genetic basis of floral isolation in a specialized pollination system

Presented by Shuqing Xu

Supervised by
Prof. Dr. Florian P. Schiestl
Institute of Systematic Botany University of Zürich

And
Prof. Dr. Alex Widmer
Institute of Integrative Biology ETH Zürich

2011 February
The ecological and genetic basis of floral isolation in a
specialized pollination system

A dissertation submitted to

ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
Shuqing Xu
MSc Cell Biology, Graduate University of Chinese Academy of Sciences
born April 12th, 1983
citizen of China

accepted on the recommendation of
Prof. Dr. Alex Widmer examiner
Prof. Dr. Florian P. Schiestl co-examiner
Prof. Dr. Salvatore Cozzolino co-examiner
Prof. Dr. Scott Hodges co-examiner
Index

Summary ................................................................................................................................. I
Zusammenfassung .................................................................................................................. III
General Introduction ............................................................................................................. 1

Chapter 1
Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids ................................................................................................................................. 21

Chapter 2
cDNA representational difference analysis reveals the genetic basis of divergent traits between closely related orchid species .............................................................................................. 67

Chapter 3
Stearoyl-ACP desaturases are associated with floral isolation in sexually deceptive orchids ................................................................................................................................. 93

Chapter 4
The genetic basis of pollinator adaptation in a specialized pollination system ................................................................................................................................. 141

Acknowledgements .............................................................................................................. 181
Curriculum vitae .................................................................................................................. 183
Summary

Floral diversity is an extraordinary feature of angiosperms. It is one of central goals in plant evolutionary biology to understand why and how this diversity has evolved. Studies have suggested that pollinator-mediated selection is the main driving force for floral diversity and plant speciation. In the process of pollinator-driven floral diversification and speciation, floral isolation, a form of pollinator-mediated pre-pollination reproductive isolation, plays a key role, because it is the link between pollinator-mediated selection and reproductive isolation. Therefore, the knowledge of the ecological and genetic bases of floral isolation is essential for understanding floral diversity and the process of plant speciation.

Floral isolation often evolves as a consequence of adaptation to different pollinators, which have different preferences and behaviours that would ultimately drive the diversification of floral characters. Once different floral characters are specialized for different classes of pollinators and reach a certain degree of exclusiveness, floral isolation is formed. The evolutionary pattern of floral isolation and its role in the process of pollinator-driven speciation depend on three key components: the strength of floral isolation; the strength of pollinator-mediated selection; and the genetic basis of floral isolation.

The aim of this study is to systematically study the key components of floral isolation in a specialized pollination system, namely sexually deceptive orchids of the genus Ophrys, from both ecological and genetic aspects:

Chapter 1: I investigated the strength of different reproductive barriers, as well as population structure among our study species experimentally. The results showed that floral odour-mediated floral isolation acts as the main reproductive isolation and could effectively isolate the species from each other in sympatry.

Chapter 2: I applied a cDNA representational difference analysis (RDA) to isolate genes underlying floral isolation among our study species. The results successfully identified several floral scent-related genes, one of which might be involved in the biosynthesis of alkenes that are directly linked to specific pollinator attraction.

Chapter 3: The function of candidate genes underlying alkene biosynthesis in our study species were characterized. The results suggested that one stearoyl-acyl-carrier-protein
**Summary**

*desaturase (SAD)* homologue, namely *SAD2*, is likely involved in specific alkene productions that contribute to floral isolation among our study species.

**Chapter 4:** I investigated the expression and evolutionary pattern of six members of the *SAD* gene family among our study species. The results indicated that four of these *SAD* homologues may be involved in the differences in alkenes production among different species. And pollinator adaptation in *Ophrys* might be primarily due to the changes of these *SAD* homologues, in terms of their expression and potentially also in terms of the enzymatic function.

The thesis presented here provides insight into the ecological and genetic basis of floral isolation in a specialized pollination system. This information is also relevant for understanding pollinator-driven speciation in other systems. Furthermore, the evolutionary patterns of floral traits under pollinator-mediated selection sheds light on plant evolution and adaptation in general.
Zusammenfassung


Blütenisolation entwickelt sich oft als eine Konsequenz aus der Anpassung an unterschiedliche Bestäuber mit verschiedenen Präferenzen und Verhalten, welche so die Diversifikation von Blütenmerkmalen antreiben. Diese Blütenisolation entwickelt sich, sobald unterschiedliche Blütenmerkmale an bestimmte Bestäuberklassen angepasst sind und einen gewissen Grad an Bestäuber-Exklusivität erreicht haben. Die evolutionären Muster der Blütenisolation und ihre Rolle in Bestäuber-getriebenen Artbildungsprozessen hängen von drei Faktoren ab: die Stärke der Blütenisolation, die Stärke der Selektion durch Bestäuber, und die genetische Grundlage der Blütenisolation.

Das Ziel dieser Dissertation ist die systematische Untersuchung der wesentlichen Faktoren der Blütenisolation in einem hochspezifischen Bestäubungssystem, nämlich der Sexualtäuschblume Ophrys, wobei sowohl ökologische also auch genetische Aspekte berücksichtigt wurden:

Kapitel 1: Die Stärke verschiedener reproduktiver Barrieren und die Populationsstruktur der untersuchten Pflanzenarten wurde experimentell untersucht. Die Resultate zeigten, dass auf Blütenduftstoffen basierende Blütenisolation die wichtigste Form reproduktiver Isolation in den untersuchten Species ist, und den Genfluss zwischen sympatrisch vorkommenden Arten effektiv unterbinden konnte.

Kapitel 2: cDNA Repräsentations-Differenz Analyse (RDA) wurde verwendet, um die der Blütenisolation zugrunde liegenden Gene zu isolieren. Einige Gene, die für die Produktion von Blütenduftstoffen relevant sind, wurden hierbei identifiziert. Eines dieser
Zusammenfassung

Gene ist wahrscheinlich in der Biosynthese von Alkenen involviert, die direkt an der Bestäuberanlockung beteiligt sind.

**Kapitel 3:** Die Funktion einiger Kandidatengene in der Alken-Biosynthese wurde charakterisiert. Ein Gen (SAD2) aus der kleinen SAD Genfamilie, die für lösliche Fettsäure-Desaturasen codiert, ist mit hoher Wahrscheinlichkeit an der Alken-Biosynthese beteiligt und trägt somit zur Blütenisolation zwischen den untersuchten Arten bei.

**Kapitel 4:** Die Gen-Expression und evolutionären Verwandtschaftsverhältnisse von sechs Mitgliedern der SAD Genfamilie wurde untersucht. Die Resultate legen nahe, dass insgesamt vier SAD Homologe zu den Unterschieden in der Alken-Produktion zwischen drei untersuchten Arten beitragen. Änderungen in der Gen-Expression und möglicherweise auch der enzymatischen Funktion dieser SAD Homologe könnten die Anpassung an spezifische Bestäuber ermöglichen.

General introduction

Floral diversity has been suggested as one the most extraordinary features of angiosperms. Flowers vary enormously in size, colour, shape, texture, and scent among different groups and species, from the tiny, white radially symmetrical flowers of Arabidopsis thaliana to the tongue-shaped, dark brown flowers of Chiloglottis that mimic female wasps and are pollinated by male wasps during mating attempts. It has become a central goal in plant evolutionary biology to understand the mechanisms and processes of floral diversification and speciation. Since Darwin, pollinators have been assumed to play a major role in floral diversification (Darwin 1862). Most angiosperm species (about 300,000) rely on animals as vectors for pollination (Waser and Campbell 2004). Pollinator-mediated selection has been suggested as the main driving force for plant speciation and floral diversification (Johnson 2006). However, the details of this process are little understood.

1. Ecological speciation in flowering plants.

The idea that species originated by natural selection was first proposed by Darwin:

“Natural selection, also, leads to divergence of character; for more living beings can be supported on the same area the more they diverge in structure, habits, and constitution […]. Therefore, during the modification of the descendants of any one species, and during the incessant struggle of all species to increase in numbers, the more diversified these descendants become, the better will be their chance of succeeding in the battle of life. Thus the small differences distinguishing varieties of same species, will steadily tend to increase till they come to equal the greater differences between species of the same genus, or even of distinct genera.” (Darwin 1859).
However, it is only recently that evolutionary biologists realized that he was correct. After investigating speciation through genetic drift for decades, evolutionary biologists seem to have mostly agreed that natural selection is indeed the primary driving force of speciation (Coyne and Orr 2004; Waser and Campbell 2004; Johnson 2006; Schluter and Conte 2009). This lead to the concept of ecological speciation, which “occurs when divergence selection on traits between populations or subpopulations in contrasting environments leads indirectly or directly to evolution of reproductive isolation” (Schluter 2001). For ecological speciation, reproductive isolation could evolve as a by-product of adaptation to different niches (indirectly), but also could evolved by selection (directly). Selection directly driving reproductive isolation generally occurs when hetero-specific mating has reduced fitness, and therefore is selected against (reinforcement). Studies on ecological speciation have mainly focused on the roles of natural selection and the evolution of reproductive isolation during the process of speciation.

Ecological speciation in flowering plants could be driven by pollinator-mediated selection. The concept of pollinator-driven speciation was first developed by Grant and Grant (1965). They implied that “pollinator climate” (both efficiency and abundance) might vary from one area to another. Therefore two populations of the same plant living in these two territories will be under different pollinator-mediated selection depending on the preference of local pollinators, and evolving to different floral characters. Once the different floral characters specialized for different classes of pollinators reach a certain stage of exclusiveness, these differences would contribute to the reproductive barrier (for example, floral isolation) and permit these two species to evolve as genetically independent lines.

Pollinator-driven speciation in flowering plants is thought to be due to pollen limitation. Many ecological studies have shown that seed production in flowering plants is mainly limited by receiving compatible pollen (Burd 1994; Ashman et al. 2004; Harder and
Aizen 2010). In this case, the individuals and alleles that overcome this constrain should be favoured by natural selection and spread in the population. Shift to either pollinators that are locally more efficient in pollen transfer, or a mating system change to selfing are the ultimate outcomes. However, the advantage of selfing may be ephemeral due to increased extinction rates associated with reduced genetic diversity and/or accumulation of deleterious mutations (Takebayashi and Morrell 2001; Bartkowska and Johnston 2009). Study by Goldberg et al. (2010) also showed that the short-term advantage of selfing could be offset by strong species selection, which favours outcrossing. Therefore, pollinator shifts may be the most common evolutionary outcome of pollen limitation. One example of such evolutionary patterns is provided by orchids, in which pollinator shifts are very common and self-pollination is rare (5% of taxa), despite strong pollen limitation in this family (Tremblay et al. 2005; Johnson 2006).

Although in theory, pollinator-driven speciation can happen with ongoing gene flow (in sympatry or parapatry), most examples of pollinator-driven speciation found involve geographic isolation (allopatry) (Coyne and Orr 2004; Johnson 2006). In the scenario of ecological speciation, whether speciation could happen or not depends on three conditions (Waser and Campbell 2004): (1) the genetic linkage of traits under selection to assortative mating; (2) the strength and consistency of disruptive selection; (3) the strength of assortative mating (reproductive isolation). For pollinator-driven speciation, the first condition is often met since pollinator-mediated selection on floral traits, e.g. floral odour, colour and spur length, is usually directly linked to assortative mating (Grant 1994). However, among closely related species, strong pollinator-mediated disruptive selection on floral traits was found to be uncommon and inconsistent (Campbell et al. 1997; Kingsolver et al. 2001; Waser and Campbell 2004; Rymer et al. 2010). This may be because different pollinators are often not stably and equally distributed (Johnson 2006). Furthermore, pollinator-mediated reproductive
isolation (floral isolation) is often incomplete (Campbell et al. 1998; Campbell et al. 2002; Lowry et al. 2008) (But see Chapter 1). Therefore, compared to an allopatric speciation scenario, pollinator-driven speciation is less likely to occur in sympatry or parapatry.

However, a system with strong pollinator-mediated disruptive selection and complete floral isolation - like the very specialized pollination system investigated in our study - would be extremely valuable for investigating the process of sympatric pollinator-mediated speciation, and would extend our understanding of speciation patterns in plants.

Ecological speciation in plants can also be driven by habitat adaptation. In this scenario, plants adapt to different environments (other than animal pollinators), like different soil moisture, temperature, chemistry, herbivore, diseases etc., and reproductive isolation evolves as a by-product. Local habitat adaptation in plants has been reported from parapatry to allopatry (Jain and Bradshaw 1966; Davies and Snaydon 1976; Turkington and Harper 1979; Bennington and Mcgraw 1995; Ronsheim 1997). The critical condition for habitat adaptation-driven speciation is the evolution of reproductive barriers. In strict allopatry, post-zygotic isolation could evolve as described in the Bateson-Dobzansky-Muller model (Coyne and Orr 2004). Several studies suggested that this might be the common route for habitat adaptation driven speciation (Waser and Campbell 2004). However, there are some cases in which reproductive barriers could also be direct pleiotropic effects of genes involved in local abiotic adaptation (Maenair and Christie 1983; Galen and Stanton 1991). For example, in *Mimulus guttatus*, the genes conferring tolerance to copper contaminated soil are lethal in F1 hybrids when crossed with plants growth in uncontaminated soil (Macnair and Christie 1983).

The process of ecological speciation in flowering plants often involves both pollinator and habitat adaptation. In the scenario of pollinator-driven speciation, as discussed above, it is more likely to happen in allopatry. Adaptation to local pollinators will play an indirect role
for facilitating adaptation to local environments by reducing pollen dispersal-mediated gene flow (Waser and Campbell 2004). Therefore, even in pollinator-driven speciation, also habitat adaptation can often be observed. On the other hand, in the scenario of habitat-driven speciation, adaptation to different habitats in parapatry or allopatry is often combined with pollinator adaptation, since the abundance and efficiency of pollinators often form 'a geographic mosaic' (Johnson 2006). Furthermore, habitat adaptation can also mediate changes in floral traits, which may affect pollinator attraction and behaviour (see few examples by (Waser and Campbell 2004), however none of these examples is fully understood how habitat adaptation mediate changes in floral traits).

Although rapid progress has been made during the last decades, the process of ecological speciation is not yet fully understood. Previous studies have shown that disruptive selection, assortative mating (reproductive isolation) as well as the linkage of selection and assortative mating are the three most important factors for speciation (Waser and Campbell 2004). However, studies quantifying all these factors in different plant pollination system are largely still missing. For example, it would be important to understand how strong pollinator-mediated disruptive selection needs to be to drive speciation in given pollination system? How strong does floral isolation have to be to allow sympatric speciation? Systematically measuring floral isolation and pollinator-mediated selection in nature, and obtaining knowledge of the genetic basis of the traits underlying selection by pollinators and assortative mating would extend our understanding of the process of ecological speciation in plants.

2. Reproductive isolation in flowering plants

Although natural selection plays a central role in the process of ecological speciation, the evolution of reproductive isolation is another key factor for speciation. Reproductive
isolation itself is not the cause of diversification *per se*, however, it can lead to the accumulation of differences between populations, and help each population reach its fitness optimum (Rieseberg and Willis 2007). In plants, multiple barriers contribute to reproductive isolation among most species, including pre-pollination barriers, post-pollination pre-zygotic barriers and post-zygotic barriers. One of the central goals of speciation research therefore is to quantitatively determine (1) the contribution of different barriers to total reproductive isolation among closely related species, (2) the evolutionary pattern of these barriers in the process of speciation, and (3) the genetic basis of reproductive barriers (Coyne and Orr 2004; Rieseberg and Willis 2007; Lowry et al. 2008; Widmer et al. 2009).

Although only a small number of studies have investigated in detail the contributions of each isolating barrier to reproductive isolation in plant, in general they indicate that pre-zygotic isolation barriers are stronger than post-zygotic barriers (Ramsey et al. 2003; Kay 2006; Martin and Willis 2007; Lowry et al. 2008; Widmer et al. 2009). The comparison across different study systems suggested that pre-zygotic isolation, on average, is twice as strong as post-zygotic isolation in flowering plants (Lowry et al. 2008). Pre-zygotic isolating barriers in plants are mostly in the form of floral isolation, which is a form of pre-pollination isolation mediated through the morphology of flowers (mechanical isolation) or the behaviour of pollinators (ethological isolation) (Kay and Sargent 2009; Schiestl and Schlüter 2009). For example, *Mimulus lewisii* and *M. cardinalis* co-occur in sympatry, but are reproductively isolated from each other by floral isolation (ethological): *M. lewisii* is visited by bees (100%), whereas *M. cardinalii* is visited by hummingbird (97.9%) (Ramsey et al. 2003). However, in most studied plant systems, pre-zygotic isolating barriers alone are rarely sufficient for complete reproductive isolation (Lowry et al. 2008; Kay and Sargent 2009) (but see Chapter 1) but act in concert with other isolating factors to reduce the homogenizing effects of gene flow.
Reproductive barriers may evolve differently during the process of species divergence. In animals, studies on the relationship between reproductive isolation and genetic distance indicated that pre-zygotic isolation evolves earlier and faster than post-zygotic reproductive isolation in general (Coyne and Orr 1989, 1997; Mendelson 2003). The evolutionary patterns of the trait depend on their genetic architecture. In animals, pre-zygotic isolation is often due to changes on few genes with large effects, and can therefore evolve rapidly and play an important role in the early stage of speciation, whereas post-zygotic barrier tend to be due to changes in many genes with minor effects, and therefore evolve more gradually (Coyne and Orr 2004).

However, the evolutionary patterns of reproductive isolation in plants are not well understood. Investigations on the relationship of reproductive isolation and genetic distance on three angiosperm genera (Glycine, Silene, Streptanthus) indicated varied evolutionary patterns among them (Moyle et al. 2004). In Silene, both pre-zygotic and post-zygotic isolation are significantly positively correlated with genetic distance, whereas in Glycine and Streptanthus, either weak or no correlation was found. Comparison of the evolutionary rate of different reproductive isolation in Silene and Streptanthus also showed no significant differences between pre-zygotic and post-zygotic isolation (Moyle et al. 2004). Comparative studies on reproductive isolation in two orchids groups with different pollination specificity indicated that pre-zygotic isolation evolves independently of genetic distance, whereas post-zygotic evolves gradually with genetic distance in orchids with less specific pollination system, but not in orchids with highly specific pollination system (Scopece et al. 2007; Scopece et al. 2008). Thus, the evolutionary pattern of reproductive isolation in plants seems to be less consistent than in animals. Whether such variable evolutionary patterns among different plant genera and different pollination systems were due to the different genetic basis
underlying isolating barriers, or due to the uncertainty of their reproductive strategy (pollinator-dependent) remains to be tested (Widmer et al. 2009).

The identification of the genetic basis of reproductive isolation is crucial for speciation studies, however, it is also challenging. Methods are especially important in this case, since many speciation studies are dealing with non-model systems.

There are two categories of approaches that have been employed for uncovering the genetic bases of reproductive isolation. The first one is a classical phenotype-genotype association approach. Once reproductive isolation has been narrowed down to one or very few specific traits, e.g., floral colour, the genetic basis of such traits can be revealed by quantitative trait locus (QTL) mapping and/or functional studies on candidate genes. The phenotype-genotype association approach has been used in many study systems and successfully identified genes involved in reproductive barriers, for example, *Mimulus* (Bradshaw et al. 1995; Bradshaw and Schemske 2003), *Aquilegia* (Hodges et al. 2002), *Cadia* (Citerne et al. 2006), and *Petunia* (Hoballah et al. 2007), etc. However, to apply this approach requires certain amount of knowledge on the traits underling reproductive isolation. Furthermore, either a large amount of genetic markers combined with a way of generating high diversity or a detailed molecular understanding of the traits are necessary for applying this approach. Although to develop enough genetic markers is possible in many systems due to the development of new sequencing technology, the information regarding the traits underlining reproductive barriers are only available in few systems, and are mainly on pre-pollination isolation barriers (Lowry et al. 2008; Kay and Sargent 2009; Rieseberg and Blackman 2010; Wolf et al. 2010). The second approach is the evolutionary genomics approach, which uses population genetic (or genomic) or subtractive cloning methods to detect genes directly under divergent selection without *a priori* knowledge on the phenotypical traits associated with the reproductive barrier. The comparative genomics
approach may not only uncover the genetic basis of reproductive barriers, but may also detect genes under natural selection. This can provide more information of the whole process of speciation, and extend our understanding of speciation genetics in detail. However, the power of this approach is often affected by many other factors, for example, population demography, selective sweeps, etc. (Wolf et al. 2010).

For both approaches, the biggest challenge is to validate the function of the identified genes / locus. Unlike some traditional genetic model systems, for example, Arabidopsis, most non-model plants have longer life cycles and bigger genomes, which make the application of typical functional validation methods (like transgenic experiments) difficult. Virus-induced gene silencing (VIGS), which is a method utilizes RNAi pathway in plants to knock-down target genes (Dinesh-Kumar S.P. et al. 2003), however hold a promise for functional studies of identified genes in non-model systems (Burch-Smith et al. 2004; Watson et al. 2005; Gould and Kramer 2007). Although specific optimizations on this method are needed for different plant systems, the introduction of VIGS to studies on the genetic basis of reproductive barriers promises to be extremely valuable (Hoballah et al. 2007).

Despite methodological challenges, significant progress on the genetic bases of reproductive barriers in plants has been made over the last years [see reviews by Lowry et al. (2008), Rieseberg and Blackman (2010)]. This progress has provided some clues for the key questions that are essential for understanding the evolutionary patterns of reproductive barriers, for example, whether reproductive barriers are built due to changes of few genes with large effects or many genes with small effects? What genetic elements contribute to reproductive barrier? What is the relative importance of gene expression changes compared to changes in enzymatic functions? Are the genes associated with reproductive barriers always are linked to the traits under selection?
The genetic architecture and features underlying reproductive barriers may differ due to the process of how they were fixed during (under strong divergent selection or not), as well as the genetic pathway underlying the traits (Widmer et al. 2009). For traits driven by strong divergent natural selection, for example, strong pollinator-mediated selection, the intermediate phenotypes normally have relatively low fitness, and a shift from one adaptive peak to another is more likely to be due to changes at few loci with large effects. By contrast, for traits evolved gradually under weak selection, fitness peaks will be bridged by intermediate phenotypes, and therefore more likely to be caused by many loci with small effects (Widmer et al. 2009).

Most traits associated with pre-pollination barriers are likely under strong pollinator-mediated selection (as discussed above). These traits have mostly been found to be controlled by very few genes of large effect (Lowry et al. 2008). For example, changes of floral colour in Petunia were found to be due to the changes of only one gene, the transcription factor ANTHOCYANIN-2 (AN2) (Hoballah et al. 2007). Genes involved in pre-pollination barriers were dominated by regulatory elements (bot cis and trans), which indicated that gene expression changes might play more important in the evolution of pre-pollination barriers (Rieseberg and Blackman 2010). However, few studies have detected selection acting on these genes. This may be because the mechanisms by which selection act on genes are not well understood.

In contrast to pre-pollination barriers, post-zygotic barriers involve two incompatible loci, as described by Bateson-Dobzhansky-Muller (BDM), that have evolved gradually. The genetic architecture of post-zygotic barriers in plants is more complex than that of pre-pollination barriers. The number of loci involved in post-zygotic barriers is often not directly comparable among plant systems, because it increases exponentially with time since the initial splitting point due to the “snow-ball” effect (Matute et al. 2010; Moyle and Nakazato
General introduction

2010). Genes involved in post-zygotic barriers were contributed by both regulatory elements and protein-coding genes (Rieseberg and Blackman 2010).

Our knowledge on the genetic basis of reproductive barrier in flowering plants is still incomplete. Firstly, most studies on pre-zygotic barriers primarily focused on floral colour and shape, whereas floral odour, which plays a very important role for pollinator adaptation (Schiestl 2010), has largely been ignored. Secondly, studies on the genetic basis of reproductive barrier have rarely been performed in a geographical and phylogenetic context, and the evolutionary pattern of genes underlying reproductive barrier was little understood. Third, interactions between the genetic basis of reproductive barrier and ecological factors have not been investigated. Although this requires the manipulation of the expression or enzymatic function of barrier genes in vivo - which is very difficult in many non-model systems - it is essential for the understanding of how reproductive isolation evolved. Such studies will extend our understanding of the process of speciation.

3. Study system

*Ophrys* L. is a Mediterranean genus of orchids, which is distributed from the Canary Islands to the Caspian Sea and from southern Scandinavia to the northern Maghreb (Delforge 2006). One of most extraordinary features of the genus *Ophrys* is their pollination system. They mimic female insects’ mating signals, and are pollinated by male insects during matting attempts (Kullenberg 1950; Paulus and Gack 1990a; Schiestl et al. 1999). This mechanism is called sexual deception. Among mimicries of different signals, chemically mimicry of sex pheromones has been suggested as the most important factor (Schiestl et al. 1999; Mant et al. 2005; Vereecken and Schiestl 2009). Since the mating signals in insects are usually very specific, the sexually deception pollination system in *Ophrys* is also highly specific (Paulus
and Gack 1990b). The species diversity in this genus is very high, despite debates on the taxonomy issues (Delforge 2006; Pedersen and Faurholdt 2007; Devey et al. 2008). Although pollinator-driven speciation has been suggested as reason for high species diversity (Paulus and Gack 1990b), the speciation and diversification process in *Ophrys* is still poorly understood.

In this thesis, three closely related and sympatrically occurring species in Southern Italy were investigated: *Ophrys sphegodes* Miller, *O. exaltata* subsp. *archipelagi* (GÖLZ & H. R. REINHARD) DEL PRETE, and *O. garganica* NELSON EX O. & E. DANESCH. It has been reported that *O. sphegodes* is pollinated by males of *Andrena nigroaenea* (Schiestl et al. 1999); *O. exaltata* is pollinated by males of *Colletes cunicularius* (Mant et al. 2005); and *O. garganica* is pollinated by *Andrena carbonaria* (Paulus and Gack 1990a).

4. **The aim of the thesis.**

The thesis presented here seeks to understand the mechanism and process of pollinator-mediated reproductive isolation (floral isolation) in a specialized pollination system (*Ophrys*), from both ecological and genetic aspects. Specifically, the following three questions are addressed: (1) What are the relative contributions of different components of reproductive isolation among the investigated species? (2) What is the genetic basis of reproductive isolation among the investigated species? and (3) What are the evolutionary patterns of the genes underlying the reproductive isolation among the species?

**Thesis outline**

**Chapter 1:** “Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids”
We measured the different components of reproductive barriers among three sympatric *Ophrys* species. Furthermore, we also performed population genetic analysis to investigate the pattern of genetic divergence among these species. The results indicated that floral isolation is the main (if not only) reproductive barrier among the investigated sympatric *Ophrys* species, and that the strength of floral isolation is sufficient to prevent significant inter-species gene flow. This highlights the role of floral isolation in the process of plant speciation, and provides a valuable example for us to understand the evolution of different reproductive barriers.

**Chapter 2:** “cDNA representational difference analysis reveals genetic basis of divergent traits among closely related sexually deceptive *Ophrys* orchids”

In this chapter, we report the use of a sensitive, powerful and economical subtractive cloning method, representational difference analysis (RDA), for identifying genes underlying divergent floral traits between closely related species. Using this method, we successfully identified and cloned several genes that are associated with different floral scents that were divergent between two *Ophrys* species. Among these genes, two might be involved in the biosynthesis pathway of alkenes, which are associated with floral isolation between studied species. Our study implies that RDA can be an effective tool for identifying genes underlying divergent traits between closely related species, when sequence information is limited.

**Chapter 3:** “Stearoyl-ACP desaturases are associated with floral isolation in sexually deceptive orchids”
We functionally characterized two candidate genes putatively encoding stearoyl-acyl-carrier-protein desaturases (SAD) in *Ophrys*, which appear to be involved in alkene biosynthesis. Gene expression showed that one *SAD* gene is associated with certain alkenes that are linked to floral isolation between two closely related species. The results indicated that SAD is associated with floral isolation between our study species. This study implies that evolution of reproductive barriers could be due to changes on very few genes in *Ophrys*.

**Chapter 4: “Genetic basis of pollinator adaptation in a specialized pollination system”**

We investigated the evolutionary pattern of six gene copies of the *SAD* gene family in three *Ophrys* species. Analysis of gene expression and sequence variation indicated that four homologs were significantly associated with the production of different alkenes among these *Ophrys* species. Two different regulation systems (*cis* and *trans*) on gene expression were found for genes underlying changes of different alkenes. Pollinator behaviour investigated by previous studies indicated that different changes in specific alkene production might be due to pollinator-mediated selection. The results shed light on the process of pollinator-mediated selection driving phenotypic changes in molecular level.

Taken all together, the presented thesis systematically investigated ecological and genetic aspects of floral isolation in a specialized pollination system. The results provided not only insightful information for us to understand how pollinator adaptation drives plant speciation in our study system, but also extended our knowledge on the ecological and genetic basis of plant adaptation and speciation process in general.

However, to fully understand the pollinator-driven speciation process in plants, taking geographical effects into account and applying theoretical modelling approaches are
necessary. Firstly, to measure the strength of floral isolation among closely related sympatric species in different geographical area and combine with population genetic studies would offer an opportunity to understand species evolutionary pattern; Secondly, detailed and systematically investigate genes under pollinator-mediated selection and their contribution to reproductive barrier in different populations are essential for understanding speciation pattern in *Ophrys*; Third, combine the ecological and genetic information in a theoretical model would help to reveal the pattern of pollinator-driven speciation and evolution in plants.
LITERATURE CITED


*Evolution* **51**:295-303.


Chapter 1

FLORAL ISOLATION IS THE MAIN REPRODUCTIVE BARRIER AMONG CLOSELY RELATED SEXUALLY DECEPTIVE ORCHIDS

Accepted for publication by EVOLUTION as:

Shuqing Xu¹,², Philipp M. Schlüter¹, Giovanni Scopece³, Hendrik Breitkopf³, Karin Gross¹, Salvatore Cozzolino³, Florian P. Schiestl¹*

1, Institute of Systematic Botany, University of Zürich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland.

2, Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, CH-8092 Zurich, Switzerland.

3, Department of Structural and Functional Biology, University of Naples Federico II, Via Cinthia, I-80126, Naples, Italy.

*, Corresponding author

Author contributions: SX, PMS, SC and FPS designed the research; SX, PMS, FPS and GS performed field experiments; SX performed floral odor analysis; GS performed cross experiments; HB performed AFLP experiments; KG performed ploidy level experiments; SX and PMS analyzed data; SX, PMS and FPS wrote the paper.
Abstract

Floral isolation is an important component of pollinator-driven speciation. However, up to now, only a few studies have quantified its strength and relative contribution to total reproductive isolation. In this study, we quantified floral isolation among three closely related, sympatric orchid species of the genus *Ophrys* by directly tracking pollen flow. *Ophrys* orchids mimic their pollinators’ mating signals, and are pollinated by male insects during mating attempts. This pollination system, called sexual deception, is usually highly specific. However, whether pollinator specialization also conveys floral isolation is currently under debate. In this study, we found strong floral isolation: among 46 tracked pollen transfers in two flowering seasons, all occurred within species. Accounting for observation error rate, we estimated a floral isolation index $\geq 0.98$ among each pair of species. Hand pollination experiments suggested that post-pollination barriers were effectively absent among our study species. Genetic analysis based on AFLP markers showed a clear species clustering and very few $F_1$ hybrids in natural populations, providing an independent evidence that strong floral isolation prevents significant inter-species gene flow. Our results provide the first direct evidence that floral isolation acts as the main reproductive barrier among closely related plant species with specialized pollination.

**Key words:** sexually deceptive orchids, *Ophrys*, pollination, floral odor, reproductive isolation, speciation.
Floral isolation is a form of pre-pollination reproductive isolation that can play an important role during the process of plant speciation (Grant 1994; Lowry et al. 2008; Kay and Sargent 2009; Schiestl and Schlüter 2009). Floral isolation can be mediated through the behavior of pollinators (ethological isolation) or the morphology of the flower (mechanical isolation) and work in concert with other, later acting reproductive barriers (Grant 1994; Fulton and Hodges 1999; Schemske and Bradshaw 1999; Ramsey et al. 2003; Aldridge and Campbell 2007; Schiestl and Schlüter 2009). The determination of the relative importance of different types of reproductive barriers among species has become a central topic in the study of speciation (Ramsey et al. 2003; Coyne and Orr 2004; Cozzolino and Scopece 2008; Lowry et al. 2008; Widmer et al. 2009). Previous studies have shown that in many plants, pre-zygotic isolation contributes more to total isolation than post-zygotic isolation (Rieseberg and Willis 2007; Lowry et al. 2008; Widmer et al. 2009). In the absence of geographic barriers to gene flow (i.e. among sympatric species), floral isolation can be the most important pre-zygotic barrier. However, the relative strength of pre-zygotic and post-zygotic isolation may differ between species, and may depend on the pollination system (Cozzolino et al. 2004; Cozzolino and Scopece 2008). In orchids, floral isolation has been suggested to be strong, since their associations with pollinators are often highly specific (Schiestl and Schlüter 2009).

*Ophrys* L. is a genus of sexually deceptive orchids, which mainly occurs in the Mediterranean area. In order to attract pollinators, these orchids mimic the olfactory, visual, and tactile signals of the females of their associated pollinator insects, and thereby induce so-called pseudocopulations in males, leading to pollination (Kullenberg 1961; Kullenberg and Bergström 1976; Paulus and Gack 1990a,b; Schiestl et al. 2000). In this pollination system, floral odor is the key factor for specific pollinator attraction (Schiestl et al. 1999; Schiestl et al. 2003; Mant et al. 2005a; Mant et al. 2005b; Peakall et al. 2010). One of the major characteristics of sexual deception is its high specificity, with each species of *Ophrys* only
attracting one or very few species of male insects as pollinator(s) (Paulus and Gack 1990b). Therefore, different *Ophrys* species, which are mostly genetically compatible and crossable, are potentially isolated from each other due to ethological floral isolation, i.e. the non-sharing of pollinator species (Ehrendorfer 1980; Paulus and Gack 1990b; Schiestl and Ayasse 2002; Scopece et al. 2007; Schiestl and Schlüter 2009). Mechanical floral isolation is also present between some *Ophrys* species, mainly between the sections *Pseudophrys* and *Ophrys* (Kullenberg 1950; Ågren et al. 1984; Borg-Karlson 1990; Cortis et al. 2009). Among these groups, different *Ophrys* species can be pollinated by the same pollinator in sympathy, because pollinia are attached to different parts of the pollinator’s body (head or abdomen), thus preventing pollen transfer between species. A recent study by Cortis et al. (2009), however, showed that cross-pollination can occur in natural population despite mechanical isolation, which indicates that mechanical isolation in *Ophrys* may not be a very strong barrier to gene flow.

Understanding the process of speciation and diversification in *Ophrys* orchids is challenging due to their high morphological variability, which can mean that it is often difficult to reliably identify species in the field. This is further complicated by the multiple and often highly divergent taxonomic treatments of the group. For example, the number of species in *Ophrys* listed by different authors ranges from 17 species (and 44 subspecies) (Sundermann 1980) or 19 species (Pedersen and Faurholdt 2007) to 250 species (Delforge 2006). Moreover, recent genetic and molecular phylogenetic studies of *Ophrys* showed low genetic divergence among species (Soliva et al. 2001; Soliva and Widmer 2003; Devey et al. 2008). The pattern of low genetic differentiation among species can be explained by two (non-exclusive) hypotheses: (1), the genus *Ophrys* may have undergone (a) recent radiation(s), or (2), there is frequent gene flow among species. Under the first scenario, *Ophrys* species-diversification is either due to pollinator shifts mediated by a change in key
floral traits (such as floral odor bouquets; Mant et al. 2005b; Schlüter et al. 2009; Vereecken et al. 2010) or habitat adaptation. However, under scenario 1, if the time since species diversification was short, neutral genetic structure would not yet be expected to show a clear separation among species (Harrison 1991; Klein 1998). Under scenario 2, it is assumed that the strength of reproductive isolation among sympatric Ophrys species is weak, perhaps due to low fidelity of pollinators, therefore resulting in frequent gene flow among species, which reduces the genetic differentiation among species after their initial divergence (Soliva and Widmer 2003; Devey et al. 2008). One fundamental difference between these two scenarios is the assumed strength of floral isolation among sympatric Ophrys species: the first scenario assumes strong floral isolation, whereas second scenario assumes relatively weak floral isolation.

The absolute pollinator specialization (i.e. the number of pollinators visiting each species) in Ophrys has previously been investigated (e.g. Paulus and Gack 1990b; Mant. et al 2005b). However, the relative pollinator specialization (pollinator sharing, ethological floral isolation) and the resulting proportion of inter-specific pollen transfer are still unknown. In this study, we directly tracked pollen flow within and among three sympatric and co-flowering, closely related Ophrys species, and quantified floral isolation as well as components of post-pollination reproductive barriers among these three species. Additionally, the genetic structure was investigated among species. Specifically, the following questions are addressed in this paper: (1) how strong is floral isolation among sympatric Ophrys species? (2) What is the contribution of pre-pollination, post-pollination pre-zygotic, and post-zygotic isolation barriers to the total reproductive isolation among sympatric Ophrys species? (3) What is the proportion of hybrids in natural populations?

Materials and Methods
STUDY SPECIES

In order to most effectively address the question of the relative importance of the different putative isolation mechanisms in *Ophrys*, a set of species with the following criteria are needed: (1), Species should occur and co-flowering in sympatry; (2), species should have the same ploidy level; (3), species should be closely related. According to these criteria, the species *Ophrys sphegodes* MILLER, *O. exaltata* subsp. *archipelagi* (GÖLZ & H.R.REINHARD) DEL PRETE, and *O. garganica* NELSON EX O. & E. DANESCH were chosen in this study, because these three species co-flower and co-occur sympatrically in Southern Italy, phylogenetic analysis indicates that these species are closely related (Devey et al. 2008), and ploidy levels of these species are expected to be same (D'Emerico et al. 2005).

PLANT MATERIAL

The species *Ophrys sphegodes*, *O. exaltata*, and *O. garganica* were identified based on floral morphology, according to criteria described by Mant et al. (2005b). At Capoiale (CAP: 41°54' N, 15°40' E), where all three species co-occur and co-flower, samples of these three species were collected in 2008 and 2009 for both scent and genetic analysis; at Marina di Lesina (MDL: 41°54' N, 15°20' E), where mostly *O. exaltata* and *O. garganica* co-occur and co-flower (and only very few individuals of *O. sphegodes* were found), these two species were collected in 2008 and 2009 only for floral scent analysis; at the more distant Foce Garigliano (FCG: 41°13' N, 13°46' E), where *O. exaltata* and *O. sphegodes* co-occur and co-flower (and *O. garganica* does not occur), these two species were collected in 2008 for both scent and genetic analysis. Each study area was all about five hectares in size, and was estimated to contain 2000-3000 flowering plant individuals (counting all three species). For each sampled plant individual, a piece of leaf tissue was collected, and placed in a plastic bag filled with silica gel (Sigma) for subsequent molecular analysis, and one labellum of an unpollinated flower was cut, placed in a 2 ml vial (Supelco) and rinsed in 500 µl hexane.
(Fluka) for one minute while gently shaking. Thereafter, the labellum was removed and all scent samples were stored at -28°C until being analyzed by gas chromatography (GC). In total, 73 *O. sphegodes* (49 from CAP, 24 from FCG), 72 *O. exaltata* (48 from CAP, 24 from FCG), and 26 *O. garganica* (all from CAP) were sampled for genetic analysis; 100 *O. exaltata* (35 from CAP, 34 from MDL, 31 from FCG); 94 *O. sphegodes* (62 from CAP, 12 from MDL, 20 from FCG) and 56 *O. garganica* (30 from CAP, 26 from MDL) were sampled for floral odor analysis.

**IN SITU MEASUREMENT OF FLORAL ISOLATION**

We used an experimental approach with a plot design to measure floral isolation. The plots were set up in the field as follows, at the same localities as naturally occurring plants. Two individuals of each species were randomly positioned in each plot (6 plants for each plot in CAP and MDL, 4 plants for each plot in FCG where *O. garganica* was absent from the natural populations). The distance between neighboring plants was 0.5 m. For each experiment, 20 plots were set up along a transect through the habitat of the orchids. The distance between neighboring plots was 20 m, since the average pollinia-carrying distance of *Colletes* pollinators was estimated to be around 5 m (Peakall and Schiestl 2004).

Plants for the plot experiments were picked from natural populations, each flower was checked for pollinia removal or pollen deposition, and pollinia were stained alternately with the dyes brilliant green, methylene aniline blue, orange G and trypan red as described previously (Peakall 1989). The color used for each species was randomized between experiments to reduce potential effects of staining color on pollinator behavior. The inflorescence was put in a water-filled 15 ml plastic tube placed in the ground. Pollinia removal and deposition of massulae were recorded three days after setting up the plots.
Because *Ophrys* massulae are relatively small, and the assessment of their presence requires some experience in the field, there is a potential for observation errors to happen. Thus, to assess the observation error rate, in a subset of the experiments, plants were checked at several time points. For about half of the plant individuals used in the plot experiments, one unpollinated flower labellum was removed from the inflorescence to collect floral odor as described above.

In 2008, two replicates of the experiment were performed at each of the following locations: CAP, MDL and FCG (in total 6 experiments, 120 plots). In 2009, two replicates of the experiment were performed at locations CAP and MDL (in total, 4 experiments, 80 plots). The experiment at FCG was not repeated in 2009 due to the relatively poor overall pollinator activity at this location in 2008 (see Table 1). At each location where experiments were performed, the pollination success of naturally occurring plants in the surrounding area (within 20 meters of the transect) was recorded at the end of the flowering season.
Table 1. Summary of plot experiments. Plot experiments were set up in 2008 and 2009. CAP, MDL and FCG refer to locations where experiments were set up. Number of total flowers refers to the total number of flowers from three species used in each plot experiment; Number of flowers with stained pollinia refers to number of flowers carrying stained pollinia – there were no flowers with unstained pollinia at the outset of the experiment, but some flowers without pollinia (either due to pollinators or handling by experimenter). Number of plants refers to total number of plants from three species in each plot experiment; Pollinated flowers refers to number of flowers receiving pollinia; Flowers pollinated by stained pollinia refers to number of flowers that received stained pollinia; Pollinia removal refers to number of pollinia removed; Inter-species pollinia transfer refers to number of inter-species pollinia movements observed in plots. Number of visited plants counted as number of plants which lost or received pollinia.

<table>
<thead>
<tr>
<th>Year</th>
<th>2008</th>
<th>2009</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot</td>
<td>CAP1</td>
<td>CAP2</td>
<td>MDL1</td>
</tr>
<tr>
<td>Number of total flowers</td>
<td>153</td>
<td>168</td>
<td>202</td>
</tr>
<tr>
<td>Number of flowers with stained pollinia</td>
<td>146</td>
<td>154</td>
<td>165</td>
</tr>
<tr>
<td>Number of plants</td>
<td>58</td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td>Pollinated flowers</td>
<td>14</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Flowers pollinated by stained pollinia</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Pollinia removal</td>
<td>38</td>
<td>16</td>
<td>76</td>
</tr>
<tr>
<td>Inter-species pollinia transfer</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Visited plants</td>
<td>20</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>Stained pollinia on O. sphegodes flowers</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stained pollinia on O. exaltata flowers</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Stained pollinia on O. garganica flowers</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
MEASUREMENT OF POST-POLLINATION ISOLATION BARRIERS

Manual crosses were performed in spring 2010 in the greenhouse of the Department of Structural and Functional Biology, University of Naples Federico II. All crossed plants were collected from sympatric natural populations of the three investigated species at CAP, where we set up the plot experiments.

To prevent uncontrolled pollinations, plants were placed in cages covered with a thin net prior to flowering. Pollination experiments were performed by removing pollinia by touching the viscidia with a plastic toothpick and placing them on the stigmas of other plants of the same species (intra-species pollinations), or of a different species (inter-species pollinations). Care was taken to pollinate no more than two flowers per individual to prevent the potential negative effects of over-pollination on fruit set and seed viability.

All possible crossing combinations among *O. exaltata*, *O. garganica*, and *O. sphegodes* were performed bidirectionally (yielding a total of 78 crossings, see Table 2). Ripe fruits were collected and stored in silica gel. Seeds were then observed under an optical microscope with 100 × magnification and assigned to two mutually exclusive categories: viable and inviable seeds, based on the presence or absence of embryos, respectively.
Table 2. Fruit set formation ratio after hand pollination within and between species pairs. The ratio (%) was calculated by $100 \times \frac{\text{number of fruit sets}}{\text{number of crosses performed}}$. No significance was found between inter-species and intra-species crosses.

<table>
<thead>
<tr>
<th>Crossing type</th>
<th>Pollen donor ($\delta$)</th>
<th>Pollen receiver ($\varphi$)</th>
<th>Number of crosses</th>
<th>Fruit sets</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-species</td>
<td>O. sphegodes</td>
<td>O. garganica</td>
<td>10</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>O. garganica</td>
<td>O. sphegodes</td>
<td>10</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>O. exaltata</td>
<td>O. sphegodes</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>O. sphegodes</td>
<td>O. exaltata</td>
<td>6</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>O. exaltata</td>
<td>O. garganica</td>
<td>10</td>
<td>9</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>O. garganica</td>
<td>O. exaltata</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Intra-species</td>
<td>O. sphegodes</td>
<td>O. sphegodes</td>
<td>8</td>
<td>6</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>O. garganica</td>
<td>O. garganica</td>
<td>10</td>
<td>9</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>O. exaltata</td>
<td>O. exaltata</td>
<td>7</td>
<td>7</td>
<td>100</td>
</tr>
</tbody>
</table>
PLOIDY LEVEL ANALYSIS

Differences in ploidy can provide an important barrier to gene flow. While we expected all three study species to be diploid, with a chromosome number of 2n = 36 having been reported from *O. sphegodes* and *O. garganica* (Greilhuber and Ehrendorfer 1975; D'Emerico et al. 2005), the ploidy level of *O. exaltata* has not previous been reported. Furthermore, ploidy may also be variable within species and/or among different populations. Therefore, we investigated the ploidy levels of all three species in our study populations. Ploidy levels of the three species were analyzed using pollinia. Two pollinia of a single flower per individual were collected in spring 2010 from CAP and MDL populations. In total, 86 *O. sphegodes*, 71 *O. exaltata*, and 6 *O. garganica* samples were analyzed. Using flow cytometry, we analyzed the relative ploidy level for each individual separately. For sample preparation and analysis, we followed a two-step protocol (Doležel et al. 2007). Two pollinia were chopped and mashed together with approximately 25 mm² leaf material of *Phaseolus coccineus* (2n, 1C = 1.01 ± 0.4 pg; Bennett and Leitch 2005), which served as internal standard (IS), with a sharp razor blade in 1 ml ice-cold Baranyi’s solution (0.1 M citric acid, 0.5% Triton X-100; Baranyi and Greilhuber 1995). After filtering the suspension through a 30 µm CellTrics® disposable filter (Partec GmbH, Münster, Germany), the filtrate was centrifuged (5 min, 380 × g, room temperature) using a Sorvall® RMC 14 centrifuge (Kendro Revco Lindberg Heraeus Sorvall, Asheville, USA). After removal of supernatant, nuclei were resuspended in 40 µl of ice-cold Baranyi’s solution. 160 µl of Otto II solution (0.4 M Na₂HPO₄) supplemented with DAPI (4’, 6-diamidino-2-phenylindole; final concentration: 4 µg ml⁻¹) was added and relative fluorescence intensity was recorded using a Cell Lab Quanta™ SC-MPL flow cytometer (Beckman Coulter, Fullerton, Canada) with a mercury arc lamp. Only samples with pollinia peaks of at least 1000 counts and a coefficient
of variation of less than 10% were analyzed. To determine relative ploidy level of the three species, the ratio between the median of pollinia peaks and the median of IS peaks was calculated.

**FLORAL ODOR ANALYSIS**

GC analysis was performed as described by Mant et al. (2005b) with 300 ng \( n \)-octadecane (C18) added to the floral extracts as an internal standard. One \( \mu l \) of each sample was injected into an Agilent 6890 GC at 50 °C, followed by opening of the split valve and heating to 300 °C at rate of 4°C/min. An HP-5 column and flame ionization detector (FID) were used, and hydrogen was used as carrier gas, with nitrogen as the make-up gas. For identification of compounds, several samples were re-analyzed by GC with a mass selective detector (GC/MSD; Agilent 5975) using the same oven and column parameters. Spectrum and retention time of compounds were compared with those of synthetic standards, i.e., alkanes: nonadecane (C19), henicosane (C21), docosane (C22), tricosane (C23), tetracosane (C24), pentacosane (C25), hexacosane (C26), heptacosane (C27), octacosane (C28), nonacosane (C29); and alkenes: (Z)-7- heneicosene [(Z)-7-C21], (Z)-9-heneicosene [(Z)-9-C21], (Z)-7-tricosene [(Z)-7-C23], (Z)-9-tricosene [(Z)-9-C23], (Z)-7-pentacosene [(Z)-7-C25], (Z)-9-pentacosene [(Z)-9-C25], (Z)-11-pentacosene [(Z)-11-C25], (Z)-12-pentacosene [(Z)-12-C25], (Z)-7-heptacosene [(Z)-7-C27], (Z)-9-heptacosene [(Z)-9-C27], (Z)-11-heptacosene [(Z)-11-C27], (Z)-12-heptacosene [(Z)-12-C27], (Z)-7-nonacosene [(Z)-7-C29], (Z)-9-nonacosene [(Z)-9-C29], (Z)-11-nonacosene [(Z)-11-C29], (Z)-12-nonacosene [(Z)-12-C29], where (Z)-number indicates the cis double-bond position. For sources of standard compounds see Mant et al (2005a). It is noted that the discrimination of (Z)-11- and (Z)-12- alkenes was not possible with the GC parameters used. The relative amount of each odor
compound was calculated as the proportion of the total amount of all alkenes and alkanes of a chain length between 18 and 30 carbons.

GENETIC DIVERGENCE AMONG SPECIES

Genomic DNA was extracted using GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich Italy). The AFLP procedure was performed as described by Vos et al. (1995), with modifications as reported in Moccia et al. (2007) using fluorescent dye-labeled primers. An initial trial using 14 different primer combinations on four individuals each of *O. sphegodes* and *O. exaltata* was conducted to identify those primers that yield the highest number of easily detectable polymorphic peaks that were different between the two species. After the screening, six selective primer combinations were chosen: FAM-EcoRI-AGC/MseI-ACAC, NED-EcoRI-ACC-/MseI-ACTG, HEX-EcoRI-AGC/MseI-ATCG, FAM-EcoRI-ATG/MseI-CGG, NED-EcoRI-AAC/MseI-CGC and HEX-EcoRI-AGC/MseI-CCAA. For the restriction digestion, the enzymes EcoRI and MseI were used on a total of 250 ng of genomic DNA. Ligation of EcoRI and MseI adapters took place in the same reaction. Two µl of the restriction-ligation product were used for a preselective PCR with primers having one selective base. For the successive selective PCR, 1 µl of a 1:10 dilution of the PCR product was used. Primers were the same as in the preselective PCR, but with three or four additional selective bases. Fragment separation and detection took place on a 3130 Genetic Analyzer (Applied Biosystems, Foster City, USA). GeneScan-500 LIZ (Applied Biosystems, Foster City, USA) was used as internal standard. Processing of the raw data and sizing of the fragments were done with Genemapper 3.7 software (Applied Biosystems, Foster City, USA). Absence or presence of AFLP bands was carefully scored by eye. To avoid artifacts, only AFLP markers that could be unambiguously scored over the whole data set were included in the binary matrix.
AFLP analysis was performed as two experiments at different dates, runs and scored independently, preventing us from merging two AFLP data sets. These two separate data sets were therefore analyzed separately: the first data set contains 58 *O. sphegodes* (34 from CAP, 24 from FCG), 55 *O. exaltata* (31 from CAP, 24 from FCG), and 26 *O. garganica* (all from CAP) individuals; the second data set contains 30 *O. sphegodes* (15 from CAP, 15 from FCG) and 32 *O. exaltata* (17 from CAP, 15 from FCG). In the second data set, both species from FCG population were the same individuals as the first data set.

**EVALUATION OF THE STRENGTH OF INDIVIDUAL ISOLATING BARRIERS**

The strength of each type of reproductive isolation barrier was calculated based on the quantitative approach suggested by Lowry et al. (2008) and Martin and Willis (2007). The floral isolation index was calculated based on the following formula:

\[ \text{RI}_{\text{floral}} = 1 - \left( \frac{\text{observed/expected inter-species pollen flow}}{\text{observed/expected intra-species pollen flow}} \right) \]

The precision of this estimate of \( \text{RI}_{\text{floral}} \) is limited by the number of observed events in our experiments and any potential observation error introduced when checking plants for pollination or pollinia removal. Therefore, \( \text{RI}_{\text{floral}} \) was re-calculated so as to account for these errors. Three of our experiments were checked twice, any error recorded, and these data were used to estimate the observation error rate. For example, on the same flower, massulae might have been recorded at the first but not on the second inspection, indicating that one out of these two data points was probably erroneous. This error rate followed a pattern, i.e. the best fitted curve decreased exponentially with the number of experiments we performed, concordant with our expectation that observation error rate decreases as the observers’ experience increases [\( \ln(\text{Error rate}) = -0.25*i \pm SD \), where \( i \) refers to the number of]
experiments performed by that time]. Based on this formula, the error rate for each experiment was estimated, and the floral isolation index recalculated by allowing observation error to occur at the estimated error rate ± SD. We repeated this estimation procedure 1000 times to obtain a simulated distribution of floral isolation index values, and used this distribution to obtain mean and 95% confidence values for RI_{floral}.

Since in orchids, female gametophyte development and fruit set formation usually happens after successful pollination with compatible pollen (Zhang and O’Neill 1993), the post-pollination pre-zygotic isolation index can be estimated as the proportion of fruit set (i.e. capsules) formed following inter-species pollinations, relative to the proportion of fruit set formed following intra-species pollinations with each parental species (Scopece et al. 2007):

$$RI_{\text{post-pollination-pre-zygotic}}=1-(\text{ratio of fruit set formed in inter-species crosses}) / \text{(average ratio of fruit set formed in parental intra-species crosses)}$$

Similarly, the post-zygotic isolation index was estimated by viable seeds and quantified based on the following formula (Scopece et al. 2007):

$$RI_{\text{post-zygotic}}=1-(\text{proportion of viable seeds in inter-species cross}) / \text{(average proportion of viable seeds in parental intra-species cross)}$$

DATA ANALYSIS

Linear discrimination analysis (LDA) was used for analysis of floral scent based on relative amount of hydrocarbons. Comparison of fruit set formation ratios among inter-species crosses and intra-species cross was performed by Fisher exact tests. The significance of different seed viability among inter-species and intra-species crosses was assessed using Student’s $t$-test, after normality testing of the data distribution by the Shapiro test (Royston 1982). Statistical analysis of AFLP data was performed in FAMD 1.25 ($\Phi_{ST}$ and PCoA).
(Schlüter and Harris 2006) and Hindex 1.41 (hybrid index; Buerkle 2005). Principal coordinate analysis (PCoA) was based on Jaccard’s similarity coefficient.

Correlations between pairwise floral odor and genetic distance was assessed in a Mantel test (10000 permutations). Here, floral odor distance was calculated as a Euclidean distance, and genetic distance was calculated as 1-Jaccard’s similarity. Except for the analysis of AFLP data, all statistical analyses in this study were carried out in R 2.11.0 (R Development Core Team 2010).

Results

POLLINATION SUCCESS

Among the natural populations, the average pollination success (defined by the percentage of pollinated flowers) was 10.8% in *O. sphegodes*, 18.8% in *O. exaltata*, and 3.2% in *O. garganica*, based on two flowering seasons (2008 and 2009). Furthermore, within each species, pollination success varied among populations (Figure 1). No significant differences were found between the two years of observations. The overall pollination success in our plot experiments was similar to the natural populations. Pollination success in plot experiments was about 7.8%, 9.2% and 4.1% for *O. sphegodes*, *O. exaltata* and *O. garganica* respectively. The population variation in pollination success was similar for natural populations and plot experiments, except for *O. sphegodes* in MDL where pollination success was much lower than in CAP (3.9%), whereas in plot experiments, the pollination success was similar to CAP (8.9%).
Figure 1. Pollination success of the three *Ophrys* species in different populations in 2008 and 2009. (A) natural population; (B) plot experiments. *O. exaltata*, *O. sphegodes* and *O. garganica* are shown in grayscale as indicated in the figure inset. Names on the axis refer to populations (see Materials and Methods). The number of plants surveyed is indicated on the top of each bar.
FLORAL ISOLATION

The flowering time of *O. sphegodes* and *O. exaltata* was similar, whereas the peak of *O. garganica* blooming was 1-2 weeks later (Xu, field observations). However, there was a broad overlap in flowering time, with around 70% of *O. garganica* flowers, and around 95% of *O. sphegodes* and *O. exaltata* flowers open during the experimental period.

Out of all 1855 flowers and 1686 stained pollinaria used in the experiments, 131 flowers (7.1%) were pollinated. Among these pollinated flowers, 46 flowers were pollinated with stained pollinia. Fifteen *O. sphegodes* flowers, 21 *O. exaltata* flowers, and 10 *O. garganica* flowers received stained pollinia (Table 1). All of the 46 pollination events with stained pollinia were within species, and we did not observe a single inter-species transfer. Thus, because no inter-species pollen flow was observed, the floral isolation index equals 1. Likewise, the simulated data incorporating observation error rates (see Methods) showed strong pre-pollination reproductive isolation among each species, $RI_{floral} \geq 0.98$ (Table 3 and Figure S1).
Table 3. Components of reproductive isolation among *Ophrys* species (numbers are isolation indices; mean ± SD). Indices of both post-pollination pre-zytoci and post-zygotic are not significantly different from zero.

<table>
<thead>
<tr>
<th>Species A</th>
<th>Species B</th>
<th>Floral isolation</th>
<th>Post-pollination pre-zygotic</th>
<th>Post-zygotic</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. exaltata</em></td>
<td>0.980 ± 0.0015</td>
<td>-0.14</td>
<td>-0.22 ± 0.53</td>
</tr>
<tr>
<td><em>O. garganica</em></td>
<td><em>O. exaltata</em></td>
<td>0.982 ± 0.0015</td>
<td>0.05</td>
<td>0.16 ± 0.41</td>
</tr>
<tr>
<td><em>O. garganica</em></td>
<td><em>O. sphegodes</em></td>
<td>0.980 ± 0.0015</td>
<td>-0.21</td>
<td>-0.04 ± 0.47</td>
</tr>
</tbody>
</table>
POST-POLLINATION ISOLATION

Post-pollination, pre-zygotic isolation was estimated as the fruit set ratio after hand pollination. Most of the inter- and intra-species crosses led to the development of capsules (Table 2). The lowest fruit set ratio was found for intra-species crosses in *O. sphegodes*, although this was not statistically significant when compared to other crosses. Thus, the post-pollination pre-zygotic isolation index among each species was estimated to be very low. For species pairs *O. sphegodes* / *O. exaltata* and *O. sphegodes* / *O. garganica*, the isolation indices were negative (however, not significantly different from zero), which might indicate that inter-species crosses performed better than intra-species crosses, and for *O. garganica* / *O. exaltata*, the isolation indices was close to zero (Table 3).

The proportion of seeds with embryos (viable seeds) was used to estimate the post-zygotic isolation index. The number of seeds analyzed for each capsule was 324 ± 103 (mean ± SD). Among all fruits, the average percentage of viable seeds was 46.8 ± 21.8% (mean ± SD). We did not find a significant difference between any inter- or intra-species crosses (Figure 2). Similar to post-pollination pre-zygotic isolation, the mean post-zygotic isolation index was also negative between species pairs *O. sphegodes* / *O. exaltata* and *O. sphegodes* / *O. garganica*, whereas for *O. garganica* / *O. exaltata*, the mean value index was slightly higher (Table 3). However, statistical analysis showed that none of these values were significantly different from zero.
Figure 2. Presence of embryos for crosses among each species (E: *O. exaltata*; G: *O. garcanica*; S: *O. sphegodes*). No difference was found between inter-species and intra-species crosses based on Student’s *t*-test, (*p* > 0.05). Error bars depict the standard error of the mean.
PLOIDY LEVEL ANALYSIS

No difference in the ploidy level was detected among the three species. The ratios of the relative inflorescence intensity between the pollinia and the IS are shown in Figure S2. All samples of these species showed similar relative genome size. It is most likely that all three species in our study are diploid, since previous studies showed that *O. sphegodes* and *O. garganica* are diploid (Greilhuber and Ehrendorfer 1975; D’Emerico et al. 2005).

FLORAL ODOR BOUQUET

The differences in floral odor bouquets among studied species were similar to those reported previously (Mant et al. 2005b). The major floral odor difference among the species was the proportion of different alkenes. By LDA, 236 out of 250 samples (94.4%) were classified as the same species as they were identified in the field based on floral morphology according to the criteria described by Mant et al (2005b). The morphological/chemical identification mismatch rate between *O. sphegodes* and *O. exaltata* was 4.1%, between *O. sphegodes* and *O. garganica* 3.3% and between *O. garganica* and *O. exaltata* 0.64%.

GENETIC DIVERGENCE AMONG SPECIES

AFLP data sets one and two contained 242 and 322 markers, respectively. Genetic divergence among population pairs, as estimated by pairwise $\Phi_{ST}$, was relatively low. For the first data set, where all three species from CAP were analyzed, the lowest $\Phi_{ST}$ value (0.044) was found between *O. sphegodes* and *O. garganica*, and the highest (0.064) was found between *O. garganica* and *O. exaltata* (see Table 4). However, the differences among species pairs were very small. For the second dataset, where *O. sphegodes* and *O. exaltata* from both CAP and FCG were analyzed, the highest $\Phi_{ST}$ (0.074) was found between *O. sphegodes* in
CAP and *O. exaltata* in FCG, whereas $\Phi_{ST}$ values between *O. sphegodes* in CAP and *O. sphegodes* in FCG, and between *O. sphegodes* in CAP and *O. exaltata* in CAP were lowest (0.059) (Table 4). Overall, $\Phi_{ST}$ values were low, within-species $\Phi_{ST}$ values being slightly lower than between-species values.

Genetic structure among species was investigated by PCoA (Figure 5). This analyses suggest that the genetic similarity between *O. sphegodes* and *O. garganica* is higher than between the species pairs *O. sphegodes* / *O. exaltata* or *O. garganica* / *O. exaltata*. Although a few outliers were found for both CAP and FCG populations, the three species formed genetically separable clusters (Figure 5A, B, C and Table 4).

For the *O. sphegodes* / *O. exaltata* species pair, floral odor showed significant correlation with genetic distance for population FCG ($r = 0.42, p = 0.0001$), but not for population CAP ($r = -0.11, p = 0.69$). For species pairs *O. sphegodes* / *O. garganica* and *O. garganica* / *O. exaltata* in CAP, significant correlations were found in both ($r = 0.28, p = 0.0026$ and $r = 0.17, p = 0.036$, respectively).

For population CAP, no obvious F$_1$ hybrids were found among the three species, as defined by a mean maximum-likelihood hybrid index estimate between 0.4 and 0.6. In contrast, for population FCG, two samples were classified as potential F$_1$ hybrids between *O. sphegodes* and *O. exaltata* according to the same criteria (Figure 4). Both samples were classified as *O. sphegodes* based on floral odor discrimination analysis (Figure 3F). Overall, for samples from both populations (146 individuals), the percentage of F$_1$ hybrids was very low (1.37%).
Table 4. Pairwise population differentiation ($\Phi_{ST}$) and overlap based on floral odor bouquet and AFLP among each species for two data sets. Data set 1 contains all three species from CAP, whereas data set 2 contains *O. exaltata* and *O. sphegodes* from CAP and FCG. The number of matches in floral odor was assigned based on the comparison between linear discrimination analysis of floral odor bouquet and species identification in the field. Matches in AFLP were assigned based on visual comparisons between principal coordinate analysis of AFLP markers and species identification in the field.

<table>
<thead>
<tr>
<th>Species A</th>
<th>Species B</th>
<th>$\Phi_{ST}$</th>
<th>Number of matches in floral odor / Total number of samples</th>
<th>Number of matches in AFLP / Total number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. garganica</em></td>
<td>0.044</td>
<td>0 / 173 = 0%</td>
<td>2 / 75 = 2.67%</td>
</tr>
<tr>
<td>CAP</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. exaltata</em></td>
<td>0.055</td>
<td>2 / 209 = 0.95%</td>
<td>1 / 97 = 1.03%</td>
</tr>
<tr>
<td>CAP</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. garganica</em></td>
<td><em>O. exaltata</em></td>
<td>0.064</td>
<td>1 / 138 = 0.72%</td>
<td>1 / 74 = 1.35%</td>
</tr>
<tr>
<td>CAP</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data set 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. exaltata</em></td>
<td><em>O. sphegodes</em></td>
<td>0.074</td>
<td>5 / 68 = 7.35%</td>
<td>0 / 39 = 0%</td>
</tr>
<tr>
<td>FCG</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. exaltata</em></td>
<td><em>O. exaltata</em></td>
<td>0.063</td>
<td>121 / 126 = 96.03%</td>
<td>1 / 41 = 2.44%</td>
</tr>
<tr>
<td>FCG</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. exaltata</em> in</td>
<td><em>O. sphegodes</em></td>
<td>0.068</td>
<td>5 / 68 = 7.35%</td>
<td>3 / 49 = 6.12%</td>
</tr>
<tr>
<td>FCG</td>
<td>FCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. exaltata</em></td>
<td>0.059</td>
<td>2 / 209 = 0.95%</td>
<td>0 / 32 = 0%</td>
</tr>
<tr>
<td>CAP</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. exaltata</em></td>
<td>0.073</td>
<td>4 / 116 = 3.44%</td>
<td>0 / 42 = 0%</td>
</tr>
<tr>
<td>FCG</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td><em>O. sphegodes</em></td>
<td>0.059</td>
<td>150 / 151 = 99.34%</td>
<td>5 / 40 = 12.5%</td>
</tr>
<tr>
<td>FCG</td>
<td>CAP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. Plots of genetic structure (PCoA) and floral odor discrimination analysis for three species. (A) genetic structure of *O. exaltata*, *O. garganica* and *O. sphegodes* in population CAP. Axis x and y represent 10.5% and 5.4% of variance respectively; (B) genetic structure of *O. exaltata* and *O. sphegodes* in population FCG. Axis x, y and z represent 12.1%, 8.1% and 7.1% of variance respectively; (C) genetic structure of *O. exaltata* and *O. sphegodes* in populations CAP and FCG. Axis x, y and z represent 9.1%, 8.6% and 5.9% variance respectively; (D) floral odor bouquet of *O. exaltata*, *O. garganica* and *O. sphegodes* in CAP, LD1 and LD2 representing 58.5% and 41.2% of trace respectively; (E) floral odor bouquet of *O. exaltata* and *O. sphegodes* in FCG, LD1 and LD2 representing 55.9% and 44.1% of trace respectively; (F) floral odor bouquet *O. exaltata* and *O. sphegodes* in FCG, CAP and MDL, LD1 and LD2 representing 55.9% and 44.1% of trace respectively. An asterisk (*) indicates outliers for which both AFLP and floral odor bouquet were analyzed. Different colors indicate species (red: *O. exaltata*; green: *O. garganica*; blue: *O. sphegodes*).
Figure 4. Hybrid index (± SD) for individuals of each sympatric species pairs. The two dashed lines indicate hybrid index $h = 0.4$ and $h = 0.6$ cutoff values for assigning putative $F_1$ hybrids. Sample names of two putative hybrids are indicated. (A) Individuals of *O. exaltata* and *O. garganica* in population CAP; (B) Individuals of *O. exaltata* and *O. sphegodes* in population CAP; (C) Individuals of *O. exaltata* and *O. sphegodes* in population CAP; (D) Individuals of *O. exaltata* and *O. sphegodes* in population FCG.
Discussion

Reproductive isolation has been a central topic in the study of speciation (Coyne and Orr 1998; Moyle et al. 2004; Rieseberg and Willis 2007; Scopece et al. 2007; Scopece et al. 2008; Schiestl and Schlüter 2009; Widmer et al. 2009). Here, we quantified three different kinds of reproductive barriers (floral isolation, post-pollination pre-zygotic isolation and post-zygotic isolation), as well as ploidy level among three sympatric sexually deceptive *Ophrys* orchids using experimental approaches. Among these potential barriers, floral isolation was found to be very strong ($RI_{floral} \geq 0.98$), whereas later acting barriers were effectively absent in our study species. Furthermore, population genetic analysis showed a clear separation between species despite low genetic divergence, with few hybrids within natural populations. Our results shed light on the role of plant-pollinator interactions in the evolution of reproductive isolation and plant speciation. We suggest that pollinator adaptation, that conveys strong floral isolation, is the main driver of speciation in this plant group with highly specialized pollination.

FLORAL ISOLATION IN PLANT SPECIATION

Floral isolation has been found in many plant-pollination systems, such as *Ipomopsis* (Grant 1992), *Mimulus* (Schemske and Bradshaw 1999; Ramsey et al. 2003), *Nicotiana* (Ippolito et al. 2004), *Petunia* (Hoballah et al. 2007), and *Silene* (Goulson and Jerrim 1997; Wälti et al. 2008), and meta-analyses indicate that floral isolation acts as a strong reproductive barrier in various families of flowering plants (Grant 1994; Lowry et al. 2008; Schiestl and Schlüter 2009; Kay and Sargent 2009; Schiestl in press). However, there are few examples where floral isolation alone is sufficient to maintain species differentiation in sympathy (Kay and Sargent 2009; Schiestl in press). In most studied cases, floral isolation acts together with other isolation barriers (post-pollination isolation, ecogeographic isolation etc.; Lowry et al. 2008). Co-occurrence of floral isolation with other isolation barriers may be
due to two reasons: (1) in most plant systems, it is unlikely that floral isolation has initially evolved in sympatry, because the shift to completely new pollinators may require changes in many floral traits (but see Bradshaw and Schemske 2003 and Hoballah et al. 2007). Therefore, geographical or habitat-associated barriers would often be involved in the evolution of floral isolation; (2) once floral isolation is established, secondary isolation barriers can build up over time (Via and West, 2008; Matute et al., 2010; Moyle and Nakazato, 2010). However, in order to better understand the contribution of floral isolation to plant speciation, as well as its evolutionary patterns, cases in which only floral isolation is involved in the speciation process are particularly valuable.

REPRODUCTIVE ISOLATION AMONG SYMPATRIC OPHRYS SPECIES

As shown in our study, _Ophrys_ may represent a case in which floral isolation is the most important barrier to gene flow among species with a large geographic overlap, sympatric occurrence, and flowering time overlap in their given habitats. The here investigated _Ophrys_ species showed strong ethological floral isolation and a lack of post-pollination isolation barriers. Our conservative estimation of the strength of floral isolation, which took into account the number of trackable pollination events and the observation error, showed that the floral isolation index among each _Ophrys_ species pair was higher than 0.98. This estimation is consistent with our AFLP data, which indicate 1.37% (two out of 146 samples) putative _F_1 hybrids between _O. sphegodes_ and _O. exaltata_. Our finding of strong floral isolation is also consistent with expectations from pollinator-behavior studies often indicating little pollinator sharing among co-flowering species pairs (Kullenberg 1961; Paulus and Gack 1990b; Mant et al 2005a, 2005b; Schlüter et al. 2009). Although we have only partially quantified post-zygotic isolation barriers, it is clear from our results that the
early acting floral isolation is the major (if not the only) reproductive barrier among closely related *Ophrys* species.

Recently, floral isolation in *Ophrys* has come under scrutiny, because studies based on genetic markers argued that floral isolation in *Ophrys* might be weak and allow for considerable gene flow across species boundaries (Soliva and Widmer 2003; Devey et al. 2008). However, the data presented in these studies only allow for an indirect inference on floral isolation. The genetic pattern observed by Soliva and Widmer (2003) estimated gene flow among species based on $F_{ST}$, however, such estimation can be misleading and should be interpreted cautiously (reviewed by Whitlock and McCauley 1999). In the study by Devey et al. (2008), based on phylogenetic analysis, the authors suggested that cross-pollination is common among species since no clear phylogenetic patterns were found based on DNA sequences (ITS and plastid markers) and AFLP markers. However, gene flow cannot be assessed based on phylogenetic analysis without proper population genetic data (Slatkin 1985). Interestingly, Devey et al. (2008) found no significant genetic differentiation between *O. exaltata* and *O. sphegodes* (based on their ITS, plastid markers, and AFLP), whereas we found a clear clustering pattern with AFLP markers and even more so with floral-odor bouquet analyses. This discrepancy is likely because our study provides a more fine-grained resolution through the analysis of multiple populations and large sample sizes. We suggest one should be careful in drawing any conclusions on gene flow from investigations using only molecular markers or morphological data (see also discussion on hybridization in *Ophrys* below). As a consequence of the suggested gene flow across putative species boundaries, and the typically high variability among individuals in *Ophrys*, some authors have lumped several species together, resulting in a classification of few species and many subspecies (Pedersen and Faurholdt 2007). In *Ophrys*, species identification based on morphological characters alone can indeed be difficult. Floral scent, however, often shows a
specific pattern among closely related, and morphologically very similar species (Mant et al. 2005b; Stökl et al. 2009), and should thus be taken into consideration when assigning individuals into species categories or testing such assignments. To better understand reproductive isolation and the implicated taxonomic consequences in *Ophrys*, we suggest that a complementary approach should be taken. This approach should incorporate the quantification of floral isolation and later acting reproductive barriers (Ramsey et al. 2003), and combine these results with the analysis of the traits under selection (e.g. floral odor) as well as neutral molecular markers.

**MECHANISMS OF FLORAL ISOLATION IN OPHRYS**

In the studied species only ethological isolation contributes to floral isolation, because all three species attach pollinia to their pollinators' heads, and hence there is no evidence of mechanical isolation. In *Ophrys*, floral odor acts as a key trait for specific pollinator attraction (Schiestl et al. 1999; Mant et al. 2005b) and is therefore likely responsible for ethological isolation among species. Our discriminant function analysis of floral odor bouquets showed a clear separation among each species (Figure 3, D E and F). It has been shown that *O. sphegodes* attracts males of *Andrena nigaoenea* by emitting a hydrocarbon mixture with high proportions of (Z)-9 and (Z)-11/12 alkenes (Schiestl et al. 1999), whereas *O. exaltata* attracts males of *Colletes cunicularius* by emitting high proportions of (Z)-7 alkenes (Mant et al. 2005a). The pollinator of *O. garganica* was reported to be *Andrena carbonaria* (Paulus and Gack 1990b); *O. garganica* emits high proportions of (Z)-9 and (Z)-11/12 alkenes with different carbon chain lengths (typically longer than in *O. sphegodes*), however, the active compounds for its pollinator have not yet been identified. Behavioral tests showed that *C. cunicularius* was only attracted by the floral odor of *O. exaltata*, but not by the other two species (Mant et al. 2005b). Recently, Vereecken and Schiestl (2009) showed that floral color differences between *O. sphegodes* and *O. exaltata* do not contribute
to species-specific pollinator attraction. Collectively, these data suggest that strong ethological isolation in *Ophrys* is primarily due to different floral odor bouquets produced by each species, which are linked to the attraction of different, highly specific pollinators.

**HYBRIDIZATION AMONG *OPHRYS* SPECIES**

Pollinator adaptation may drive floral diversification and speciation in *Ophrys*, however, hybridization has sometimes been considered to be common among the *Ophrys* species (Devey et al. 2008). In contrast, putative *Ophrys* F₁ hybrids (as identified by morphology) were often found to be solitary, with large number of plants from the parental species surrounding them (Stebbins and Ferlan 1956), suggesting hybridization may not happen frequently. In accordance with this, we found only two putative F₁ hybrids out of 146 samples (1.37%) between *O. sphegodes* and *O. exaltata*. Both were found in one population (FCG). Those two putative F₁ hybrids produced a floral odor bouquet similar to *O. sphegodes*. Possible (non-exclusive) reasons for hybrids found in natural population could be the following: (1) the strength of floral isolation may be variable among populations due to variable specificity in the responses of pollinators to floral odor bouquets; (2) Changes of floral odor in *Ophrys* may happen through occasional changes in scent genes, leading to a break-down of floral isolation. To test the first possibility, floral isolation should be assessed in various populations. We found consistently strong floral isolation in two adjacent populations, but could not precisely estimate floral isolation in the more distant population FCG, since the total number of pollination events observed in this population was small (only three in total). The study by Vereecken et al. (2010) suggested that floral isolation among *Ophrys* species pollinated by *Colletes cunicularius* and *Andrena nigroaenea* can break down in some populations, although the frequency of hybrids in that study was always much lower than the parental species. Break-down of floral isolation was also found in some populations...
among *Ophrys* species pollinated by other *Andrena* species (Stökl et al. 2008; Cortis et al. 2009). Varying strengths of floral isolation would suggest a geographical mosaic, with merging of populations through hybridization in some areas and divergence through strong floral isolation in other areas. This geographic mosaic may help to explain the phylogenetic pattern of *Ophrys* species observed in previous studies (Soliva et al. 2001; Devey et al. 2008). However, further investigations about geographical variation in pollinator behavior and floral isolation are needed to evaluate this hypothesis.

A second reason for hybrids occasionally found in nature may be the genetic basis of floral odor changes in *Ophrys*. Since changes in floral odor production in *Ophrys* may be brought about by few genetic changes (Schlüter and Schiestl 2008, Schlüter et al. submitted), one would expect that some individuals of one species could stochastically evolve the same floral odor as another species through mutation or recombination. This would eventually lead to hybridization in natural populations, considering that floral odor is the major attractant for specific pollinators in this system and post-pollination barriers are effectively absent. This hypothesis is consistent with the genetic and floral odor analyses in this study. Among the samples investigated here, a few plant individuals showed mismatches among their assignments from genetic and odor data (samples S07, G10, G14, S47 in Figure 3, summary in Table 4). In other words, these samples have the neutral genetic background of one species, but an odor phenotype of another species, possibly due to changes in few genes controlling floral odor production. However, to further test this hypothesis, detailed studies on the genetic basis of floral odor components in *Ophrys*, and their consequence for pollinator attraction are needed.
As a scenario for speciation in *Ophrys*, we propose that incipient *Ophrys* species adapt to different pollinators by changing floral traits, especially floral odor, that conveys strong floral isolation and induces the speciation process. In *Ophrys*, pollination success is relatively low due to pollen limitation (compare pollination success of natural populations and fruit set rate from hand pollination); a sexually deceptive pollination mechanism may thus induce negative density-dependent selection: high population density may lead to low pollination success since pollinators are more likely to learn and avoid the deceptive flowers. Therefore, a shift in pollinators mediated by a change in floral scent genes may convey a selective advantage by increasing pollination success in the initially few novel genotypes. Furthermore, as shown in our study, different pollinators in *Ophrys* are associated with strong floral isolation, which is sufficient to prevent significant gene flow in sympatry. Changes of floral odor bouquets may be based on changes in few genes involved in the biosynthesis (or regulation) of pollinator-attractive floral compounds (Schlüter and Schiestl 2008, Schlüter et al. submitted). Therefore, speciation in *Ophrys* could happen rapidly, even in sympatry. For example, Vereecken et al. (2010) showed that novel floral odor bouquets in *Ophrys* could evolve rapidly (after only one generation of hybridization), and directly lead to pollinator shifts in sympatry. The remarkable plant-pollinator interaction in *Ophrys* orchids provides a particularly interesting system to study pollinator adaptation directly involved in species divergence, a process that may be important in several other, highly specific pollination systems (Schiestl in press).

Conclusions

By *in situ* tracking pollen flow and experimental hand pollination, we found floral isolation to be very strong among closely related, sympatric *Ophrys* species, whereas later-acting barriers to gene flow were effectively absent. Our results provide direct evidence that
the reproductive barrier among these closely related plant species with specialized pollination consists mostly of floral isolation. In such a system, pollinator adaptation could directly lead to floral isolation and speciation. This offers a particular opportunity to study the role of floral isolation during the evolution of reproductive isolation and speciation. However, further studies that systematically combine neutral traits (such as molecular markers), traits under selection (such as floral odor) and their genetic basis, pollinator behavior, as well as quantification of floral isolation in natural habitats will be helpful to better understand speciation processes in plants with specialized pollination systems.

ACKNOWLEDGMENTS

We wish to thank Myles Menz for his help in the field and Ed Connor for assistance in the GC analysis as well as for providing comments on the manuscript. We thank two anonymous reviews for their constructive comments. We are grateful for financial support by the ETH Zürich (TH0206-2 to FPS), the Austrian Science Fund (FWF, J2678-B16 to PMS), and PRIN program 2010 (to SC).
LITERATURE CITED


Foundation for Statistical Computing, Vienna, Austria.

between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*
57:1520-1534.


Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral

Schiestl, F.P. (in press). Animal pollination and speciation in plants: general mechanisms and
examples from the orchids. In *Evolution of Plant Pollinator Interactions*, S. Patiny, ed.
(Cambridge: Cambridge University Press).

Schiestl, F. P., and M. Ayasse. 2002. Do changes in floral odor cause speciation in sexually


2000. Sex pheromone mimicry in the early spider orchid (*Ophrys sphegodes*): patterns of
hydrocarbons as the key mechanism for pollination by sexual deception. *J. Comp. Physiol.*
186:567-574.


Supplementary information

Figure S1. Observation error rate estimation and simulated floral isolation index among species. (A) Observation error rate estimation. X-axis refers to the number of experiments performed. Squares represent observed errors. The solid line represents the simulated mean value; the dashed lines represent the 95% confidence interval. (B) Simulated floral isolation index for each species pair.
Figure S2. Ploidy levels of the three *Ophrys* species. No differences in ploidy level were found among the three species. Each data point refers to a plant individual. Lower and upper dashed lines refer to the expected relative ratio to internal standard for samples with half or double the genome size, respectively.
cDNA representational difference analysis reveals the genetic basis of divergent traits between closely related orchid species

Manuscript for Annals of Botany:

Shuqing Xu¹,²*, Philipp M. Schlüter¹, Florian P. Schiestl¹, Ueli Grossniklaus³

1, Institute of Systematic Botany, University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

2, Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, CH-8092 Zürich, Switzerland

3, Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

*, Corresponding author

Authors contrubutions: SX, FPS and UG designed research; SX performed experiments; SX analyzed data; SX, PMS and FPS wrote the paper.
ABSTRACT

Background: The study of the genetic changes underlying divergent traits among closely related species is of great interest to many evolutionary biologists, because such traits may often be under divergent selection and/or directly involved in reproductive isolation. However, the identification of the relevant genes is often challenging (or prohibitively expensive) since most of the study objects are non-model species. Species of the Mediterranean orchid genus Ophrys mimic insect mating signals and are pollinated during the pollinator males’ mating attempts. In previous studies, it was suggested that floral odour produced among closely related Ophrys species may be under pollinator-mediated selection. Furthermore, different compositions of floral odour is directly linked to species-specific pollinator attraction, and therefore also linked to reproductive isolation. The genes responsible for different floral odour composition among closely related species might thus be the key for understanding their speciation and evolutionary patterns. However, the relatively large genome and lack of sequence data makes the identification of such genes difficult in Ophrys.

Methods: In this study, representational difference analysis (RDA), a PCR based subtractive cloning method, was applied on cDNA from flower tissue of both species. Rapid amplification of cDNA ends (RACE) was employed to obtain the full length of candidate reproductive isolation genes. Semi-quantitative reverse transcription polymerase chain reaction (RT-PCR) experiments were carried out to determine the expression level of the candidate genes. RT-PCR products were cloned and sequenced.

Key Results: In total 47 unique-genes were obtained from the RDA library. The functional annotation showed that the majority of differentially expressed (or divergent) genes were
involved in terpenoid and fatty acid biosynthesis, which is thought to play a role in floral odour production.

Conclusions: Combining our RDA results with odour data suggest that stearoyl-acyl-carrier-protein desaturase (SAD) and Geranylgeranyl pyrophosphate synthase (GGPPS) homologues may be involved in the synthesis of floral odour compounds that related with different pollinator-attractions: alkenes and linalool. These genes may therefore be important for floral isolation between O. sphegodes and O. exaltata, and might be under pollinator-mediated selection. Our study implies that RDA can be an effective tool for identifying genes underlying divergent traits between closely related species, when sequence information is limited.

Key Words: Sexually deceptive orchids; Ophrys; Pollinator-mediated selection; Reproductive isolation; Floral odour; cDNA representational difference analysis (RDA); Stearoyl-ACP desaturase (SAD);
INTRODUCTION

The identification of the genetic basis of reproductive isolation and traits under natural selection are crucial for understanding the process of speciation and evolution (Waser and Campbell, 2004, Noor and Feder, 2006, Wolf et al., 2010, Schluter and Conte, 2009, Coyne and Orr, 2004). This often involves uncovering the genetic basis of traits that are divergent among closely related species, because these traits were likely under divergent selection and/or are directly associated with reproductive isolation (Waser and Campbell, 2004, Boughman, 2001, Schluter, 2001, Coyne and Orr, 2004). However, this step is often challenging since the sequence information are limited for most of non-model species, which were the subjects in many evolutionary studies.

In plants, floral traits, e.g., flower shape, reward, colour scent etc., directly integrate with pollinators, which could impose selection on these traits, since they mediated reproductive success. Because morphology, sensory preferences, and behaviour differ among different pollinators, selection imposed by them may lead to floral trait diversification (Stebbins, 1970, Sargent and Otto, 2006, Gomez et al., 2008, Kay and Sargent, 2009, Schiestl and Schlüter, 2009). Furthermore, adaptation to different pollinators often conveys floral isolation, which reduces gene flow among populations/species (Kay and Sargent, 2009, Schiestl and Schlüter, 2009, Grant, 1994). As a consequence of pollinator adaptation, floral trait divergence among closely related species may also be linked with reproductive isolation. The identification of the genetic basis of such traits is the key for understanding pollinator-mediated plant evolution and speciation process.

Species of the Mediterranean orchid genus Ophrys mimic mating signals of their pollinators and are pollinated by sexual deception (Kullenberg, 1961, Paulus and Gack, 1990). Pollination by sexual deception is a good example of a highly specialized pollination system, in which floral odour is the most important signal for attracting pollinators (Schiestl
et al., 2000). It has been shown that, the divergence of floral odour composition produced by closely related *Ophrys* species were under pollinator-mediated selection (Mant et al., 2005b, Stökl et al., 2009, Vereecken and Schiestl, 2008). Furthermore, differences in floral odour among closely related species are the key for establishing floral isolation (Schiestl and Ayasse, 2002, Xu et al., submitted). Therefore, to identify genetic basis of different floral scent composition among closely related species in *Ophrys* is essential for understanding pollinator-mediated evolution and speciation process in this genus (Schlüter and Schiestl, 2008). However, mainly due to the relatively large genome (mean 1C DNA amount, [i.e the DNA amount contained in a haploid nucleus] = 10.20 pg, or approximately 9.98 Gbp) (Leitch et al., 2009) and scarcity of sequence data available, it is difficult to identify these floral odour-related genes in *Ophrys*.

Representational difference analysis (RDA) is a sensitive and powerful PCR-mediated subtraction method for the identification of DNA fragments that are the most different between two samples. Although the method was first developed by Lisitsyn et al. (1993) for identifying the differences between two complex genomes, Hubank et al. (1994) later adapted the method for use with cDNA to identify differentially expressed genes. We here use an optimized cDNA RDA method as described by Pastorian et al. (2000). In the RDA method, genes that are either differentially expressed or divergent in their sequences are enriched by PCR after successive rounds of subtractive hybridization, whereas non-differentially expressed/non-divergent sequences (which are of no interest) are suppressed and effectively removed. In this study, RDA was applied to cDNAs from flower tissues of two closely related *Ophrys* species, *O. sphegodes* and *O. exaltata*, which are known to differ in their pollinators and floral odour bouquets (Mant et al., 2005c). The aim of the study was to identify candidate genes underlying species differences and, in particular, pollinator
attraction and reproductive isolation between the two orchid species, as well as test the usefulness of cDNA RDA for evolutionary studies.

**MATERIALS AND METHODS**

**Sample collection**

Samples of *Ophrys sphegodes* Miller (n=11) and *O. exaltata* subsp. *archipelagi* (GÖLZ & H. R. REINHARD) DEL PRETE (n=11) were collected in southern Italy, at populations “Monte Gargano” (41°54’N, 15°40’E) (four *O. sphegodes* and seven *O. exaltata*) and “Foce Garigliano” (41°13’N, 13°46’E) (seven *O. sphegodes* and three *O. exaltata*), where both species co-occur and co-flower. Species were identified based on floral morphology, according to the criteria described by Mant et al. (2005a). For each plant individual, the whole inflorescence was collected in the field and stored in a water-filled falcon tube placed in a cooled box for transport to the laboratory, where one labellum of an unpollinated flower was collected and flash frozen in liquid nitrogen, and stored at -80 °C. Another unpollinated flower labellum from the same inflorescence was collected and washed with 500 µl hexane for one minute to extract floral odour. Floral extracts were stored at -20 °C until gas chromatographic analysis.

**Analysis of floral odour**

Floral scent was analysed using gas chromatography (GC). Before GC analysis, 100 ng of *n*-octadecane were added into to each floral extract as an internal standard. The GC analysis was performed similarly as in a previous study (Mant et al., 2005a, Xu et al., submitted). One microlitre of sample was injected into an Agilent 6890 GC at 50 °C, followed by opening of the split valve and heating to 300 °C at rate of 4 °C/min. An HP-5 column (30 m × 0.32 mm diameter × 0.25 µm film thickness) and flame ionisation detector
(FID) were used for GC, and hydrogen was used as carrier gas. For identification of compounds, several samples were re-analysed on a GC with mass selective detector (GC/MSD; Agilent 5975) using the same oven and column parameters (Xu et al., submitted, Mant et al., 2005a). The relative amount of each odour compound was calculated as the proportion of all alkenes and alkanes of a chain length greater than 18 and lower than 30 carbons.

**RNA extraction and cDNA synthesis**

Total RNA was extracted using Trizol (Invitrogen) following the manufacturer’s instructions. RNA quality and quantity were assessed by agarose gel electrophoresis and spectrophotometry on a NanoDrop ND-1000 (Witec AG Littau). To obtain a sufficient amount of RNA for RDA, 11 *O. sphegodes* total RNAs were extracted and pooled together as an *O. sphegodes* pool (in total ~135 µg total RNA). Likewise 10 *O. exaltata* total RNAs were pooled together as an *O. exaltata* pool (in total, ~145 µg total RNA). RNA was treated with DNase I (Fermentas) for 1 h at 37 °C in the presence of RNAase inhibitor (Fermentas). Subsequently, RNAs were purified and extracted with phenol:choloroform:isoamyl alcohol (25:24:1). Poly (A) RNAs were extracted from total RNA using Qiagen Oligotex Kit (Qiagen Switzerland), which yielded 639 ng mRNA for the *O. sphegodes* pool, and 734 ng mRNA for *O. exaltata*.

cDNA was synthesized immediately after poly (A) RNA extraction. First-strand cDNA was synthesized by RevertAid M-MuLV H’ Reverse Transcriptase (Fermentas) with oligo-dT and random hexamer primers (1:1 ratio) (Fermentas), following the supplier’s instructions. Second-strand cDNA was synthesized using *E.coli* DNA Polymerase I (Fermentas), with RNase H (Fermentas) treated first-strand cDNA as a template, following the supplier’s instructions.
Chapter 2

Representational difference analysis (RDA)

Representational difference analysis was performed following the protocol by Hubank & Schatz (1994), and optimized by Pastorian et al. (2000). The detailed laboratory protocol was kindly provided by Dr. Leo Hawel. For the two species-specific cDNA pools, double-stranded cDNA was digested by DpnII (New England Biolabs). The digested products were purified with NucleoSpin Extract II kit (Macherey Nagel, Düren, Germany), and ligated with adapter [Driver12Mer (5'→3'): GAT CAA TAA CTA, Driver24Mer (5'→3'): TGA CGG ACC GGT TGC GTA GTT ATT] using T4 DNA ligase (Fermentas). The ligation products were amplified using Driver24Mer primer for 24 cycles (95 °C 1 min, 72 °C 4 min) to obtain sufficient starting material for RDA. The amplified fragments, here referred to as driver, were purified with phenol:chloroform:isoamyl alcohol (25:24:1). An aliquot of amplified fragments of each sample was digested with DpnII to remove the adapter sequence. Those fragments were purified with NucleoSpin Extract II kit, and ligated with another adapter [TesterRnd1-12Mer (5'→3'): GAT CTT ATG GCT; TesterRnd1-24Mer (5'→3'): AGA CAG TGC CGG ATG TAG CCA TAA]. After ligation, these products, here referred to as tester, were purified with NucleoSpin Extract II kit. To isolate O. sphegodes-specific or highly expressed genes, we mixed tester cDNA from O. sphegodes and driver cDNA from O. exaltata at an 1:12 ratio (500 ng tester : 6 µg driver, measured by NanoDrop ND-1000). Likewise, to isolate O. exaltata-specific or highly expressed genes, we mixed tester cDNA from O. exaltata and driver cDNA from O. sphegodes using the same tester to diver ratio and absolute amounts.

The first round of hybridization was performed to generate differential products in 5 µl (95 °C 5 min, 67 °C 25 h). The hybridization products were diluted to 50 µl with Tris-EDTA (TE) buffer (pH 7.0). The first PCR amplification of subtracted tester cDNA was performed with TesterRnd1-24Mer primer for 7 cycles (95 °C 45 s, 72 °C 4 min). The products of the
first amplification were diluted 1:10, and used as template for the second round of PCR amplification for 33 cycles (95 °C 45 s, 72 °C 4 min). All PCRs in this study were performed on a T1 Thermolycler (Biometra Germany). REDTaq® ReadyMix™ PCR Reaction Mix (Sigma) was used for all RDA PCRs according the supplier’s protocol, scaled to 15 µl total volume.

Cloning and sequencing

PCR products were purified with NucleoSpin Extract II kit, and ligated into pDrive vector (Qiagen) using the manufacturers’ protocol. Competent E. coli DH5α cells (Invitrogen) were used for chemical transformation. The transformed E. coli cells were grown on Luria-Bertani (LB)-agar plates containing 80 µg/ml isopropyl β-D-1-thiogalactopyranoside (IPTG) and 50 µm/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) overnight. White colonies were picked and screened for positive inserts by PCR using M13 universal primers. For each RDA library, 2 × 96 positive clones were randomly picked and screened by PCR. The PCR products were purified using the method of Werle et al. (1994), and the purified products sequenced using BigDye 3.1 on an ABI 3130XL Genetic Analyzer (Applied Biosystems), following the manufacturer’s recommendations.

RACE and RT-PCR

To obtain the full-length coding sequence of candidate genes, 5’ rapid amplification of cDNA ends (RACE) was performed as described by Scotto-Lavino et al. (2006b), whereas 3’ RACE was performed using classic RACE (Scotto-Lavino et al., 2006a). First-strand cDNA was synthesized as described before. Gene-specific primers used for RACE are listed in supplemental Table S1. JumpStart™ REDAccuTaq™ LA DNA Polymerase (Sigma-Aldrich) was used for RACE PCR [95 °C 3 min; 34 cycles of (95 ºC 30 s, 58 °C 30 s, 68 °C 1 min 30 s) final extension at 72 °C for 10 min]. The amplified fragments were cloned and sequenced as described before. At least four positive clones were sequenced for each fragment.
To test gene expression of the candidate genes identified, eight individuals of each species were selected from populations. Total RNA extraction and first-strand cDNA synthesis was performed as described before, and gene-specific primers (supplemental Table S1) used to amplify each gene. Each PCR was performed as 95 °C 3 min; 34 cycles of [95 °C 30 s, 58 °C 30 s (62 °C for chalcone synthase), 72 °C 1 min 30 s]; final extension at 72 °C for 10 min using REDTaq® ReadyMix™ PCR Reaction Mix as described above. For all RT-PCRs, an Ophrys gapC homologue (Schlüter et al., submitted), putatively encoding glycolytic glyceraldehyde-3-phosphate dehydrogenase (G3PDH) was used as control.

Data analysis

Sequences obtained from the RDA library were pre-processed by PHRED (version 0.020425) for base calling with the Q20 parameter, and CROSSMATCH to remove vector sequences (Ewing and Green, 1998). All sequences were assembled by using PHRAP (version 0.990329) (http://www.phrap.org/) with default parameters. Sequences more than 100 bp in length were selected for the annotation process. All sequence annotated as ribosomal RNA or ribosomal protein were removed. Functional annotation of RDA library sequences was performed by comparing to the NCBI nr database (www.ncbi.nlm.nih.gov) using the BLASTX program (Altschul et al., 1997). Gene orthology and pathway annotation were carried out by comparing to the KEGG orthology and pathway database from KEGG (www.kegg.jp) using BLASTX. For all blast comparisons, E-values were filtered using a threshold of 1e-5, and scores with more than 100 hits were selected. The local RDA library data analysis pipeline was developed in Perl version 6 (www.perl.org). The pipeline script is available from the first author upon request. The relative amount of gene expression from RT-PCR was analysed in Image J 1.42 (Abramoff et al., 2004). Statistical analysis using a one-way analysis of variance (ANOVA) was performed in R version 2.7 (R Development Core Team, 2008).
Results

*Differences in floral odour between O. sphegodes and O. exaltata*

Floral odour of these two species showed a similar pattern as previous studies (Mant et al., 2005c): *O. exaltata* individuals produced significantly (p<0.05) higher amounts of (Z)-7-C21, (Z)-7-C23 and (Z)-7-C25 alkenes than *O. sphegodes*, whereas *O. sphegodes* produced significantly (p<0.05) higher amount of the (Z)-9 alkenes, (Z)-9-C27, (Z)-9-C29 (Z)-11/12-C25, (Z)-11/12-C27, (Z)-11/12-C29 than *O. exaltata* (Figure 1).
Figure 1. Differences in floral odour between the *O. sphegodes* (dark grey) and *O. exaltata* (light grey) accessions used in this study ($N=10$ and $N=11$, respectively). Each bar represents the relative amount of compounds in floral odour. Data shown represent mean value ± standard error. Significant differences between species for given compounds are indicated (one-way ANOVA, *** $p \leq 0.01$, * $0.01 < p \leq 0.05$)
Transcript differences among *O. sphegodes* and *O. exaltata* identified by RDA

Two subtractive hybridizations were performed, one with *O. sphegodes* as driver and one with *O. exaltata* as driver. After the first round of subtractive hybridization, the fragments that amplified differentially between these two species showed no obvious size differences between the two fragments pools (which was confirmed further by sequencing data; not shown). In total, 192 clones from each library were sequenced. From these $2 \times 192$ clones, 100 and 98 clean sequences of $>100$ bp length were selected from the *O. sphegodes* and *O. exaltata* RDA library respectively and analysed. In total 47 unique-genes were obtained from the two libraries.

Functional annotation using the KEGG gene orthology database indicated that these transcripts are putatively involved in: biosynthesis of secondary metabolites, lipid metabolism, fatty acid biosynthesis *etc.* (Table 1). Further pathway annotation using the KEGG pathway database showed that, for both RDA libraries, the most abundant sequences were involved in terpenoid biosynthesis (Table 1): 20 out of 100 (20.0%) and 28 out of 98 (28.6%) transcripts from *O. sphegodes* and *O. exaltata* involved in the terpenoid biosynthesis, respectively.
Table 1. Functional classification of all transcripts found in the RDA library. The number of transcripts refers to the number of sequenced clones contributing to a given category. In total 122 transcripts were mapped to KEGG Orthology functional classes. The most abundant transcripts are involved in terpenoid biosynthesis, and glutathione metabolism. Several transcripts are involved in flavonoid biosynthesis, carbon fixation, and fatty acid biosynthesis pathways.

<table>
<thead>
<tr>
<th>KEGG Orthology Annotation</th>
<th>Number of Transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Biosynthesis of terpenoids</td>
<td>48</td>
</tr>
<tr>
<td>Glutathione metabolism</td>
<td>28</td>
</tr>
<tr>
<td>Carbon fixation in photosynthetic organisms</td>
<td>9</td>
</tr>
<tr>
<td>Flavonoid biosynthesis</td>
<td>8</td>
</tr>
<tr>
<td>Ascorbate and aldarate metabolism</td>
<td>8</td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td>7</td>
</tr>
<tr>
<td>Other enzymes</td>
<td>6</td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>6</td>
</tr>
<tr>
<td>Glycerophospholipid metabolism</td>
<td>2</td>
</tr>
</tbody>
</table>
Candidate genes underlying divergent floral traits between O. sphegodes and O. exaltata

In our RDA library, two stearoyl-ACP desaturase (SAD) homologues (SAD1 and SAD2), one Geranylgeranyl pyrophosphate synthase (GGPPS) homologue, and two Chalcone synthase (CHS) homologues (CHS1 and CHS2) were found.

One of the most frequently sequenced transcripts was a fragment of GGPPS. This GGPPS homologue was found in both species’ RDA library of both species. However, comparison to known GGPS sequences suggests there were approximately 120 bp at the 5’ end of this homologue that we were not be able to obtain by our 5’ RACE experiment. Based on this partial sequence (~89% of expected full-length), we found the most closely related known GGPPS gene in Ricinus (73% identity on the nucleic acid level). RT-PCR results indicated that this GGPPS homologue is expressed in both orchid species. However, the expression level of this homologue is significantly higher in O. sphegodes than in O. exaltata (Figure 2).
Figure 2. Gene expression of five candidate genes. (A): Gene expression in four representative *O. sphegodes* and *O. exaltata* individuals as shown by RT-PCR. The candidate genes analysed are: stearoyl-ACP desaturase 1 (*SAD1*) and 2 (*SAD2*), chalcone synthase 1 (*CHS1*) and 2 (*CHS2*), and geranylgeranyl pyrophosphate synthase (*GGPPS*). The putative housekeeping gene *glyceraldehyde-3-phosphate dehydrogenase* (*G3PDH*) was used as a control. (B): Relative gene expression level of each candidate gene normalized to *G3PDH* in all 8 individuals of each species. Data shown represent mean value ± standard error and significant differences between species are indicated (one-way ANOVA, *** $p$ <0.01, * $0.01 < p < 0.05$, $N$=8 for each species).
Two SAD homologues were found, which are similar to two homologues (SAD1 and SAD2, (accession number: FR688108; FR688109) that had been previously been cloned and sequenced based on a candidate gene approach (Schlüter et al., submitted). Full-length sequence similarity of those two homologues is ~90% at the nucleic acid level, and ~83% at the amino acid level. Gene expression analyses of those two SAD homologues by RT-PCR indicate that, in the flower labellum, SAD2 is expressed in both species whereas SAD1 is only expressed in *O. sphegodes* (Figure 2).

Two chalcone synthase homologues (CHS1 and CHS2) were found in our library and full-length coding sequence isolated by RACE. The similarities between these two homologues are 98.5% on the nucleic acid level, and 94.4% on the amino acid level. These two CHS homologues belong to the orchid-specific ‘second CHS group’ (Han et al., 2006). RT-PCR showed that both CHS1 and CHS2 were highly expressed in *O. exaltata* and *O. sphegodes* (Figure 2), without a clear difference among species.

**Discussion**

In this study, we used cDNA RDA to identify genetic basis underlying floral traits that were divergent between two closely related species orchid species, *O. sphegodes* and *O. exaltata*. This approach does not require prior sequence knowledge and is relatively sensitive and efficient, and may thus be promising for identifying differential genes in closely related species of non-model organisms. Based on this approach, we identified genes involved in biosynthesis pathway of floral odours and colours. Among these genes, we found three that were associated with floral odour were significantly differentially expressed between our study species. Furthermore, two of these two genes, SAD1 and SAD2 are potentially involved
in alkenes production of *Ophrys*, which is related to specific pollinator attraction and reproductive isolation.

*Candidate genes underlying floral traits divergence among O. sphegodes and O. exaltata*

Floral scents constitute a major floral trait important for the pollinator attraction and reproductive isolation between *O. sphegodes* and *O. exaltata* (Mant et al, 2005b; Xu et al, submitted). Transcripts that are putatively involved in floral odour biosynthetic pathways might responsible for floral odour divergences. The first set of floral odour related genes found in our RDA library likely encodes an acyl-ACP (acyl carrier protein) desaturases, which are putatively involved in fatty acid biosynthesis. From the comparison of floral odour bouquets between *O. sphegodes* and *O. exaltata*, it is apparent that the main difference lies in the double-bond position of alkenes. According to the hypothesis proposed by Schlüter and Schiestl (2008), these alkenene double-bond position differences may be linked to differences in the activity of desaturase enzymes, which are responsible for introducing the double-bonds into fatty acids [the probable precursors of hydrocarbons (Kunst and Samuels, 2005, Perera et al., 2010)] in many plant species (Shanklin and Cahoon, 1998). In our RDA library, two *SAD* homologues, putatively encoding acyl-ACP desaturases (EC: 1.14.19.2), were found, both of which showed significantly different levels of gene expression among *O. sphegodes* and *O. exaltata*. Furthermore, sequence comparison of the two expressed homologues also shows amino acid differences between these two species. The different gene expression and sequence variation among those two homologues, along with their suspected role in alkene biosynthesis, makes them good candidate genes for differences in the odour bouquets among the two orchid species. Functional and gene expression studies confirmed that these *SAD* homologues are strongly associated with some floral odour compounds (Schlüter et al. submitted).
The second floral odour related gene found in our RDA library was GGPPS, which is putatively involved in the terpenoid biosynthesis pathway. Known GGPPS enzymes catalyse the condensation of dimethylallyl pyrophosphate (DMAPP) with isopentenyl pyrophosphate (IPP) to produce geranylgeranyl pyrophosphate (GGPP), which is the key precursor for several holoterpenoids (Okada et al., 2000). The GGPPS substrates IPP and DMAPP also act as substrates of geranyl pyrophosphate synthase, which converts DMAPP and IPP into geranyl pyrophosphate (GPP; Figure 3). GPP in turn is the direct precursor of all monoterpenes, which are typically produced in many plants as floral scents and are potentially involved in plant-insect communication (Schiestl, 2010). Among these monoterpenes, S-linalool is used as a pollinator attractant in many plant species (Chen and Song, 2008, Pichersky et al., 1995, Theis, 2006). In Ophrys, S-linalool has been reported in O. exalatata, but appears to be absent from O. sphegodes (Mant et al., 2005b). Since the expression level of GGPPS is significantly higher in O. sphegodes than in O. exaltata, one may hypothesize that the presence (or higher amounts of) GGPPS protein (and thus presumably higher enzyme activity) in O. sphegodes may divert metabolic pathway flux away from linalool synthesis. Thus, differences in GGPPS expression may be linked to differences in S-linalool production between these two species. However, there is currently no experimental support for this hypothesis: S-linalool was not detected in headspace of the samples of either species, which were collected from population “Monte Gargano” in 2009 (data not shown). The different results of S-linalool found by our data and Mant et al (2005b) may due to variation among the different study populations. Further functional studies and further investigations of S-linalool production in different Ophrys species and population will be required to test this hypothesis.
Figure 3. The function of GGPPS in terpenoid biosynthesis. GGPPS catalyses the consecutive condensation of dimethylallyl pyrophosphate (DMAPP) with isopentenyl pyrophosphate (IPP) to produce geranylgeranyl pyrophosphate (GGPP). The substrates of GGPPS, IPP and DMAPP are also the substrate of geranyl pyrophosphate synthase, which converts DMAPP and IPP into geranylpyrophosphate (GPP), which is the direct precursor for all of terpenoids. Figure redrawn and modified after (Shimon & Efraim, 2006).
Although floral colour is unlikely to be directly involved in pollinator-attraction in *O. sphegodes* and *O. exaltata* (Vereecken and Schiestl, 2009), the colour of labella were slightly different between these two species. Labella of *O. sphegodes* are dark brown, while labella of *O. exaltata* are light brown. Genes involved in floral colour biosynthesis could be associated with this colour difference. In our RDA library, we found two *chalcone synthase* (*CHS*) homologues, which might be involved in the biosynthesis of floral anthocyanin pigments. However, both homologues fall into the ‘orchid *CHS2* group’ (Han et al., 2006), which probably lacks CHS enzyme activity because of an amino acid substitution in the CHS active site. The detailed function of this ‘orchid *CHS2*’ group is still unknown, although expression of one member of this group in transgenic tobacco has been reported to change floral morphology (Han et al., 2005). Furthermore, gene expression of these two *CHS* homologues was variable in both species, without any significant differences between them. Collectively, the *CHS* homologues found in our RDA library might not be associated with different floral labella colour between species.

The application of RDA.

To identify traits that are divergent among closely related species is of great interest for evolutionary biologists (Wolf et al., 2010), since these traits might be directly or indirectly involved in evolution and speciation process. However, this step is often limited by lack of publically available data on genome or EST sequences, since the study systems in many evolutionary researches are non-model organisms. The genic view of speciation predicts that the majority of the genome will show little difference between two closely related species except for a few genes or gene networks related to reproductive isolation and/or natural selection (Wu, 2001, Lexer and Widmer, 2008). Furthermore, these genetic differences were
often contributed by changes of expression or mutations on specific genes (Wolf et al., 2010, Rieseberg and Blackman, 2010). Currently, the identification of these genetic differences can be accomplished mainly by the following approaches: (1) quantitative trait locus (QTL) mapping and map-based cloning, (2) whole genome or cDNA/RNA profile sequencing or (3) subtractive cloning approaches, like RDA. QTL-based approaches are however very laborious and time-consuming, especially for organisms - such as orchids - which have a long life cycle. The second method, sequencing of whole genomes, transcriptomes, or cDNA profiles (Gupta, 2008, Wang et al., 2009) is becoming more attractive nowadays due to the dramatically reduced costs for large scale sequencing provided by next-generation sequencing technology (Shendure and Ji, 2008). Nonetheless, the cost of comparative sequencing of whole genomes (or even transcriptomes) are still high enough to be prohibitive in many situations, especially for species with a large genome. Subtractive cloning methods have the advantage that they focus on the differences among species and are therefore more cost-effective. RDA identifies differentially expressed or divergent genes and has been shown to be a relatively sensitive and efficient subtractive cloning method (Sperotto et al., 2007, Park et al., 2007, Weil et al., 2007, Weil et al., 2009). This study successfully used cDNA RDA to identify candidate genes underlying different floral odour produced by two closely related orchids species *O. sphegodes* and *O. exaltata*, demonstrating its utility for this purpose. Our study implies that RDA can be an effective tool for identifying genetic basis underlying traits that were divergent among closely related species, when resources are limited.
ACKNOWLEDGEMENTS

We thank Dr. Leo Hawel for providing a detailed cDNA RDA protocol, Professor Salvatore Cozzolino for sample collection in the field, Dr. Stephen Schauer for helping on RACE experiments.

FUNDING

We are grateful for financial support by the ETH Zurich (TH0206-2 to FPS), Austrian Science Fund (FWF; Schrödinger fellowship J2678-B16 to PMS).
LITERATURE CITED


Xu S, Schlüter PM, Scopece G, Breitkopf H, Gross K, Cozzolin S, Schiestl FP. accepted for publication. Floral isolation is the main reproductive barrier among closely related sexually deceptive Orchids. *Evolution*.
Chapter 3

Stearoyl-ACP desaturases are associated with floral isolation in sexually deceptive orchids

Submitted to PNAS as:

Philipp M. Schlüter\textsuperscript{a,b,1}, Shuqing Xu\textsuperscript{a,b,c}, Valeria Gagliardini\textsuperscript{b}, Edward J. Whittle\textsuperscript{d}, John Shanklin\textsuperscript{d}, Ueli Grossniklaus\textsuperscript{b}, Florian P. Schiestl\textsuperscript{a}.

\textsuperscript{a}Institute of Systematic Botany & Zürich-Basel Plant Science Center, University of Zürich, Zollikerstr. 107, CH-8008 Zürich, Switzerland;

\textsuperscript{b}Institute of Plant Biology & Zürich-Basel Plant Science Center, University of Zürich, Zollikerstr. 107, CH-8008 Zürich, Switzerland;

\textsuperscript{c}Institute of Integrative Biology & Zürich-Basel Plant Science Center, Swiss Federal Institute of Technology Zürich, Universitätsstr. 16, CH-8092 Zürich, Switzerland;

\textsuperscript{d}Department of Biology, Brookhaven National Laboratory, 50 Bell Avenue, Upton, NY 11973, USA.

\textsuperscript{1}To whom correspondence should be addressed. E-mail: philipp.schlueter@systbot.uzh.ch

Authors contributions: P.M.S., J.S., U.G. and F.P.S. designed research; P.M.S., S.X., V.G., E.J.W., J.S. and F.P.S. performed research; P.M.S., S.X., and V.G. analyzed data; P.M.S. wrote the paper.
Non-standard Abbreviations

ACP - acyl carrier protein
FA - fatty acid
FAME - fatty acid methyl ester
SAD - stearoyl-ACP desaturase
VLCFA - very-long-chain fatty acid

Shorthand notation for fatty acids and their derivatives is given in C:D form, where C specifies the number of carbon atoms and D the number of double-bonds. The position x of a cis double-bond in the carbon chain is indicated by $\Delta^x$ when counted from the substituted end (if applicable), or by $\omega$-x when counting from the unsubstituted end.
Abstract

The orchids *Ophrys sphegodes* and *O. exaltata* are reproductively isolated from each other by the attraction of two different, highly specific pollinator species. For pollinator attraction, flowers chemically mimic the pollinators’ sex pheromones, the key components of which are alkenes with different double-bond positions. This study identifies genes likely involved in alkene biosynthesis, encoding stearoyl-acyl carrier protein (ACP) desaturase (SAD) homologs. The expression of two isoforms, *SAD1* and *SAD2*, is flower-specific and broadly parallels alkene production during flower development. *SAD2* shows a significant association with alkene production and *in vitro* assays show that *O. sphegodes* SAD2 has activity both as an 18:0-ACP Δ⁹ and a 16:0-ACP Δ⁴ desaturase. Downstream metabolism of the SAD2 reaction products would give rise to alkenes with double-bonds at position 9 or position 12, matching double-bond positions observed in alkenes in the odor bouquet of *O. sphegodes*. *SAD1* and *SAD2* show evidence of purifying selection prior to, and positive or relaxed purifying selection after gene duplication. By contributing to the production of species-specific alkene bouquets, SAD2 is suggested to contribute to differential pollinator attraction and reproductive isolation among these species. Taken together, these data are consistent with the hypothesis that *SAD2* is a florally expressed barrier gene of large phenotypic effect, and possibly a genic target of pollinator-mediated selection.
Introduction

Reproductive isolation is a central topic in the study of evolution, its origin and maintenance being critical for the process of speciation. This is especially true for ecological speciation, in which divergent selection pressures on key traits drive the establishment of reproductive isolation even in the absence of geographic barriers to gene flow (1). This fits the genic view of speciation, in which only few loci of large effect may be responsible for species differentiation, while gene flow is possible throughout the rest of the genome (2, 3).

In practice, the challenge in studying these processes is identifying the traits under divergent selection and their genetic basis (1). In plants with strong pollinator-mediated reproductive isolation (floral isolation), however, key floral traits are direct targets of selection (1, 4). By identifying the molecular mechanisms underlying these traits, genes directly involved in reproductive isolation (so-called ‘barrier’ or ‘isolation’ genes) or even speciation can be identified (3-5).

Strong floral isolation and high pollinator specificity make sexually deceptive orchids an excellent system for identifying barrier genes (4, 6). Rewardless orchids of the genus *Ophrys* attract male insects by sexual mimicry, inducing mating attempts of pollinators with flowers, whereby pollen is transferred. The key component to this system is the chemical mimicry of the pollinator female’s sex pheromone (7, 8), a blend of substances consisting mostly of cuticular hydrocarbons, e.g. alkanes and alkenes. Alkenes (unsaturated hydrocarbons) are of special importance, and a different proportion of alkenes was found to be the major odor difference among two closely related *Ophrys* species attracting different pollinators (9). In *Ophrys*, speciation by pollinator shift has been hypothesized and there is evidence both for pollinator-driven genetic differentiation and selection on floral hydrocarbon profiles (4, 6, 9, 10, and references therein). In particular, specific pollinators mediate strong floral isolation among the co-flowering closely related species *O. sphegodes*.
and *O. exaltata* by effectively preventing gene flow, while other reproductive barriers are largely absent (11). These species differ mainly in the double-bond position of their major alkenes (9), implying that the genes underlying this alkene difference may be barrier genes (6).

While alkanes are common components of the wax layer covering the aerial parts of plants (12), alkenes are rare. Alkanes are synthesized from fatty acyl-coenzyme A (CoA) intermediates that undergo several rounds of chain elongation from the carboxyl terminus. These fatty acid (FA) intermediates undergo reduction to aldehydes and decarbonylation to form alkanes, mostly producing odd-numbered alkanes from even-numbered very-long-chain fatty acid (VLCFA) intermediates (12, 13). Alkenes are thought to follow the same synthesis scheme, except for the introduction of double-bonds in an additional desaturation step (6).

Notably, biosynthesis of the alkenes in insect sex pheromones is likely very different from that in plants. Although insect acyl-CoA desaturases (which introduce the double-bond into alkene precursors) were identified as putative speciation genes (3, 5, and references therein), plant acyl-acyl carrier protein (ACP) desaturases that are responsible for the conversion of saturated to unsaturated FAs are mostly unrelated to their animal counterparts (14). Specifically, plant homologs of the animal integral membrane acyl-CoA desaturases act mostly on acyl-lipid intermediates. In contrast, soluble, plastid-localized stearoyl-ACP desaturases (SAD) carry out the ubiquitous desaturation of 18:0 (saturated C\(_{18}\)) to 18:1 (monounsaturated C\(_{18}\)) FA intermediates (14). Such SADs are candidates for the insertion of a double-bond into the precursors of alkenes in plants (6, 10). Double-bond insertion at position Δ\(^9\) of 18:0-ACP’s carbon chain (counting from the substituted end) by a Δ\(^9\)-SAD would yield 18:1Δ\(^9\)-ACP. This could be elongated, to e.g. 28:1Δ\(^9\)-CoA (double-bond at position Δ\(^9\) = ω-9, with ω counting from the unsubstituted end), leading to the production of 27:1Δ\(^9\) alkenes upon decarbonylation. Therefore, species differences in alkene composition
might result from changes in gene expression and/or enzyme activity of specific SAD-encoding genes among species, implying that such genes are candidate barrier genes in *Ophrys* orchids.

Here we report the isolation of SAD homologs from *O. sphegodes* and *O. exaltata* and discuss their potential role as barrier genes. Specifically, we address the following questions: (i) are there any differences among species regarding SAD gene expression or protein structure, (ii) are such differences associated with alkene production, (iii) are SAD proteins functional desaturases, and (iv) is there any evidence for selection on these enzymes?

**Results**

*Gene cloning of Ophrys stearoyl-ACP desaturase homologs*

Putative SAD-encoding transcripts were cloned by homology to *Arabidopsis thaliana SSI2* (*SUPPRESSOR OF SA-INSENSITIVITY2*; At2g43710), the main Δ⁹-SAD encoding gene of *Arabidopsis*. Three putative homologs, named *SAD1–3* (Fig. S1A), were identified from cDNA of *Ophrys* flower labella and their full coding sequence obtained by RACE. *SAD1* was only identified from *O. sphegodes*, whereas the *SAD2* and *SAD3* genes were cloned from both species. *SAD3* showed only silent substitutions between the *O. sphegodes* and *O. exaltata* alleles (hereafter, denoted by *Os* and *Oe* prefixes). In contrast to *SAD3*, *OsSAD2* and *OeSAD2* differed at the amino acid level (Fig. S1A).

*Evolutionary analysis*
Homologs of *A. thaliana SSI2* (Table S1) were identified in public sequence databases and used to construct a Bayesian inference phylogeny of plant acyl-ACP desaturases (Fig. 1, S2A). There was only one group of monocot desaturases, with *Ophrys SAD1* and *SAD2* occupying a position separate from *SAD3*. This indicated that the gene duplication events associated with plant desaturase diversification occurred after the split of monocots and eudicots. Furthermore, the *SAD1/SAD2* dichotomy is more recent than the split of proto-*SAD1/2* and *SAD3*. To test for the signature of selection, a maximum likelihood-based analysis of synonymous mutations (*d_S*; preserving the amino acid sequence) versus non-synonymous mutations (*d_N*; altering the amino acid sequence) was performed. This analysis revealed no indication of selection for *SAD3*. However, significant purifying selection (*p*=0.002) was found on the *SAD1/SAD2* clade prior to the split of *SAD1* and *SAD2*, and significant positive or relaxed purifying selection (all *p*<0.001) thereafter (Fig. 1, S2, Table S2). A more conservative exact test of synonymous and non-synonymous sites is consistent with this interpretation (Table S2).
Fig. 1. Phylogenetic analysis of SAD homologs, showing monocot clade. Bayesian phylogeny with branch lengths from BaseML, numbers indicating posterior probabilities (where >0.5) next to branches. Selected branches for orchid desaturases are labeled and the respective $d_N/d_S$ ratios (from CodeML free-ratio model) indicated in the inset. An asterisk marks branches A, B, and C, for which $d_N/d_S$ ratios are significant ($p<0.01$) among 1- and 2-ratio models (Table S2).
Cuticular hydrocarbons and gene expression

As high levels of alkenes were found on flowers, but not on leaves of *Ophrys* (7), the occurrence of hydrocarbons and *SAD* expression in different *O. sphegodes* and *O. exaltata* tissues and floral developmental stages was investigated (Fig. 2, S3, S4). Mature flowers of the two species differed significantly in the levels of different alkenes, with *O. sphegodes* producing high levels of 9-alkenes and 12-alkenes (strictly speaking, 11/12-alkenes; see Methods), and high levels of 7-alkenes in *O. exaltata* (Fig. S3B,C). Expression of *SAD1* and *SAD2* (but not *SAD3*) differed among mature labella from the two species (Fig. 2A). Together with the finding that *SAD3* was expressed in leaf tissue lacking alkenes, this suggests that *SAD1* and/or *SAD2* are involved in species-specific differences in alkene production. While alkanes were found in all tissues, most alkenes were barely detectable in leaves/bracts, sepals/petals, and labella from the smallest buds. The relative amount of alkenes, however, increased throughout flower development (Fig. 2B, S3F-O). As judged by semi-quantitative reverse transcriptase (RT)-PCR, *SAD1* and *SAD2* expression broadly paralleled alkene occurrence, but only *SAD2* expression could significantly explain the presence of several 9- and 12-alkenes, which are different among species and detectable by pollinators (Fig. 2C, S3). Although *SAD3* showed a significant association with one species-specific 9-alkene (C25), its lack of species-specific expression pattern makes it unlikely to be a causative factor.
Fig. 2. **SAD expression and hydrocarbons.** (A) Mean relative expression of SAD1–3 in *O. sphegodes* (*Os*) and *O. exaltata* (*Oe*), normalized to G3PDH control, in flower labella (top), and leaves (bottom). *, *p*<0.05 (1-way ANOVA). (B) Normalized expression of SAD1–3 (top) and relative amounts (%) of major hydrocarbon classes (bottom) in *O. sphegodes* flower labella of different developmental stages (-4, smallest bud; 0, flower at anthesis), mature sepals/petals (SP) and leaves (L). (C) Correlation of normalized *O. sphegodes* SAD2 expression with relative alkene amount after \( f(x) = \arcsin x^{0.5} \) transformation, for 27:1\( \Delta^9 \) alkene (top; adjusted \( R^2=0.48, p=2.8\cdot10^{-5} \)), and 27:1\( \Delta^{12} + 29:1\Delta^{12} \) alkenes (bottom; adjusted \( R^2=0.32, p=0.0009 \)), showing regression lines.
**Protein predictions**

As common stearoyl-ACP desaturases are plastid localized (14), we checked if a plastid transit peptide was predicted for *Ophrys* SAD proteins. For SAD1 and SAD3 (but not SAD2), the presence of a transit peptide was predicted (Table S3). However, moderate prediction scores and N-terminal sequence divergence from the well-characterized *Ricinus communis* plant SAD (RcSAD) indicated that care is needed when postulating the subcellular localization of the *Ophrys* SADs. Using a crystal structure of RcSAD as a template, structural homology models were generated for OsSAD1, OsSAD2, OeSAD2 and OsSAD3 (which is identical in sequence to OeSAD3). These models were in good overall agreement, with differences among protein backbones localized mainly to one loop region (Fig. S1B). Geometry around the active site and substrate-binding pocket appeared to be mostly conserved among *Ricinus* and *Ophrys* proteins, and a canonical stearic acid (18:0) substrate modeled into RcSAD fitted into *Ophrys* structures similarly well (Fig. S1C). The most prominent difference between RcSAD and OsSAD2 is Leu123 at the aliphatic end of the hypothetical substrate (Fig S1C). Between *Ophrys* SAD1 and SAD2s, there were several amino acid changes, mainly on the protein surface (Fig. S1D), and there was a marked difference in isoelectric point (Table S3). Hypothetically substrate-interacting regions were mostly similar among OsSAD2 and OeSAD2, but OsSAD1 showed some amino acid differences near the aliphatic end of the substrate (Fig. S1E). Overall, homology models suggested *Ophrys* proteins to be functional desaturases, although differences among the proteins indicated they might not be functionally equivalent.
**SAD functional characterization**

Protein function of putative *Ophrys* desaturases OsSAD1, OsSAD2 and OeSAD2 was investigated in transgenic *Arabidopsis* and by *in vitro* assays of enzyme activity. The *Ophrys* SAD coding sequences were heterologously expressed in *Arabidopsis* under the control of the Cauliflower mosaic virus 35S RNA promoter. None of the transgenic plant lines complemented the dwarf phenotype of homozygous ssi2 mutants [see Supporting Information (SI)], indicating that orchid transgenes could not fully functionally replace the *A. thaliana* desaturase SSI2. However, the presence of the OsSAD2 transgene was significantly associated with changes in unsaturated C₁₈ and C₁₆ FA levels in *Arabidopsis* leaf lipids, suggesting that OsSAD2 has enzymatic activity in *Arabidopsis* (Fig. S5).

To uncover the specific reaction catalyzed by each *Ophrys* SAD, recombinant proteins were assayed for desaturase activity *in vitro*, using acyl-ACP from regiospecifically deuterated fatty acids. For OsSAD1, no product was detectable by gas chromatography coupled to mass spectroscopy (GC/MS), and lack of soluble OeSAD2 expression precluded its analysis. However, desaturase activity was observed for OsSAD2. Consistent with the lack of complementation of *Arabidopsis* ssi2 mutants, *in vitro* OsSAD2 activity was low. This may reflect a requirement for specific ACP or ferredoxin proteins different from those present in enzyme assays or in *Arabidopsis* (cf. 15, 16). OsSAD2 was active both on 18:0 and 16:0 substrates, producing 18:1Δ⁹ and 16:1Δ⁴ products, respectively, as confirmed by MS of fatty acid methyl esters (FAMEs) of reaction products and their pyrrolidine derivatives (Fig. 3). Considering fatty acid elongation from the carboxyl end, these desaturation products would be expected to give rise to 9-alkenes and 12-alkenes.
Fig. 3. GC/MS analysis of OsSAD2 desaturase assay. (A, C, E) 12,12-\textsuperscript{2}$\text{H}_2$-18:0-ACP substrate; (B, D, F) 7,7,8,8-\textsuperscript{2}$\text{H}_4$-16:0-ACP substrate. (A, B) GC trace showing assay without desaturase (control; left) and with desaturase (right), with retention times (min) indicated. Left peak, substrate; right peak, background FA; second peak (with desaturase only), specific reaction product. (C, D) MS fragmentation patterns of specific FAME peaks marked by a gray arrow in (A) and (B), showing mass ion and depicting the molecular structure. (E, F) MS fragmentation patterns of pyrrolidine derivatives of FAMEs in (C, D). Arrows indicate ions that are diagnostic for the double-bond positions inferred. This analysis confirms $\Delta^9$ and $\Delta^4$ positions for 18:1 and 16:1 reaction products, respectively.
Discussion

Reproductive isolation between *O. sphegodes* and *O. exaltata* depends upon the attraction of two different, highly specific pollinator species by chemical mimicry of their sex pheromones. This specificity is due to the presence of alkenes with different double-bond positions (7-9). During development, these alkenes accumulate in the labella of *Ophrys* flowers. This is in marked contrast to the ubiquitous presence of alkanes on orchid surfaces, suggesting that alkene production is indeed tissue- and stage-specific. Among the three putative orchid desaturases, *SAD3* showed a relatively constant expression without obvious species difference, consistent with a function as a housekeeping desaturase rather than a factor linked to alkene production. By contrast, *SAD1* and *SAD2* expression broadly paralleled alkene production, and *SAD2* showed a significant association with 9- and 12-alkene levels in *O. sphegodes*, supporting a functional link. *SAD1/2* probably originated by gene duplication, forming a lineage distinct from *SAD3*. Purifying selection prior to this duplication event suggests a conserved role of the ancestral protein. The higher rate of amino acid change after duplication may indicate a partial release from functional constraints, although, considering that alkenes are likely under divergent selection (9), it is also possible that selection drove the divergence of protein function. Taken together, these results implicate *Ophrys SAD2* as a desaturase-encoding gene associated with the biosynthesis of alkenes in the floral pseudo-pheromones.

*OsSAD2* is a functional desaturase capable of producing 18:1Δ⁹ (ω-9) and 16:1Δ⁴ (ω-12) FA intermediates from which 9-alkenes and 12-alkenes could be synthesized (Fig. 4). However, since housekeeping desaturase activity should be ubiquitous and not restricted to alkene-producing tissues, other proteins must be involved to ensure that desaturation products
enter the VLCFA elongation pathway in flowers. For example, changes in the activities of acyl-ACP thioesterase or acyl-CoA synthetase isoforms (12, 13) would be potential candidates. Several orchid genera related to *Ophrys* produce low levels of alkenes, which might have served as a preadaptation for sexual deception in *Ophrys* (17). If so, changes in the relevant proteins should be present in both *Ophrys* and related genera.

**Figure 4. Model summarizing SAD2 involvement in floral isolation among *O. sphegodes* and *O. exaltata*.** SAD2 activity is higher in *O. sphegodes* (blue arrows) than in *O. exaltata* (red arrows), due to expression (and possibly functional) differences. SAD2 reaction products are elongated and converted to 9- and 12-alkenes, the levels of which are higher in *O. sphegodes* than in *O. exaltata*. The exact source of high levels of 7-alkenes in *O. exaltata* is currently unknown. Floral alkenes are detected by pollinators, with 9- and 12-alkenes functioning as attractants to the bee *Andrena nigroaenea* (the pollinator of *O. sphegodes*). Conversely, the bee *Colletes cunicularius* (the pollinator of *O. exaltata*) is attracted by 7-alkenes, whereas 9-alkenes reduce this attraction. Overall, different alkene blends in the two species lead to differential pollinator attraction associated with reproductive isolation.

Different *Ophrys* species produce different alkenes and double-bond differences will ultimately be due to desaturation reactions. Although several different mechanisms could potentially explain differences in desaturation among species, it appears that the higher expression of *SAD2* in *O. sphegodes* contributes to higher 9- and 12-alkene levels in this species. As OeSAD2 hardly differs from OsSAD2 around the active site and putative
substrate-binding pocket (Fig. S1E), it is likely that both enzymes catalyze the same reaction. There are, however, amino acid changes on the surface of SAD2 (Fig. S1D), so that an additional activity change due to different interactions with reaction partners (e.g., specific ACP or ferredoxin isoforms) (15, 16) cannot be ruled out. This may explain why only OsSAD2 affected unsaturated FA levels in transgenic Arabidopsis. SAD1 differs from SAD2 by both changes on the protein surface and changes in the substrate-binding pocket, especially where the aliphatic end of the substrate is expected to bind. However, two lines of evidence suggest that SAD1 is not a functional desaturase: First, SAD1 expression was not associated with alkene production. Second, no evidence of SAD1 activity was detected in either in vitro assays or transgenic Arabidopsis.

The species-specific alkene differences associated with SAD2 are biologically relevant (Fig. 4). Electrophysiological studies with the two specific pollinators, the solitary bees Andrena nigroaenea (for O. sphegodes) and Colletes cunicularius (for O. exaltata) showed that both detect 9-alkenes (C_{23}, C_{25}, C_{27}, C_{29}) and some 12-alkenes (Andrena: C_{27}, C_{29}; Colletes: C_{29}) (7, 8). Moreover, O. sphegodes alkene blends induced mating behavior in A. nigroaenea (7), whereas O. exaltata alkene blends containing 7- and 9-alkenes were less effective than only 7-alkenes for C. cunicularius (8), indicating that 9-alkenes may repel this pollinator. Taken together, this suggests that the alkenes linked to SAD2 activity are directly involved in the specificity of pollinator attraction, and thus reproductive isolation among the two orchid species.

In conclusion, our data are consistent with the proposal that the SAD2 desaturase underlies the phenotypic difference in 9- and 12-alkenes among O. sphegodes and O. exaltata and thereby contributes to differential pollinator attraction and reproductive isolation among these species. SAD2 therefore represents a barrier gene of large phenotypic effect on pollinator attraction by orchid flowers.
Methods

Plant material

Plants of *Ophrys sphegodes* Miller and *Ophrys exaltata* Tenore subsp. *archipelagi* (Gölz & Reinhard) Del Prete were grown in a greenhouse at the Botanic Garden of the University of Zürich. For developmental stage-specific analysis of hydrocarbons and gene expression, inflorescences were taken on the first day of anthesis of the first flower of a given plant, flowers and buds dissected, and the first open flower used as a reference point.

Gene cloning and expression analysis

Total RNA was extracted from flash-frozen orchid tissue using Trizol reagent (Invitrogen) according to the manufacturer’s instructions, followed by assessment of RNA quality and quantity by agarose gel electrophoresis and spectrophotometry using an ND-1000 (NanoDrop Technologies). Where necessary, RNA was further purified by LiCl precipitation (18). Total RNA was treated with DNAse I (Fermentas) and reverse-transcribed into cDNA using RevertAid M-MuLV H’ Reverse Transcriptase (Fermentas), an anchored oligo-dT primer and the supplier's protocol. Locus-specific and/or semi-quantitative PCR was carried out using RedTaq ReadyMix (Sigma), the supplier's protocol scaled to 10 – 20 µl with cDNA from 1 ng/µl total RNA as a template. For primers and cycling conditions, see Table S4 and SI. Initial amplification of orchid *SAD* fragments used a nested degenerate primer approach. PCR products were cloned into pDRIVE (Qiagen), positive clones identified, and Sanger-sequenced using BigDye 3.1 and a 3130XL Genetic Analyzer (Applied Biosystems), as recommended by the manufacturers. Full-length coding sequence was isolated as detailed in SI, deposited in GenBank (accessions FR688105-FR688110), and amplified essentially as before (but reactions also containing 0.015 u/µl *Pfu* DNA polymerase; Promega) with modified PCR primers (Table S4) to engineer flanking *attB* sequences during PCR, as
recommended by Invitrogen. AttB-site containing PCR products of OsSAD1, OsSAD2 and OeSAD2 were cloned into pDONR207 by BP recombination (Invitrogen) to give pENTR207-SAD, followed by selection on LB agar containing 10 µg/ml gentamicin, plasmid isolation, and sequence confirmation as described before.

**GC and GC/MS analyses**

Cuticular hydrocarbons were extracted by washing plant tissue in 0.5 ml n-hexane for one minute, adding 100 ng of n-octadecane as an internal standard. GC was carried out as described previously (9), except for the use of a lower heating rate of 4°C/min. Retention times were compared against those of synthetic hydrocarbon standards run with the same settings. The standards were: C_{19} and C_{21} – C_{29} n-alkanes and odd-chain (Z)-7-C_{21} – C_{25}, (Z)-9-C_{21} – C_{29}, (Z)-11-C_{25} – C_{29} and (Z)-12-C_{25} – C_{27} n-alkenes. Several samples were re-analyzed on an Agilent 5975 GC/MS with the same oven and column settings. Discrimination of (Z)-11/12 alkenes is not possible with these parameters. However, double-bond positions have previously been determined: both study species contain 11- and 12-alkenes, with 12-alkenes as the predominant isomer (19, 20). FAMEs extracted from Arabidopsis lines were analyzed by GC/MS using the same settings. FAMEs from desaturase assays were analyzed as in (21).

**Plant expression of desaturases and biochemical activity assay**

To create 2x35S:SAD expression vectors, pENTR207-SAD entry clones were recombined with the binary plant expression vector pMDC32 (22) by LR recombination (Invitrogen) and selected on kanamycin. Plasmids were isolated, sequenced and transformed into Agrobacterium tumefaciens strain LBA4404, which was in turn used to transform A. thaliana line SALK_036854 (23) using the floral dip method (24). This line carries a T-DNA insertion in SSI2, associated with a recessive dwarf phenotype (see SI). Transgenic
*Arabidopsis* plants were selected on MS (25) agar containing 0.05 % Plant Preservative Mixture (Plant Cell Technology) and 25 µg/ml hygromycin. Selected independent transgenic lines in an *ssi2/ssi2* background were tested for complementation: 35S:*OsSAD1* (*N*=2), 35S:*OsSAD2* (*N*=5), and 35S:*OeSAD2* (*N*=2). Transgene expression (Fig. S5B) and sequence were confirmed by RT-PCR and Sanger sequencing as described above. FAMEs were prepared by BCl3/methanol extraction (26).

Different constructs for protein expression in *Escherichia coli* were made and evaluated as detailed in SI. Expression clones containing N-terminally modified orchid desaturases in the pET9d (Novagen) expression vector were chosen for functional analysis. In these clones, amino acids 2-5 (ELHL) were deleted to remove part of the putative chloroplast transit peptide. Proteins were purified and assayed as previously described (21), with minor modifications: only 7,7,8,8-²H₄-16:0-ACP and 12,12-²H₂-18:0-ACP substrates were used in assays containing 100 µg of desaturase, incubated for 2 h at 22 °C. FAMEs were suspended in 50 µl hexane for GC/MS analysis.

**Bioinformatic and statistical analyses**

Molecular mass and isoelectric point of proteins were predicted using the ExPASy Server (27) and the presence of a chloroplast transit peptide predicted using the ChloroP 1.1 server (28). Homology modeling was performed using the SWISS-MODEL server (29) and the 2.4 Å crystal structure 1OQ4 (chain A) (30) of RcSAD as a template. Validation and quality checking of the models were done using the ProSA-web server (31) and Procheck software (32).

Homologs of the *Arabidopsis SSI2* desaturase gene (Table S1) were extracted from public sequence databases as detailed in SI and aligned based on amino acid sequence using PRANK 0.91 (33). Poorly alignable regions were excluded from downstream analysis. The
GTR+I+Γ nucleotide substitution model was selected using MrModeltest 2.2 (34) and phylogenetic analysis conducted in MrBayes 3.1.2 (35), discarding results prior to apparent convergence of analysis chains (burn-in 1M out of 30M generations). Branch lengths of the resulting consensus tree were optimized with BaseML and used as input for CodeML, both part of the PAML 4.3 (36) package. Different models of sequence evolution were calculated with CodeML and compared by likelihood ratio testing. Fisher’s exact tests were done on (non-)synonymous site counts (37) using CodeML output. Statistical analyses were performed in Microsoft Excel and R 2.11.0 (38).

Acknowledgments

We thank A. Bolaños, A. Boyko, S. Cozzolino, M. Curtis, S. Kessler, M. & S. Schauer and H. Zheng for providing laboratory materials, help or source code, and M. Anisimova and M. & S. Schauer for discussions and comments. This work was supported by the Austrian Science Fund (fellowship no. J2678-B16 to P.M.S.), the Swiss Federal Institute of Technology Zürich (grant no. TH 02 06-2 to F.P.S.), the University of Zürich (U.G. and F.P.S.) and the Office of Basic Energy Sciences of the United States Department of Energy (J.S. and E.J.W.).
References


11. Xu S, et al. (subm.) Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids.


Supporting Information

Supplementary Methods

Details of expression analysis and RACE experiments

RT-PCR reactions were set up as described in the main text, with a typical cycling program as follows: 95 °C 3 min, 37 × (95 °C 30 s, 58 °C 30 s, 72 °C 1 min 30 s), 72 °C 10 min. *Pfu* DNA polymerase (Promega) was added to all reactions to be sequenced or re-cloned at a final concentration of 0.015 u/µl.

5' and 3' RACE were carried out using GeneRacer kit (Invitrogen) and as described elsewhere (1, 2), except for the use of a different RNA oligonucleotide (Table S4; a generous gift of Dr S. Schauer) for 5' RACE. For RACE, PCRs were carried out with JumpStart REDAccuTaq LA DNA Polymerase (Sigma) or Advantage GC 2 Polymerase (Clontech) and a PCR extension step at 68 °C. RACE was complemented by PCR-walking on genomic DNA extracted with Nucleon PhytoPure (GE Healthcare) using the DNA Walking SpeedUp Kit (Seegene). Products were cloned into pDRIVE (Qiagen) and sequenced. In addition to the *SAD* genes, a full-length *Ophrys G3PDH* cDNA (putatively encoding a glycolytic glyceraldehyde 3-phosphate dehydrogenase) was isolated as a control.

Locus-specificity of amplification was initially determined by gradient PCRs (as outlined above) using 1.5 ng/µl *SAD* plasmid as a template and locus-specific full-length primers (Table S4). *SAD2*-specific primers did not result in bands visible on agarose gel from *SAD1* plasmid at annealing temperatures (T\_A) of 60 °C or higher. Conversely, *SAD1*-specific primers did not amplify *SAD2* at T\_A ≥ 63 °C annealing temperature. Since using the same high stringency of PCR conditions reduced the amplification from cDNA, T\_A was lowered to 58 °C in RT-PCR.

For analysis of tissue- and stage-specific expression levels, RT-PCR was carried out at T\_A = 58 °C and *attB*-containing primers allowing recombination cloning of PCR products. This experiment was conducted for 80 tissue samples (different floral developmental stages.
from flowers and buds at different positions relative to the first open flower; labella, sepal and petal, and green leaves or bracts) from 6 *O. sphegodes* and 4 *O. exaltata* individuals, for which GC data from the same tissues were available. For every plant individual in the experiment, PCR products from open flowers were cloned into pDONR221 by BP recombination (Invitrogen) and 5-10 clones sequenced completely. All clones were found to be from the correct *SAD* locus. To measure expression level, RT-PCR from 0.8 ng/µl template was carried out independently at two cycle numbers for each gene investigated (29 and 33, except for *G3PDH*, where 29 and 31 cycles were used). For *G3PDH*, the primer combination used (1f/2r; Table S4) amplifies a larger, intron-containing region from genomic DNA and thus served as a control for genomic DNA contamination in RT-PCR experiments (appropriate controls were included). 5 µl of each PCR product were loaded on 0.8% TAE (Tris-acetate-EDTA)-agarose gel containing 0.1 µg/ml ethidium bromide and electrophoresed for 40 min at 95 V. Loading ensured that PCR products of different genes of the same biological samples were loaded beside each other. Digital photographs of gels were taken under UV at three constant exposure settings for each experiment and band intensity quantified using Image J 1.43u software (3). *SAD* expression levels were normalized against *G3PDH* and measurements averaged for each sample.

**Statistical analysis of hydrocarbon and expression data**

Since hydrocarbons were analyzed from buds of different developmental stages, use of comparable amounts of tissue could not be ensured. Hydrocarbons were thus analyzed as relative alkane and alkene amounts (as percentages of the total amount of alkanes and alkenes per sample). Data analysis was performed in R 2.11.0 (4). Among-species hydrocarbon differences were tested in a Wilcoxon signed-rank test. The pairwise correlation of hydrocarbons was investigated by clustering based on a correlation distance (pvclust method) (5). For this cluster analysis only, additional hydrocarbon data were added (*N*=62+96 additional data points from 7 *O. exaltata* and 12 *O. sphegodes* individuals, respectively, for which no RNA was available). The effect of *SAD* expression on relative alkane and alkene proportions (arcsine square-root transformed) was assessed in a generalized linear model (GLM) and a linear mixed-effect model (LME). The GLM used a Gaussian error distribution and modeled hydrocarbon levels with the explanatory variables species, tissue type and
Chapter 3

SAD1, SAD2 and SAD3 expression. The LME used the same explanatory variables and added biological individual as a random-effect factor in a maximum likelihood model. The models were simplified using stepwise factor removal by AIC (Akaike information criterion; stepAIC method).

Characterization of Arabidopsis T-DNA insertion line SALK_036854

Arabidopsis line SALK_036854 (6) carries a T-DNA insertion in the SSI2/FAB2 (SUPPRESSOR OF SA-INSENSITIVITY2/FATTY ACID BIOSYNTHESIS 2) locus (At2g43710). The position of the insertion was determined by sequencing a PCR product amplified with primers SSI2Dn1R and LB (Table S4) and was found to be between nucleotide positions 6 and 7 of the second intron of At2g43710. The insertion line displayed a recessive dwarf phenotype similar to other loss-of-function ssi2 alleles (7-9). These phenotypic dwarves exhibited a poor transformation rate and a high mortality rate. Genotyping of plants was done by PCR using primer combinations SSI2Up1F/SSI2Dn1R (wild-type allele) and LB/SSI2Dn1R (T-DNA allele), or using the combination SSI2Up2F/LB/SSI2Dn1R for concomitant amplification of both alleles. Further, for confirmation of transgenic plants carrying orchid SAD, primer combinations B1-SAD1f/B2-SAD1r (SAD1) or B1-SAD2f/B2-SAD2r (SAD2) were used on genomic DNA (for genotyping) or cDNA (for confirmation of expression). For RT-PCR, an intron-containing region of ACT1 (ACTIN 1; At2g37620; primers ACT1f/ACT1r) was used as a control.

Orchid desaturase constructs for bacterial expression

Three sets of desaturase constructs (pRSET-SAD, pET9d-MAX-SAD and pET9d-RcX-SAD constructs) were made and evaluated for protein expression and feasibility of desaturase assays, as described below:

pRSET-GW and pRSET-SAD constructs. The plasmid pRSET-B (Invitrogen) for bacterial protein expression with an N-terminal His6-tag was modified by replacing the multi-
cloning site by a Gateway cassette. This was done by digesting the plasmid with BamHI and EcoRI, blunting by T4 DNA polymerase treatment (10), ligation of Gateway conversion cassette C.1 (Invitrogen) and transformation into *Escherichia coli* DB3.1 (Invitrogen), followed by selection on 25 µg/ml chloramphenicol and 50 µg/ml ampicillin. Plasmids were isolated and their sequence confirmed, selecting only plasmids in the correct orientation, and the resulting plasmids called pRSET-GW. pRSET-GW was functionally validated by recombination with a pENTR-GUS control plasmid (Invitrogen). pRSET-SAD constructs were made by LR recombination among pRSET-GW and pENTR207-SAD constructs (see main text), followed by transformation into *E. coli* TOP10 (Invitrogen) and selection on 50 µg/ml ampicillin. All plasmids were confirmed by sequencing. pRSET-SS1 (OsSAD1), pRSET-SS2 (OsSAD2) and pRSET-ES2 (OeSAD2) constructs were transformed into *E. coli* BL21 (DE3) cells (Invitrogen), their sequence validated, protein expressed as recommended by Invitrogen, and purified using His SpinTrap columns (GE Healthcare) under non-denaturing conditions as recommended by the manufacturer. Since no desaturase activity was detected in initial enzyme assays (as described in the main text, but using unlabeled fatty acids from Sigma-Aldrich, and *E. coli* ACP and acyl-ACP synthetase from Invitrogen), we reasoned that the presence of additional amino acids at the N-terminus (tag and possibly transit peptide) might adversely affect enzyme activity. Therefore, additional constructs were made and evaluated.

**pET9d-MAX-SAD constructs.** Untagged bacterial expression constructs in pET9d (Novagen) were made deleting amino acid residues 2-5 (ELHL) to remove part of the putative chloroplast transit peptide from the *Ophrys* desaturases. OsSAD1 (SS1), OsSAD2 (SS2) and OeSAD2 (ES2) were re-amplified from pENTR207 clones by PCR, introducing an *Nco*I site before the translation start site at the 5’ end and a *Bgl*II site after the stop codon at the 3’ end of the gene, using PCR primers *Nco*I-MA-SS1f/SS2f (forward) and *Bgl*II-SS1r/SS2r/ES2r (reverse; Table S4). PCR products were purified, cloned into pJET1.2 (Fermentas), chemically transformed into *E. coli* DH5α-T1R (Invitrogen) and transformants selected using 100 µg/ml ampicillin. Colonies were screened for positive clones by PCR using the primers recommended by the supplier, and plasmids extracted by alkaline lysis and sequenced (10). As the plasmids still contained an internal undesired *Nco*I site, this site was removed by site-directed mutagenesis (10), using oligonucleotides SAD1-XNcoI-f/r and SAD2-XNcoI-f/r, Jumpstart Red AccuTaq LA DNA polymerase mix (Simga-Aldrich) and DpnI digestion to select against methylated source plasmids. Resulting plasmids were
transformed into DH5α cells and colonies screened by PCR and digestion with Neol. The desaturase sequences were transferred to pET9d by digestion with Neol and BglII and ligation into Neol/BamHI-digested pET9d, followed by transformation into DH5α cells, selection on 50 µg/ml kanamycin, colony PCR using T7 primers, plasmid extraction and sequence verification, to yield pET9d-MAX-SS1, pET9d-MAX-SS2 and pET9d-MAX-ES2.

**pET9d-RcX-SAD constructs.** Additional constructs were made for Ophrys SAD2 to remove the putative chloroplast transit peptide completely and replace the N-terminal end of the Ophrys desaturases, up to half of the first predicted α-helix, with the corresponding sequence from the Ricinus communis Δ^9^-SAD enzyme. The first amino acid residue of SS2/ES2 sequence in these constructs was Trp52. The front-end of the Ricinus enzyme was PCR amplified from a pET9d plasmid containing the mature Ricinus wild-type 18:0-ACP Δ^9^ desaturase (11) using primers Neol-MastL-RcSADF and a chimaeric RcSAD/Ophrys SAD2 primer NRcOSAD2r carrying a BspTI site. The purified PCR product was cloned into pDRIVE, transformed into DH5α cells and selected on 50 µg/ml kanamycin. Resulting plasmids were mixed with pET9d-MAX-SAD constructs, digested with Neol and BspTI, ligated and transformed into DH5α cells, selected on kanamycin and transformants screened for the correct insert by colony PCR. The resulting plasmids pET9d-RcX-SS2 and pET9d-RcX-ES2 were confirmed by sequencing.

pET9d-MAX-SAD and pET9d-RcX-SAD constructs were transformed into *E. coli* BL21 (DE3) cells and assayed for protein expression under different conditions. For OsSAD1 and OsSAD2 desaturase assays, pET9d-MAX-SAD cultures were induced with IPTG for 4 hours at room temperature, protein extracted and enriched to >90% purity by 20CM cation exchange chromatography (Applied Biosystems) as previously described (12). Soluble expression of OeSAD2 could not be attained.

*Identification of desaturase homologs in public sequence databases.*

*Arabidopsis* SSI2 protein sequence (NP_850400) was chosen as a seed for retrieval of homologous sequences from public sequence databases using BLAST searches (13). SSI2 protein sequence was first compared against the NCBI NR database (http://www.ncbi.nlm.nih.gov/) by BLASTP (BLAST+ executables from the NCBI C++
Toolkit; http://www.ncbi.nlm.nih.gov/). All protein sequences matching at least 65% of SSI2 sequence in BLASTP were retrieved. From this set, all sequences with at least 98% pairwise amino acid sequence identity to any other sequence in the data set were considered redundant and only one of the redundant sequences kept. The resulting sequences were used for tBLASTN comparison to NCBI genome and NCBI cDNA databases (both from http://www.ncbi.nlm.nih.gov/), the *Populus trichocarpa* (http://genome.jgi-psf.org/) and *Vitis vinifera* genomes (http://www.cns.fr/vitis/), and the PlantGDB EST and genome database (http://www.plantgdb.org/), in their most current versions as of 29 July 2009. *Ophrys* sequences from this study were also included at this step, which was carried out using the FGF scripts of (14), kindly provided by Hongkun Zheng, using the default parameters (i.e., a minimum of 50% of sequence match at the amino acid level). Resulting nucleic acid sequences were aligned to the input protein sequences and open reading frames identified using GeneWise 2.2 (15). Redundant sequences were removed from the resulting set of sequences if pairwise identity was at least 96% at the nucleic acid level (excepting *Ophrys* sequences), keeping only the best possible hit, to yield the final sequence set for phylogenetic analysis. Perl scripts and data for Table S1 are available upon request.
Table S1. List of stearoyl-ACP desaturase homologs identified. Gene symbols are reported with the species, where such symbols were available in the source databases. In other cases, different homologs in a species are simply numbered (e.g., #1). The ID or accession number within a source database set refers to the best partial match of a given sequence identified by our script.

<table>
<thead>
<tr>
<th>Species and gene symbol</th>
<th>Source</th>
<th>ID or accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana SAD1</td>
<td>TAIR</td>
<td>At5g16240</td>
</tr>
<tr>
<td>Arabidopsis thaliana SAD2</td>
<td>TAIR</td>
<td>At3g02610</td>
</tr>
<tr>
<td>Arabidopsis thaliana SAD3</td>
<td>TAIR</td>
<td>At5g16230</td>
</tr>
<tr>
<td>Arabidopsis thaliana SAD4</td>
<td>TAIR</td>
<td>At3g02620</td>
</tr>
<tr>
<td>Arabidopsis thaliana SAD5</td>
<td>TAIR</td>
<td>At3g02630</td>
</tr>
<tr>
<td>Arabidopsis thaliana SAD6</td>
<td>TAIR</td>
<td>At3g02630</td>
</tr>
<tr>
<td>Arabidopsis thaliana SSI2</td>
<td>TAIR</td>
<td>At2g43710</td>
</tr>
<tr>
<td>Arachis hypogaea FAD1</td>
<td>NCBI</td>
<td>AF172728.1</td>
</tr>
<tr>
<td>Asclepias syriaca</td>
<td>NCBI</td>
<td>U60277.1</td>
</tr>
<tr>
<td>Bassia scoparia #1</td>
<td>NCBI</td>
<td>AF315599.1</td>
</tr>
<tr>
<td>Bassia scoparia #2</td>
<td>NCBI</td>
<td>AF315600.1</td>
</tr>
<tr>
<td>Brassica juncea fab2</td>
<td>NCBI</td>
<td>AF153420.1</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>NCBI</td>
<td>AY642537.1</td>
</tr>
<tr>
<td>Brassica napus #1</td>
<td>NCBI</td>
<td>DQ886528.1</td>
</tr>
<tr>
<td>Brassica napus #2</td>
<td>NCBI</td>
<td>X60978.1</td>
</tr>
<tr>
<td>Carica papaya</td>
<td>NCBI</td>
<td>AY451852.1</td>
</tr>
<tr>
<td>Carthamus tinctorius</td>
<td>NCBI</td>
<td>M61109.1</td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii FAB2</td>
<td>NCBI</td>
<td>XM_001691545.1</td>
</tr>
<tr>
<td>Cinnamomum lancepaniculatum SAD</td>
<td>NCBI</td>
<td>EU131523.1</td>
</tr>
<tr>
<td>Coriandrum sativum</td>
<td>NCBI</td>
<td>AY770499.1</td>
</tr>
<tr>
<td>Cucumis sativus</td>
<td>NCBI</td>
<td>M59858.1</td>
</tr>
<tr>
<td>Descurainia sophia</td>
<td>NCBI</td>
<td>EF524186.1</td>
</tr>
<tr>
<td>Glycine max SACPD</td>
<td>NCBI</td>
<td>L34346.1</td>
</tr>
<tr>
<td>Glycine max SACPD A</td>
<td>NCBI</td>
<td>AY885234.2</td>
</tr>
<tr>
<td>Glycine max #3</td>
<td>NCBI</td>
<td>AK245255.1</td>
</tr>
<tr>
<td>Glycine max #4</td>
<td>PlantGDB</td>
<td>Gm13A</td>
</tr>
<tr>
<td>Glycine max #5</td>
<td>PlantGDB</td>
<td>Gm13B</td>
</tr>
<tr>
<td>Species</td>
<td>Repository</td>
<td>Accession</td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Gossypium hirsutum sad1</td>
<td>NCBI</td>
<td>AJ132636.1</td>
</tr>
<tr>
<td>Gymnadenia conopsea</td>
<td>NCBI</td>
<td>EF051334.1</td>
</tr>
<tr>
<td>Haematococcus pluvialis</td>
<td>NCBI</td>
<td>EF586860.1</td>
</tr>
<tr>
<td>Helianthus annuus #1</td>
<td>NCBI</td>
<td>U91339.1</td>
</tr>
<tr>
<td>Helianthus annuus #2</td>
<td>NCBI</td>
<td>U91340.1</td>
</tr>
<tr>
<td>Hordeum vulgare #1</td>
<td>PlantGDB</td>
<td>PUT-157a-Hordeum_vulgare-46041</td>
</tr>
<tr>
<td>Hordeum vulgare #2</td>
<td>PlantGDB</td>
<td>PUT-157a-Hordeum_vulgare-65964</td>
</tr>
<tr>
<td>Hordeum vulgare #3</td>
<td>NCBI</td>
<td>151419245</td>
</tr>
<tr>
<td>Jatropha curcas SAD</td>
<td>NCBI</td>
<td>DQ084491.1</td>
</tr>
<tr>
<td>Kosteletzkya virginica SAD</td>
<td>NCBI</td>
<td>FJ750952.1</td>
</tr>
<tr>
<td>Linum usitatissimum sad2</td>
<td>NCBI</td>
<td>AJ006958.1</td>
</tr>
<tr>
<td>Lotus japonicus</td>
<td>NCBI</td>
<td>DQ020280.1</td>
</tr>
<tr>
<td>Lupinus luteus Lidd3A20</td>
<td>NCBI</td>
<td>AF139377.1</td>
</tr>
<tr>
<td>Macfadyena unguis-cati fad1</td>
<td>NCBI</td>
<td>AF051134.1</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>NCBI</td>
<td>BT052760.1</td>
</tr>
<tr>
<td>Olea europaea</td>
<td>NCBI</td>
<td>U58141.1</td>
</tr>
<tr>
<td>Ophrys exaltata subsp. archipelagi SAD2</td>
<td>This study</td>
<td>FR688110</td>
</tr>
<tr>
<td>Ophrys sphegodes SAD1</td>
<td>This study</td>
<td>FR688108</td>
</tr>
<tr>
<td>Ophrys sphegodes SAD2</td>
<td>This study</td>
<td>FR688109</td>
</tr>
<tr>
<td>Ophrys sphegodes SAD3</td>
<td>This study</td>
<td>FR688106</td>
</tr>
<tr>
<td>Oryza sativa #1</td>
<td>TIGR</td>
<td>Os01g0880800</td>
</tr>
<tr>
<td>Oryza sativa #2</td>
<td>TIGR</td>
<td>Os01g0919900</td>
</tr>
<tr>
<td>Oryza sativa #3</td>
<td>TIGR</td>
<td>Os02g0504800</td>
</tr>
<tr>
<td>Oryza sativa #4</td>
<td>TIGR</td>
<td>Os03g0423300</td>
</tr>
<tr>
<td>Oryza sativa #5</td>
<td>TIGR</td>
<td>Os04g0379900</td>
</tr>
<tr>
<td>Oryza sativa #6</td>
<td>TIGR</td>
<td>Os08g0199400</td>
</tr>
<tr>
<td>Oryza sativa #7</td>
<td>TIGR</td>
<td>Os08g0200100</td>
</tr>
<tr>
<td>Ostreococcus lucimarinus</td>
<td>NCBI</td>
<td>XM_001417261.1</td>
</tr>
<tr>
<td>Pavlova viridis fadA</td>
<td>NCBI</td>
<td>EU000382.1</td>
</tr>
<tr>
<td>Pelargonium x hortorum</td>
<td>NCBI</td>
<td>U40344.1</td>
</tr>
<tr>
<td>Persea americana</td>
<td>NCBI</td>
<td>AF116861.1</td>
</tr>
<tr>
<td>Phaeodactylum tricornatum</td>
<td>NCBI</td>
<td>XM_002177381.1</td>
</tr>
<tr>
<td>Phyllostachys edulis</td>
<td>NCBI</td>
<td>FP091892.1</td>
</tr>
<tr>
<td>Species</td>
<td>Database</td>
<td>Accession</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #1</td>
<td>NCBI</td>
<td>168039273</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #2</td>
<td>NCBI</td>
<td>168040944</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #3</td>
<td>NCBI</td>
<td>168044588</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #4</td>
<td>NCBI</td>
<td>168065984</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #5</td>
<td>PlantGDB</td>
<td>scaffold</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #6</td>
<td>PlantGDB</td>
<td>scaffold</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em> #7</td>
<td>PlantGDB</td>
<td>scaffold</td>
</tr>
<tr>
<td><em>Picea sitchensis</em></td>
<td>NCBI</td>
<td>EF085288.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #1</td>
<td>NCBI</td>
<td>XM_002307442.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #2</td>
<td>NCBI</td>
<td>XM_002311231.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #3</td>
<td>NCBI</td>
<td>XM_002316135.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #4</td>
<td>NCBI</td>
<td>XM_002330664.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #5</td>
<td>NCBI</td>
<td>XM_002321290.1</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em> #6</td>
<td><em>Populus genome</em></td>
<td>LG_X</td>
</tr>
<tr>
<td><em>Ricinus communis</em> SAD</td>
<td>NCBI</td>
<td>M59857.1</td>
</tr>
<tr>
<td><em>Ricinus communis</em> SAD2</td>
<td>NCBI</td>
<td>EU878554.1</td>
</tr>
<tr>
<td><em>Ricinus communis</em> SAD3</td>
<td>NCBI</td>
<td>EU878553.1</td>
</tr>
<tr>
<td><em>Ricinus communis</em> SAD4</td>
<td>NCBI</td>
<td>FJ643486.1</td>
</tr>
<tr>
<td><em>Saussurea involucrata</em></td>
<td>NCBI</td>
<td>DQ516384.1</td>
</tr>
<tr>
<td><em>Sesamum indicum</em></td>
<td>NCBI</td>
<td>D49832.1</td>
</tr>
<tr>
<td><em>Simmondsia chinensis</em></td>
<td>NCBI</td>
<td>M83199.1</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em> #1</td>
<td>NCBI</td>
<td>AK325946.1</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em> #2</td>
<td>NCBI</td>
<td>BT014263.1</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em> #3</td>
<td>NCBI</td>
<td>BT014387.1</td>
</tr>
<tr>
<td><em>Solanum tuberosum</em></td>
<td>NCBI</td>
<td>M91238.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #1</td>
<td>PlantGDB</td>
<td><em>Sorghum chromosome 1</em></td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #2</td>
<td>NCBI</td>
<td>XM_002464685.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #3</td>
<td>NCBI</td>
<td>XM_002465071.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #4</td>
<td>NCBI</td>
<td>XM_002460450.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #5</td>
<td>NCBI</td>
<td>XM_002453832.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #6</td>
<td>NCBI</td>
<td>XM_002446233.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #7</td>
<td>NCBI</td>
<td>XM_002443958.1</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> #8</td>
<td>NCBI</td>
<td>XM_002437426.1</td>
</tr>
<tr>
<td>Species</td>
<td>Database</td>
<td>Accession</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Spinacia oleracea sadA</td>
<td>NCBI</td>
<td>FJ418167.1</td>
</tr>
<tr>
<td>Thalassiosira pseudonana DES7</td>
<td>NCBI</td>
<td>XM_002297328.1</td>
</tr>
<tr>
<td>Thunbergia alata pTAD2</td>
<td>NCBI</td>
<td>U07597.1</td>
</tr>
<tr>
<td>Thunbergia alata pTAD3</td>
<td>NCBI</td>
<td>U07605.1</td>
</tr>
<tr>
<td>Thunbergia alata pTAD4</td>
<td>NCBI</td>
<td>U09269.1</td>
</tr>
<tr>
<td>Triadica sebifera</td>
<td>NCBI</td>
<td>EF079655.1</td>
</tr>
<tr>
<td>Triticum aestivum #1</td>
<td>PlantGDB</td>
<td>PUT-163b-Triticum_aestivum-014567</td>
</tr>
<tr>
<td>Triticum aestivum #2</td>
<td>PlantGDB</td>
<td>PUT-163b-Triticum_aestivum-019534</td>
</tr>
<tr>
<td>Triticum aestivum #3</td>
<td>NCBI</td>
<td>AK331076.1</td>
</tr>
<tr>
<td>Triticum aestivum #4</td>
<td>PlantGDB</td>
<td>PUT-163b-Triticum_aestivum-766165441</td>
</tr>
<tr>
<td>Triticum aestivum #6</td>
<td>PlantGDB</td>
<td>PUT-163b-Triticum_aestivum-157676</td>
</tr>
<tr>
<td>Tropaeolum majus</td>
<td>NCBI</td>
<td>AY188688.1</td>
</tr>
<tr>
<td>Vernicia montana SAD</td>
<td>NCBI</td>
<td>EU072353.1</td>
</tr>
<tr>
<td>Vitis vinifera #1</td>
<td>NCBI</td>
<td>225434117</td>
</tr>
<tr>
<td>Vitis vinifera #2</td>
<td>NCBI</td>
<td>225434123</td>
</tr>
<tr>
<td>Vitis vinifera #3</td>
<td>NCBI</td>
<td>XR_077944.1</td>
</tr>
<tr>
<td>Vitis vinifera #4</td>
<td>Vitis genome</td>
<td>chr5</td>
</tr>
<tr>
<td>Vitis vinifera #5</td>
<td>NCBI</td>
<td>XM_002274672.1</td>
</tr>
<tr>
<td>Vitis vinifera #6</td>
<td>NCBI</td>
<td>XM_002264759.1</td>
</tr>
<tr>
<td>Vitis vinifera #7</td>
<td>NCBI</td>
<td>XM_002265150.1</td>
</tr>
<tr>
<td>Zea mays #1</td>
<td>NCBI</td>
<td>EU946999.1</td>
</tr>
<tr>
<td>Zea mays #2</td>
<td>NCBI</td>
<td>EU966614.1</td>
</tr>
<tr>
<td>Zea mays #3</td>
<td>NCBI</td>
<td>NM_001139285.1</td>
</tr>
<tr>
<td>Zea mays #4</td>
<td>NCBI</td>
<td>BT063291.1</td>
</tr>
<tr>
<td>Zea mays #5</td>
<td>NCBI</td>
<td>NM_001157868.1</td>
</tr>
<tr>
<td>Zea mays #6</td>
<td>NCBI</td>
<td>226506507</td>
</tr>
<tr>
<td>Zea mays #7</td>
<td>NCBI</td>
<td>BT018751.1</td>
</tr>
<tr>
<td>Zea mays #8</td>
<td>NCBI</td>
<td>NM_001157610.1</td>
</tr>
<tr>
<td>Zea mays #9</td>
<td>PlantGDB</td>
<td>PUT-163a-Zea_mays-103945</td>
</tr>
</tbody>
</table>
### Table S2. Analysis of selection using PAML

Top, likelihood values \((\ln L)\) and parameter maximum likelihood estimates (MLEs) for substitution models (names reflect naming in PAML output). Middle, likelihood ratio tests for differences among substitution models, where \(2\delta = 2(\ln L_1 - \ln L_2)\) is compared against a \(\chi^2\) distribution with the degrees of freedom (d.f.) corresponding to the difference in the number of free model parameters. Bottom, Fisher’s exact tests (one-sided) comparing the inferred number of non-synonymous \((N)\) and synonymous \((S)\) changes at the branch tested to the expected total number of (non-)synonymous sites \((N_t\) and \(S_t)\), all obtained from PAML output. This follows the “test of positive selection” (16) to test for deviations from neutral expectation (“Neutral”); to test for relaxation of purifying selection (“Relaxed”), \(N\) and \(S\) were compared to the total numbers of inferred (non-)synonymous changes, which represent overall purifying selection. Letters after models M2 refer to branches as labeled in Fig. 1.

<table>
<thead>
<tr>
<th>Substitution model</th>
<th>(\ln L)</th>
<th>Parameter MLEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0 (one (\omega) ratio)</td>
<td>-60531.7</td>
<td>(\omega_0=0.101)</td>
</tr>
<tr>
<td><strong>Site models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1a (nearly neutral)</td>
<td>-60211.2</td>
<td>(\omega_0=0.097, \omega_1=1.0) ((p_0=0.896, p_1=0.104))</td>
</tr>
<tr>
<td>M2a (positive selection)</td>
<td>-60211.2</td>
<td>(\omega_0=0.097, \omega_1=1.0, \omega_2=1.0) ((p_0=0.896, p_1=0.033, p_2=0.071))</td>
</tr>
<tr>
<td>M7 ((\beta))</td>
<td>-58747.8</td>
<td>(p=0.794, q=4.738)</td>
</tr>
<tr>
<td>M8 ((\beta+\omega&gt;1))</td>
<td>-58745.1</td>
<td>(p_0=0.991, p=0.813, q=5.059) ((p_1=0.009)), (\omega=1.055)</td>
</tr>
<tr>
<td>M8a ((\beta+\omega=1))</td>
<td>-58745.1</td>
<td>(p_0=0.991, p=0.813, q=5.057) ((p_1=0.009)), (\omega=1.0)</td>
</tr>
<tr>
<td><strong>Branch models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1 (free-ratio)</td>
<td>-59625.5</td>
<td>(\omega_A=0.004, \omega_B=85.215, \omega_C=0.206, \omega_D=0.071, \omega_E=0.060, ) etc.(^1)</td>
</tr>
<tr>
<td>M2 (2-ratio), A: SAD1/SAD2</td>
<td>-60527.0</td>
<td>(\omega_0=0.102, \omega_1=\omega_A=0.009)</td>
</tr>
<tr>
<td>M2 (2-ratio), B: SAD1</td>
<td>-60506.9</td>
<td>(\omega_0=0.100, \omega_1=\omega_B=999)</td>
</tr>
<tr>
<td>M2 (2-ratio), C: SAD2</td>
<td>-60515.7</td>
<td>(\omega_0=0.101, \omega_1=\omega_C=0.210)</td>
</tr>
<tr>
<td>M2 (2-ratio), D: SAD3/Gymnadenia SAD</td>
<td>-60531.0</td>
<td>(\omega_0=0.102, \omega_1=\omega_D=0.063)</td>
</tr>
<tr>
<td>M2 (2-ratio), E: SAD3</td>
<td>-60529.8</td>
<td>(\omega_0=0.102, \omega_1=\omega_E=0.060)</td>
</tr>
<tr>
<td>M2 (3-ratio), A,B</td>
<td>-60503.0</td>
<td>(\omega_0=0.101, \omega_1=\omega_A=0.0098, \omega_2=\omega_B=999)</td>
</tr>
<tr>
<td>M2 (3-ratio), A,C</td>
<td>-60509.6</td>
<td>(\omega_0=0.101, \omega_1=\omega_A=0.008, \omega_2=\omega_C=999)</td>
</tr>
<tr>
<td>M2 (3-ratio), B,C</td>
<td>-60505.1</td>
<td>(\omega_0=0.100, \omega_1=\omega_B=999, \omega_2=\omega_C=0.210)</td>
</tr>
</tbody>
</table>
Chapter 3

M2 (4-ratio), A,B,C  -60501.0  \( \omega_0=0.101, \omega_1=\omega_A=0.009, \omega_2=\omega_B=999, \omega_3=\omega_C=0.218 \)

<table>
<thead>
<tr>
<th>Likelihood ratio test</th>
<th>2( \delta )</th>
<th>d.f.</th>
<th>( p )</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1a - M2a</td>
<td>4.56·10^4</td>
<td>2</td>
<td>&gt;0.999</td>
<td>ns</td>
</tr>
<tr>
<td>M7 - M8</td>
<td>5.424</td>
<td>2</td>
<td>0.066</td>
<td>ns</td>
</tr>
<tr>
<td>M8 - M8a</td>
<td>0.040</td>
<td>1</td>
<td>0.841</td>
<td>ns</td>
</tr>
<tr>
<td>M0 - M1</td>
<td>1812.276</td>
<td>245</td>
<td>7.58·10^{11}</td>
<td>****</td>
</tr>
<tr>
<td>M0 - M2, A</td>
<td>9.451</td>
<td>1</td>
<td>0.002</td>
<td>**</td>
</tr>
<tr>
<td>M0 - M2, B</td>
<td>49.541</td>
<td>1</td>
<td>1.942·10^{12}</td>
<td>****</td>
</tr>
<tr>
<td>M0 - M2, C</td>
<td>31.998</td>
<td>1</td>
<td>1.543·10^{8}</td>
<td>****</td>
</tr>
<tr>
<td>M0 - M2, D</td>
<td>1.426</td>
<td>1</td>
<td>0.232</td>
<td>ns</td>
</tr>
<tr>
<td>M0 - M2, E</td>
<td>3.676</td>
<td>1</td>
<td>0.055</td>
<td>ns</td>
</tr>
<tr>
<td>M2, A - M2, A,B</td>
<td>47.987</td>
<td>1</td>
<td>4.29·10^{12}</td>
<td>****</td>
</tr>
<tr>
<td>M2, A - M2, A,C</td>
<td>34.758</td>
<td>1</td>
<td>3.734·10^{9}</td>
<td>****</td>
</tr>
<tr>
<td>M2, B - M2, A,B</td>
<td>7.897</td>
<td>1</td>
<td>0.005</td>
<td>**</td>
</tr>
<tr>
<td>M2, B - M2, B,C</td>
<td>3.612</td>
<td>1</td>
<td>0.057</td>
<td>ns</td>
</tr>
<tr>
<td>M2, C - M2, A,C</td>
<td>12.211</td>
<td>1</td>
<td>4.752·10^{4}</td>
<td>***</td>
</tr>
<tr>
<td>M2, C - M2, B,C</td>
<td>21.155</td>
<td>1</td>
<td>4.235·10^{9}</td>
<td>****</td>
</tr>
<tr>
<td>M2, A,B - M2, A,B,C</td>
<td>3.937</td>
<td>1</td>
<td>0.047</td>
<td>*</td>
</tr>
<tr>
<td>M2, A,C - M2 A,B,C</td>
<td>17.166</td>
<td>1</td>
<td>3.425·10^{7}</td>
<td>****</td>
</tr>
<tr>
<td>M2, B,C - M2 A,B,C</td>
<td>8.222</td>
<td>1</td>
<td>0.004</td>
<td>**</td>
</tr>
<tr>
<td>M2, B - M1</td>
<td>1762.735</td>
<td>244</td>
<td>5.58·10^{28}</td>
<td>****</td>
</tr>
<tr>
<td>M2, C - M1</td>
<td>1780.278</td>
<td>244</td>
<td>2.86·10^{21}</td>
<td>****</td>
</tr>
<tr>
<td>M2, A,B,C - M1</td>
<td>1750.901</td>
<td>242</td>
<td>1.27·10^{26}</td>
<td>****</td>
</tr>
</tbody>
</table>

Fisher’s exact tests

<table>
<thead>
<tr>
<th>Branch/Gene</th>
<th>Model</th>
<th>Test</th>
<th>( N )</th>
<th>( S )</th>
<th>( N_f - N )</th>
<th>( S_f - S )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/SAD1</td>
<td>M2, B</td>
<td>Neutral</td>
<td>61</td>
<td>0</td>
<td>1227</td>
<td>404</td>
<td>4.18·10^{-8}</td>
</tr>
<tr>
<td>B/SAD1</td>
<td>M1</td>
<td>Neutral</td>
<td>55</td>
<td>0</td>
<td>1233</td>
<td>404</td>
<td>2.30·10^{-7}</td>
</tr>
<tr>
<td>C/SAD2</td>
<td>M2, C</td>
<td>Neutral</td>
<td>47</td>
<td>0</td>
<td>1241</td>
<td>405</td>
<td>2.14·10^{-9}</td>
</tr>
<tr>
<td>C/SAD2</td>
<td>M1</td>
<td>Neutral(^1)</td>
<td>32</td>
<td>49</td>
<td>1256</td>
<td>355</td>
<td>5.40·10^{-13}</td>
</tr>
<tr>
<td>C/SAD2</td>
<td>M1</td>
<td>Relaxed</td>
<td>32</td>
<td>49</td>
<td>9944</td>
<td>175745</td>
<td>&lt;2.20·10^{-16}</td>
</tr>
</tbody>
</table>

\(^1\)This model has different \( \omega \) ratios (where \( \omega = d_N/d_S \)) for every branch in the tree (not listed).

\(^2\)ns, not significant (\( p \geq 0.05 \)); *, \( p < 0.05 \); **, \( p < 0.01 \); ***, \( p < 0.001 \); ****, \( p < 0.0001 \).
In this case, the test outcome signifies purifying rather than positive selection. This branch/model combination was therefore subsequently tested for relaxation of purifying selection.
Supplementary References


Supplementary Figures

Figure S1. Orchid SAD homologs and protein structures. (A) Alignment of amino acid sequences of *Ricinus communis* SAD (RcSAD) and *Arabidopsis thaliana* SSI2 (AtSSI2) with *Ophrys* SADs, where gaps and identities are denoted by ‘-’ and ‘.’, respectively. Background color indicates α-helix (red), β-sheet (dark green), conserved Fe-coordinating residues (yellow) and sites in orchid SADs within 6 Å of substrate position that differ from RcSAD (magenta). (B) Comparison of protein backbones among the RcSAD crystal structure (27) and the superimposed *Ophrys* SAD homology models. The RcSAD structure is shown, with
magenta and orange strands indicating α-helices and β-sheets, respectively, and a thick gray line indicating the protein backbone elsewhere. Active site Fe atoms are shown as yellow spheres. The backbones of *Ophrys* SAD models are indicated by colored lines wherever their positions deviate from the RcsAD structure by at least 1.0 Å. (C) OsSAD2 homology model, showing sites within 6 Å of 18:0 substrate or Fe atoms, indicating conserved Fe-binding sites (yellow) and differences from RcsAD within 4 Å of substrate (magenta). (D) Differences among OsSAD1, OsSAD2 and OeSAD2 homology models, showing OsSAD2 as a gray wireframe and the active site Fe atoms (yellow) and modeled 18:0 substrate (from 27). All atoms deviating by > 1.0 Å in their positions among structures are highlighted in color. (E) Amino acid differences among OsSAD1, OsSAD2 and OeSAD2 homology models within 6.0 Å of active site Fe and modeled 18:0 substrate. Where there are no differences, the OsSAD2 structure is shown in gray; amino acid differences are superimposed in color. (B,D,E) Parts specific to OsSAD1 are shown in green, OsSAD2 in red, OeSAD2 in orange and OsSAD3 in dark blue.
Figure S2. Phylogenetic trees of plant stearoyl-ACP desaturase homologs. (A) Phylogenetic tree with topology from Bayesian inference (BI) reconstruction of sequences in Table S1, with BaseML-optimized branch lengths. Values near branches indicate statistical support as given by BI posterior probabilities (where >0.5). A subsection of this tree is presented and annotated in Fig. 1. (B-C) Same tree topology with branch lengths from CodeML’s free-ratio model M1, indicating rates of non-synonymous site change $d_N$ (B) or synonymous site change $d_S$ (C).
Figure S3. Details of *SAD* expression and hydrocarbons. (A) Ward clustering of hydrocarbons by pairwise correlation across different tissues and flower stages of *O.*
sphegodes and O. exaltata. Labels above branches represent AU (Approximately Unbiased) p-value percentages. (B) Color-coded summary of statistical tests and compound detectability by pollinators, showing the significance of correlations of SAD1, SAD2 and SAD3 expression with hydrocarbons in GLM and LME, the electro-antennographic detection (EAD) of compounds by pollinators Andrena nigroaenea (An) and Colletes cunicularius (Cc) (from 28, 29) and SD, the species difference in compounds in a Wilcoxon signed-rank test. (C) Relative amounts and s.e.m. of hydrocarbons in mature labella of O. sphegodes and O. exaltata. This pattern is concordant with previous reports (30). (D,E) Example gel pictures of RT-PCRs from O. exaltata (D) and O. sphegodes (E) showing expression of SAD genes (33 cycles) and G3PDH (31 cycles). The O. exaltata individual shown, together with a second individual in this experiment, shows no detectable SAD2 expression at 33 cycles, although SAD2 expression for the same individuals was detectable at higher cycle numbers (≥38). (F-O) Gene expression and hydrocarbon abundance in different flower stages and tissues, bars showing s.e.m., in O. exaltata (left) and O. sphegodes (right). (F,G) Normalized expression of SAD1, SAD2 and SAD3 relative to G3PDH. (H-O) Relative amounts (%) of 7-alkenes (H,I), 9-alkenes (J,K), 12-alkenes (L,M) and alkanes (N,O). Labella are labeled 0 (anthetic flower) to -X (bud, X positions before flower 0). Pooled sepal/petal samples are labeled sp followed by flower position, or SP for mature sepal/petal, and leaves are labeled L.
Figure S4. Correlations of hydrocarbons with gene expression. Normalized gene expression of SAD1, SAD2, and SAD3 versus $f(x) = \arcsin x^{0.5}$ transformed relative amounts of alkanes and alkenes from all tissues and floral stages, plotted individually for each comparison. O. exaltata is shown in red diamonds and O. sphegodes in blue circles. Regression lines are plotted for significant correlations ($p<0.05$, after correction of $p$-values for multiple testing using Holm's method). Where correlations were significantly different for the two species (1-way ANOVA; $p<0.05$), separate regression lines are shown for them. $R^2$ values are shown, where E and S subscripts denote O. exaltata and O. sphegodes, respectively, and their significance indicated: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$. Note that this correlation analysis is less powerful than the analyses summarized in Fig. S3B.
Figure S5. Comparison of *Arabidopsis* lines. (A) Photograph of two *Arabidopsis* plants before bolting. Both plants carry the 2x35S:OsSAD2 transgene. Plants heterozygous for the *ssi2* T-DNA insertion (left) have a wild-type (WT) phenotype, whereas *ssi2/ssi2* homozygous plants (right) have a dwarf phenotype that is not complemented by *Ophrys* SAD transgenes. (B) Expression of *Ophrys* SAD transgenes (and *ACT1* control) in *Arabidopsis ssi2/ssi2* plants, as observed after RT-PCR (T_A=56 °C), with untransformed WT plant and water as controls. (C) Relative amounts of C_{16} and C_{18} FAMEs from leaves of *Arabidopsis* plant lines (or whole plants, without roots, in the case of phenotypic dwarves), error bars indicating the s.e.m. The
sample number $N$ shown refers to pooled samples (typically 5 individuals/sample) analyzed by GC/MS. The levels of FAs in untransformed plants are similar to those listed in (31). The effect of any transgene on $C_{16}$ and $C_{18}$ unsaturated FA levels was tested in a generalized linear model and found to be significant ($p<0.05$) only for the OsSAD2 transgene (indicated by an asterisk; $p=0.012$ and $p=0.004$ for unsaturated $C_{16}$ and $C_{18}$ FAs, respectively). Effects of 35S:OsSAD2 include e.g. slight increases in 16:1 and 18:1 FA levels in SS12/ssi2 and ssi2/ssi2 plants, respectively.
Chapter 4

The genetic basis of pollinator adaptation in a specialized pollination system

Manuscript:

Shuqing Xu\textsuperscript{1,2}, Philipp M. Schlüter\textsuperscript{1}, Ueli Grossniklaus\textsuperscript{3}. Florian P. Schiestl\textsuperscript{1}

1, Institute of Systematic Botany, University of Zürich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland.

2, Institute of Integrative Biology, ETH Zürich, Universitätsstrasse 16, CH-8092 Zurich, Switzerland.

3, Institute of Plant Biology and Plant Science Center, University of Zürich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland.

Keywords: Pollinator adaptation | ecological speciation | sexually deceptive orchids | alkene biosynthesis | acyl-ACP desaturase | gene expression

Author contributions: SX, PMS, UG and FPS designed the research, SX performed experiments, SX, PMS analyzed data, SX, PMS and FPS wrote the paper.
Summary

In plants, pollinator adaptation has been considered to be one of major driving forces for floral diversification and speciation. However, the genetic basis of pollinator adaptation is little understood. *Ophrys* orchids mimic their pollinators’ mating signals and are pollinated by male insects during mating attempts. In this pollination system, chemical mimicry of the pollinators’ pheromones, especially of alkenes with different double bonds positions, plays a key role for specific pollinator attraction. Different alkenes produced in different species have been suggested to be a consequence of pollinator adaptation. In this study, we identify genes, which are likely involved in alkene biosynthesis, encoding stearoyl-acyl carrier protein (ACP) desaturase (SAD) enzymes in three closely related *Ophrys* species, *O. garganica*, *O. sphegodes* and *O. exaltata*. Combing floral odour data and gene expression patterns, two SAD homologues (SAD1/2) showed significant association with the production of (Z)-9- and (Z)-12-alkenes that were abundant in *O. garganica* and *O. sphegodes*, whereas the two other homologues (SAD5/6) were significantly associated with (Z)-7-alkenes that were highly abundant only in *O. exaltata*. Furthermore, the evolutionary pattern of these homologues showed species-specific divergence pattern. The expression patterns of these two genes in F1 hybrids indicated that species-specific expression differences of SAD1/2 are likely due to cis regulation, whereas differences in SAD5/6 expression are likely due to trans regulation. Taken together, we report the evidence on the genetic mechanism of pollinator-mediated divergent selection that drives changes in floral alkene biosynthesis involved in floral isolation and the barrier to gene flow between *Ophrys* species.
Chapter 4

Introduction

Understanding the genetic basis of adaptation is one of the central topics in evolutionary biology. For more than a century, it has been under debate whether adaptations are by caused by a large number of mutations with small phenotypic effects or very few genetic changes with a large phenotypic effect [1-3]. To address this question, it is necessary to study systems in which both the genetic basis and the ecological significance of adaptive traits can be identified.

Pollinator-mediated selection on floral traits has been considered to be a major driving force of floral diversity and speciation in plants [4-9]. Closely related species featuring distinct floral traits, such as floral colour, odour, nectar, spur and display are widely believed to be a consequence of pollinator adaptation [5, 7-11]. Furthermore, pollinator adaptation often conveys floral isolation (that is, pollinator-mediated pre-pollination reproductive isolation) [12, 13], which reduces gene flow among populations, and may thus contribute to the origin of novel species. Therefore, floral traits associated with pollinator adaptation are of special interest for understanding plant speciation and evolution.

Floral scent attracts pollinators in many plants [14-16]. This is especially true in specialized pollination systems, in which floral scent is usually the most important floral signal for attracting specific pollinators [16-22]. Therefore, in such systems, floral scent is often directly under pollinator-mediated selection, since it is linked to plant reproductive success.

*Ophrys* orchids mimic their pollinators’ mating signals and are pollinated by male insects during mating attempts with the flower. This pollination by sexual deception is very specific, and one species of plant typically only attracts males of very few species of insects [23, 24]. Specific pollinator attraction has been reported to be the main reproductive barrier in *Ophrys* [24-26]. The key factor for specific pollinator attraction is the chemical mimicry of
the female’s sex pheromone [11, 19, 21, 27-29], a blend of cuticular hydrocarbons, namely alkanes and alkenes. Among those hydrocarbons, alkenes with different double-bond position are particularly important for species-specific pollinator attraction [11, 27, 29]. This suggests that the genes specifying alkene double-bond positions may be directly associated with pollinator adaptation and reproductive isolation.

In plants, alkenes are rare components of the cuticular wax layer [30-33]. Their double-bonds are likely introduced by desaturases such as stearoyl- or acyl-ACP (acyl carrier protein) desaturases (SAD) [31-34]. During wax biosynthesis, SAD can insert a double-bond into a saturated fatty acid intermediate, such as 16:0-ACP1 or 18:0-ACP, to produce an unsaturated fatty acid such as 16:1ω7-ACP (ω-7 counting from the aliphatic end = double bond at position Δ9 from substituted end) or 18:1ω9-ACP. These unsaturated fatty acid (FA) intermediates may be elongated at ACP end, and lead to the production of (Z)-7 or (Z)-9 alkenes2 [32-37]. Therefore, changes in gene expression and/or enzyme activity of SAD-encoding genes could result in alkenes with different double-bond position among species [36].

In this study, we focused on three closely related sympatric species of Ophrys: O. garganica, O. sphegodes, and O. exaltata, which are reproductively isolated from each other due to the attraction of different pollinators [26]. Among these three species, both O. garganica and O. sphegodes flowers produce a high proportion of 9- and 11/12-alkenes (with differences in carbon chain length), whereas flowers of O. exaltata produce high amounts of 7-alkenes. We have previously shown that in O. sphegodes, 9- and 12-alkene production is likely associated with the geneSAD2, which encodes a functional desaturase [35]. However, the genetic basis of 7-alkene biosynthesis in O. exaltata is still unknown. Furthermore, the

1 C:D denotes a fatty acyl group of C=16 carbons length with D=0 double-bonds.
2 Since all alkenes in this study are cis (Z), therefore only double bond position will be indicated for alkenes in the following text.
mechanisms and reasons for the evolution of different alkene composition among species are unknown. Here we investigate gene expression and evolutionary pattern of the SAD gene family members among different natural populations of our study species, and discuss the possible genetic basis of 7-alkenes as well as the role of pollinator-mediated selection on changing alkene production among species. Specifically, we address the following questions: (1) which desaturase gene/enzyme is responsible for 7-alkene biosynthesis? (2) what is the role of pollinator-mediated selection in the evolution of alkene production? (3) how was different alkenes composition regulated in molecular level among different Ophrys species?

**Results**

*Stearoyl-ACP-desaturase (SAD) homologues in Ophrys.*

We identified six SAD homologues, which we termed SAD1-6, from high-throughput transcriptome sequencing of *O. sphegodes* flowers (Xu, unpublished data). Three of them has been described (SAD1-3) previously [35]. Phylogenetic analysis of plant *Stearoyl-ACP-desaturase* indicated that all six homologues likely evolved via gene duplication after the split of monocots and eudicots, forming three distinct lineages, SAD1/2, SAD3, and SAD4/5/6. The split of SAD1/2 from SAD4/5/6 was more recent than the split from SAD3 (Figure 1).

A codon-based analysis of selection, which compares the rates of synonymous versus non-synonymous mutations, revealed a significant signal indicative of positive selection or relaxed purifying selection (both *p* <0.001) after the split of SAD1 and SAD2 (Figure 1). However, no indication of positive selection was found for any other SAD locus or clade. Significant purifying selection was found on branches of SAD3 and before the split of SAD4/5/6, however, neither positive nor purifying selection was detected on the branches after split of SAD4/5/6 (Figure 1).
Figure 1. Bayesian inference phylogenetic tree of monocot SAD homologues. Numbers indicate posterior probabilities (where >0.5) above branches. SAD homologues found in Ophrys and their assignment to different allele-groups are highlighted. \( d_N/d_S \) ratio for the branches of interest were indicated. Black upward triangle indicates positive selection was detected, while gray downward triangle indicates purifying selection was detected for the branch.
The expression of SAD homologues and alkenes production in different floral tissues and different floral developmental stages.

Among the six SAD homologues that were investigated for tissue- and developmental-stage specific expression, five (SAD1-5) were found to be expressed in the tested individuals. Four of these homologues showed flower-specific expression; only SAD3 was expressed in all tested tissues (leaf, sepals/petals and labella) (Figure 2). The expression levels of two SAD homologues, SAD2 and SAD5, were found to be significantly associated with the presence of the alkenes. SAD2 expression was significantly associated with specific 9- and 12-alkenes across different tissues and floral developmental stages (and Figures 2 & 3). Additionally, we found a significant association between the expression of SAD5 and 7-alkenes in O. exaltata. Floral tissues of O. exaltata, which produce high levels of 7-alkenes, expressed SAD5 highly, whereas SAD5 was not expressed in leaf tissue, which produces hardly any alkenes. Furthermore, expression of SAD5 was significantly correlated ($p < 0.001$) with 7-alkene amounts in different floral developmental stages (Figure 3).
Figure 2. Gene expression and floral odour in different plant tissues. (A), relative amount (as proportion of hydrocarbons) of different alkenes in floral labella of *O. sphegodes* and *O. exaltata*; (B), relative amount of alkenes in floral sepals/petals of *O. sphegodes* and *O. exaltata*; (C), relative amount of different alkenes in leaf tissue of *O. sphegodes* and *O. exaltata*; (D), relative gene expression of all five *SAD* homologues in floral labella of *O. sphegodes* and *O. exaltata*; (E), relative gene expression of all five *SAD* homologues in floral sepals/petals of *O. sphegodes* and *O. exaltata*; (F), relative gene expression of all five *SAD* homologues in leaf tissue of *O. sphegodes* and *O. exaltata*. Error bars in all figures show standard error. Asterisks indicate significant differences between species.
Figure 3. Regression of normalized desaturase expression with alkenes. Relative amount of alkenes were used after \(f(x) = \arcsin x^{0.5}\) transformation in time course dataset (A-C). Absolute amount of alkenes (in µg) were used after \(f(x) = \ln(x + 0.01)\) transformation in population dataset (D-F). For both datasets, normalized SAD expression were used after \(f(x) = x^{0.5}\) transformation. Adjusted R\(^2\) is indicated in each figure. A, Correlation of all 7-alkenes (C21-C29) relative amount with SAD5 expression in O. exaltata; B, Correlation of 9-C27 + 9-C29 relative amount with SAD2 expression in O. sphegodes; C, Correlation of all 12-alkenes (C25-C29) relative amount with SAD2 expression in O. sphegodes; D, Correlation of all 7-alkenes (C21-C29) absolute amount with SAD5 expression among all three species; E, Correlation of 9-C27 + 9-C29 alkenes absolute amount with SAD2 expression among all three species; F, Correlation of all 12-alkenes (C25-C29) absolute amount with SAD2 expression among all three species.
Allelic variation of SAD homologues among species and populations.

Among our study species, four SAD homologues (SAD1/2/5/6) showed species-specific patterns of allelic variation, whereas SAD3 and SAD4 did not (Figure 4 and Figures S1-S5). Two allele groups were found for SAD1, one (SAD1-A) was shared between O. gagarica and O. sphegodes, one (SAD1-B) was shared among all three species (Table 1 and Figure S1). In SAD2, four allele groups were found, SAD2-A and SAD2-B were shared between O. gagarica and O. sphegodes, whereas SAD2-C and SAD2-D were found only in two O. exaltata populations (MDL and CAP) respectively (Figure 4 and Figure S2). Interestingly, at the beginning of SAD2-D exon, we found a repetitive sequence insertion; two allele groups were found for each, SAD5 and SAD6. Both SAD5 allele groups were mostly found in O. exaltata, but not in the other two species. However, for SAD6, one (SAD6-A) was only found in one O. exaltata population (SPF), while the other one (SAD6-B) was shared among all three species (Figure S5). This pattern of allelic variation in our study species and populations was also confirmed via allele group-specific RT-PCR for all individuals (Table 1).

Among all SAD allele groups, two (SAD2-D and SAD6-B) contained significantly higher proportions of alleles with frame-shift mutations (including the insertion of repetitive sequence) and/or premature stop codons than expected by chance (Table 2), indicating that most alleles from these groups are probably non-functional. Furthermore, biochemical activity assays suggest that SAD1-B alleles may not encode functional desaturases, and that SAD2-A does contain functional alleles [35].
**Table 1.** Allelic gene expression among species. The number in each cell refers to the number of individuals that showed expression of a certain allele group. N refers to the total number of individuals assessed. E, G and S refer to *O. exaltata*, *O. garganica* and *O. sphegodes*, respectively. E × S refers to F1 hybrids among *O. sphegodes* and *O. exaltata* (identified by Xu et al., accepted, according to molecular maker data; direction of the cross unknown).

<table>
<thead>
<tr>
<th>Species</th>
<th>N</th>
<th>SAD 1-A</th>
<th>SAD 1-B</th>
<th>SAD 2-A/B</th>
<th>SAD 2-C</th>
<th>SAD 2-D</th>
<th>SAD 3</th>
<th>SAD 4</th>
<th>SAD 5-A</th>
<th>SAD 5-B</th>
<th>SAD 6-A</th>
<th>SAD 6-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>16</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>G</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>14</td>
<td>9</td>
<td>14</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>14</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E × S</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2. Number of alleles that frame-shift or stop codon mutations were found for each allele group. Numbers in each cell refer to the number of sequences obtained from each allele groups. All sequences were classified into two groups: putatively functional and non-functional, based on whether frame-shifts/stop codons were observed in sequences or not. Asterisks indicate the proportion of non-functional sequences was significantly higher than average, by using Fisher’s exact test \([p\text{ cut-off value 0.05}]\).

<table>
<thead>
<tr>
<th></th>
<th>SAD 1-A</th>
<th>SAD 1-B</th>
<th>SAD 2-A</th>
<th>SAD 2-B</th>
<th>SAD 2-C *</th>
<th>SAD 2-D *</th>
<th>SAD 3</th>
<th>SAD 4</th>
<th>SAD 5-A</th>
<th>SAD 5-B</th>
<th>SAD 6-A</th>
<th>SAD 6-B *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional</td>
<td>21</td>
<td>77</td>
<td>50</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>64</td>
<td>80</td>
<td>16</td>
<td>4</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>Non-functional</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>81</td>
<td>51</td>
<td>27</td>
<td>3</td>
<td>6</td>
<td>67</td>
<td>83</td>
<td>18</td>
<td>5</td>
<td>9</td>
<td>31</td>
</tr>
</tbody>
</table>
Figure 4. Allelic distribution of SAD gene among three species. Red colour refers to O. exaltata, blue colour refers to O. sphegodes and green colour refers to O. garganica. N indicates number of sequences analysed. An asterisk indicates that an allele group that is putatively non-functional.
Allelic gene expression of SAD homologues among species.

Five SAD copies (all except SAD3) showed a species-specific allelic divergence in gene expression (Figure S6). Among the two allele groups of SAD1, one (SAD1-A) was highly expressed in O. sphegodes and O. gargarica, whereas the other one (SAD1-B) was highly expressed in O. sphegodes and O. exaltata. Among the SAD2 allele groups, SAD2-A/B were highly expressed in both O. sphegodes and O. gargarica, whereas the expression of SAD2-C/D was low in all study species. The expression of SAD4 was found to be high in O. gargarica and O. exaltata, but low in O. sphegodes. All allele groups of SAD5 and SAD6 were highly expressed only in O. exaltata (Figure S7).

The allelic expression of SAD homologues in the F1 hybrids of O. sphegodes and O. exaltata showed similar pattern to O. sphegodes. Two SAD1 allele groups (SAD1-A/B), two alleles of SAD2 (SAD2-A/B) that was most likely inherited from O. sphegodes (Figure S2), SAD3 and SAD4 were found highly expressed in these F1 hybrids, whereas none of SAD5/6 alleles was expressed.

Statistical analysis showed strong correlation between expression of specific SAD allele groups and alkenes production among species. The expression level of SAD1-A was found to be significantly ($p < 0.001$) positively correlated with (Z)-9-C29, (Z)-12-C27, (Z)-12-C29 alkenes (Figure S7); SAD2-A was found to be significantly ($p < 0.001$) positive correlated with (Z)-9-C27, (Z)-9-C29, (Z)-12-C25, (Z)-12-C27, (Z)-12-C29; while SAD5-A/B were significantly correlated with all 7-alkenes (Figure 3 D, E and F).

Discussion

Differences in alkene production are associated with SAD gene expression.

Our data indicated that, the alkene biosynthesis is associated with the expression of certain SAD homologues in Ophrys. SAD2 appears to be responsible for introducing a
double-bond at position 9 of 18:0-ACP and position 4 of 16:0-ACP, producing 18:1\textsuperscript{ω-9}-ACP and 16:1\textsuperscript{ω-12}-ACP, which would eventually lead to the production of 9- and 12-alkenes (Schlüter et al. submitted). The significant correlation of SAD2 expression level with the amount of 9- and 12-alkenes among individuals of different populations lends further support to this hypothesis. Although functional assays suggested that SAD1 (from allele group SAD1-B) is not catalytically active (Schlüter et al. submitted), this may not be true for other alleles of SAD1 (SAD1-A) found in both O. sphegodes and O. garganica, since SAD1-A and SAD1-B differ by 14% at the amino acid sequence level. Furthermore, SAD1-A was significantly correlated with certain 9- and 12-alkenes. This indicates that, like SAD2, SAD1 might also contribute to 9- and/or 12-alkene biosynthesis in natural population.

Tissue-specific gene expression of SAD5 and the significant correlation between its expression and amount of 7-alkenes among species and populations (Figure 3) suggested that SAD5 is likely involved in 7-alkenes synthesis. Although SAD6-A was highly expressed in one population (population SPF), the higher than 95% amino acid homology between SAD6-A and SAD5 indicated that both might have the same function. We hypothesize that SAD5/6 may introduce a double-bond at position 11 into 18:0-ACP, producing 18:1\textsuperscript{ω-7}-ACP, or at carbon position 9 of 16:0-ACP, producing 16:1\textsuperscript{ω-7}-ACP. Further functional studies are required to test this hypothesis. These unsaturated metabolites could then be elongated from the ACP end, to produce 7-alkenes. Therefore, changes in the expression of SAD5/6 could directly lead to different amounts of 7-alkenes in different Ophrys species.

**SAD homologues are likely under pollinator-mediated selection**

Our data suggests that SAD homologues evolve under pollinator-mediated selection. The expression of SAD1/2 was high in both O. garganica and O. sphegodes, but, was very low in most O. exaltata individuals. The very few individuals of O. exaltata that highly
expressed $SAD2$ either had a repetitive sequence inserted in its first exon ($SAD2$-$D$ allele group in population MDL), or had amino acid substitutions located on the surface of the protein ($SAD2$-$C$ allele group), which have been suggested may reduce the function of $SAD2$ (Schlüter et al. submitted). This indicates that disruptive selection might act on $SAD2$ (in terms of gene expression or overall enzymatic activity) to maintain (in $O.\ sphegodes$ and $O.\ garganica$) or reduce (in $O.\ exaltata$) the production of 9 and/or 12-alkenes in $Ophrys$ floral odour. Evidence of positive selection on $SAD1/2$ detected by $d_{SN}/d_{S}$ ratio is consistent with this hypothesis (Figure 1). Furthermore, disruptive selection on 9/12-alkenes among $O.\ sphegodes$ and $O.\ exaltata$ was suggested in a previous study (Mant et al. 2005a). While 9/12-alkenes act as the main attractants of $O.\ sphegodes$ to its pollinator (Schiestl et al. 1999), these compounds actually reduce the attractiveness of the odour bouquet of $O.\ exaltata$ to its pollinator, $Colletes\ cunicularius$ (Mant et al. 2005a). Therefore, pollinator imposed selection in $O.\ exaltata$ may reduce the proportion of 9/12-alkenes in the total bouquet, which can be achieved by changing the expression or enzymatic activity of $SAD1/2$.

However, for $SAD5/6$, the opposite pattern was observed. $SAD5/6$ were highly expressed in $O.\ exaltata$, but hardly expressed in $O.\ sphegodes$ and $O.\ garganica$. A significantly higher frequency of frame-shifts or premature stop codons was found in $SAD6$ of $O.\ sphegodes$ and $O.\ garganica$, indicating that $SAD6$ alleles in these two species may be released from purifying selection such that loss-of-function mutations can accumulate. Neither positive selection, nor purifying selection was detected on $SAD5/6$ suggesting that (Figure 1), if there are pollinator-mediated selection on 7-alkenes in $O.\ sphegodes$, it may act on the expression level of $SAD5/6$ in $O.\ sphegodes$ and $O.\ garganica$. Indeed, selection against 7-alkenes in $O.\ sphegodes$ was suggested by behavioural test with its pollinators, which indicated that adding 7-alkenes in $O.\ sphegodes$ floral scent again reduces attraction of its pollinator (Schiestl, unpublished data). Therefore, it is likely that, pollinator-mediated
selection in *O. sphegodes* may act to suppress the expression of *SAD5/6* to suppress their expression and thereby reduce 7-alkene production. In contrast to these genes likely associated with alkene production, *SAD3* and *SAD4*, which were not significantly correlated with alkene occurrence, showed no divergence of sequences or expression levels among species (Figure 4 and Figure S7).

**Regulatory mechanisms controlling alkene differences among Ophrys species**

The changes in 7- and 9/12-alkene production due to changes in the expression of *SAD* genes might be explained by the action of cis- or trans- acting elements. The expression of *SAD1/2*, which is associated with 9/12-alkene production, was different among *O. exaltata* (weak expression) and *O. sphegodes* (strong expression) (Figure S6). However, two putative F₁ hybrids only expressed the alleles expected to be inherited from *O. sphegodes*, but not *O. exaltata* (Table 1 and Figure S6). This indicated that down-regulation of *SAD1/2* expression in *O. exaltata* might be due to changes in a cis regulatory element (such as the promoter). In contrast, although differences in expression of *SAD5/6*, which is associated with 7-alkene production, were found between *O. sphegodes* and *O. exaltata*, the putative F₁ hybrids did not express either allele from the parental species (Table 1 and Figure S6). This suppression of expression of *SAD5/6* in F₁ hybrids indicated that - while additional cis regulatory changes cannot be ruled out - a trans-acting factor is likely involved in the different *SAD5/6* gene expression among species. This suggests the presence of a dominant suppressor of *SAD5/6* expression in *O. sphegodes* (e.g., a transcriptional repressor or a miRNA reducing *SAD5/6* mRNA levels) that is absent or inactive in *O. exaltata*.

In conclusion, our data suggested that pollinator adaptation in *Ophrys* might primarily be due to changes in *SAD1/2* and *SAD5/6*, in terms of their expression and potentially also in terms of the function of their gene products. This indicates that pollinator adaptation in
specialized pollination system might be due to pollinator-mediated selection effecting few changes in the genome, with a large phenotypic effect. Moreover, it is likely that both cis and trans regulation of gene expression contribute to this process.

**Experimental Procedures**

**Plant material**

Samples of *Ophrys sphegodes* MILLER (n=12); *O. exaltata* subsp. *archipelagi* (GÖLZ & H.R. REINHARD) DEL PRETE (n=13), and *O. garganica* NELSON EX O. and E. DANESCH (n=5) were collected in southern Italy (Table S1), at the same locations as described in Xu et al. (accepted): populations “Capoiale” (CAP); “Foce Garigliano” (FCG), and “Marina di Lesina” (MDL), where species co-occur and co-flower. Two F₁ hybrids between *O. exaltata* and *O. sphegodes* from the FCG population were previously identified based on AFLP markers (Xu et al. accepted). In addition, three *O. exaltata* samples from population “San Pietro in Fine” (SPF: 41°25' N, 13°58' E) and two samples from population “Campozillone” (CPZ: 41°23' N, 14°00' E) were included in this study. Species were first identified based on floral morphology, according to the criteria described by Mant et al. (2005a), and species assignments were confirmed by floral scent analysis (see below). For each plant individual, one labellum of an unpollinated flower was used for floral odour extraction as described in Xu et al. (accepted), and the same labellum tissue was immediately flash frozen in liquid nitrogen, and stored at -80 °C until RNA extraction. Floral extracts were stored at -20 °C until gas chromatographic analysis. For developmental stage-specific and tissue-specific analysis of hydrocarbon and gene expression, five *O. exaltata* and six *O. sphegodes* individuals grown in a greenhouse at the Botanic Garden of the University of Zürich were used, as described in Schlüter et al. (submitted).
**GC and GC/MS analysis**

Before GC analysis, 300 ng of n-octadecane (C18) were added into each floral extract as an internal standard. GC and GC/MS analysis, identification and quantification of compounds were performed as described by Mant et al. (2005b) with modifications (Xu et al. accepted for publication). One µl of each sample was injected into an Agilent 6890 GC at 50 °C, followed by opening of the split valve and heating to 300 °C at rate of 4°C/min. Discrimination and identification of alkenes with different double bonds was achieved by comparison with synthetic reference standards (Mant et al. 2005a). Discrimination of 11 and 12-alkenes was not possible based on retention times with the GC parameters used, however, previous study that determined double bond position of alkenes has shown that 12-alkenes is the predominant isomer in *O. sphegodes* and *O. exaltata* (Erdmann 1996; Schulz 2005). The absolute amount of alkenes and alkanes with a carbon chain length from 21 to 29 were calculated based on the internal standard. For tissue/stage-specific samples, the relative amount of alkenes (proportions) was calculated since floral size varies among different developmental stages and the use of comparable amounts of tissue could not be ensured.

**RNA extraction, cDNA synthesis, RACE and RT-PCR**

Total RNA was extracted using Trizol (Invitrogen) following the manufacturer’s instructions, and RNA quality and quantity were assessed by agarose gel electrophoresis and spectrophotometry on a NanoDrop ND-1000 (Witec AG Littau). One µg RNA was treated with one unit DNase I (Invitrogen) for 1 h at 37 °C in the presence of 20 units RiboLock™ RNAase inhibitor (Fermentas). Thereafter, RevertAid M-MuLV H Reverse Transcriptase (Fermentas) was used to synthesize first strand cDNA with oligo-dT primer following the manufacturer’s instructions. After first strand cDNA synthesis, the reaction was treated with one unit RNase H (Fermentas) for 20 minutes at 37°C.
To obtain the full-length coding sequence of candidate genes, 5’ rapid amplification of cDNA ends (RACE) was performed as described by Scotto-Lavino et al. (2006b), with minor modifications (Schlüter et al. submitted), and 3’ RACE was performed using classic RACE (Scotto-Lavino et al. 2006a). Gene-specific primers used for RACE are listed in supplemental Table S2. Advantage® GC 2 DNA Polymerase DNA Polymerase (Clontech) was used for RACE PCR with a touchdown program: 96°C 15 s; 3 cycles of [94°C 20 s, 68°C 3 min 30 s]; 7 cycles of [94°C 20 s, 67°C (1°C decrease per cycle) 30 s, 68°C 3 min 30 s]; 30 cycles of [94°C 25 s, 55°C 30 s, 68°C 3 min 30 s]; final extension at 68°C for 10 min. The amplified fragments were cloned into pDRIVE vector (Qiagen, Switzerland), following the provided protocol.

Desaturase homologues were amplified for all cDNA samples, using gene specific primers containing attB adapter sequence (Table S2). RT-PCR was performed in 15 µl reaction volume containing 15 ng cDNA template as follows: 95 °C 3 min; 33 cycles of [95 °C 30 s, 58-60 °C 30 s (see Table S2 for different annealing temperatures of each genes), 72 °C 1 min 30 s]; final extension at 72 °C for 10 min, using REDTaq® ReadyMix™ (Sigma, Switzerland) mix supplemented with 0.6 units Pfu polymerase (Promega, Switzerland). Three µl PCR product were loaded on agarose gel to confirm amplification.

**Cloning and sequencing**

The amplified PCR products of each population of each species were pooled together, thereafter, purified with Wizard® SV Gel and PCR Clean-Up kit (Promega Switzerland), and recombined into Gateway cloning vector pDONR221 (Invitrogen, Switzerland) using the manufacturers’ protocol. Competent *E. coli* One Shot® TOP10 cells (Invitrogen, Switzerland) were used for chemical transformation. The transformed *E. coli* cells were grown on Luria-
Bertani (LB)-agar plates with kanamycin (50 µg/ml). Colonies were picked and screened for positive inserts by PCR using M13 universal primers. In order to recover all possible alleles, number of clones that was at least three times of all possible alleles were picked and screened by PCR, for each cloning library. For example, since all study species are diploids (Xu et al., accepted for publication), when 5 PCR products were pooled for one library, the maximum number of alleles was 10, and therefore we sequenced at least 30 clones from this library. Colony PCR products were purified using the method of Werle et al. (1994), and the purified products sequenced using BigDye 3.1 with both forward and reverse M13 primers on an ABI 3130XL Genetic Analyzer (Applied Biosystems), following the manufacturer’s recommendations.

**Sequence analysis and allele group assignment.**

The forward and reverse sequences of each clone were assembled and manually edited in SeqMan v7.1.0 (lasergene DNASTar). For each SAD homologue, the assembled sequences of each clone was aligned by using Clustal W (Thompson et al. 1994). Sequences with less than two nucleotide differences were assumed to be the same allele with PCR or sequencing error, and were merged into one consensus sequence. The consensus sequences and all singleton sequences, which differed by more than two nucleotides, were used for assignment to allele groups. To do so, a dendrogram was constructed for each SAD homologue in MEGA 4.0 (Tamura et al. 2007), using UPGMA and a pair-wise distance, with pair-wise deletion of gaps, and a homogeneous substitution pattern among lineages and sites. Bootstrap analysis was conducted using 1000 pseudo-replicates. Allele groups were assigned based on tree topology (Figures S1-S5).
**Measuring gene expression by semi-quantitative PCR**

Gene expression of desaturase genes was assessed by semi-quantitative RT-PCR with allele-specific primers (Table S1). Each PCR was performed as: 95 °C 3 min; 29 cycles of [95 °C 30 s, 58-60 °C 30 s (see Table S1 for different primer annealing temperatures), 72 °C 1 min 30 s]; final extension at 72 °C for 10 min using REDTaq® ReadyMix™ (Sigma). PCR was performed in 10 µl reaction volume with cDNA from 12 ng total RNA as a template. For all RT-PCRs, the putative *Ophrys* housekeeping gene *G3PDH* (Schlüter et al. submitted) was used as control. Five µl of each PCR product were loaded on 0.8% agarose gel, recorded and quantified using ImageJ (1.42q) (Abramoff et al. 2004) as described in Schlüter et al. (submitted).

**Phylogeny reconstruction and selection analysis.**

All monocot *SAD* sequences were taken from the plant *SAD* homologue dataset of (Schlüter et al. submitted). All sequences were re-aligned based on amino acid sequence using Muscle 3.8.31(Edgar 2004). Poorly alignable sequences and regions were excluded for downstream analysis. GTR+I+Γ was estimated to be the best substitution model by MrModeltest 2.3 (Nylander 2004), and used for phylogenetic analysis in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), discarding results prior to apparent convergence of analysis chain (burn-in 13M out of 40M generations). Branch lengths of the resulting consensus tree were optimized with BaseML for use with a codon-based model of sequence evolution, and used as input for CodeML, both from the PAML 4.4 (Yang 2007) package. Likelihood ratio tests were performed to determine the significance of different CodeML models in Microsoft Excel.
Statistical analysis

The significance of different amounts of floral odour as well as gene expression among species was assessed by ANOVA after normality testing of the data distribution by the Shapiro test (Royston 1982). The association between floral scent and gene expression was assessed using a generalized linear model (GLM) and a linear mixed-effect model (LME) with population as random factor. These models were simplified by stepwise removal of factors using the stepAIC method (Venables and Ripley 2002). For the tissue/stage-specific dataset, the relative amount of each floral scent compound was used (arcsine square-root transformed) as described in (Schlüter et al. submitted) since the size of floral labella varies in different developmental stages. For population dataset, the absolute amount of each floral scent compound was used. The significance of presence of non-functional alleles in different allele groups was tested using the Fisher’s exact test. All statistical analysis were performed in R 2.11.0 (R Development Core Team 2010).

Acknowledgements

We thank Salvatore Cozzolino for his assistance with sample collections, Ed Connor for helping on GC analysis, Jurriaan de Vos for helping on phylegetic analysis, Liu Zhongjian for whole transcriptome sequencing of Ophrys flower. We also thank Alex Widmer, Thomas Staedler, and Stephen Schauer for stimulating discussions and comments on the manuscript. This work was supported by ETH Zurich (TH0206-2 to FPS) and Austrian Science Fund (FWF; Schrödinger fellowship J2678-B16 to PMS).
References


52. R Development Core Team (2010). R: A Language and Environment for Statistical Computing. 2.11.0 Edition. (Vienna, Austria.).
Supplemental information

Table S1. Plant samples collected in this study. Numbers in cells refer to number of individuals. Hybrid refers to F1 hybrids between *O. sphegodes* and *O. exaltata* as assigned based on AFLP data in previous study (Xu et al. accepted for publication). Location information of each population is given in the main text.

<table>
<thead>
<tr>
<th>Species / Population</th>
<th>CAP</th>
<th>FCG</th>
<th>MDL</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>O. exaltata</em></td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td><em>O. garganica</em></td>
<td>5</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>O. sphegodes</em></td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>/</td>
</tr>
<tr>
<td>Hybrid</td>
<td>/</td>
<td>2</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
Table S2. Oligonucleotides used in this study. For primers compatible with Gateway (Invitrogen) cloning, the full attB sites were introduced as described in Invitrogen’s manuals. 

\( T_A \) refers to the annealing temperature used for PCR reactions.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Gene</th>
<th>Sequences (5'-&gt;3')</th>
<th>( T_A ) (°C)</th>
<th>Experiment</th>
<th>Purpose/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>SADN5-538F</td>
<td>SAD4/5/6</td>
<td>GGTGGACTGCCGAGAGAGACC</td>
<td>55</td>
<td>3'RACE 1\textsuperscript{st} PCR</td>
<td></td>
</tr>
<tr>
<td>SADN5-575F</td>
<td>SAD4/5/6</td>
<td>TGCTGAACGATCTGATCTTC</td>
<td>55</td>
<td>3'RACE 2\textsuperscript{nd} PCR</td>
<td></td>
</tr>
<tr>
<td>SADN5-698R</td>
<td>SAD4/5/6</td>
<td>CCGCTCTTGGAGGAAGGTGTAG</td>
<td>55</td>
<td>5'RACE 1\textsuperscript{st} PCR</td>
<td></td>
</tr>
<tr>
<td>SADN5-678R</td>
<td>SAD4/5/6</td>
<td>AGGTGATAGAATCCATGAG</td>
<td>55</td>
<td>5'RACE 2\textsuperscript{nd} PCR</td>
<td></td>
</tr>
<tr>
<td>FAttB1-SAD1f-fs*</td>
<td>SAD1</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+T</td>
<td>58</td>
<td>RT-PCR</td>
<td>* refers Phosphorothioate modification</td>
</tr>
<tr>
<td>FAttB2-SAD1r-fs</td>
<td>SAD1</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttB1-SAD2f-fs*</td>
<td>SAD2</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td>* refers Phosphorothioate modification</td>
</tr>
<tr>
<td>FAttB2-SAD2r-fs</td>
<td>SAD2</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb1-SAD3f-fs</td>
<td>SAD3</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb2-SAD3r-fs</td>
<td>SAD3</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb1-SAD4f-fs</td>
<td>SAD4</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb2-SAD4r-fs</td>
<td>SAD4</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb1-SAD5-141F</td>
<td>SAD5</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>FAttb2-SAD5r-fs</td>
<td>SAD5</td>
<td>GGGGACACACCTTTTGATACAAGAAA GCGGGTCACGTACGCCGACCTTCACC TGGCACTAC+G</td>
<td>58</td>
<td>RT-PCR</td>
<td></td>
</tr>
<tr>
<td>SAD1-AlleleGroup pA_F</td>
<td>SAD1</td>
<td>CAGCAGTATCATCTATCATCAAC ACT</td>
<td>60</td>
<td>Allelic gene expression</td>
<td>SAD1AlleleGroup pA_F+SAD1AlleleGroup pA_R: SAD1-A allele</td>
</tr>
<tr>
<td>SAD1-AlleleGroup pB_F</td>
<td>SAD1</td>
<td>AACAGCAGTAACACCACTCAAC ACT</td>
<td>60</td>
<td>Allelic gene expression</td>
<td>SAD1AlleleGroup pB_F+SAD1AlleleGroup pB_R: SAD1-B allele</td>
</tr>
<tr>
<td>Gene</td>
<td>Allele Group</td>
<td>Primer Sequence</td>
<td>Allelic Gene Expression</td>
<td>Allele</td>
<td>Comment</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>-----------------</td>
<td>-------------------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>SAD1</td>
<td>A_R</td>
<td>CATCTCGCCGTAGTCTCTGACA</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD1</td>
<td>A_F</td>
<td>CATCTCGCTGTAGTCTCGACA</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD2</td>
<td>A_F</td>
<td>AAAAAGCAGGCTAAATGGAACTTCACCTTGCACTACG</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD2</td>
<td>B_F</td>
<td>CCTTGCACTACGTGTTATGCTTA</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD2</td>
<td>C_R</td>
<td>TAGATGATCCACACAAGGCTGT</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD5</td>
<td>A_F</td>
<td>TGATAAGTTCTACGAGCAAACGA</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD5</td>
<td>A_F</td>
<td>TGATAAGTTCTACGAGCAAACGA</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD5</td>
<td>B_F</td>
<td>GATTAAGGAGCTTCGGAAGCG</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD5</td>
<td>C_F</td>
<td>AGGAGCTTCGAAAGGCTAAAG</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>SAD5</td>
<td>C_F</td>
<td>ATCAAGCATTGCTAGTGCGGAGAG</td>
<td>Allelic gene expression</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>
Figure S1. Phylogenetic tree of SADI. Symbol colour refers to the species that sequence come from, while symbol shape indicates the source population. Blue indicates O. sphegodes; red indicates O. exaltata, green indicate O. garganica, pink indicates F₁ hybrids of O. sphegodes and O. exaltata, whereas black refers to consensus sequences from all three species. Numbers on branches were bootstrap values. An asterisk (*) indicates stop codon or frame-shift mutations occurred in the sequence. Sequences included for PAML analysis were also indicated by “#”. 
**Figure S2.** Phylogenetic tree of *SAD2*. Symbol colour refers to the species that sequence come from, while symbol shape refers to population information. Blue indicates *O. sphegodes*; red indicates *O. exaltata*, green indicate *O. garganica*, pink indicates F₁ hybrids of *O. sphegodes* and *O. exaltata*. Numbers on branches were bootstrap values. An asterisk (*) indicates stop codon or frame-shift mutations occurred in the sequence. Sequences included for PAML analysis were also indicated by “#”.
Figure S3. Phylogenetic tree of SAD3. Symbol colour refers to the species that sequence come from, while symbol shape refers to population information. Blue indicates *O. sphegodes*; red indicates *O. exaltata*, green indicate *O. garganica*, whereas black refers to consensus sequences from all three species. Numbers on branches were bootstrap values. An asterisk (*) indicates stop codon or frame-shift mutations occurred in the sequence. Sequences included for PAML analysis were also indicated by “#”. 
Figure S4. Phylogenetic tree of SAD4. Symbol colour refers to the species that sequence come from, while symbol shape refers to population information. Blue indicates O. sphegodes; red indicates O. exaltata, green indicate O. garkanica, pink indicates F1 hybrids of O. sphegodes and O. exaltata, whereas black refers to consensus sequences from all three species. Numbers on branches were bootstrap values. An asterisk (*) indicates stop codon or frame-shift mutations occurred in the sequence. Sequences included for PAML analysis were also indicated by “#”. 
**Figure S5.** Phylogenetic tree of SAD5/6. Symbol color refers to the species that sequence come from, while symbol shape refers to population information. Blue indicates *O. sphegodes*; red indicates *O. exaltata*, green indicate *O. garkanica*, whereas black refers to consensus sequences from all three species. Numbers on branches were bootstrap values. An asterisk (*) indicates stop codon or frame-shift mutations occurred in the sequence. Sequences included for PAML analysis were also indicated by “#”.

---

**SAD5/6**

- ▲ *O. sphegodes* FCG
- ▲ *O. sphegodes* CAP
- ▲ *O. exaltata* CAP
- ○ *O. exaltata* MDL
- ▲ *O. exaltata* SPF
- ▲ *O. garkanica* CAP
- ■ Consensus sequence of *O. exaltata* from different populations
- ■ Consensus sequence of *O. exaltata* from different populations
- ■ Consensus sequence from *O. exaltata*, *O. garkanica* and *O. sphegodes*

* Stop codon or frame shift mutations were found

# Sequences used for PAML analysis
Figure S6

Allelic gene expression of all six SAD homologues in the study species. The height of each bar indicates mean normalized expression of each allele, and error bars indicate standard error. Letters on each bar indicate statistical significance comparing among species within each allele group.
Figure S7. Statistical summary of association between normalized allelic gene expressions of \textit{SAD1-SAD6} versus natural logarithms transformed absolute amount of each alkenes. GLM and LME indicate different statistical methods: Generalized Linear Model and Linear Mixed-Effects model, respectively.
Acknowledgments

First of all, I would like to thank my supervisor Prof. Florian Schiestl and Prof. Alex Widmer for giving me this opportunity to work on this Ph.D thesis. Their scientific assistances, knowledge, experiences and advices allowed me to complete this thesis. They introduced me into the research field of speciation and fascinating orchids world.

I am indebted to Dr. Philipp Schlüter who guided me during my Ph.D. In last four years, his patience, knowledge, assistances and kindness made it possible for me to finish my thesis. During the last four years, I enjoyed working with him, not only in the lab, but also in the field. I also enjoyed countless discussions with him during the daily coffee break time, when we developed many new ideas about our projects, and some turned into real data in the thesis.

Many thanks to Prof. Salvatore Cozzolino for his help in the filed and comments on the manuscript. And I also thank him agree to be one of my co-examiner. Thanks to Prof. Ueli Grossniklaus for his suggestions on experiments and comments on manuscript. Thanks to Dr. Giovanni Scopece, Hendrik Breitkopf, Karin Gross who helped me a lot in Southern Italy during fieldwork, and their contributions to the manuscript. I thank Dr. Ed Connor and Mimi Sun for their help on language improvements for part of my thesis and the manuscript.

I thank Prof. Scott Hodges, who contributed his time and efforts on reviewing the thesis and being the external examiner.

I want to thank all colleagues and friends from the Institute of Systematic Botany University of Zürich and Institute of Integrative Biology ETH Zürich: Stefan Abrahamczyk, Dr. Daniele Cicuzza, Prof. Elena Conti, Prof. Peter Endress, Daniel Gerasi, Alok Gupta, Serge Haemmerli, Dr. Colin Hughes, Dr. Aelys Humphreys, Dr. Barbara Keller, PD Dr. Michael Kessler, Birgit Lemcke, Anita Lendel, Prof. Peter Linker, Sara Manafzadeh, Dr. Aria Minder, Dr. Maria Mocccia, Paul Page, Dr. Joelle Rahmé, Manuel Rey, John Spillmann, Dr. Yannick Städler, Lukas Taxböck, Dr. Nicolas Vereecken, Jurriaan M. de Vos. They created such a friendly atmosphere in the institute, and it was always possible to find a hand offered when I needed help. I thank Katharina Rentsch and Corinne Burlet for their help on dealing with administration issues with ETH Zürich and University of Zürich.

Thanks to Prof. Luo Yibo from Institute of Botany, Chinese Academy of Science and Prof. Liu Zhongjian from National Orchid Conservation and Research Center of China for their contributions for sequencing whole transcriptome of Ophrys flower.

I thank all friends I met in switzerland who helped me in personal during my Ph.D. I would like to thank Meret Huber for her friendship, love and encouragements.

I am very grateful for funding by Swiss Federal Institute of Technology (ETH Zürich), University of Zürich, and Swiss Academy of Sciences (SCNAT).

Last but not least, I would like to thank my parents and my sister for their ongoing priceless support and love.
Curriculum vitae

First Name: Shuqing        Last Name: Xu
Date of Birth: 1983 April 12        Nationality: CHINA

Education

2007 - 2011  Ph.d student in Institute of Integrative Biology, ETH Zurich, Switzerland. Supervised by Prof. A. Widmer and Prof. F. P. Schiestl.


2002 – 2004  Undergraduate student in Department of Animal Science, China Agricultural University. Obtained Bachelor degree in Animal Genetics

2000 – 2002  Undergraduate student in Department of Applied Physics, China Agricultural University.

Research and project experiences

2007 – 2011  The ecological and genetic basis of floral isolation in a specialized pollination system.

2005 – 2007  Comparative genomics on plants, animals and virus.
Publications:

- Xu S, et al. Floral isolation is the main reproductive barrier among closely related sexually deceptive orchids. *Evolution*. Accepted for publication.


