universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„The use of phylogenomic tools to investigate diploid and polyploid evolution in *Dactylorhiza* and other orchids“

verfasst von / submitted by

Marie Kristine Brandrud

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie/Biology
Betreut von / Supervisor:	Ass.-Prof. Dipl.-Ing. Dr. Ovidiu Paun

ACKNOWLEDGMENTS

I would like to thank Ass. Prof. Ovidiu Paun for allowing me to do this PhD here in Vienna, including me in his research group and being my supervisor these past 3 and a half years. I would also like to thank all our collaborators, and especially Prof. Mikael Hedrén, for sharing all his knowledge and time.

Colleagues has come and gone during my time here, but I have been very lucky to always have a good group of PhD students around. I could never have imagined going through this PhD without my steady and silly office colleague Thomas Wolfe, that started his PhD at the same time as me and has been going through the same phases in parallel. Our project working on "The time capsule analysis of ancient variation stuck in polyploids" is now coming to an end. Our Post-Doc, when I started, Emiliano Trucchi as well as Lucia Castello and Sonia Mentrida was good support in the beginning to settle in in the lab in Vienna and I am very thankful for the support and coffee that always would be found in Jacky Heckenhauer's office. Jacky is also especially acknowledged for helping to translate the abstract for this thesis. Furthermore, it has been wonderful to share experiences, valuable discussions, lunch and excellent coffee with Mimmi Eriksson, Aglaia Szukala, Gil Yardeni, Marylaure de La Harpe, Huijing Shang, David Field, Juliane Baar, Jamie McCann, Camille Criste, Rita Tonin, Nico Criscuolo and Jacky Hess.

Ásgeir Trausti Einarsson, Mammút and Ólafur Arnalds are acknowledged for providing inspiration throughout this PhD. From the outside world it has been wonderful to live in a really great flat share and having flatmates accepting an at times messy PhD student. Chiara Rudel is also acknowledged for looking over the translation of the abstract and always being there to listen. I would like to thank my family and especially my parents for always supporting, being enthusiastic and believing in me. My dad especially for nice discussion and comments on the project and my mum for amazing photographs in the field and for always believing there is a solution. I also do not know how I would have managed without my Austrian family - I am very grateful to have Hannah and Clemens Tussetschläger in my life and I always enjoy every weekend trip in Innsbruck. It has also been nice to stay in touch with fellow students from my home university – Oslo, José Cerca de Oliveira and Siri Birkeland to mention a few. Moral support has also always been there from my friends back home. Bjørn Midthun is acknowledged for IT help (e.g. computer's refusal of the new battery while in the field and restoring, what should have been, eliminated files) and I am indebted to almost the whole Grech family

for technical support on the crucial printing and scanning of a form that refused to print correctly.

And lastly, David Grech. As much for scientific discussions, stupid laughs and distractions, pushing me forward and being especially caring, supportive and understanding in this last phase. David is also acknowledged for helping with the proofreading of the thesis.

I would also like to thank the reviewers for accepting to review the thesis.

This project was funded by the Austrian Science Fund (FWF, Project Y661-B16, a grant given to Ass. Prof. Ovidiu Paun).

"tiny changes to earth" – Scott Hutchison

TABLE OF CONTENTS

ABSTRACT

Disentangling phylogenetic relationships is challenging in recently evolved groups, especially in the presence of ongoing hybridization. Allopolyploids originate from the merging of two or more entire genomes and are in most cases immediately isolated from their parents. The complexity of the genome ancestry is increased by hybridization with other allopolyploids as well and backcrossing with their parental species. *Dactylorhiza* (Orchidaceae: Orchidinae) is a genus much affected by allopolyploid speciation and reticulate phylogenetic relationships. Here we use the genetic variation at tens of thousands of genomic positions to trace evolutionary history in *Dactylorhiza* by first delimiting the diploid taxa and establishing their relationship and then grouping 16 allopolyploids by maximum affiliation to their putative parents. The two sibling allopolyploids, *D. majalis* and *D. traunsteineri* were studied in more depth to investigate genetic structure and demography. The diploid parents throughout their European distribution were used to create two artificial genomic references for which two pipelines were used to align the polyploid reads and call genotypes/infer genotype likelihood. Two independent origins are suggested for *D. traunsteineri* and *D. majalis*, though gene flow does occur. Two other boreal sister genera *Nigritella* and *Gymnadenia* are morphologically distinct, though single-marker studies have previously given contradictory results on the inclusion/exclusion of *Nigritella* in *Gymnadenia*. Thousands of genetic markers were used to unravel their phylogenetic status. Lastly, evolutionary and ecological questions were addressed within the genus *Epipactis* on speciation via transitions from allogamy to autogamy. Genomic inferences based on RADseq (i.e. restriction-site associated DNA sequencing) data have become widely applied to answer evolutionary questions in diploid organisms and is included in all four chapters of this PhD, also showing its application to polyploid organisms. We obtain highly congruent evolutionary relationships inferred both with genotype-based and genotype-free analytical methods and develop bioinformatic approaches that can be applied to other naturally occurring, non-model, polyploid complexes.

ZUSAMMENFASSUNG

Die Untersuchung phylogenetischer Beziehungen von sich kürzlich entwickelten Gruppen stellt eine große Herausforderung dar, insbesondere bei fortlaufender Hybridisierung. Allopolyploide Organismen stammen aus der Verschmelzung von zwei oder mehreren Genomen und werden in den meisten Fällen sofort von ihren Eltern isoliert. Die Komplexität der genomischen Abstammung wird durch die Hybridisierung mit anderen Allopolyploiden sowie durch Rückkreuzung mit ihren möglicherweise vorkommenden Elternarten erhöht. Die Gattung *Dactylorhiza* (Orchidaceae: Orchidinae) ist stark von allopolyploider Artbildung und retikulären phylogenetischen Beziehungen betroffen. In dieser Arbeit verwenden wir genetische Variation, um die Evolutionsgeschichte in *Dactylorhiza* zu rekonstruieren, indem wir zunächst die diploiden Taxa abgrenzen, ihre Verwandtschaftsbeziehung festlegen und dann 16 allopolyploide Taxa nach maximaler Zugehörigkeit zu ihren mutmaßlichen Eltern gruppieren. Die beiden Geschwister Allopolyploide, *D. majalis* und *D. traunsteineri*, wurden untersucht, um die genetische Struktur und Demografie zu untersuchen. Die diploiden Eltern wurden in ihrer gesamten europäischen Verbreitung dazu verwendet, zwei künstliche genomische Referenzen zu erstellen um die polyploiden Werte abzugleichen und Genotypen zu ermitteln bzw. Genotyp-Wahrscheinlichkeit zu nennen. Für *D. traunsteineri* und *D. majalis* wurden zwei unabhängige Ursprünge ermittelt, obwohl ein Genfluss zwischen beiden Arten herrscht. Zwei weitere boreale Schwestergattungen *Nigritella* und *Gymnadenia* sind morphologisch unterschiedlich, obwohl Einzelmarkerstudien zuvor widersprüchliche Ergebnisse zum Einschluss / Ausschluss von *Nigritella* in *Gymnadenia* erbracht haben. Tausende genetischer Marker wurden verwendet, um ihren phylogenetischen Status aufzuklären. Schließlich wurden evolutionäre und ökologische Fragen in der Gattung *Epipactis* zur Artbildung durch einen Übergang von Allogamie zur Autogamie angesprochen. RAD-Daten (Restriktionsstellen Assoziierte DNA) werden in weitem Umfang zur Beantwortung von evolutionären Fragestellungen diploider Organismen angewendet. Die vier Kapitel dieser Doktorarbeit zeigen den Nutzen von RAD-Daten für die Arbeit mit polyploiden Organismen. Aus genotypbasierten und genotypfreien Analyseverfahren resultierende Verwandtschaftsbeziehungen sind kongruent. Außerdem wurden bioinformatische Ansätze entwickelt, die für die Arbeit mit anderen natürlich vorkommende polyploiden nicht-model Organismen verwendet werden können.

INTRODUCTION

Speciation and reticulate evolution

Speciation and evolutionary relationships have often been illustrated as phylogenetic trees where the entities, e.g. species, constitute the leaves on the branches. Embedded in the species tree lie the individual gene trees, each with its own history, which (often) can be discordant with the history of the species itself (Maddison 1997; Degnan and Rosenberg 2006). There are several reasons why such discordance may occur, e.g. horizontal gene transfer, gene duplication/loss and incomplete lineage sorting (Mallo and Posada 2016). Horizontal gene transfer is an exchange of genetic material without reproduction and is typically important in the evolution of unicellular organisms, but can affect multicellular organisms in symbiosis or interactions like those between parasites, pathogens, epiphytes, endophytes and their hosts (Gao 2014). When a gene duplication occurs, a set of paralogs are created that will start evolving individually. In order to estimate meaningful phylogenetic relationships it is necessary to compare orthologs, which are the homologous loci separated by speciation events in different species. If they are not distinguished from paralogs, this could result in confusing signals (Degnan and Rosenberg 2009). If ancestral polymorphism is retrained in the descending branches, but randomly lost later, i.e. as deep coalescence or incomplete lineage sorting, this will also lead to discordance between gene trees and the species tree.

Phylogenetic analyses are usually performed expecting that the phylogenetic signal of the majority of the genes will reflect the species tree. Two schools of methodologies have normally been proposed that both have their drawbacks and advantages, but when applied in parallel should give sound results: (i) concatenated models, where all gene trees are treated as the same variable and (ii) coalescent models, where each gene is considered separately backwards in time, for how they can be traced back to the common ancestor, i.e. the coalesce point (Liu et al. 2015). The latter methodology unites population genetics, by integrating changes in allele frequencies over time within the phylogenetic framework. To study the species demographical history one needs to look at the different dynamics and forces acting on populations to alter their allele frequencies. These forces can be both random (e.g. genetic drift) and/or non-random forces (selection).

Moving beyond the standard speciation process and bifurcating trees, we find that evolution is often acting in a more reticulate manner as a result of hybridization and polyploidization, i.e. whole genome duplication (Mallet 2007). Such reticulate speciation histories are often shaped as networks rather than as bifurcating trees. When the reproductive barriers are not complete,

gene flow between species can lead to inconsistent phylogenies and as with gene duplications mentioned earlier, a whole genome duplication, i.e. polyploidization, might have quite an impact on and complicate phylogenetic inference.

Polyploidization can lead to instantaneous or saltational speciation because the neopolyploid will come with a high likelihood of reproductive isolation towards its parents (Vogt et al. 2015) as well as possible transgressive traits, i.e. if hybridization is involved (Dittrich-Reed and Fitzpatrick 2013). The latter form of polyploidization has the prefix allo-, whereas polyploidization within the same species is termed auto- (Ramsey and Schemske 1998). Even though many organisms are functionally diploid, it is now known that they are actually ancient polyploids that have gone through a number of cycles of polyploidization followed by diploidization by which the redundant genes gradually received one of the following fates: gene non-functionalization, sub-functionalization, neo-functionalization (Soltis et al. 1993; Soltis and Soltis 1999; Wendel 2000; Ma and Gustafson 2005; Wendel 2015).
The advancement of molecular methods has brought forward the frequency and importance of evolution and speciation by reticulation (Van de Peer 2004; Soltis 2005). In animals at least two palaeopolyploidization events have been estimated at the origin of vertebrates (Dehal and Boore 2005) and another correlated with the radiation of bony fish (Amores et al. 1998). In plants one polyploidization event coincided with the origin of seed plants and another one with the origin of flowering plants (Jiao et al. 2011). Polyploidization is especially frequent in flowering plants (Angiosperms) where we see that recurrent polyploid formation followed by later hybridization and introgression has contributed to build up complicated genomes with interesting histories. The genetic diversity found in polyploid species will thus be the result of the number of independent origins as well as the population genetic and demographical history they have been through (Soltis and Soltis 1999; Beck et al. 2011).

The hybridizing nature of orchids

Orchids (Orchidaceae) is an incredibly diverse plant group and one of the two largest families within the Angiosperms with ~ 25 000 species affiliated to five subfamilies; Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae and Vanilloideae (Chase 2005). Because of the broad diversity in morphology, growth form, life history and often strict habitat specialization, they can exhibit various physiological properties, making them suitable for studying different aspect of evolution and speciation (Zhang et al. 2018). Deception is a regual phenomenon orchids as one third is estimated to have pollination by deceit (Cozzolino and Widmer 2005) and all depend on resources of mycorrhizal fungi, i.e. orchid mycorrhiza (Sathiyadash et al. 2012). The fungal component is especially required for establishment and early development due to need of nutrients, because their seeds are lacking an endosperm. Due to their mycorrhiza facilitated germination, they can produce millions of minute seeds (lacking an endosperm) that make way for spreading far as well as potentially creating hybrids.

The large, mainly Eurasian genus *Dactylorhiza* Necker ex Nevski (subfamily Orchidoideae, tribe Orchidinae) of preferentially boreal-montane species are known for comprising many interspecific hybrids accompanied by genome duplication; allopolyploids. The evolutionary history of *Dactylorhiza* appears to be complex because of multiple origins of polyploids distributed through time and space, secondary hybridization between the polyploids and with

their parental lineages. The genus thus offers a great opportunity for studying polyploid evolution and speciation by studying sibling allopolyploids that have arisen from independent crosses of the same parental lineages (Hedrén et al. 2011). Common features of these geophytes are that they are perennial herbs with an unbranched erect stem and spiral arrangement of leaves and terminal inflorescence (Averyanov 1990). The tuber/root (ρίζα "rhiza") can look like a hand with fingers (δάκτυλος "daktylos"), thereof the name “dactylorhiza”. They are food deceptive, which has been proposed as a strategy to avoid self-fertilization (Nilsson 1992), and pollinated by bumblebees, bees and flies.

The species of *Dactylorhiza*, commonly known as marsh orchids, grow in more or less calcareous open or semi-open habitats, some preferring moist sites, from calcareous swamps, fens to (less calcareous) *Sphagnum* bogs, others preferring dry grasslands, shrublands, sand-dunes, some even open forest types (Tutin 1980; Streeter et al. 2009). Some species might occur along a wide gradient of moisture conditions, but with more strict requirements as to calcareous/basiphilous soil types. Many species seem to be favoured by certain semi-natural conditions, being abundant in habitats kept open by traditional, extensive grazing or mowing. With the cessation of traditional management types, these habitats are declining across Europe, and many of the species are apparently declining (Hedrén 2001a; Blinova and Uotila 2012).

The first observations of ploidy variation in *Dactylorhiza* came from chromosome counts in Hagerup (1938), Heuser (1938) and Vermeulen (1938). They were able to distinguish diploid *D. fuchsii* and tetraploid *D. maculata*. Furthermore, Vermeulen (1938) proposed that tetraploids in the complex could have arisen by hybridization and genome duplication between diploid species. Heslop-Harrison (1953) suggested that a species similar to *D. fuchsii* was one of the parents of the tetraploids after studying the meiotic behaviour in triploid hybrids. Further support that the European allotetraploids have originated from crosses between the *D. fuchsii-maculata* clade and *D. incarnata-euxina* clade was detected in allozyme studies (Hedrén 1996a, 1996b, 1996c, 2001b; Pedersen 1998, 2002; Hedrén et al. 2007; Pedersen and Hedrén 2010), but allozymes were not variable enough to reveal divergence between allopolyploids and between some of the parents, e.g. *D. fuchsii* and *D. maculata*. Amplified fragment length polymorphism analysis, i.e. AFLP, on the other hand, could reveal a separation between *D. fuchsii* and *D. maculata* and some degree of differentiation between allopolyploids of separate geographic origins (Hedrén et al. 2001, 2007).

Further details on the origin of allopolyploid *Dactylorhiza* were obtained by studies of the plastid genome. *Dactylorhiza majalis* and *D. cordigera* have plastid elements no longer present in this exact form in extant members of the maternal *D. maculata* clade, suggesting that such allopolyploids might be of relatively greater age (Hedrén et al. 2007; Pillon et al. 2007). Plastid results have also indicated that many well-known polyploid taxa, such as *D. traunsteineri*, may have had multiple origins (Devos et al. 2003; Nordström and Hedrén 2008) or at least multiple independent colonization events northwards that has given rise to the variation found currently in its populations (Hedrén and Olofsson 2018; Hedrén et al. 2018c). Nuclear microsatellites have been used to track origins of allopolyploid *Dactylorhiza* and to study relationships among allopolyploids. Still, the taxa may not be easy to differentiate (Balao et al. 2016), and large numbers of specimens and samples are necessary to obtain reliable results (Hedrén et al. 2011).

The most commonly studied nuclear region was the rDNA tandem repeat regions containing the variable ITS/ETS portions (Pridgeon et al. 1997; Bateman et al. 2003). The diploids *D. incarnata* and *D. fuchsii* have been found to be well separated by ITS, while diploid *D. fuchsii*, *D. saccifera* and the autotetraploid *D. maculata* are not (Devos et al. 2005; Pillon et al. 2007). Even though allopolyploids inherit ITS/ETS from both parents, the homogenization of rDNA has been shown to be an example of fast concerted evolution that will increase the relative proportion inherited from one parent (Wendel et al. 1995; Kovarik et al. 2005; Volkov et al. 2007), and it has been suggested that the progress of ITS gene conversion can be used to estimate the age of the allopolyploid; for instance, *D. majalis* may be older than *D. traunsteneri* because the former is completely converted and the latter is not (Pillon et al. 2007). It has also been demonstrated that chimeric ITS sequences created by recombination can be produced in allopolyploids, combining portions from different parental lineages in the same repeat variant (Devos et al. 2006). The markers used so far have contributed well to build the treads in this web of a complex genus. With all methods having their upsides and downsides there are, however, bits and pieces still missing which this study has tried to fill in.

In the same tribe as *Dactylorhiza*, the boreal-montane(-alpine) genera *Nigritella* represent another complex system with known recurrent polyploidizations, but in contrast to *Dactylorhiza,* the polyploids are predominantly apomictic (asexuals also with odd ploidal levels) and assumed to have originated from parents that might not be extant anymore (Bateman et al. 2003; Hedrén et al. 2018a; Hedrén et al. 2018b). The extant diploids in *Nigritella* are not ap-

omictic and it is likely that the parents of the polyploids were also sexual. Phylogenetic estimations from ITS, plastid rpl16 and mitochondrial coxI have given conflicting results to whether *Gymnadenia*, which is closely related to *Nigritella,* should be its sister genus or if the two should be merged to a single genus (Bateman 1997; Pridgeon et al. 1997; Bateman et al. 2003; Stark et al. 2011; Inda et al. 2012). *Gymnigritella* is an (intergeneric) allopolyploid between *Nigritella* and *Gymnadenia*.

Reproductive switches are also found in another temperate-boreal genus, *Epipactis* (subfamily Epidendroideae, tribe Neottieae), namely transitions from cross-fertilization to self-fertilization (Hollingsworth et al. 2006; Tałałaj and Brzosko 2008). Both clonality and self-fertilization are excellent mechanisms under difficult conditions to preserve energy (e.g. many examples in the arctic) or in an establishment phase to spread fast (e.g. in interglacials, or after a polyploidization event), however, if not possible to switch back to cross-fertilization one might think these strategies could be an evolutionary dead-end (Igic and Busch 2013; Hand and Koltunow 2014).

How to handle polyploid data with genome-wide markers

The search for the tree of life has been revolutionized by the emergence of genome-scale data (Medina 2005; Simonson et al. 2005). Sequencing the whole genome of organisms with large genomes is still very costly and in need of powerful computing resources, however, for many purposes the whole genome is not necessary and using a reduced representation of the genome is sufficient (Hohenlohe et al. 2012). Restriction site associated DNA sequencing (RADseq) is a NGS-based, i.e. next generation sequencing, reduced representation of the genome that has the advantages of (i) being applicable to any non-model group with or without a reference available, (ii) being capable of investigating polymorphism across a large number of homologous loci across multiple accessions, (iii) providing sufficient discriminatory information to be useful among closely related species as well as infraspecific populations, and (iv) having codominant properties, thereby allowing calling of homozygotes and heterozygotes. As a consequence of the variability of cut sites at a broader phylogenetic scale, the proportion of homologous sequences obtained by RADseq will decrease with phylogenetic distance, which may be problematic when distantly related taxa are combined in the same analysis. However, RADseq has been shown to be useful for divergences projected backward to estimated dates as old as 63 million years (Cariou et al. 2013), while also being applicable to recently radiated groups (e.g. Paun et al. 2016; Bateman et al. 2018).

To investigate evolutionary questions both on a phylogenetic and a population genetic scale in this thesis, thousands of markers from RADseq were analysed with a plant sampling aimed to cover most of the distribution of *Dactylorhiza* species of Europe and adjacent Asia including (closely) related taxa and some deeper population sampling of *D. majalis*, *D. traunsteineri*. The processing and analyses of polyploid data have, however, still no standard optimized way of being carried out compared to diploid systems, for which phylogenetic and population genetic theory have been extensively developed. A polyploid-specific issue that needs to be taken into consideration before processing polyploid data is segregation (disomic versus polysomic), which will have an impact on result of the genotyping, i.e. occurrence of partial heterozygotes is expected with polysomic inheritance, whereas disomic inheritance should lead to a distinction between homoeologs (Dufresne et al. 2014; Meirmans et al. 2018). Other implications when it comes to the interpretation of polyploid data are that there might, or might not, be admixture by gene flow between ploidal levels, and the polyploids might have a switch in reproductive system, i.e. often self-fertilization and clonality (Dufresne et al. 2014).

Theory and tools for analysing polyploid data already started to develop for traditional markers (Baker 2008; Clark and Jasieniuk 2011; Pembleton et al. 2013; Hardy 2016). This applies in particular to (auto)polyploids with polysomic inheritance, i.e. non-preferential pairing of the chromosomes to several homologs, where we would expect that at the molecular level the homologs cannot be distinguished. An approach for such situations would normally be to call genotypes allowing for more than two alleles and either have a way of inferring the correct heterozygote constellation (Aaaa, AAaa, AAAa) or allowing for that uncertainty. Programs handling NGS data and developed after the autopolyploid model include PolyRAD that uses the uncertainty to calculate the probability of the genotype (Clark et al. 2018), updog that performs genotyping taking sequencing errors, allelic bias and overdispersion into account (Gerard et al. 2018), ANGSD can be used completely genotype free and output the genotype likelihoods directly, which then can be applied in downstream analysis (Korneliussen et al. 2014) and the "Polyploid Genotyping Approach" described by Blischak et al. (2017). The latter comprises both an approach for autopolyploids and for allopolyploids with disomic inheritance, i.e. preferential pairing of the homologs (note that the homoeologs are not pairing with each other).

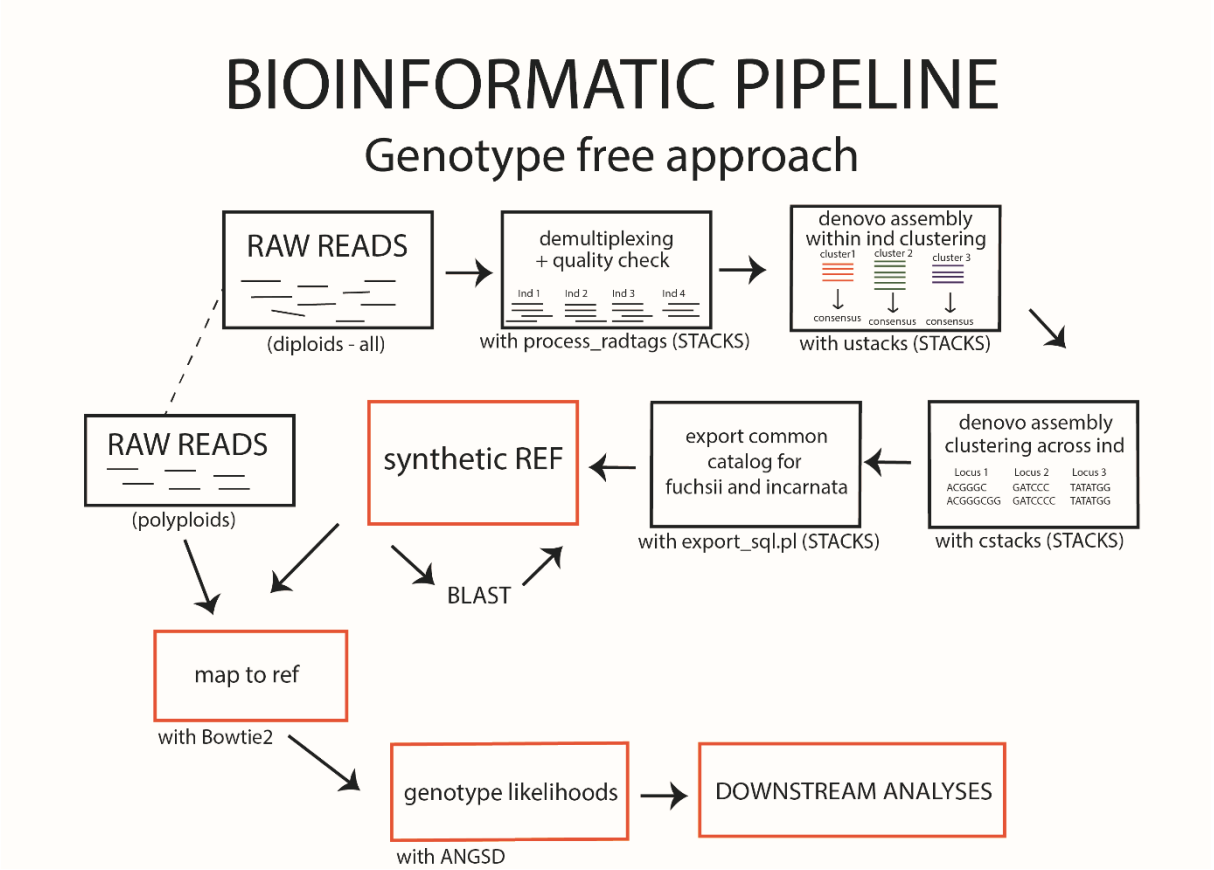


Figure 1: Illustrating pipeline I applied in Chapter 2. Built til create a combined synthetic genomic reference (organge boxes) from the two parents; *D. fuchsii* and *D. incarnata*. Optimized to pick up loci that cannot be distinguished to the two subgenomes in the allopolyploids. A genotype free approach is applied, using the genotype likelihoods to allow for partial heterozygotes.

On the molecular level we would expect that the homoeologs in an allopolyploid should be possible to distinguish if long enough haplotypes would be analysed and, if the parental genomes are available, the optimal approach would be to separate the polyploid genome into the two subgenomes (corresponding to the two parental genomes) and analyse them as diploid genomes. Assuming that the parents are available the "Polyploid Genotyping Approach" uses the allele frequencies expected at each subgenome for the allopolyploids for the genotyping (Blischak et al. 2017). There are, however, different approaches to split the subgenomes. For non-neutral evolving data like RNAseq fixed positions in the parents are then traced in the polyploids to split back to the respective subgenome (Page et al. 2013; Duchemin et al. 2015). For data assumed to be neutral evolving, like RADseq, there won't necessarily be as many of

these fixed differences between the parents, but there should still be an overall sequence difference.

Depending on the divergence between the parents of the allopolyploid and the time past since the polyploidization event (diploidization will gradually happen over time; Mandáková et al. 2016), one can also expect to find parts of the genome that are possible to distinguish between the two parental genomes, and other parts that are not possible to distinguish. For such cases a combination of approaches might be beneficial. In this thesis two approaches were applied, one following the autopolyploid model and the other following the allopolyploid model.

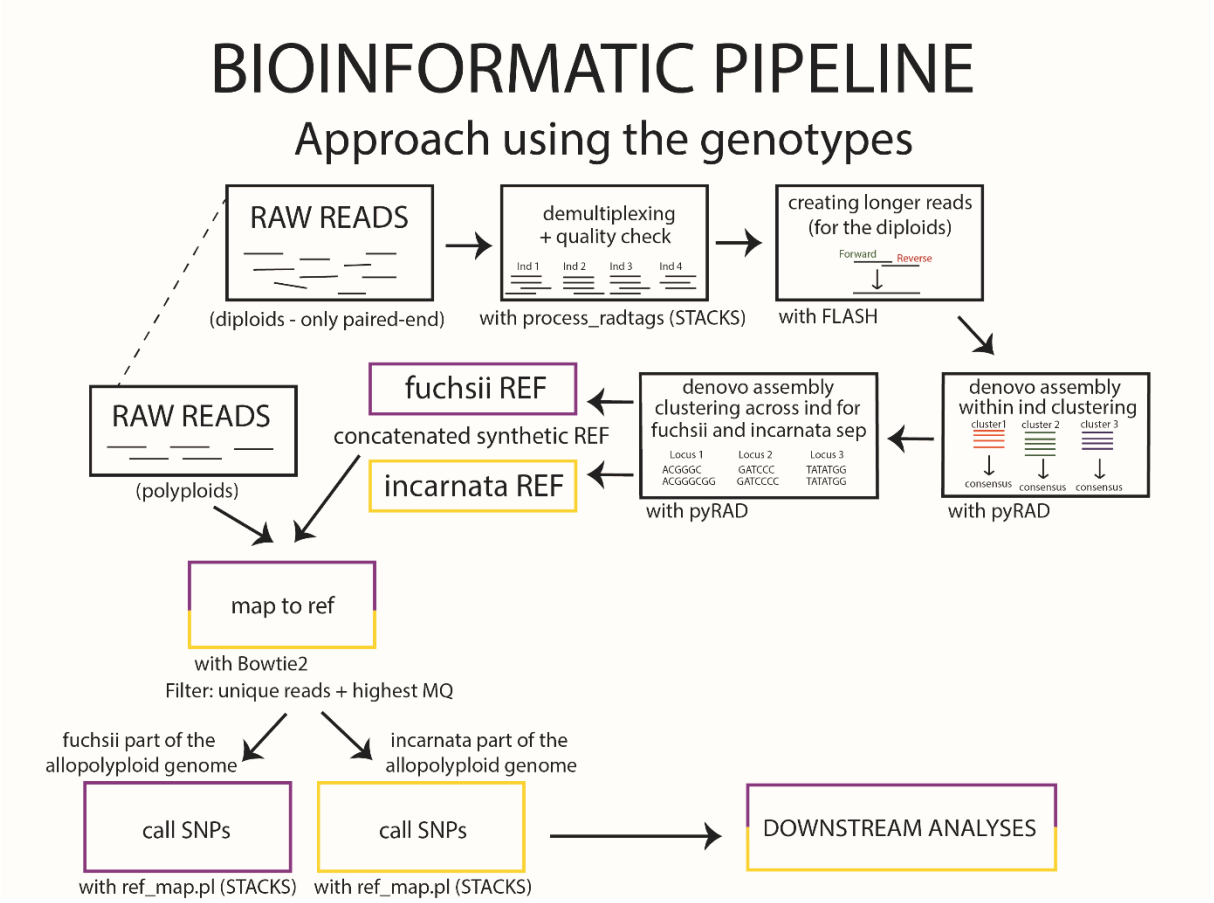


Figure 2: Illustrating pipeline II applied in Chapter 2. Built til create two separate synthetic genomic reference from the two parents; *D. fuchsii* (purple) and *D. incarnata* (yellow). Optimized to separate the loci in the allopolyploids to the subgenomes and then proceeeding downstream analyses with diploid genotypes.

THESIS QUESTIONS

Chapter 1. Phylogenetic tree reconstruction and parental inference in *Dactylorhiza. Dactylorhiza* has had a complex evolutionary history and it currently includes more allopolyploid taxa than diploid species. The allopolyploids are believed to hybridize with each other. We made use of a European-wide sampling and a genome-wide SNPs datasets derived from RADseq to resolve the following questions:

1.1. How many diploid species are there in *Dactylorhiza* and what is their circumscription?

1.2 What is the phylogenetic history between diploid species (i.e., within the species tree) and their relationships to related genera?

1.3. Can the paternal and maternal lineages and the relative age of *Dactylorhiza* polyploids be inferred using RADseq and a plastid reference genome?

Chapter 2. Genetic structure and origin of the allopolyploids *D. traunsteineri* and *D. majalis* and the methodology for treating allopolyploid RADseq data. The allopolyploid group originating from the cross of *D. fuchsii* and *D. incarnata* contain allopolyploids of different age, but all relatively young and not yet fully reproductively isolated from its sibling allopolyploids or from its parents. With a population sampling across their distribution ranges and making use of numerous representatives of their parents, we applied a dataset of genome-wide SNPs derived from RADseq to answer the following questions for these allopolyploids:

2.1 What is the best methodology to process and treat allopolyploid RADseq data?

2.2 What is the genetic structure of the two allopolyploids and their parents?

2.3 How many times and where did the sibling *Dactylorhiza* allopolyploids originate each? Can local origins be inferred from regional diploid parents *D. fuchsii* and *D. incarnata*?

2.4 How old are the two allopolyploids? Is the hypothesis of different age of the sibling allopolyploids supported by high-throughput genomics data?

Chapter 3. Resolving phylogenetic relationship between *Nigritella* and *Gymnadenia* and parental inference for *Gymnigritella.* Previous phylogenetic studies have been contradictory on the relationship between *Nigritella* and *Gymnadenia* and the intergeneric hybrid that exist between them. Genome-wide SNPs derived from RADseq was applied to elucidate the following questions

3.1. Are *Nigritella* and *Gymnadenia* forming independent monophyletic clades? Should they be treated as different genera?

3.2. What are the phylogenetic relationships between the diploid species of *Gymnadenia* and *Nigritella*?

3.3. Can the parental species of the intergeneric allopolyploid *Gymnigritella runei* be inferred?

3.4. What is the genetic structure of the polyploid aggregate in *Nigritella*?

Chapter 4. Investigating the iterative allogamy-autogamy transitions during the evolutionary radiation within *Epipactis*, focusing on section *Eupipactis*. *Epipactis* is becoming a model for the study of speciation by transitions from allogamy to autogamy, but still much of its phylogenetic history is unknown. We made use of genome-wide SNPs datasets derived from RADseq to answer the following questions:

4.1. How many *bon fide* species can be recognized in *Epipactis*? What is their genetic structure and phylogenetic relationships?

4.2. Are transitions from allogamy to autogamy found to be iterative in *Epipactis*?

4.3. Can the hypothesized breeding system (i.e., autogamy vs. allogamy) be tested and confirmed with the genetic data?

4.4. Are there examples of autogams that form new autogams or even examples of autogams reverse transitions to form allogams in *Epipactis*?

LITERATURE CITED

Amores A., Force A., Yan Y.-L., Joly L., Amemiya C., Fritz A., Ho R. K., Langeland J.,Prince V., Wang Y.-L. 1998. Zebrafish hox clusters and vertebrate genome evolution.Science 282:1711-1714.

Averyanov L. 1990. A review of the genus *Dactylorhiza*. Orchid Biol. Rev. Persp., V. Timber Press: Portland, Oregon: 159-206.

Baker, P. 2008. polySegratio: An R library for autopolyploid segregation analysis. PolySegratio Library. Accessed January 9: 2008.

Balao, F., Tannhäuser M., Lorenzo M. T., Hedrén M., Paun O. 2016. Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. Heredity 116: 351.

Bateman R.M. 1997. Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 2. Intrageneric relationships and reclassification to achieve monophyly of Orchis sensu stricto. Lindleyana 12:113-141.

Bateman R.M., Hollingsworth P.M., Presto J., Yi-Bo L., Pridgeon A.M., Chase M.W. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). Bot. J. Linnean Soc. 142:1-40.

Bateman R.M., Sramkó G., Paun O. 2018. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. Ann. Bot. 121:85-105.

Beck J. B., Windham M. D., Pryer K. M. 2011. Do asexual polyploid lineages lead short evolutionary lives? A case study from the fern genus Astrolepis. Evolution: International Journal of Organic Evolution 65:3217-3229.

Blinova I. V., Uotila P. 2012. *Dactylorhiza traunsteineri* (Orchidaceae) in Murmansk Region (Russia). Memoranda Societatis pro Fauna et Flora Fennica.

Blischak, P. D., Kubatko L. S., Wolfe A. 2017. SNP genotyping and parameter estimation in polyploids using low-coverage sequencing data. Bioinformatics 34:407-415.

Cariou M., Duret L., Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and evolution 3:846-852.

Chase M. W. 2005. Classification of Orchidaceae in the age of DNA data. Curtis's Botanical Magazine 22:2-7.

Clark L. V., Jasieniuk M. 2011. POLYSAT: an R package for polyploid microsatellite analysis. Molecular Ecology Resources 11:562-566.

Clark L. V., Lipka A. E., Sacks E. J. 2018. polyRAD: Genotype calling with uncertainty from sequencing data in polyploids and diploids. bioRxiv: 380899.

Cozzolino S., Widmer A. 2005. Orchid diversity: an evolutionary consequence of deception? Trends in Ecology & Evolution 20:487-494.

Degnan J. H., Rosenberg N. A. 2006. Discordance of species trees with their most likely gene trees. PLoS genetics 2:e68.

Degnan J. H., Rosenberg N. A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332-340.

Dehal P., Boore J. L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biology 3:e314.

Devos N., Oh S.-H., Raspé O., Jacquemart A.-L., Manos P.S. 2005. Nuclear ribosomal DNA sequence variation and evolution of spotted marsh-orchids (*Dactylorhiza maculata* group). Molec. Phylogenet. Evol. 36:568-580.

Devos N., Raspé O., Oh S.-H., Tyteca D., Jacquemart A. L. 2006. "The evolution of *Dactylorhiza* (Orchidaceae) allotetraploid complex: Insights from nrDNA sequences and cpDNA PCR-RFLP data. Molecular Phylogenetics and Evolution 38:767-778.

Devos N., Tyteca D., Raspé O., Wesselingh R.A., Jacquemart A.-L. 2003. Patterns of chloroplast diversity among western European *Dactylorhiza* species (Orchidaceae). Plant Syst. Evol. 243:85-97.

Dittrich-Reed D. R., Fitzpatrick B. M. 2013. Transgressive hybrids as hopeful monsters. Evolutionary biology 40:310-315.

Duchemin W., Dupont P.-Y., Campbell M. A., Ganley A. R., Cox M. P. 2015. HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. BMC bioinformatics 16:8.

Dufresne F., Stift M., Vergilino R., Mable B.K. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molec. Ecol. 23:40-69.

Gerard D., Ferrão L. F. V., Garcia A. A. F., Stephens M. 2018. Genotyping Polyploids from Messy Sequencing Data. Genetics 210:789-807.

Gao C., Ren X., Mason A. S., Liu H., Xiao M., Li J., Fu D. 2014. Horizontal gene transfer in plants. Functional & integrative genomics, *14*:23-29.

Hagerup O. 1938. Studies on the significance of polyploidy. II. Orchis. Hereditas 24:258-264.
Hand M. L., Koltunow A. M. 2014. The genetic control of apomixis: asexual seed formation. Genetics 197:441-450.
Hardy O. J. 2016. Population genetics of autopolyploids under a mixed mating model and the estimation of selfing rate. Molecular Ecology Resources 16:103-117.
Hedrén M. 1996a. The allotetraploid nature of *Dactylorhiza praetermissa* (Druce) Soó (Orchidaceae) confirmed. Watsonia 21:113-118.
Hedrén M. 1996b. Electrophoretic evidence for all otetraploid origin of *Dactylorhiza purpurella* (Orchidaceae). Nordic Journal of Botany 16:127-134.
Hedrén M. 1996c. Genetic differentiation, polyploidization and hybridization in Northern European *Dactylorhiza* (Orchidaceae): evidence from allozyme markers. Plant Systematics and Evolution 201:31-55.
Hedrén M. 2001a. Conservation priorities in Dactylorhiza, a taxonomically complex genus. Lindleyana 16:17-25.
Hedrén M. 2001b. Systematics of the *Dactylorhiza euxina*/*incarnata*/*maculata* polyploid complex (Orchidaceae) in Turkey: evidence from allozyme data. Plant Systematics and Evolution 229:23-44.
Hedrén M., Fay M. F., Chase M. W. 2001. Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidaceae). American Journal of Botany 88:1868-1880.
Hedrén M., Lorenz R., Ståhlberg D. 2018a. Evidence for bidirectional hybridization between *Gymnadenia* and *Nigritella*. Journal Europäischer Orchideen 50:43-60.
Hedrén M., Lorenz R., Teppner H., Dolinar B., Giotta C., Griebl N., Hansson S., Heidtke U., Klein E., Perazza G. 2018b. Evolution and systematics of polyploid *Nigritella* (Orchidaceae). Nordic Journal of Botany.
Hedrén M., Nordström S., Bateman R. M. 2011. Plastid and nuclear DNA marker data support the recognition of four tetraploid marsh orchids (*Dactylorhiza majalis* sl, Orchidaceae) in Britain and Ireland, but require their recircumscription. Biological Journal of the Linnean Society 104:107-128.
Hedrén M., Nordström S., Hovmalm H. A. P., Pedersen H. Æ., Hansson S. 2007. Patterns of polyploid evolution in Greek marsh orchids (*Dactylorhiza*; Orchidaceae) as revealed by allozymes, AFLPs, and plastid DNA data. American Journal of Botany 94:1205-1218.
Hedrén M., Olofsson S. N. 2018. High levels of genetic diversity in marginal populations of the marsh orchid *Dactylorhiza majalis* subsp. *majalis*. Nordic Journal of Botany
Hedrén M., Olofsson S. N., Paun O. 2018c. Orchid colonization: multiple parallel dispersal events and mosaic genetic structure in *Dactylorhiza majalis* ssp. *lapponica* on the Baltic island of Gotland. Annals of botany 122:1019-1032.
Heslop-Harrison J. 1953. Microsporogenesis in some triploid *dactylorchid* hybrids. Annals of botany 17:539-550.
Heuser K. 1938. Chromosomenverhältnisse bei schweizerischen basitonen Orchideen.
Hollingsworth P. M., Squirrell J., Hollingsworth M. L., Richards A. J., Bateman R. M. 2006. Taxonomic complexity, conservation and recurrent origins of self-pollination in *Epipactis* (Orchidaceae). Current taxonomic research on the British and European flora. London: Botanical Society of the British Isles: 27-44.
Hohenlohe P. A., Catchen, J., Cresko, W. A. 2012. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Data production and analysis in population genomics. Humana Press, Totowa, NJ. 235-260
Igic B., Busch J. W. 2013. Is self-fertilization an evolutionary dead end? New Phytologist 198:386-397.

Inda L.A., Pimentel M., Chase M.W. 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. Annals Bot. 110:71-90.
Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., dePamphilis C.W.2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97-100.
Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. BMC Bioinf. 15:356.
Kovarik A., Pires J. C., Leitch A. R., Lim K. Y., Sherwood A., Matyasek R., Rocca J., Soltis D., Soltis P. 2005. Rapid Concerted Evolution of Nuclear Ribosomal DNA in two Tragopogon Allopolyploids of Recent and Recurrent Origin. Genetics.
Liu L., Xi Z., Wu S., Davis C.C., Edwards S.V. 2015. Estimating phylogenetic trees from genome-scale data. Annals New York Acad. Sci. 1360:36-53.
Ma X.-F., Gustafson J. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenetic and genome research 109:236-249.
Maddison W. P. 1997. Gene trees in species trees. Systematic biology 46:523-536.
Mandáková T., Gloss A. D., Whiteman N. K., Lysak M. A. 2016. How diploidization turned a tetraploid into a pseudotriploid. American Journal of Botany, 103:1187-1196.
Mallet J. 2007. Hybrid speciation. Nature 446:279.
Mallo D., Posada D. 2016. Multilocus inference of species trees and DNA barcoding. Phil. Trans. R. Soc. B 371.1702:20150335.
Medina M. 2005. Genomes, phylogeny, and evolutionary systems biology. Proceedings of the National Academy of Sciences 102:6630-6635.
Meirmans P. G., Liu S., van Tienderen P. H. 2018. The analysis of polyploid genetic data. Journal of Heredity 109:283-296.
Nilsson L. A. 1992. Orchid pollination biology. Trends in Ecology & Evolution 7:255-259.
Nordström S., Hedrén M. 2008. Genetic differentiation and postglacial migration of the *Dactylorhiza majalis* ssp. *traunsteineri/lapponica* complex into Fennoscandia. Plant Systematics and Evolution 276:73-87.
Page J. T., Gingle A. R., Udall J. A. 2013. PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3: Genes, genomes, genetics 3:517-525.
Paun O., Turner B., Trucchi E., Munzinger J., Chase M.W., Samuel R. 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. Syst. Biol. 65:212-227.
Pedersen H.Æ., Hedrén M. 2010. On the distinction of *Dactylorhiza baltica* and *D. pardalina* (Orchidaceae) and the systematic affinities of geographically intermediate populations. Nordic J. Bot. 28:1-12.
Pedersen H. Æ. 1998. Allozyme variation and genetic integrity of *Dactylorhiza incarnata* (Orchidaceae). Nordic Journal of Botany 18:15-21.
Pedersen H. Æ. 2002. *Dactylorhiza majalis* sl (Orchidaceae) in acid habitats: variation patterns, taxonomy, and evolution. Nordic Journal of Botany 22:641-658.
Pembleton L. W., N. O. Cogan, Forster J. W. 2013. St AMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources 13:946-952.
Pillon Y., Fay M.F., Hedrén M., Chase M.W. 2007. Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). Taxon 56:1185-1185.

Pridgeon A., Bateman R., Cox A. V., Hapeman J. R., Chase M. W. 1997. Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences: 1. Intergeneric relationships and polyphyly of Orchis sensu lato. Lindleyana 12:89-109.
Ramsey J., Schemske D. W. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual review of ecology and systematics 29:467-501.
Sathiyadash K., Muthukumar T., Uma E., Pandey R. R. 2012. Mycorrhizal association and morphology in orchids. Journal of plant interactions 7:238-247.
Simonson A. B., Servin J. A., Skophammer R. G., Herbold C. W., Rivera M. C., Lake J. A. 2005. Decoding the genomic tree of life. Proceedings of the National Academy of Sciences 102:6608-6613.
Soltis D. E., Soltis P. S., Rieseberg L. H. 1993. Molecular data and the dynamic nature of polyploidy. Critical reviews in plant sciences 12:243-273.
Soltis D. E., Soltis P. S. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution 14:348-352.
Soltis P. S. 2005. Ancient and recent polyploidy in angiosperms. New Phytologist 166:5-8.
Stark C., Michalski S. G., Babik W., Winterfeld G., Durka W. 2011. Strong genetic differentiation between *Gymnadenia conopsea* and *G. densiflora* despite morphological similarity. Plant Systematics and Evolution 293:213-226.
Streeter D., Hart-Davies A., Hardcastle A., Cole F., Harper L. 2009. Collins flower guide: the most complete guide to the flowers of Britain and Ireland, Harper Collins Publishers Ltd, London.
Tałałaj I., Brzosko E. 2008. Selfing potential in *Epipactis palustris*, *E. helleborine* and *E. atrorubens* (Orchidaceae). Plant Systematics and Evolution 276:21-29.
Tutin T. G. 1980. Flora Europea; Vol. 5. Alismataceae to Orchidaceae.
Van de Peer Y. 2004. Computational approaches to unveiling ancient genome duplications. Nature reviews genetics 5:752.
Vermeulen P. 1938. Chromosomes in Orchis. Chron. Bot 4: 107-108.
Vogt G., Falckenhayn C., Schrimpf A., Schmid K., Hanna K., Panteleit J. Helm M., Schulz R., Lyko F. 2015. The marbled crayfish as a paradigm for saltational speciation by autopolyploidy and parthenogenesis in animals. Biology open: bio. 014241.
Volkov R. A., Komarova N. Y., Hemleben V. 2007. Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. Systematics and Biodiversity 5:261-276.
Wendel J. F. 2000. Genome evolution in polyploids. Plant molecular evolution, Springer: 225-249.
Wendel J. F. 2015. The wondrous cycles of polyploidy in plants. American Journal of Botany 102: 1753-1756.
Wendel J. F., Schnabel A., Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). Proceedings of the National Academy of Sciences 92:280-284.
Zhang S.-B., Yang Y.-J., Li J.-W., Qin J., Zhang W., Huang W., Hu H. 2018. Physiological diversity of orchids. Plant Diversity.

— CHAPTER 1 —

Phylogenomic Relationships of Diploids and the Origins of Allotetraploids in *Dactylorhiza* (Orchidaceae): RADseq Data Track Reticulate Evolution

---

Marie K. Brandrud, Juliane Baar, Maria T. Lorenzo,

Richard M. Bateman, Mark W. Chase, Ovidiu Paun, Mikael Hedrén

Status: Under review in Systematic Biology

Contribution: Data processing, analyses and visualization, writing manuscript

Running head: PHYLOGENOMICS OF *DACTYLORHIZA*

Phylogenomic Relationships of Diploids and the Origins of Allotetraploids in *Dactylorhiza* (Orchidaceae): RADseq Data Track Reticulate Evolution

MARIE K. BRANDRUD[1], JULIANE BAAR[1], MARIA T. LORENZO[1], ALEXANDER ATHANASIADIS[1], RICHARD M. BATEMAN[2], MARK W. CHASE[2,3], MIKAEL HEDRÉN[4]*, AND OVIDIU PAUN[1]*

[1] *Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria*

[2] *Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3AB, UK*

[3] *School of Biological Sciences, University of Western Australia, Crawley, Perth, Australia*

[4] *Department of Biology, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden*

*These authors contributed equally to this work.

Corresponding authors: Ovidiu Paun, University of Vienna, Rennweg 14, A-1030 Vienna, Austria, ovidiu.paun@univie.ac.at; Mikael Hedrén, University of Lund, Sölvegatan 37, SE-223 62, Lund, Sweden, mikael.hedren@biol.lu.se

Abstract—Disentangling phylogenetic relationships proves challenging for groups that have evolved recently, especially for those involving ongoing reticulation. Although they are in most cases immediately isolated from diploid relatives, sets of sibling allopolyploids often hybridize with each other and backcross with local representatives of their parental species, thereby increasing the complexity of an already intricate situation. *Dactylorhiza* (Orchidaceae: Orchidinae) is a genus much affected by allopolyploid speciation and reticulate phylogenetic relationships. Here we use the genetic variation at tens of thousands of genomic positions to trace evolutionary history in *Dactylorhiza* by using coalescent and maximum likelihood methods to first investigate circumscription and relationships of diploid species in the genus, and then to group 16 allopolyploids by maximum affiliation to putative parentals. We obtain highly congruent evolutionary relationships inferred both with genotype-based and genotype-free analytical methods. The direction of origin was inferred for each allopolyploid using RADseq loci that localized to the plastid genome. Starting from age estimates of parental lineages, the relative age of different allopolyploid groups was further estimated by quantifying their relatedness to diploids and their numbers of private alleles compared with sibling allopolyploids. Whereas NW Europe is dominated by young allopolyploids of postglacial origins, comparatively old allopolyploids are distributed further south, where climatic conditions remained relatively stable during Pleistocene glaciations. Our bioinformatics approach should prove useful for the study of other naturally occurring, non-model, polyploid complexes.

Keywords: Allopolyploidy, coalescent, *Dactylorhiza*, genotype-free inference, phylogenomics, RADseq, reticulate evolution, speciation.

Disentangling phylogenetic relationships often proves challenging in groups that have diverged recently and/or rapidly (Mallo and Posada 2016; Fernández-Mazuecos et al. 2017). Problems encountered include insufficient morphological differentiation, phenotypic convergence, slow accumulation of genetic divergence, widespread incomplete lineage sorting, and/or permeable barriers to interspecific gene flow. Molecular phylogenetic approaches traditionally propose to investigate as many genomic regions as possible by combining them, with the general expectation that the history of the majority of genes will accurately reflect the true species history (Ebersberger et al. 2007). However, due to coalescent stochasticity along the genome, individual gene trees are likely to be incongruent with each other and with the species tree (Degnan and Rosenberg 2006; Liu et al. 2015; Pease et al. 2016). Hence, it has become clear that, although reasonably accurate when there is little heterogeneity among gene trees, concatenation methods have limitations due to issues such as long-branch attraction, heterogeneity in rates of substitutions among otherwise related lineages, and limited taxon sampling. Multispecies coalescent approaches have recently been suggested as appropriate alternatives, but they are computationally intensive, sensitive to inaccurate species circumscriptions and vulnerable to incongruence originating from interspecific gene flow (Liu et al. 2015; Mallo and Posada 2016) – a putatively omnipresent biological process in plants.

Understanding evolutionary histories is further greatly complicated by hybrid and/or polyploid taxa that exhibit multiple alleles from different ancestors and mixed inheritance patterns (Dufresne et al. 2014; Meirmans et al. 2018). Hybridization and polyploidy are, however, pervasive evolutionary processes driving speciation and adaptation (Ramsey and Schemske 1998; Adams and Wendel 2005; Van de Peer et al. 2017). Allopolyploids integrate the entire nuclear content of distinct parents but can become instantly reproductively isolated from their ancestors (Soltis et al. 2014). Allopolyploid genotypes often originate recurrently starting from the same parents, examples having been demonstrated for *Tragopogon mirus* and *T. miscellus* (Soltis et al. 2004), and for *Mimulus peregrinus* (Vallejo-Marin et al. 2015). As the sequentially produced genotypes subsequently interbreed, this recurrent process further increases genetic diversity and genome complexity of the resulting allopolyploids (Soltis and Soltis 1999; Soltis et al. 2014). In other cases, such recurrent origins establish independent species that remain genetically distinct, even in sympatry. Such examples include *Achillea alpina* and *A. wilsoniana* (Guo et al. 2013), *Asplenium* ferns (Perrie et al. 2010), *Leucaena involucrata* and *L. pallida* (Govindarajulu et al. 2011) and *Oryza minuta* and *O. malampuzhaensis* (Zou et al. 2015).

The parental taxa of allopolyploids typically belong to the same genus or, rarely, to closely related genera, and may have been isolated from each other for a relatively long period (e.g., Paun et al. 2009). However, due to genomic heterogeneity in parental divergence, gene conversion and illegitimate recombination, most allopolyploids will present a mixed inheritance (Stift et al. 2008). This, together with a slow development of population genetic theory for polyploids, and difficulties in clearly identifying homoeologs and allele dosage, have impaired widespread accurate investigations of polyploid evolution (Meirmans et al. 2018).

Genotype-free methods that estimate the probability of distribution of sample allele frequencies directly from aligned next generation sequencing (NGS) data have been shown to drastically improve accuracy when analyzing intermediate to low coverage data (Nielsen et al. 2011; Han et al. 2014). Whereas accounting for uncertainty in NGS data (Korneliussen et al. 2014), genotype-free methods can also reduce bias in allele frequency estimates at biallelic single nucleotide polymorphisms (SNPs) for polyploids, by retaining genotype information as probabilities for further analyses.

In this paper, we explore the utility of restriction site associated DNA sequencing (RADseq; Baird et al. 2008) for inferring evolutionary patterns in a complex orchid genus (*Dactylorhiza* Necker ex Nevski) that comprises numerous allopolyploids, an autopolyploid and their putative diploid progenitors. As a consequence of the variability of restriction sites at a broader phylogenetic scale, the proportion of homologous sequences obtained by RADseq will decrease with phylogenetic distance, which may be problematic when distantly related taxa are investigated together. However, RADseq has been shown to be useful for resolving divergences projected backward to estimated dates as old as 63 Ma (Cariou et al. 2013), while also providing resolution to recently radiated groups (e.g., Cruaud et al. 2014; Paun et al. 2016; Bateman et al. 2018b).

The present study focuses on the temperate (-boreal) terrestrial orchid genus *Dactylorhiza*, which has its main distribution in Europe with few species in Asia, N America and N Africa (Averyanov 1990; Pillon et al. 2007). The genus has not previously been assessed in a phylogenomic context with such a broad sampling of both diploid and polyploid accessions, but also of genomic loci. Due to morphological heterogeneity and common interspecific hybridization, *Dactylorhiza* is considered taxonomically controversial (Pillon et al. 2006), and the number of recognized species by different authors varies between six and 75 (reviewed by Pedersen 1998; see also Supplementary Table S2). Published molecular analyses, which in general included few *Dactylorhiza* accessions within a broader phylogenetic context, have employed from one to only a few molecular markers (e.g., ITS – Bateman et al. 2003 and Pil-

lon et al. 2007; ITS and ETS – Devos et al. 2006; *cox1* spacer – Inda et al. 2010; ITS, a plastid and a mitochondrial intron – Inda et al. 2012) and generally resulted in low resolution and often conflicting topologies. Greater taxonomic controversy afflicts the numerous *Dactylorhiza* allopolyploids, which predominantly originated from hybridization between two broadly defined parental groups, the *D. fuchsii-maculata* and the *D. incarnata-euxina* clades (Heslop-Harrison 1954; Hedrén 1996, 2001; Hedrén et al. 2007; Pedersen 2004). However, due to the high reticulation in the *Dactylorhiza* polyploid complex, the success of previous attempts to resolve polyploid origins based on restricted marker data (e.g., Pillon et al. 2007) has been limited.

We summarize here a comprehensive approach to obtain detailed insights into diploid and polyploid evolution within this complex genus with RADseq data, in order to assess its value for studying reticulate evolution in other polyploid complexes. The current study first employs thousands of SNPs to delimit with coalescent methods diploid *Dactylorhiza* species and estimate their phylogenetic relationships. The delimited diploid species and a synthetic reference based on diploids only are then employed with traditional and genotype-free methods of estimating allele frequencies to elucidate the many independent origins of allopolyploids within the genus, including distinguishing the direction of the hybridization events that produced them. We finally explore the influence of a long-hypothesized gene flow among various diploid and polyploid lineages on the evolutionary relationships within this intricate genus.

MATERIALS AND METHODS

Plant Material

Our comprehensive sampling covers most of the diploids in the genus (apart from a couple of poorly documented diploids in the Himalayas and China) and 16 established allopolyploids. The autopolyploid *D. maculata* has been regarded as a putative parent for some of the allopolyploids (Devos et al. 2003; Hedrén 2003; Pillon et al. 2007), and it is also included in our sampling. Altogether, we included 94 diploid, 18 autopolyploid, and 95 allopolyploid accessions of *Dactylorhiza* (Figs. 1 and 2, Supplementary Table S1). When possible, we included several accessions of each putative *Dactylorhiza* species, though for the diploid *D. aristata* and two allopolyploids only single individuals were available. Additionally, 18 diploids from the related genera *Gymnadenia*, *Nigritella*, *Galearis*, *Pseudorchis* and *Platanthera* were sampled. For each analyzed species vouchers have been deposited in the herbaria of Lund University (LD), of the University of Vienna (WU) or of Royal Botanic Gardens Kew

(K) (Supplementary Table S1). Maps of the sampling locations for the parental species and the allopolyploids were generated using QGIS v. 2.4.071 (QGIS Development Team 2018), constructed on a map layer extracted from GADM version 1.0 (available from www.gadm.org).

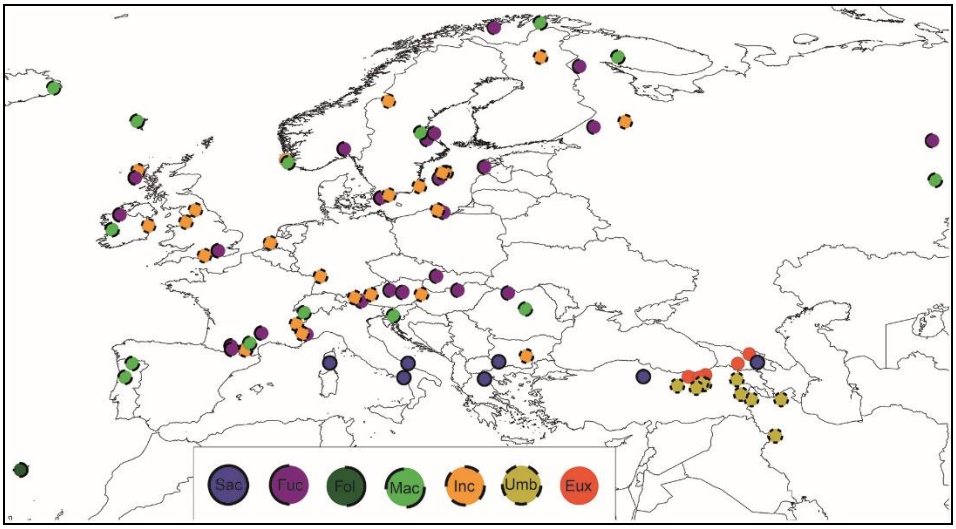


FIGURE 1. Map showing the sampling locations of 85 diploid and 18 autotetraploid *Dactylorhiza* individuals that have been analyzed here as representatives of potential parents of allopolyploids. The study further includes nine diploid individuals of *D. aristata*, *D. iberica*, *D. sambucina*, and *D. viridis*, whose sampling locations are not displayed here. Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata*, Sac = *D. saccifera*, Umb = *D. umbrosa*. The map layer was extracted from GADM version 1.0 (available from www.gadm.org). Exact coordinates are given in Supplementary Table S1.

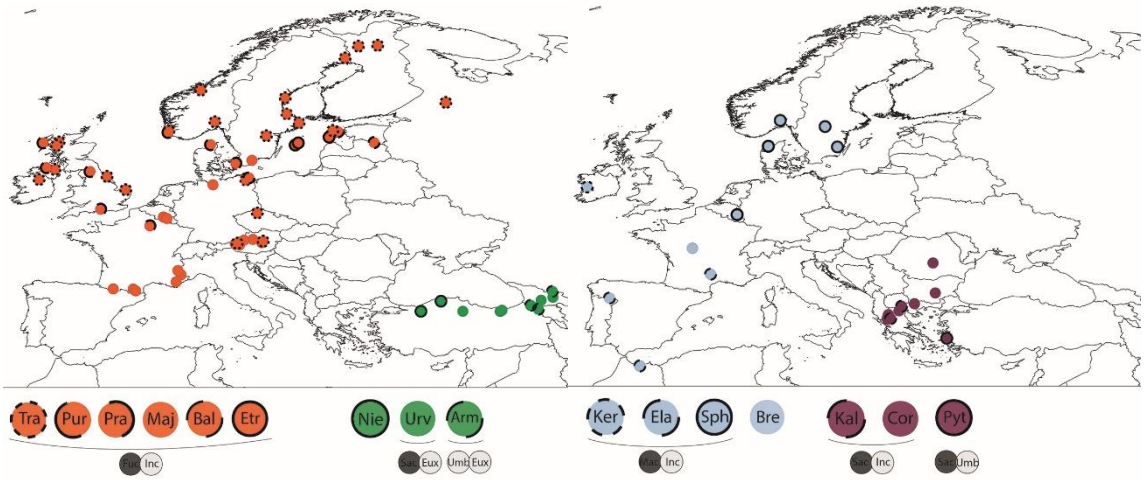


FIGURE 2. Maps showing the sampling locations of 95 allotetraploid *Dactylorhiza* individuals analyzed using RADseq. Arm = *D. armeniaca*, Bal = *D. baltica* (incl. *D. ruthei*), Bre = *D. brennensis*, Cor = *D. cordigera*, Ela = *D. elata*, Etr = *D. elatior*, Kal = *D. kalopissii* incl. *D.*

macedonica, Ker = *D. kerryensis*, Nie = *D. nieschalkiorum*, Maj = *D. majalis*, Pra = *D. praetermissa*, Pur = *D. purpurella*, Pyt = *D. pythagorae*, Sph = *D. sphagnicola* (incl. *D. calcifugiens*), Tra = *D. traunsteineri*, Urv = *D. urvilleana*. Lineages involved in the origin of the allopolyploids are also indicated: Eux = *D. euxina*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata*, Sac = *D. saccifera*, Umb = *D. umbrosa*. The map layer was extracted from GADM version 1.0 (available from www.gadm.org). Exact coordinates are given in Supplementary Table S1.

Depending on their degree of morphological, ecological, reproductive, and molecular distinctiveness, and on the taxonomic criteria applied, *Dactylorhiza* allopolyploids have been in the past variously treated as species, subspecies, varieties, or formae (Supplementary Table S3). In the following text, for simplicity, all polyploids are referred to with a formal name but without assignment to a particular taxonomic rank.

DNA Extraction, Library Preparation, and Sequencing

Total DNA was isolated from silica-dried leaves or flowers with bracts by following a cetyl trimethylammonium bromide (CTAB) procedure (Doyle 1990) or the DNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands). DNA was purified with the Nucleospin gDNA clean-up kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. RADseq libraries of 30–72 individuals per library, including repeats of individuals when necessary, were prepared following the protocol detailed in Paun et al. (2016), with the following modifications. Depending on the library, for each sample 100–400 ng DNA was used, except for the repeated individuals where only 50–100 ng was used. In general, only half the number of individuals per library and twice the amount of DNA was used for tetraploid accessions in comparison with diploids. The DNA content of samples was normalized within each library. The DNA was sheared with a Bioruptor Pico using 0.65 ml tubes (Diagenode) and three cycles of 30 sec ON and 60 sec OFF. The inline and index barcodes used differed from each other by at least three sequence positions. All RADseq libraries were sequenced as single-end 100 bp reads on an Illumina HiSeq platform at VBCF NGS Unit (https://www.vbcf.ac.at/ngs), Vienna, Austria.

In addition, one individual of *D. fuchsii* from Austria (accession 2144) was used to construct a whole genome sequencing library using a TruSeq DNA PCR-Free Library Kit (Illumina Inc) according to manufacturer's instructions. This was sequenced as a spike-in pair-end 125 bp reads on an Illumina HiSeq at VBCF NGS Unit and used to build a reference plastid genome with Fast-Plast v. 1.2.6 (available at: https://github.com/mrmckain/Fast-Plast) em-

ploying Asparagales in the --bowtie_index option. Plastome annotations were performed online using GeSeq (Tillich et al. 2017), guided with NCBI reference annotations for the orchidoid orchids *Habenaria pantlingiana* and *H. radiata*.

Filtering SNPs from RADseq Data

To allow for phylogenomic investigations at the genus level across non-model diploids and polyploids (i.e., in the absence of a reference genome), we optimized a bioinformatics pipeline by building *de novo* a synthetic reference from all diploid accessions, later mapping both diploids and polyploids to this reference. Finally, we called and filtered variants across all samples and compared genotype-based results to those from genotype-free methods of allele frequency estimations.

The raw reads were first demultiplexed to sublibraries based on index reads using BamIndexDecoder v. 1.03 (included in Picard Illumina2Bam package, available from http://gq1.github.io/illumina2bam/). These were further processed with STACKS v. 1.44 (Catchen et al. 2013), starting with demultiplexing of individuals based on inline barcodes via PROCESS_RADTAGS. The simultaneous quality filtering was performed to remove any read with uncalled bases or average low quality scores, as well as to rescue barcodes and cut sites with a maximum of one mismatch relative to expectations. RADseq loci were initially built *de novo* for the set of diploid individuals with DENOVO_MAP.PL in STACKS. The settings used in this step were first optimized to maximize the likelihood of ortholog recovery across species while avoiding the collapse of paralogs (i.e., maximize the number of polymorphic loci, that contain maximum ten SNPs and are covered in at least 90% of individuals), following Paun et al. (2016). The final settings used required at least six reads to create a stack (m), allowing for maximum one mismatch when merging the loci within individuals (M) and also among individuals when building the catalog (n). The setting allowing for indels was tested, but it did not improve significantly the number of loci recovered and was abandoned. Next we produced a whitelist of polymorphic loci with up to 15 SNPs and present in at least 50% of individuals with EXPORT_SQL.PL in STACKS. A consensus for each locus has been retained as individual chromosomes to produce a reference for further analysis.

In the next step, the raw reads of diploids and polyploids were mapped back to this diploid-derived reference using BOWTIE2 v. 2.2.6 (Langmead and Salzberg 2012) with default settings. Two approaches have been used for further analyses. First, genotypes were called with REF_MAP.PL and POPULATIONS in STACKS using default settings. VCFTOOLS v. 0.1.14 (Danecek et al. 2011) was further used to filter the resulting dataset to retain SNPs

covered in at least 75%, 90%, or 95% of the individuals, depending on the downstream analysis (see below and Supplementary Table S4). The filtered vcf file was converted to other formats with PGDSpider v. 2.0.8.2 (Lischer and Excoffier 2011).

As an alternative, genotype posterior probabilities have been calculated in ANGSD v. 0.921-26 (Korneliussen et al. 2014) based on the SAMtools model starting from the mapped bam files on the diploid-derived reference. The settings used allowed for a maximum of 50% missing data in order to process a site, a mapping quality threshold of 20 (minMapQ 20), a base quality threshold of 20 (minQ = 20), and retained only variable positions with a high likelihood (p-value < 1e-6).

Finally, the demultiplexed RADseq reads were also mapped to the *de novo* assembled plastid reference genome of *Dactylorhiza fuchsii* with the MEM algorithm of BWA v. 0.7.12-r1039 (Li and Durbin 2009). After sorting the aligned SAM files by coordinates, adding read groups with Picard tools v. 2.9.2 (available from http://broadinstitute.github.io/picard), and realigning around indels with the Genome Analysis Toolkit v. 3.7.0 (GATK; McKenna et al. 2010), variants were further called using the GATK HaplotypeCaller in the ERC GVCF mode with a sample ploidy of one. Finally, joint genotyping was performed on the resulting gVCF files with the GenotypeGVCFs tool of GATK.

Diploid Phylogenomic Analyses

Based on a high-quality SNPs dataset for diploid and autotetraploid individuals (i.e., *D. maculata*), we constructed phylogenetic trees with RAxML v. 8.2.9 (Stamatakis 2014) using an algorithm with 1,000 rapid bootstrap replicates while searching for the best-scoring ML tree under the GTR (general time reversible) model of nucleotide substitutions, disabled rate heterogeneity among sites (i.e., the GTRCAT model) and an ascertainment bias correction of the likelihood following the method proposed by Lewis (2001) for datasets of concatenated SNPs, and 1,000 alternative runs on distinct starting trees. The obtained relationships within the *D. fuchsii-maculata* clade made us investigate further a potential effect of gene flow between *D. fuchsii* and *D. maculata* in Central Europe (Ståhlberg and Hedrén 2009, 2010) on the recovered topology. Hence, RAxML analyses have also been performed with the same settings on a matrix that included diploids plus only the western autopolyploid *D. maculata.* The trees were visualized with FIGTREE V.1.4.2 (available from http://tree.bio.ed.ac.uk/software/figtree/). To further explore relationships between diploid *Dactylorhiza* individuals plus autotetraploid *D. maculata*, a FINERADSTRUCTURE v. 0.2 coancestry analysis (Malinsky et al. 2018) was performed on haplotype information of loci

found in at least 95% of the individuals. This input file was created by using EXPORT_SQL.PL in STACKS.

For two major clades in the RAxML phylogenetic tree that contain putative parents for allopolyploids, we separately performed Bayesian species delimitation analyses (Leaché et al. 2014). The analyses have been run separately in order to decrease computational costs. These two analyses focused on the *D. fuchsii-maculata* clade containing *D. fuchsii*, *D. saccifera*, and *D. foliosa* (plus *D. incarnata* as sister to the rest) and, respectively, on the *D. incarnata-euxina* clade containing *D. incarnata*, *D. umbrosa*, and *D. euxina* (plus *D. fuchsii* as sister to the rest). For this purpose, two smaller datasets were created including only a single biallelic SNP per RAD locus and allowing for a maximum of 10% missing data per locus. To further minimize computational cost, we included in these analyses only a few representative individuals, but ensured that at least three individuals per smallest tested group were present (Supplementary Table S1). The vcf files were converted with PGDSpider to nexus format, from which xml files were created with BEAUti v. 2.4.5 (Bouckaert et al. 2014) and edited to a path sampling analysis. Bayesian species delimitation analyses were performed in SNAPP v. 1.2.5 (Bryant et al. 2012) using 12 initialization steps and one million chain-lengths for each model.

To build a species tree, SNAPP analyses further included all diploid *Dactylorhiza* species (as defined with the species delimitation analyses above) with available data for at least two individuals (i.e., all but *D. aristata*), based on a dataset including single SNPs for each RAD tags with a chain-length of ten million, saving a tree every 1,000th generation. Convergence of the SNAPP analysis was evaluated from the log-file with TRACER v. 1.6 (Rambaut et al. 2018). We summarized the trees from SNAPP and calculated posterior probabilities of each clade with TREEANNOTATOR v. 1.8.3. After removing the first 10% of trees as burn in, the SNAPP trees were visualized as a cloudogram with DENSITREE v. 2.2.6 (Bouckaert and Heled 2014). To infer ages for divergence events, the species tree was calibrated with a general mutation rate from *Arabidopsis* (7×10^{-9} base substitutions/site/generation; Ossowski et al. 2010) and an average generation time for *Dactylorhiza* of 5.8 years (Øien and Moen 2002). The results were rescaled according to the total length of investigated sites within the included loci and the total number of polymorphic sites across this length.

Finally, for each diploid individual, a relative measure of inbreeding *F*, derived from a method of moments, was calculated with VCFTOOLS (--het option) and visualized for each species as vioplots with R v. 3.2.3 in RSTUDIO v. 1.0.44 (RStudio Team 2015). Resembling the classic population genetic measure F_{IS}, this per-individual estimate of *F* can take values

from -1 (maximum outcrossing) to 1 (maximum inbreeding), but should be regarded only as a relative measure of inbreeding, as it is based on the heterozygosity of variable loci within the dataset (i.e., it derives from a vcf file) and is not calculated over populations.

Allopolyploid Analyses

To assess relationships between allopolyploids, genotype posterior probability estimated in ANGSD were used to perform a PCA analysis with PCAngsd v. 0.95 (Meisner and Albrechtsen 2018). A PCoA was also constructed based on the called genotypes with the dartR package (Gruber et al. 2018) based on a Euclidean distance in R allowing for a maximum of 10% missing data. Further, a phenetic network for all allopolyploid accessions was constructed with SPLITSTREE4 (Huson and Bryant 2005) using the called SNPs and Jukes Cantor distances. To track the parentage of each polyploid lineage, we estimated the pairwise unadjusted Ajk statistic (--relatedness option; Yang et al. 2010) in VCFTOOLS based on called genotypes. Relatedness coefficients, in contrast to distance metrics, do not present any bias with respect to ploidy and can be used on a mixed ploidy dataset (Meirmans et al. 2018). Ajk is expected to be one if an individual is compared to itself, be around zero between individuals that are part of the same population, and take negative values for other comparisons. Here we inferred the diploid/autotetraploid lineages with the highest relatedness to be the most likely parents. The distribution of relatedness between each polyploid group and each of the potential parentals was presented as vioplots in R. The calculations were repeated with PCAngsd to incorporate genotype uncertainty when estimating a covariance matrix between diploid and polyploid accessions based on genotype posterior probabilities. To test the significance of the observed difference between overlapping relatedness distributions, statistical tests were performed with Mann-Whitney-Wilcoxon tests in R (command wilcox.test()) because distributions were not necessarily normally distributed. To determine from which side each allopolyploid received the plastid genome (i.e. the maternal parent), uncorrected-p pairwise distances between the allotetraploids and their putative parents were estimated with SPLITSTREE based on the plastid vcf file. These were transformed into similarity measures and the parent with the highest plastid similarity was inferred to be maternal.

As an estimate of relative age of each allopolyploid lineage, the number of private alleles per allopolyploid compared with its sibling lineages was calculated based on vcf files with POPULATIONS in STACKS. The estimates of private alleles were finally corrected for small sample sizes by multiplication with a factor ((n+1)/n), where n represents the sample size of the respective lineage.

RESULTS

After demultiplexing and filtering the raw reads, our data averaged 1.6 million high-quality reads per diploid individual (s.d. 1.3 million) and 1.8 million (s.d. 1.0 million) per polyploid individual. These data have been deposited in the NCBI Short Reads Archive (BioProject ID PRJNA489792, SRA Study ID SUB4486615). After parameter optimization in the *de novo* assembly pipeline of STACKS and filtering as described above, we retained 2,696 polymorphic loci of 94 bp each in the 'diploid' reference that was further used for mapping. On average, the diploid and polyploid reads had a mapping success to this reference of 21.5% and 23%, respectively, to a final average coverage after running the STACKS pipeline of 194x and 215x respectively. The raw SNP dataset obtained from REF_MAP.PL for all individuals allowing for 50% missing data contained 47,220 SNPs and no indels (Supplementary Table S4). After filtering, the direct-estimation approach implemented in ANGSD yielded information retained on 17,935 variable sites for the allotetraploids only, and 25,354 in a dataset that also included the putative parentals.

The size of the *de novo* assembled plastid genome for *D. fuchsii* is 154,007 bp (GenBank Accession number XXX). On average, 6,354 RADseq reads per accession mapped to the plastome. After variant calling, 518 SNPs were retained across the plastid dataset that included the allopolyploids and their potential parents.

Diploid Phylogenomic Analyses

The RAxML phylogenetic analyses performed on a wide array of Eurasian *Dactylorhiza* diploids and the autotetraploid *D. maculata* distinguishes nine well-supported terminal *Dactylorhiza* clades, together with one less supported clade (represented by western *D. saccifera* individuals) and the single analyzed individual of *D. aristata* on a separate branch (Fig. 3 and Supplementary Fig. 1). Four outgroup branches representing related taxa confirmed previous assertions of a monophyletic *Dactylorhiza*. A major clade, here termed the *D. fuchsii-maculata* clade, including the diploids *D. fuchsii*, *D. saccifera*, and *D. foliosa*, was highly supported (bootstrap support [BS] 100%). When only the accessions of *D. maculata* from Western Europe are included in the analysis (i.e., excluding *D. maculata* from other parts of the species' distribution), *D. fuchsii* appears sister to a paraphyletic *D. saccifera* (Fig. 3). *Dactylorhiza saccifera* splits into an eastern clade, containing individuals sampled around the Black Sea, and a western clade comprising individuals from France and Italy. When all *D. maculata* accessions are included in the RAxML analysis (Supplementary Fig. S1), *D. fuchsii*

appears sister to a *D. maculata/D. foliosa* clade, but this relationship has very low support (BS 53%). The *D. incarnata-euxina* clade (including also *D. umbrosa*) is monophyletic and fully supported (Fig. 3), and appears as sister to the *D. fuchsii-maculata* clade but with low support (BS 50%). Finally, *D. viridis* is placed as sister to all other *Dactylorhiza* species, which form a monophyletic clade albeit tentatively, with 73% BS (Fig. 3).

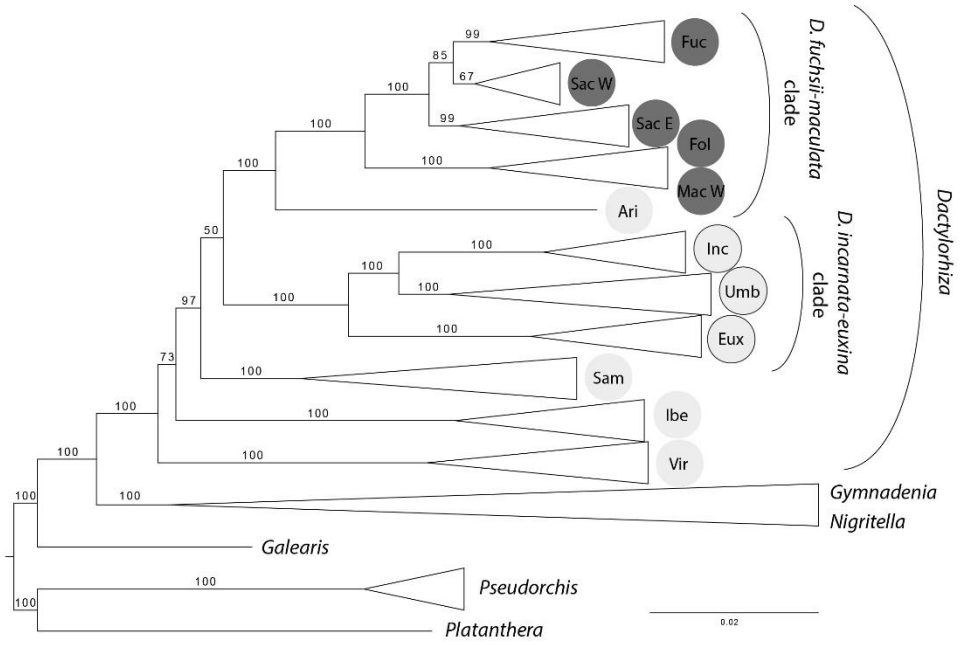


FIGURE 3. Best-scoring maximum likelihood phylogenetic tree of a dataset of 20,713 SNPs, comprising 89 diploid *Dactylorhiza* individuals. Nine autotetraploid *D. maculata* accessions (Mac W) from the western part of its distribution and 18 outgroup individuals from closely related genera are also included. Ari = *D. aristata*, Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Ibe = *D. iberica*, Inc = *D. incarnata*, Sac E = eastern *D. saccifera*, Sac W = western *D. saccifera*, Sam = *D. sambucina*, Umb = *D. umbrosa*, Vir = *D. viridis*.

The FINERADSTRUCTURE heatmap (Supplementary Fig. S2) complements well the RAxML results, pointing to a clear distinction between the *D. fuchsii-maculata* clade and the *D. incarnata-euxina* clade. However, the heatmap shows the *D. incarnata-euxina* clade to share higher coancestry with *D. sambucina*, *D. iberica*, and *D. viridis* than with the *D. fuchsii-maculata* clade. In general, *D. viridis*, *D. iberica*, *D. euxina*, *D. umbrosa*, *D. incarnata*, and *D. maculata-D. foliosa* each show high within-lineage coancestry. Further high between-species coancestry, viewed as signal for a relatively high gene flow, can also be identified, for example between *D. fuchsii* and *D. maculata*, but also between *D. umbrosa* and *D. incarnata*, the latter rather pointing to a recent divergence between these lineages, given their largely allopatric distribution.

From the species delimitation analysis (Fig. 4a), the highest marginal likelihood and best model concerning the *D. fuchsii-maculata* clade (including 11 individuals and 1,224 independent SNPs) was the split model distinguishing the three diploids: *D. fuchsii*, *D. foliosa*, and *D. saccifera*. The highest marginal likelihood and the best model concerning the *D. incarnata-euxina* clade (including 12 individuals and 1,321 SNPs, each on a different RADtag) was when *D. incarnata*, *D. umbrosa*, and *D. euxina* were recognised (Fig. 4a).

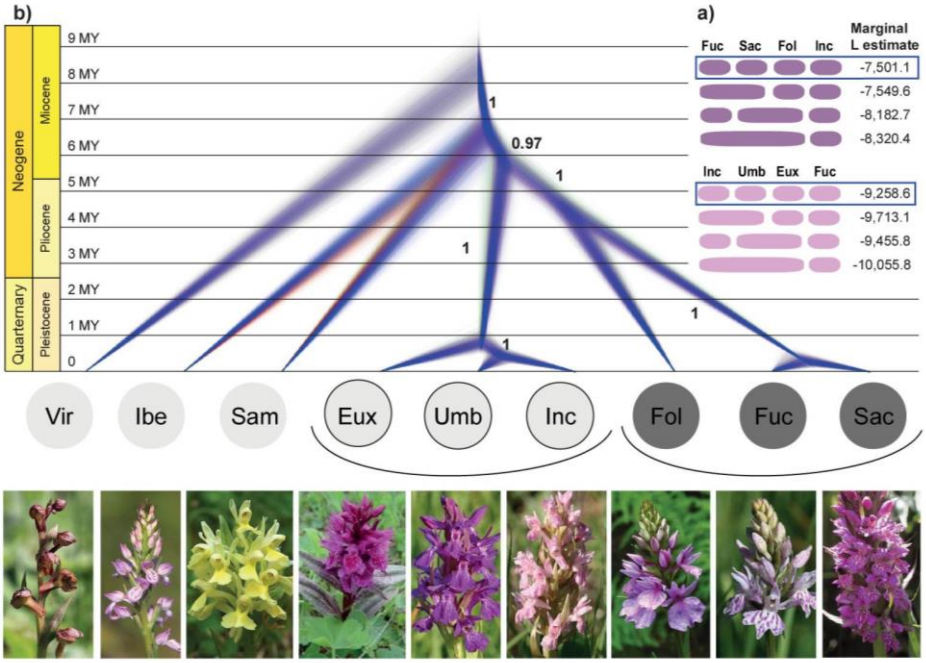


FIGURE 4. Results of coalescent-based phylogenetic analyses in SNAPP (Bryant et al. 2012) for diploid *Dactylorhiza*. **(a)** Species delimitation models tested for two main *Dactylorhiza* clades: *D. fuchsii-maculata* and *D. incarnata-euxina*, respectively. The preferred models are boxed. **(b)** Cloudogram of 9,000 trees obtained for nine *Dactylorhiza* diploid species for which at least two accessions were available. Posterior probabilities higher than 0.95 are indicated for the relevant clades. The age estimates were obtained by rescaling substitutions per site with an *Arabidopsis* mutation rate and an average generation time for *Dactylorhiza* of 5.8 years (for more details see Methods). Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Ibe = *D. iberica*, Inc = *D. incarnata*, Sac = *D. saccifera*, Sam = *D. sambucina*, Umb = *D. umbrosa*, Vir = *D. viridis*. Photos: Sven Birkedal and Mikael Hedrén.

In the SNAPP species tree constructed for diploid *Dactylorhiza* (Fig. 4b), a similar overall topology was found as that recovered by the RAxML tree (Fig. 3). Upon analyzing the SNAPP .log file in TRACER, all ESS values proved to exceed 200 except one. Although the exception scored 143, visual inspection of its trace suggested that it also had converged. The analysis of the SNAPP tree file with the TREE SET ANALYZER indicated that the 95%

highest posterior densities (HPD) contained only three main tree topologies out of a total of 15 recovered. These three main topologies all placed *D. viridis* as sister to the rest *Dactylorhiza*, but differed in the positions of *D. sambucina* and *D. iberica*; 39.6% of trees place *D. iberica* as subsequent sister to the rest, including *D. sambucina* as sister to all others, whereas 29.8% of the trees have *D. sambucina* as subsequent sister to the rest, including *D. iberica* as sister to all remaining *Dactylorhiza*. Finally, 27.9% of trees show *D. sambucina* and *D. iberica* forming a clade that is sister to all remaining *Dactylorhiza* except *D. viridis*. The dated SNAPP tree indicates that the extant lineages in the crown group of *Dactylorhiza* started to diverge from one another in the late Miocene (Fig. 4b). The Mediterranean-Pontic *D. iberica* appears to have started to diverge around the period of the Messinian salinity crisis, a period when the Mediterranean Sea was closed and nearly dried out. Whereas the *D. fuchsii-maculata* clade seems to have split from the *D. incarnata-euxina* clade in the early Pliocene, most of the splits within each of these major clades are inferred as having occurred within the last million years.

The highest relative per-individual inbreeding (Supplementary Fig. S3) was found in *D. incarnata* and the Madeiran endemic *D. foliosa*, whereas the autopolyploid *D. maculata* and diploids *D. umbrosa*, *D. euxina*, *D. sambucina*, and *D. viridis* showed lower (but still mostly positive) values for inbreeding within the polymorphic positions retained in the vcf file.

Allopolyploid Analyses

The genotype-free PCA of all allopolyploid individuals (Fig. 5a) generally separates on the first two axes (explaining 24% of the total variation) the different allopolyploid lineages, apart from those produced by *D. fuchsii* × *D. incarnata* (orange symbols in Fig. 5). When plotting only the allopolyploids derived from *D. fuchsii* and *D. incarnata* (Fig. 5b; Supplementary Fig. S4a), the taxa become better separated. A highly similar pattern is observed in the genotype-based PCoAs drawn with dartR (Supplementary Fig. S4b-c) and in the phenetic network produced by SPLITSTREE (Supplementary Fig. S5). The SPLITSTREE analysis separates the allopolyploids produced by *D. fuchsii* and *D. incarnata* by shorter distances than other sibling allopolyploids. Some *D. traunsteineri* individuals from the Alps are positioned within the cluster of *D. majalis*, in close proximity to Alps accessions of the latter. *Dactylorhiza traunsteineri* shows a clustering that follows its geographical distribution, separated into two main groups - a northwestern group in Britain and W Norway, and a central-eastern containing the other populations.

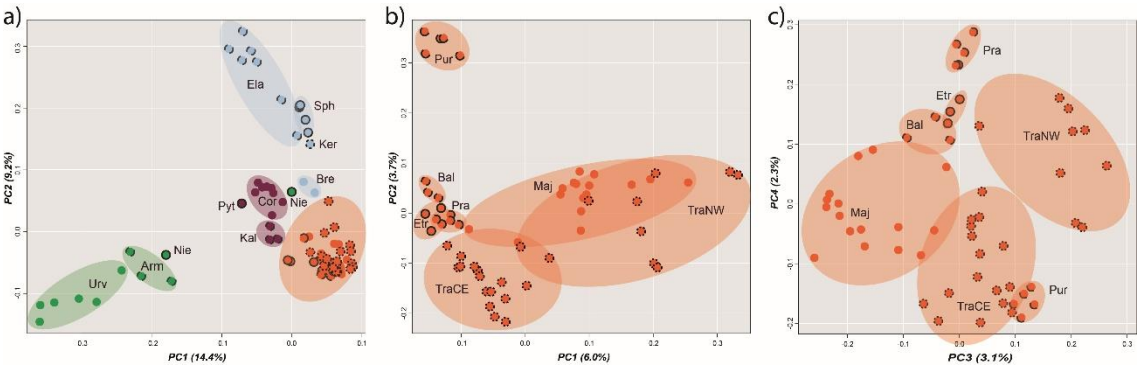


FIGURE 5. PCAs performed on genotype likelihoods at 17,935 variable sites using PCAngsd (Meisner and Albrechtsen 2018) on allopolyploid *Dactylorhiza* individuals. (a) Clustering of all allopolyploid individuals. (b-c) Detailed analysis on the allopolyploids produced by *D. fuchsii* and *D. incarnata*. Arm = *D. armeniaca*, Bal = *D. baltica* (incl. *D. ruthei*), Bre = *D. brennensis*, Cor = *D. cordigera*, Ela = *D. elata*, Etr = *D. elatior*, Kal = *D. kalopissii* (incl. *D. macedonica*), Ker = *D. kerryensis*, Nie = *D. nieschalkiorum*, Maj = *D. majalis*, Pra = *D. praetermissa*, Pur = *D. purpurella*, Pyt = *D. pythagorae*, Sph = *D. sphagnicola* (incl. *D. calcifugiens*), Tra = *D. traunsteineri* (cont. = continental, brit. = British), Urv = *D. urvilleana*. Colors indicate different parental pairs for the allopolyploids. The symbols and colors follow Fig. 2.

The results of our relatedness analyses are also highly consistent between estimates with genotype-based (Table 1, Fig. 6, and Supplementary Figs. 6 and 7) and genotype-free (results not shown) methods. They indicate the following relationships between parental and derived lineages:

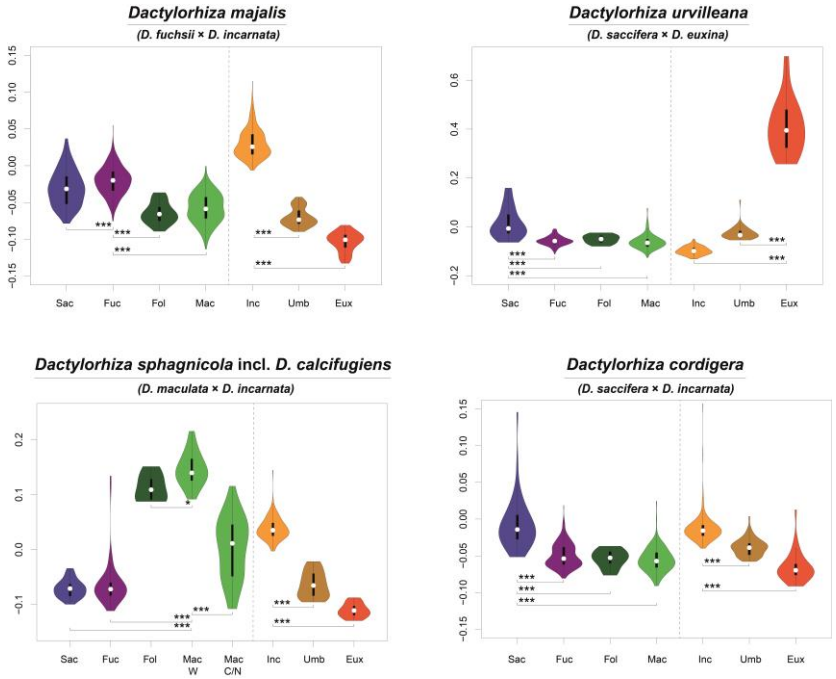


FIGURE 6. Examples of vioplots of relatedness of allopolyploids to potential ancestral genomes (diploid or autotetraploid). Plots of relatedness of the allopolyploids to members of the *D. fuchsii-maculata* clade are shown to the left of dashed vertical bars, and to members of the *D. incarnata-euxina* clade to the right. Stars indicate significantly different distributions (* $p < 0.05$; *** $p < 0.001$). Colors follow Fig. 1. Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata*, Sac = *D. saccifera*, Umb = *D. umbrosa*. Plots for the rest of the allopolyploids are given as Supplementary Figures S5 and S6.

(i) *Dactylorhiza fuchsii* and *D. incarnata* (or their ancestors) produced several central-northwestern European allopolyploids, including *D. baltica* (incl. *D. ruthei*), *D. elatior, D. majalis*, *D. praetermissa*, *D. purpurella*, and *D. traunsteineri*. *Dactylorhiza fuchsii* was the maternal parent in all of these crosses.

(ii) *Dactylorhiza maculata* (always as maternal lineage) and *D. incarnata* (or their ancestors) hybridized to form two northwestern European polyploids (*D. sphagnicola* incl. *D. calcifugiens*, *D. kerryensis*) and one southwestern-central European polyploid (*D. elata*).

(iii) *Dactylorhiza saccifera* and *D. incarnata* (or their ancestors) acted as parents for the southeastern European *D. cordigera* and *D. kalopissii* incl. *D. macedonica. Dactylorhiza saccifera* acted as the maternal parent in these crosses.

(iv) *Dactylorhza saccifera,* as a maternal parent, hybridized also with *D. umbrosa* to form *D. pythagorae*, another southeastern allopolyploid, endemic to Samos.

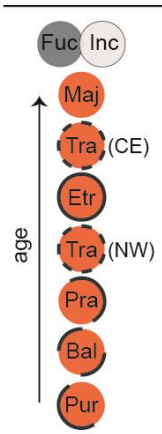
(v) *Dactylorhiza euxina* and *D. umbrosa* (or their ancestors) formed the endemic Turkish-Caucasian polyploid *D. armeniaca*, where *D. umbrosa* was the maternal parent.

(vi) Finally, *Dactylorhiza urvilleana*, with a distribution from N Turkey to N Iran, was produced by *D. euxina* and *D. saccifera* (or their ancestors), *D. saccifera* acting as the maternal parent.

The estimated private alleles, corrected for small sample size for each allopolyploid strictly against its sibling lineages (Table 1), range between 1,620 and 221 for the group *D. fuchsii* × *D. incarnate*, and between 7,454 and 860 for the group *D. maculata* × *D. incarnata*. Within the larger family of *D. fuchsii* × *D. incarnata* the number of private alleles correlates significantly with both the relatedness values towards *D. fuchsii* ($p < 0.001$) and towards *D. incarnata* ($p < 0.01$).

TABLE 1. Number of analyzed individuals (N), private alleles corrected for small sample size (P) calculated against sibling allopolyploids, average relatedness values of each allopolyploid

towards each of its parents (R) and average plastid similarity (S). Arm = *D. armeniaca*, Bal = *D. baltica* (incl. *D. ruthei*), Cor = *D. cordigera*, Ela = *D. elata*, Etr = *D. elatior*, Kal = *D. kalopissii* (incl. *D. macedonica*), Ker = *D. kerryensis*, Maj = *D. majalis*, Pra = *D. praetermissa*, Pur = *D. purpurella*, Pyt = *D. pythagorae*, Sph = *D. sphagnicola* (incl. *D. calcifugiens*), Tra = *D. traunsteineri*, and Urv = *D. urvilleana*. The lineages involved in the origin of the allopolyploids are also indicated: Eux = *D. euxina*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata*, Sac = *D. saccifera*, Umb = *D. umbrosa*.



	N	P	R_{Fuc} / R_{Inc}	S_{Fuc} / S_{Inc}
Maj	13	1620	-0.020 / 0.030	0.921 / 0.905
Tra (CE)	13	1535	-0.008 / 0.030	0.917 / 0.894
Etr	3	605	0.001 / 0.040	0.929 / 0.916
Tra (NW)	7	434	-0.025 / 0.034	0.925 / 0.901
Pra	5	365	0.008 / 0.037	0.920 / 0.899
Bal	3	264	0.006 / 0.051	0.913 / 0.897
Pur	5	221	0.003 / 0.052	0.912 / 0.889

	N	P	R_{MacW} / R_{Inc}	S_{MacW} / S_{Inc}
Ela	7	7454	0.039 / -0.020	0.940 / 0.910
Sph	5	1620	0.146 / 0.040	0.929 / 0.895
Ker	1	860	0.125 / 0.035	0.934 / 0.900

	N	P	R_{Sac} / R_{Inc}	S_{Sac} / S_{Inc}
Cor	7	7706	-0.008 / -0.013	0.949 / 0.911
Kal	3	1567	0.037 / 0.033	0.917 / 0.893

	N	P	R_{Sac} / R_{Umb}	S_{Sac} / S_{Umb}
Pyt	1	-	-0.006 / 0.034	0.955 / 0.915

	N	P	R_{Sac} / R_{Eux}	S_{Sac} / S_{Eux}
Urv	6	-	0.016 / 0.408	0.939 / 0.898

	N	P	R_{Umb} / R_{Eux}	S_{Umb} / S_{Eux}
Arm	2	-	0.094 / 0.406	0.932 / 0.902

DISCUSSION

Effectiveness of the RADseq Method for Resolving Reticulate Relationships

Despite the common occurrence and great evolutionary importance now attributed to hybridization and polyploidy (Linder and Rieseberg 2004; Mallet 2007; Giraud et al. 2008; Van de Peer et al. 2017), the subsequent restructuring of the genome complicates greatly phyloge-

netic inference. Several protocols have been recently proposed to process phylogenetic data that include polyploid and hybrid accessions (e.g., Rothfels et al. 2017; Gregg et al. 2017; Morales-Briones et al. 2018). We take here a phylogenomic approach to analyze a thoroughly sampled NGS dataset for *Dactylorhiza*, one of the most complex diploid-tetraploid plant genera examined to date. Our approach included building a reference database of RADseq loci from diploids only, and using it to further process both diploid and polyploid accessions. Directly estimating allele frequencies without calling genotypes has been shown to improve inference from whole genome sequencing data (Nielsen et al. 2011; Han et al. 2014). Retaining genotype uncertainties in further analyses, genotype-free estimates of allele distributions may avoid allele drop-out in polyploids, otherwise caused by low dosage (i.e., one of four alleles). Such alleles are more likely to be maintained in the data through genotypic probabilities specific for genotype-free methods. However, our results obtained from genotype-based and genotype-free methods are highly congruent, likely due to the high-coverage data available for this study.

The tens of thousands variable positions derived from RADseq helped us to distinguish among ten Eurasian *Dactylorhiza* diploids, one autopolyploid and 16 allopolyploids, and to disentangle the putative parents of the latter. The RADseq phylogenetic trees of the diploids and autopolyploid *Dactylorhiza* correspond largely with prevailing taxonomies based on morphology and low-density molecular studies, but significantly improve phylogenetic resolution and thus the estimated divergence times. Additionally, the RADseq data on the allopolyploids provides much new and refined data on recurrent allopolyploidization events in the genus, plus a deeper understanding of phylogeographic patterns and relative ages of allopolyploidization events. The present results are compared with former single gene/low-density molecular studies in Supplementary Table S2.

Diploid Phylogenomic Analyses

The phylogenetic analyses performed on the putative parental Eurasian *Dactylorhiza* diploids distinguish two major clades (i.e., the *D. fuchsii-maculata* clade and the *D. incarnata-euxina* clade) and the subsequent sister lineages. The phylogenetic trees obtained confirm that *D. viridis* and *D. iberica* are the earliest diverging species within *Dactylorhiza* (Bateman et al. 2018a). As in Brassac and Blattner (2015), the phylogenetic analysis in this study was performed both on a concatenated data set as well as treating each marker independently with a coalescent approach (Figs. 3-4). Both of these methods have their shortcomings. In the case of concatenated datasets, inconsistencies arise in presence of incomplete lineage sorting and hy-

bridization, sometimes artificially inflating statistical support for incorrect topologies as the data set grows larger (Degnan 2013; Liu et al. 2015). In the case of coalescent methods, the choice of markers and computational time are complicating factors (Liu et al. 2015; Fernandez-Mazuecos et al. 2017). However, good support for the relationships inferred here among the diploid taxa is given by the consistency of results between the contrasting approaches used.

The European flora has experienced dramatic contraction-expansion phases associated with Pleistocene glaciation cycles (Hewitt 1996, 2000), and plants distributed in northern Europe often exhibit lower levels of genetic diversity compared with plants in the south that persisted locally throughout the cold periods (Taberlet et al. 1998). Within the *D. incarnata*/*D. euxina* lineage, the fact that *D. umbrosa* and *D. euxina* reside in the southeast, in the Near East and the Caucasus, and are genetically variable, whereas the European *D. incarnata* is genetically depauperate (Fig. S3), suggests that *D. incarnata* has a southeastern origin. *Dactylorhiza incarnata* may have experienced strong genetic bottlenecks already during its relatively recent (Fig. 4), initial expansion from Asia into Europe (Hedrén 2009). It is also possible that the restricted diversity of *D. incarnata* can be explained by more recent loss of genetic diversity either in southern European refugia during the last ice age, or during post-glacial recolonization of northern parts of Europe (Fig. S3; Balao et al. 2016, 2017). The restricted diversity in *D. incarnata* is reinforced by comparatively high levels of inbreeding and a patchy/localized distribution pattern (Pedersen 2009; Hedrén and Nordström 2009; Naczk et al. 2016). The fact that *D. euxina* has hybridized with *D. umbrosa* to form an allopolyploid with disomic inheritance, *D. armeniaca* (Hedrén 2003), shows that the two diploid genomes are still structurally differentiated in spite of their recent divergence (Fig. 4).

We estimated the split between the *D. fuchsii/D. saccifera* lineage and the *D. maculata/D. foliosa* lineage to have occurred at approximately four Ma. In the former, *D. saccifera* has a southern distribution stretching from the central Mediterranean area to the Caucasus. As these southern regions experienced relatively stable conditions throughout the Pleistocene glaciations, regional populations of *D. saccifera* may have persisted in their present areas for long periods without much of gene exchange with other regional populations. In agreement with such a scenario, we found that accessions from Corsica and Italy were divergent from accessions from the Balkans and Asia Minor. In spite of its wide distribution in central and northern Europe, accessions of *D. fuchsii* were more coherent than those of *D. saccifera*. The paraphyletic appearance of *D. saccifera* in the ML tree (Fig. 3), and the finding that *D. fuchsii* is closest to western *D. saccifera*, suggests that *D. fuchsii* has been derived from a *saccifera*-like

ancestor in the west. Its relatively strong coherence and lack of geographic structure can be explained by a rapid expansion from a southern refugium after the last ice age. *Dactylorhiza fuchsii* exhibits high levels of genetic diversity (Fig. S3), a pattern consistent with low inbreeding levels, a relatively even distribution, and highly efficient gene flow between populations across its wide distribution.

Dactylorhiza foliosa is the only extant diploid member of the *D. maculata/D. foliosa* lineage. It has a deviating morphology and is today restricted to the island of Madeira. It is possible that it colonized Madeira soon after the emergence of the island ca. five million years ago, when the island was still linked to the Iberian peninsula by a chain of now submerged volcanic islands that could have acted as stepping stones for the dispersal from the Eurasian continent (Geldmacher et al. 2000; Fernández-Palacios et al. 2011). The endemic nature of *D. foliosa* is reflected in its restricted genetic diversity (Fig. S3). *Dactylorhiza maculata* is invariably tetraploid, but exhibits considerable geographically correlated variation. The western European population is genetically coherent and closely related to *D. foliosa* (Fig. S6), suggesting that they have originated from the same ancestral stock. In contrast, central and northern European accessions of *D. maculata* are more genetically divergent and could include autotetraploids derived from slightly different diploid ancestors. Their wider diversity could also be explained by occasional hybridization and introgression from *D. fuchsii*, a process that may have been facilitated by the presence of some autotetraploid populations of *D. fuchsii* in Central European mountain regions, especially in the Alps (Ståhlberg and Hedrén 2009, 2010). In contrast with *D. foliosa*, *D. maculata* displays high genetic diversity (Fig. S3), as expected in an established autopolyploid with polysomal inheritance (Otto 2007). Due to the close relatedness of *D. foliosa* to western *D. maculata*, the allopolyploids formed with western *D. maculata* as parent also show high relatedness values with *D. foliosa.* However, we propose that the widespread *D. maculate*, or a main land ancestor is the more likely parent for those non-Madeiran allopolyploids (e.g., *D. elata*; Fig. S6).

Allopolyploid Evolution

Allopolyploidy is of fundamental importance for understanding early angiosperm diversification (Stebbins 1980; de Bodt et al. 2005; Jiao et al. 2011; Van de Peer et al. 2017). However, few genomic studies have been performed in medium-aged polyploid complexes, such as the *Dactylorhiza* complex studied here. Available studies have been focused on (very) young polyploid complexes to study the genomic effects of the allopolyploidization process (Song et al. 1995; Soltis et al. 2014), or to estimate numbers of origins (Soltis et al. 2004).

However, such allopolyploids may not yet be established in natural habitats and should still contain two full genomes minimally affected by intergenomic recombination (Otto 2007). Other studies have been focused on ancient polyploid complexes with allopolyploids that have experienced millions of years of evolution, including karyotype evolution, climatic oscillations, differential gene loss, functional gene diversification, and concerted evolution (Adams and Wendel 2005). For instance, the *Gossypium hirsutum* group in the New World originated by an allopolyploidization event one to two Ma (Hu et al. 2015), and has itself given rise to multiple genetically distinct species.

The age of the allopolyploids analyzed in this study may range from those that have evolved in previously glaciated areas after the last ice age and which we can trace back to their precise parents, to those that evolved long before the last ice age and whose exact parents may no longer be extant (but nonetheless related to the parental lineages that we have successfully identified). The times of origins for several current polyploids appears to be associated to recent glacial cycles also in other genera (Novikova et al. 2018). Given that the allopolyploids analyzed here are typically more closely similar to either of the *D. fuchsii/D. saccifera* and *D. foliosa/D. maculata* lineages and that these lineages diverged from each other at about 4 Ma, none of the allopolyploids is likely to be older than the split between these lineages. Moreover, several allopolyploids revealed a closer similarity to either *D. saccifera* or *D. fuchsii* as one of the parents, or to either *D. incarnata* or *D. umbrosa* as the other parent. As these closely related parental lineages in both cases diverged less than one million years ago, it is unlikely that any allopolyploid is older than that age. However, some of the allopolyploids are characterized by having common and geographically widespread plastid genomes that have not been identified in any extant member of the parental lineages (Pillon et al. 2007; Nordström and Hedrén 2008, 2009). We found here that such allopolyploids are also characterized by larger proportions of private alleles than allopolyploids that share plastid genomes with present-day members of the parental lineages. Good examples of potentially relatively old allopolyploids include *D. elata* in SW Europe and NW Africa and *D. cordigera* in SE Europe. Since these species are most closely related to parental lineages in the same geographic areas, they probably originated in vicinity of the regions where they are still found today. These regions were less affected by climate changes during Pleistocene glaciations, and it is possible that *D. elata* and *D. cordigera* date back through multiple glaciation cycles.

In contrast, some of the studied allopolyploids express low numbers of private alleles compared with sibling lineages, have high similarity values to particular parental species found within their distribution areas, and have their entire distributions restricted to parts of

Europe that were glaciated during the last ice age. We conclude that such allopolyploids are probably of postglacial origins, a conclusion supported by findings that they contain no major allozyme, plastid haplotype, or ITS variants other than those present in their putative extant parents (e.g., Hedrén 1996; Devos et al. 2003; Pillon et al. 2007). Good examples are provided by *D. purpurella* (*D. fuchsii* × *D. incarnata*) with an Atlantic distribution in NW Europe and *D. sphagnicola* (*D. maculata* × *D. incarnata*) with distribution from NW continental Europe to central Scandinavia. Species with more restricted distributions, such as *D. elatior* (*D. fuchsii* × *D. incarnata*) in the central Baltic area, may even have arisen within the past thousand years (Hedrén et al. 2012a).

The wide range of similarity values between allopolyploids and putative parents suggest that allopolyploids may be of different ages and that the exact ancestors of some of them may already have gone extinct or further evolved significantly themselves. Considering allopolyploids with origins from the *D. maculata* lineage, the genetically variable allopolyploid *D. elata*, confined to NW Africa and SW Europe, could have arisen from now extinct diploid members of the *D. maculata* lineage that must have been present in the same general area and that are still represented by the Madeiran endemic *D. foliosa*. In contrast, NW allotetraploids, including *D. sphagnicola* and *D. kerryensis*, originated from within the western European type of the autotetraploid *D. maculata*.

On basis of ITS studies, it has been shown that the NW European *D. praetermissa* contains an ITS type shared only with the southern *D. saccifera*, suggesting that the latter is genetically close to the maternal parent of *D. praetermissa* (Pillon et al. 2007). However, our data show that *D. praetermissa* has a stronger overall similarity to *D. fuchsii* than to *D. saccifera* (Fig. 6), suggesting that *D. praetermissa* also has a post-glacial origin within its present distribution. In agreement with this modified scenario, more recent population-based studies show that *D. praetermissa* is variable in ITS, and that the ITS type formerly considered diagnostic is also present at low frequency in NW *D. fuchsii* (Pedersen and Hedrén 2010; Hedrén et al. 2011).

Young allopolyploids are not only restricted to previously glaciated areas. Also the southern *D. kalopissii* (*D. saccifera* × *D. incarnata*) and *D. armeniaca* (*D. euxina* × *D. umbrosa*), express high similarity values (Fig. 6) and agree in plastid haplotype with present-day members of the parental lineages (Hedrén et al. 2007).

Hybridization between Independently Formed Allopolyploids

Previous studies of the allopolyploid *D. brennensis*, which occupies a small area in central France, have revealed that it features highly divergent plastid genomes (Hedrén et al. 2012b). These genomes are not known from any present-day representatives of diploid parental lineages, but one of them is also identified in the regional population of *D. elata* in S France, and another one in *D. praetermissa* and *D. majalis*, in which it is widespread. Here, we confirmed the close relationship of *D. brennensis* to *D. elata* and a cluster including *D. majalis* and associated taxa (Figs. 5 and S4). These findings strongly suggest that *D. brennensis* has originated by hybridization between distinct allopolyploids with independent origins, a conclusion also in agreement with ITS/ETS sequence data (Devos et al. 2005).

Since *D. brennensis* is restricted to a small area and the parental groups are clearly divergent and easy to identify, *D. brennensis* may be of recent origin. However, similar hybridization events may also have taken place between other independently derived allopolyploids in the past, or repeatedly, and may explain the high internal diversity and the difficulty experienced in matching some of the investigated allopolyploids to their exact parents. One example is given by *D. majalis*, which has incorporated plastid genomes from both *D. fuchsii* and *D. maculata*, even some genomes that are not matched with those found in extant parents (Nordström and Hedrén 2009). *Dactylorhiza majalis* overlaps strongly with *D. traunsteineri* in genetic markers (Balao et al. 2016), which points towards extensive gene flow and suggests that the divergence between these two species is largely upheld by selection for specific habitats (Paun et al. 2010, 2011). Furthermore, a scenario of hybridization and gene flow between these species is also in agreement with the finding that *D. traunsteineri* exhibits a geographically structured pattern of variation (Figs. 5 and S4), populations from the European continent being somewhat differentiated from those in the NW (separated as *D. traunsteinerioides* by some authors; Bateman et al. 2011). Such a pattern could have arisen as a consequence of hybridization with different related polyploid species in different parts of the total distribution area. Other possible causes for a geographic variation pattern could be isolation by distance, secondary hybridization and introgression from local representatives of the parental lineages, or artificial aggregation of allopolyploids of independent origins into the same species due to morphological convergence.

Given a complex of multiple allopolyploid derivatives of several parental combinations and of different ages, and a widespread potential for hybridization and gene flow between them, this will drive the accumulation of genetic diversity, until the polyploids hold as much diversity as any of the parental lineages. Such a diverse polyploid complex may continue to

evolve, just like any diploid lineage and produce more-or-less well separated sibling daughter taxa.

CONCLUSIONS

In the present study we have analyzed allopolyploids ranging from perhaps a few thousand years old to allopolyploids that possibly evolved long before the last glaciation. Polyploids from such a time range present a unique opportunity to investigate recurrent allopolyploidization, which might be especially important in the establishment of young polyploids (Soltis and Soltis 1999; Mallet 2007). We find that crosses between the same parental species have occurred recurrently during the course of time to produce multiple independent allopolyploids. Similar patterns of recurrent polyploidization have been found in other polyploid complexes, such as *Achillea* (Guo et al. 2013), *Asplenium* (Perrie et al. 2010), *Leucaena* (Govindarajulu et al. 2011), and *Oryza* (Zou et al. 2015). Although certain features of genome evolution may follow common paths from repeated polyploidizations between the same parental combinations (Song et al. 1995; Soltis et al. 2009), stochasticity plays a major role in producing allopolyploid derivatives that differ in morphology and various traits of adaptive importance. The polyploid complex in *Dactylorhiza*, with a dozen independently derived allopolyploids, provides an excellent example of this process. The amount of diversity and relative age of the allopolyploids correlates well with predictions based on the postglacial history of the European flora (Hewitt 1996, 2000; Taberlet et al. 1998). The least diverse and obviously youngest allopolyploids are found in the northwestern and northern parts of the distribution (Bateman 2011), which were covered by ice sheets during the last glaciation, whereas more diverse and obviously older allopolyploids are distributed in more southern regions, which were less affected by Pleistocene glaciations.

We observed that with increasing age, an allopolyploid species will gradually acquire increased numbers of unique (i.e., private) alleles compared with their younger siblings, and also to their parents (i.e., decreasing relatedness values). We identified multiple possible causes to this pattern: (*i*) accumulation of novel alleles within allopolyploids by mutation; (*ii*) gradual change of ancestral lineages by accumulation of mutations such that the exact ancestors of the allopolyploid no longer exist; (*iii*) hybridization and merger with other independent allopolyploids with slightly different evolutionary backgrounds; and (*iv*) introgressive hybridization with the diploid, related lineages. These processes may occur partly at random, but may also be linked to the environmental selection pressures acting in the diverse habitats now occupied by the allopolyploids.

FUNDING

This research was funded by a START grant [Y661-B16] from the Austrian Science Fund (FWF) to OP, grants from the Royal Physiographic Society Lund and Nilsson-Ehle foundation to MH, and travel grants from the Botanical Society of Britain and Ireland and the UK Botanical Research Fund to RB.

ACKNOWLEDGMENTS

We thank Francisco Balao, Emiliano Trucchi, Thomas Wolfe, Gerald Schneeweis and Tobias Kaiser for their contribution to the results presented here. For assistance with collection and provision of samples we thank Maia Akhalkatsi, Kath Barrett, P. Batoušek, Sven Birkedal, Hugo de Boer, Andrew Bystrushkin, Nils Cronberg, Salvatore Cozzolino, Ian Denholm, R. Dundr, Walter Foelsche, Abdolbaset Ghorbani, Barbara Gravendeel, Sven Hansson, Jana Jersáková, Bengt Jonsell, Yong-No Lee, Richard Lorenz, Richard Manuel, Ruud van der Meijden, Sofie Nordström Olofsson, Henrik Pedersen, Tarmo Pikner, Paula Rudall, Fred Rumsey, Brendan Sayers, David Ståhlberg, Jess Stone, Åke Svensson, Sebastian Szepanski, Barry Tattersall, Daniel Tyteca, Erik Westberg, Finn Wischmann, and Shahin Zarre.

REFERENCES

Adams K.L., Wendel J.F. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8:135–141.

Averyanov L. 1990. A review of the genus *Dactylorhiza*. Orchid Biol. Rev. Persp., V. Timber Press: Portland, Oregon: 159-206.

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., Johnson E.A. 2008. Rapid snp discovery and genetic mapping using sequenced rad markers. PLoS One 3:e3376.

Balao F., Tannhäuser M., Lorenzo M.T., Hedrén M., Paun O. 2016. Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. Heredity 116:351-361.

Balao F., Trucchi E., Wolfe T., Hao B.-H., Lorenzo M.T., Baar J., Sedman L., Kosiol C., Amman F., Chase M.W., Hedrén M., Paun O. 2017. Adaptive sequence evolution is driven by biotic stress in a pair of orchid species (*Dactylorhiza*) with distinct ecological optima. Molec. Ecol. 26:3649-3662.

Bateman R.M., Hollingsworth P.M., Presto J., Yi-Bo L., Pridgeon A.M., Chase M.W. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). Bot. J. Linnean Soc. 142:1-40.

Bateman R.M. 2011. Glacial progress: do we finally understand the narrow-leaved marsh-orchids? New J. Bot. 1:2–15.

Bateman R.M., Murphy A.R.M., Hollingsworth P.M., Hart M.L., Denholm I., Rudall P.J. 2018a. Molecular and morphological phylogenetics of the digitate-tubered clade within subtribe Orchidinae s.s. (Orchidaceae: Orchideae). Kew Bull. in press.

Bateman R.M., Sramkó G., Paun O. 2018b. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. Ann. Bot. 121:85-105.
Bouckaert R., Heled J. 2014. DensiTree 2: Seeing trees through the forest. bioRxiv 012401.
Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.C., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 10:e1003537.
Brassac J., Blattner F.R. 2015. Species-level phylogenety and polyploid relationships in *Hordeum* (Poaceae) inferred from next-generation sequencing and *in silico* cloning of multiple nuclear loci. Syst. Biol. 64:792-808.
Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molec. Biol. Evol. 29:1917-1932.
Cariou M., Duret L., Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3:846-852.
Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013. Stacks: An analysis tool set for population genomics. Molec. Ecol. 22:3124-3140.
Cruaud A., Gautier M., Galan M., Foucaud J., Sauné L., Genson G., Dubois E., Nidelet S., Deuve T., Rasplus J.-Y. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Molec. Biol. Evol. 31:1272–1274.
Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., Group G.P.A. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.
De Bodt S., Maere S., Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. Trends Ecol. Evol. 20:591-597.
Degnan J. H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.
Degnan J.H. 2013. Anomalous unrooted gene trees. Syst. Biol. 62:574–590.
Devos N., Oh S.-H., Raspé O., Jacquemart A.-L., Manos P.S. 2005. Nuclear ribosomal DNA sequence variation and evolution of spotted marsh-orchids (*Dactylorhiza maculata* group). Molec. Phylogenet. Evol. 36:568-580.
Devos N., Raspé O., Jacquemart A.-L., Tyteca D. 2006. On the monophyly of *Dactylorhiza* Necker ex Nevski (Orchidaceae): Is *Coeloglossum viride* (l.) Hartman a *Dactylorhiza*? Bot. J. Linnean Soc. 152:261-269.
Devos N., Tyteca D., Raspé O., Wesselingh R.A., Jacquemart A.-L. 2003. Patterns of chloroplast diversity among western European *Dactylorhiza* species (Orchidaceae). Plant Syst. Evol. 243:85-97.
Doyle J.J. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.
Dufresne F., Stift M., Vergilino R., Mable B.K. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molec. Ecol. 23:40-69.
Ebersberger I., Galgoczy P., Taudien S., Taenzer S., Platzer M., von Haeseler A. 2007. Mapping human genetic ancestry. Molec. Biol. Evol. 24:2266-2276.
Fernández-Mazuecos M., Mellers G., Vigalondo B., Sáez L., Vargas P., Glover B.J. 2017. Resolving recent plant radiations: Power and robustness of genotyping-by-sequencing. Syst. Biol. 67:250-268
Fernández-Palacios J.M., de Nascimento L., Otto R., Delgado J.D., García-del-Rey E., Arévalo J.R., Whittaker R.J. 2011. A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. J. Biogeo. 38:226-246.

Geldmacher J., van den Bogaard P., Hoernle K., Schmincke H.-U. 2000. The 40Ar/39Ar age dating of the Madeira Archipelago and hotspot track (eastern North Atlantic). Geochem. Geophys. Geosyst. 1:1008

Giraud T., Refrégier G., Le Gac M., de Vienne DM., Hood ME. 2008. Speciation in fungi. Fungal Genet. Biol. 45:791–802.

Govindarajulu R., Hughes C.E., Alexander P.J., Bailey C.D. 2011. The complex evolutionary dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). Amer. J. Bot. 98:2064-2076.

Gregg W.C.T., Ather S.H., Hahn M.W. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. Syst. Biol. 66:1007–1018.

Gruber B., Unmack P.J., Berry O.F., Georges A. 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molec. Ecol. Resour. 18:691–699.

Guo Y.-P., Tong X.-Y., Wang L.-W., Vogl C. 2013. A population genetic model to infer allotetraploid speciation and long-term evolution applied to two yarrow species. New Phytol. 199:609-621.

Han E., Sinsheimer J.S., Novembre J. 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. Molec. Biol. Evol. 31:723–735.

Hedrén M. 1996. Genetic differentiation, polyploidization and hybridization in Northern European *Dactylorhiza* (Orchidaceae): evidence from allozyme markers. Plant Syst. Evol. 201:31-55.

Hedrén M. 2001. Systematics of the *Dactylorhiza euxina/incarnata/maculata* polyploid complex (Orchidaceae) in Turkey: evidence from allozyme data. Plant Syst. Evol. 229:23-44.

Hedrén M. 2003. Plastid DNA variation in the *Dactylorhiza incarnata/maculata* polyploid complex and the origin of allotetraploid *D. sphagnicola* (Orchidaceae). Molec. Ecol. 12:2669-2680.

Hedrén M. 2009. Plastid DNA haplotype variation in *Dactylorhiza incarnata* (Orchidaceae): evidence for multiple independent colonization events into Scandinavia. Nordic J. Bot. 27:69-80.

Hedrén M., Nordström S. 2009. Polymorphic populations of *Dactylorhiza incarnata* s.l. (Orchidaceae) on the Baltic island of Gotland: morphology, habitat preference and genetic differentiation. Annals Bot. 104:527-542.

Hedrén M., Nordström S., Bateman R.M. 2011. Plastid and nuclear DNA marker data support the recognition of four tetraploid marsh orchids (*Dactylorhiza majalis* s.l., Orchidaceae) in Britain and Ireland, but require their recircumscription. Biol. J. Linnean Soc. 104:107-128.

Hedrén M., Nordström S., Persson Hovmalm H.A., Pedersen H.Æ., Hansson S. 2007. Patterns of polyploid evolution in Greek marsh orchids (*Dactylorhiza*; Orchidaceae) as revealed by allozymes, AFLPs, and plastid DNA data. American J. Bot. 94:1205-1218.

Hedrén, M., Nordström S., Pedersen H.Æ., Ståhlberg D. 2012a. Systematics and conservation genetics of *Dactylorhiza majalis* ssp. *elatior* (Orchidaceae) on Gotland. Nordic J. Bot. 30:257-272.

Hedrén, M., Olofsson S., Nordström S. 2012b. Geographical variation and systematics of the tetraploid marsh orchid *Dactylorhiza majalis* subsp. *sphagnicola* (Orchidaceae) and closely related taxa. Bot. J. Linnean Soc. 168:174-193.

Heslop-Harrison J. 1953. Microsporogenesis in some triploid dactylorchid hybrids. Annals Bot. 17:539-549.

Hewitt G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linnean Soc. 58:247–276.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

Hu G., Koh J., Yoo M.-J., Chen S., Wendel J.F. 2015. Gene-expression novelty in allopolyploid cotton: a proteomic perspective. Genetics 200:91-104.
Huson D.H., Bryant D. 2005. Application of phylogenetic networks in evolutionary studies. Molec. Biol. Evol. 23:254-267.
Inda L.A., Pimentel M., Chase M.W. 2010. Chalcone synthase variation and phylogenetic relationships in *Dactylorhiza* (Orchidaceae). Bot. J. Linnean Soc. 163:155-165.
Inda L.A., Pimentel M., Chase M.W. 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. Annals Bot. 110:71-90.
Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., dePamphilis C.W.2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97-100.
Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. BMC Bioinf. 15:356.
Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R. 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63:534-542.
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913-925.
Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
Linder C.R., Rieseberg L.H. 2004. Reconstructing patterns of reticulate evolution in plants. Amer. J. Bot. 91:1700–1708.
Lischer H.E., Excoffier L. 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.
Liu L., Xi Z., Wu S., Davis C.C., Edwards S.V. 2015. Estimating phylogenetic trees from genome-scale data. Annals New York Acad. Sci. 1360:36-53.
Mallet J. 2007. Hybrid speciation. Nature 446:279–283.
Malinsky M., Trucchi E., Lawson D.J., Falush D. 2018. RADpainter and fineRADstructure: population inference from RADseq data. Molec. Biol. Evol. 35:1284-1290.
Mallo D., Posada D. 2016. Multilocus inference of species trees and DNA barcoding. Phil. Trans. R. Soc. B 371:20150335.
McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 9:1297–1303.
Meirmans P.G., Liu S., van Tienderen P.H. 2018. The analysis of polyploid genetic data. J. Heredity 109:283-296.
Meisner J., Albrechtsen A. 2018. Inferring population structure and admixture proportions in low depth NGS data. Genetics in press. doi:10.1534/genetics.118.301336
Morales-Briones D.F., Liston A., Tank D.C. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). New Phytol. 218:1668-1684.
Naczk A.M., Chybicki I.J., Ziętara M.S. 2016. Genetic diversity of *Dactylorhiza incarnata* (Orchidaceae) in northern Poland. Acta Soc. Bot. Pol. 85:3496.
Nielsen R., Paul J.S., Albrechtsen A., Song Y.S. 2011. Genotype and SNP calling from next-generation sequencing data. Nature Rev. Genet 12:443-451.

Nordström S., Hedrén M. 2008. Genetic differentiation and postglacial migration of the *Dactylorhiza majalis* ssp. *traunsteineri/lapponica* complex into Fennoscandia. Plant Syst. Evol. 276:73-87.
Nordström S., Hedrén M. 2009. Genetic diversity and differentiation of allopolyploid *Dactylorhiza* (Orchidaceae) with particular focus on the *Dactylorhiza majalis* ssp. *traunsteineri/lapponica* complex. Biol. J. Linnean Soc. 97:52-67.
Novikova P.Y., Hohmann N., Van de Peer Y. 2018. Polyploid *Arabidopsis* species originated around recent glacial maxima. Curr. Opinion Plant Biol. 42:8-15.
Øien D.-I., Moen A.2002. Flowering and survival of *Dactylorhiza lapponica* and *Gymnadenia conopsea* in the Sølendet Nature Reserve, Central Norway. Pp. 3-22 in: Kindlmann P., Willems J.H., Whigham D.F. (eds.), Trends and Fluctuations and Underlying Mechanisms in Terrestrial Orchid Populations. Backhuys Publishers, Leiden.
Ossowski S., Schneeberger K., Lucas-Lledó J.I., Warthmann N., Clark R.M., Shaw R.G., Weigel D., Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92-94.
Otto S.P. 2007. The evolutionary consequences of polyploidy. Cell 131:452-462.
Paun O., Bateman R.M., Fay M.F., Hedrén M., Civeyrel L., Chase M.W. 2010. Stable epigenetic effects and adaptation in allopolyploid orchids (*Dactylorhiza*: Orchidaceae). Molec. Biol. Evol. 27: 2465-2473.
Paun O., Bateman R., Fay M.F., Luna J.A., Moat J., Hedrén M., Chase M.W. 2011. Altered gene expression and ecological divergence in sibling allopolyploids of *Dactylorhiza* (Orchidaceae). BMC Evol. Biol. 11:113
Paun O., Forest F., Fay M.F., Chase M.W. 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol. 182:507-518.
Paun O., Turner B., Trucchi E., Munzinger J., Chase M.W., Samuel R. 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. Syst. Biol. 65:212-227.
Pease J.B., Haak D.C., Hahn M.W., Moyle L.C. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 14:e1002379.
Pedersen H.Æ. 1998. Species concept and guidelines for infraspecific taxonomic ranking in *Dactylorhiza* (Orchidaceae). Nordic J. Bot. 18:289-310.
Pedersen H.Æ. 2004. *Dactylorhiza majalis* s.l. (Orchidaceae) in acid habitats: variation patterns, taxonomy, and evolution. Nordic J. Bot. 22:641-658.
Pedersen H.Æ. 2009. Apochromic populations of *Dactylorhiza incarnata* s.l. (Orchidaceae): diversity and systematic significance as revealed by allozyme markers and morphology. Bot. J. Linnean Soc. 159:396-407.
Pedersen H.Æ., Hedrén M. 2010. On the distinction of *Dactylorhiza baltica* and *D. pardalina* (Orchidaceae) and the systematic affinities of geographically intermediate populations. Nordic J. Bot. 28:1-12.
Perrie L.R., Shepherd L.D., De Lange P.J., Brownsey P.J. 2010. Parallel polyploid speciation: Distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. Molec. Ecol. 19:2916-2932.
Pillon Y., Fay M.F., Hedrén M., Chase M.W. 2007. Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). Taxon 56:1185-1185.
Pillon Y., Fay M.F., Shipunov A.B., Chase M.W. 2006. Species diversity versus phylogenetic diversity: a practical study in the taxonomically difficult genus *Dactylorhiza* (Orchidaceae). Biol. Cons. 129:4-13.
QGIS Development Team 2018. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901-904.
Ramsey J., Schemske D. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Ann. Rev. Ecol. Syst. 29:467–501.
Rothfels C.J., Pryer K. M., Li F. 2017. Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. New Phytol. 213:413-429.
RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com.
Song K.M., Lu P., Tang K., Osborne T.C. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci., USA, 92:7719-7723.
Soltis D.E., Buggs R.J.A., Barbazuk W.B., Schnable P.S., Soltis P.S. 2009. On the origins of species: does evolution repeat itself in polyploid populations of independent origin? Cold Spring Harb. Quant. Biol. LXXIV:215-223.
Soltis D.E., Soltis P.S. 1999. Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. 14:348-352.
Soltis D.E., Soltis P.S., Pires C.J., Kovarik A., Tate J.A., Mavrodiev E. 2004. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. Biol. J. Linnean Soc. 82:485–501.
Soltis D.E., Visger C. J., Soltis P.S. 2014. The polyploidy revolution then…and now: Stebbins revisited. Amer. J. Bot. 101:1057–1078.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
Ståhlberg D., Hedrén M. 2009. Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid *Dactylorhiza maculata* s.l. (Orchidaceae). Evol. Ecol. 23:295.
Ståhlberg D., Hedrén M. 2010. Evolutionary history of the *Dactylorhiza maculata* polyploid complex (Orchidaceae). Biol. J. Linnean Soc. 101:503-525.
Stebbins G.L. 1980. Polyploidy in plants: unsolved problems and prospects. Pp. 495-520 in: Lewis W.H. (ed.) Polyploidy. Biological Relevance. Plenum Press, New York and London.
Stift M., Berenos C., Kuperus P., van Tienderen P.H. 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: A general procedure applied to *Rorippa* (yellow cress) microsatellite data. Genetics 179:2113-2123.
Taberlet P., Fumagalli L., Wust-Saucy A.-G., Cosson J.-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molec. Ecol. 7:453–464.
Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S, Fischer A., Bock R., Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45:W6–W11.
Vallejo-Marín M., Buggs R.J.A., Cooley A.M., Puzey J.R. 2015. Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. Evolution 69:1487–1500.
Van der Peer Y., Mizrachi E., Marchal K. 2017. The evolutionary significance of polyploidy. Nature Rev. Genet. 18:411-424.
Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., Visscher P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genet. 42:565-569.

Zou X.-H., Du Y.-S., Tang L., Xu X.-W., Doyle J.J., Sang T., Ge S. 2015. Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). Scientific Reports 5:14876.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.[NNNN]

TABLE S1. Collection data for the accessions included in the present study. Inclusion of each accession in the different analysis is marked with X; and with parentheses if included in the RAxML analysis shown in Supplementary Fig. S1. Some individuals were excluded from some of the analyses based on degree of missing data content or their potential to be introgressed. For each species, representative vouchers have been deposited in LD, WU, or K.

Accession number	Species	Country	Latitude	Longitude	DNA bank	RAxML	SNAPP	FINERAD	REL
2770	*D. saccifera*	Turkey	40°37´N	35°32´E	KEW	X	X	X	X
5551	*D. saccifera*	Bulgaria	41°52´N	23°20´E	MH	X	X	X	X
2723	*D. saccifera*	Italy	40°32´N	15°20´E	MH	X	X	X	X
7775	*D. saccifera*	France	41°46´N	09°05´E	MH	X		X	X
6917	*D. saccifera*	Georgia	41°49´N	45°09´E	MH	X		X	X
196	*D. saccifera*	Greece	40°25´N	22°08´E	MH	X		X	X
7752	*D. saccifera*	Italy	41°44´N	15°41´E	MH	X		X	X
5622	*D. fuchsii*	The Netherlands	51°53´N	04°02´E	MH	X	X	X	X
1396	*D. fuchsii*	Russia	61°40´N	31°22´E	MH	X	X	X	X
1203	*D. fuchsii*	Austria	47°47'E	15°14'N	OP	X	X	X	X
719	*D. fuchsii*	England	51°14´N	00°19´W	MH	X		X	X
5237	*D. fuchsii*	Estonia	58°17´N	22°08´E	MH	X		X	X
7798	*D. fuchsii*	France	44°11´N	07°09´E	MH	X		X	X
10377	*D. fuchsii*	France	44°17´N	03°20´E	MH	X		X	X
8675	*D. fuchsii*	Ireland	54°16´N	08°37´W	MH	X		X	X
12696	*D. fuchsii*	Italy	46°58´N	11°46´E	MH	X		X	X
5962	*D. fuchsii*	Norway	70°03´N	22°57´E	MH	X		X	X
12418	*D. fuchsii*	Norway	59°49´N	10°18´E	MH	X		X	X
10815	*D. fuchsii*	Romania	47°40´N	24°08´E	MH	X		X	X
7028	*D. fuchsii*	Russia	60°31´N	59°55´E	MH	X		X	X
5914	*D. fuchsii*	Russia	66°47´N	30°09´E	MH			X	X
1381	*D. fuchsii*	Sweden	55°41´N	13°23´E	MH	X		X	X
7223	*D. fuchsii*	Sweden	60°37´N	17°17´E	MH	X		X	X
406	*D. fuchsii*	Sweden	63°28´N	15°20´E	MH	X		X	X
3477	*D. fuchsii*	Czech Republic	49°05´N	18°05´E	MH	X		X	X
2144	*D. fuchsii*	Austria	47°43'N	15°13'E	OP	X		X	X
1158	*D. fuchsii*	Austria	47°54'N	14°10'E	OP	X		X	X
1159	*D. fuchsii*	Austria	47°54'N	14°10'E	OP	X		X	X
1001	*D. fuchsii*	France	43°12'N	0°46'E	OP	X		X	X
1710	*D. fuchsii*	France	42°54'N	1°56'E	OP	X		X	X

1739	*D. fuchsii*	France	42°55'N	0°48'E	OP	X		X	X
1565	*D. fuchsii*	Hungary	47°54'N	19°53'E	OP	X		X	X
1566	*D. fuchsii*	Hungary	47°54'N	19°53'E	OP			X	X
1187	*D. fuchsii*	Britain	57°23'N	7°19'W	OP	X		X	X
1409	*D. fuchsii*	Sweden	57°20´N	18°19´E	OP	X		X	X
2139	*D. fuchsii*	Sweden	60°37´N	17°37´E	OP	X		X	X
1410	*D. fuchsii*	Sweden	57°20´N	18°19´E	OP	X		X	X
12240	*D. foliosa*	Portugal	32°44´N	16°53´W	MH		X	X	X
12262	*D. foliosa*	Portugal	32°45´N	17°00´W	MH	X	X	X	X
8625	*W D. maculata*	Portugal	40°34´N	08°11´W	MH	X		X	X
12743	*W D. maculata*	Portugal	41°43´N	07°35´W	MH	X		X	X
8376	*W D. maculata*	Island	65°02´N	14°21´E	MH	X		X	X
1466	*W D. maculata*	Faeroes	62°09´N	07°10´W	MH	X		X	X
10372	*W D. maculata*	France	43°26´N	02°20´E	MH	X		X	X
8650	*W D. maculata*	Ireland	53°01´N	09°17´W	MH	X		X	X
12366	*W D. maculata*	Norway	58°39´N	05°34´E	MH	X		X	X
8635	*W D. maculata*	Sweden	57°49´N	18°35´E	MH	X		X	X
3758	*W D. maculata*	England	50°52´N	01°35´W	MH	X		X	X
1658	*D. maculata*	Norway	70°27´N	26°50´E	MH			X	X
5986	*D. maculata*	Norway	70°27´N	26°50´E	MH			X	X
14541	*D. maculata*	France	41°45´N	09°10´E	MH			X	X
5827	*D. maculata*	Romania	46°19´N	25°35´E	MH			X	X
5896	*D. maculata*	Russia	67°34´N	33°23´E	MH			X	X
7044	*D. maculata*	Russia	57°11´N	60°09´E	MH			X	X
8684	*D. maculata*	France	46°01´N	06°55´E	MH			X	X
3531	*D. maculata*	Slovenia	45°45´N	14°27´E	MH			X	X
14395	*D. maculata*	Sweden	61°11´N	16°45´E	MH			X	X
6129	*D. incarnata*	England	53°39´N	03°03´W	MH	X	X	X	X
8820	*D. incarnata*	Poland	54°39´N	18°10´E	MH	X	X	X	X
12457	*D. incarnata*	Norway	58°44´N	05°31´E	MH	X	X	X	X
2864	*D. incarnata*	Finland	67°34´N	26°52´E	MH	X		X	X
8028	*D. incarnata*	Sweden	63°51´N	14°02´E	MH	X		X	X
8712	*D. incarnata*	Ireland	53°20´N	06°12´W	MH	X		X	X
10218	*D. incarnata*	France	44°15´N	6°45´E	MH	X		X	X
6248	*D. incarnata*	Bulgaria	41°81´N	25°40´E	MH	X		X	X
2369	*D. incarnata*	France	45°03´N	06°17´E	MH	X		X	X
6954	*D. incarnata*	Germany	49°04´E	08°18´E	MH	X		X	X
6035	*D. incarnata*	The Netherlands	51°53´N	04°05´E	MH	X		X	X
12484	*D. incarnata*	Norway	59°49´N	10°18´E	MH	X		X	X
6453	*D. incarnata*	Romania	46°19´N	25°35´E	MH	X		X	X
6461	*D. incarnata*	Russia	67°34´N	33°23´E	MH	X		X	X
2830	*D. incarnata*	Russia	62°07´N	34°00´E	MH	X		X	X
8772	*D. incarnata*	Sweden	57°49´N	18°35´E	MH	X		X	X
12469	*D. incarnata*	Norway	58°39´N	05°34´E	MH	X		X	X
10576	*D. incarnata*	Sweden	56°40´N	16°33´E	MH	X		X	X
6103	*D. incarnata*	England	50°49´N	01°29´W	MH	X		X	X
6105	*D. incarnata*	England	50°49´N	01°29´W	MH	X		X	X
8819	*D. incarnata*	Poland	54°39´N	18°10´E	MH	X		X	X

1664	*D. incarnata*	Austria	47°32'N	12°34'E	OP	X		X	X
1665	*D. incarnata*	Austria	47°32'N	12°34'E	OP			X	X
1585	*D. incarnata*	Austria	47°17'N	11°11'E	OP	X		X	X
1797	*D. incarnata*	Britain	54°40'N	2°15'W	OP	X		X	X
1735	*D. incarnata*	France	42°52'N	1°59'E	OP	X		X	X
1771	*D. incarnata*	Hungary	47°33'N	16°49'E	OP	X		X	X
1873	*D. incarnata*	Britain	57°41'N	7°12'W	OP			X	X
1282	*D. incarnata*	Sweden	55°55´N	14°04´E	OP	X		X	X
1904	*D. incarnata*	Sweden	57°20´N	18°19´E	OP	X		X	X
1907	*D. incarnata*	Sweden	57°49´N	18°53´E	OP	X		X	X
6644	*D. umbrosa*	Turkey	40°22´N	43°25´E	MH	X	X	X	X
531	*D. umbrosa*	Turkey	39°51´N	38°24´E	MH	X	X	X	X
4537	*D. umbrosa*	Iran	38°41´N	47°09´E	MH	X	X	X	X
550	*D. umbrosa*	Turkey	40°03´N	40°26´E	MH	X		X	X
544	*D. umbrosa*	Turkey	39°58´N	40°41´E	MH	X		X	X
4575	*D. umbrosa*	Iran	38°41´N	44°39´E	MH	X		X	X
4760	*D. umbrosa*	Iran	35°37´N	46°40´E	MH	X		X	X
551	*D. umbrosa*	Turkey	40°03´N	40°26´E	MH	X		X	X
561	*D. umbrosa*	Turkey	39°39´N	40°00´E	MH	X		X	X
6630	*D. umbrosa*	Turkey	39°08´N	43°45´E	MH	X		X	X
3312	*D. euxina*	Georgia	41°43´N	43°30´E	MH	X	X	X	X
6517	*D. euxina*	Georgia	42°31´N	44°28´E	MH	X	X	X	X
546	*D. euxina*	Turkey	40°41´N	40°42´E	MH	X	X	X	X
590	*D. euxina*	Turkey	40°36´N	39°20´E	MH	X		X	X
763	*D. euxina*	Turkey	40°33´N	40°17´E	MH	X		X	X

Accession number	**Species**	**Country**	**Latitude**	**Longitude**	**DNA bank**	**PCoA**	**SPLIT**	**REL**	**PRIV**
3358	*D. armeniaca*	Georgia	42°38´N	44°39´E	MH	X	X	X	X
565	*D. armeniaca*	Turkey	40°49´N	43°13´E	MH	X	X		
756	*D. armeniaca*	Turkey	41°13´N	42°25´E	MH	X	X	X	X
3068	*D. baltica*	Estonia	57°43´N	26°30´E	MH	X	X	X	X
3090	*D. baltica*	Estonia	58°48´N	22°50´E	MH	X	X	X	X
2948	*D. elatior*	Estonia	58°16´N	21°55´E	MH	X	X	X	X
2170	*D. elatior*	Sweden	57°28´N	18°28´E	MH	X	X	X	X
10169	*D. elatior*	Sweden	57°41´N	18°47´E	MH	X	X	X	X
10038	*D. majalis*	France	42°53´N	00°03´W	MH	X	X	X	X
10431	*D. majalis*	France	42°41´N	02°14´E	MH	X	X	X	X
8527	*D. majalis*	France	44°45´N	06°32´E	MH	X	X		
14108	*D. majalis*	Belgium	50°10´N	05°03´E	MH	X	X	X	X
14119	*D. majalis*	Belgium	50°01´N	05°25´E	MH	X	X	X	X
7671	*D. majalis*	France	44°19´N	06°51´E	MH	X	X	X	X
3883	*D. majalis*	Germany	53°26´N	10°06´E	MH	X	X	X	X
10112	*D. majalis*	France	43°36´N	06°21´E	MH	X	X		
1943	*D. majalis*	Austria	47°43'N	15°13'E	OP	X	X	X	X
1134	*D. majalis*	Austria	47°34'N	15°04'E	OP	X	X	X	X
1149	*D. majalis*	Austria	47°54'N	14°10'E	OP	X	X	X	X
1529	*D. majalis*	Austria	47°20'N	12°48'E	OP	X	X	X	X
1655	*D. majalis*	Austria	47°27'N	12°22'E	OP	X	X	X	X

1716	*D. majalis*	France	42°52'N	1°59'E	OP	X	X	X	X
1761	*D. majalis*	Sweden	55°55´N	14°04´E	OP	X	X	X	X
1697	*D. majalis*	Austria	47°49'N	13°15'E	OP	X	X		
6093	*D. praetermissa*	England	50°58´N	01°20´W	MH	X	X	X	X
14015	*D. praetermissa*	France	49°17´N	03°41´E	MH	X	X	X	X
8194	*D. praetermissa*	Denmark	55°37´N	12°26´E	MH	X	X	X	X
6088	*D. praetermissa*	England	50°58´N	01°20´W	MH	X	X	X	X
14011	*D. praetermissa*	France	49°17´N	03°41´E	MH	X	X	X	X
8221	*D. purpurella*	Denmark	57°30´N	09°51´E	MH	X	X	X	X
14505	*D. purpurella*	Ireland	55°10´N	06°52´W	MH	X	X	X	X
12553	*D. purpurella*	Norway	58°44´N	05°31´E	MH	X	X	X	X
1785	*D. purpurella*	Britain	54°47'N	2°25'W	OP	X	X	X	X
1868	*D. purpurella*	Britain	57°41'N	7°12'W	OP	X	X	X	X
10126	*D. ruthei*	Germany	54°08´N	13°46´E	MH	X	X	X	X
14479	*D. traunsteineri*	Ireland	53°58´N	07°36´W	MH	X	X	X	X
4175	*D. traunsteineri*	Finland	67°29´N	24°54´E	MH	X	X	X	X
8509	*D. traunsteineri*	Norway	63°02´N	08°52´E	MH	X	X	X	X
3016	*D. traunsteineri*	Russia	61°47´N	33°48´E	MH	X	X	X	X
223	*D. traunsteineri*	Sweden	58°24´N	15°32´E	MH	X	X	X	X
12936	*D. traunsteineri*	Sweden	62°15´N	17°27´E	MH	X	X	X	X
4204	*D. traunsteineri*	Estonia	58°56´N	22°17´E	MH	X	X	X	X
10119	*D. traunsteineri*	Germany	53°55´N	13°26´E	MH	X	X	X	X
12542	*D. traunsteineri*	Norway	58°53´N	05°36´E	MH	X	X	X	X
6357	*D. traunsteineri*	Czech Republic	50°35´N	14°37´E	MH	X	X		
4147	*D. traunsteineri*	Finland	67°34´N	26°52´E	MH	X	X	X	X
10120	*D. traunsteineri*	Germany	53°55´N	13°26´E	MH	X	X	X	X
14507	*D. traunsteineri*	Ireland	54°59´N	05°59´W	MH	X	X		
12536	*D. traunsteineri*	Norway	58°53´N	05°36´E	MH	X	X	X	X
12612	*D. traunsteineri*	Norway	59°49´N	10°18´E	MH	X	X	X	X
4084	*D. traunsteineri*	Sweden	66°17´N	23 33´E	MH	X	X	X	X
10935	*D. traunsteineri*	England	52°56´N	01°13´E	MH	X	X	X	X
14225	*D. traunsteineri*	Sweden	59°42´N	18°52´E	MH	X	X		
1940	*D. traunsteineri*	Austria	47°43'N	15°13'E	OP	X	X	X	X
1671	*D. traunsteineri*	Austria	47°31'N	12°34'E	OP	X	X		
1233	*D. traunsteineri*	Britain	54°15'N	0°41'W	OP	X	X	X	X
1806	*D. traunsteineri*	Britain	54°17'N	0°41'W	OP	X	X	X	X
1815	*D. traunsteineri*	Britain	57°25'N	5°49'W	OP	X	X	X	X
1811	*D. traunsteineri*	Britain	57°46'N	5°34'W	OP	X	X	X	X
1918	*D. traunsteineri*	Sweden	60°37´N	17°37´E	OP	X	X	X	X
1068	*D. elata*	Morocco	ca 35°00´N	04°00´W	KEW	X	X	X	X
1069	*D. elata*	Morocco	ca 35°00´N	04°00´W	KEW	X	X	X	X
15123	*D. elata*	Morocco	ca 35°00´N	04°00´W	KEW	X	X	X	X
15127	*D. elata*	Morocco	ca 35°00´N	04°00´W	KEW	X	X	X	X
10089	*D. elata*	France	44°14´N	03°06´E	MH	X	X	X	X
3201	*D. elata*	Portugal	41°50´N	07°03´W	MH	X	X	X	X
14534	*D. elata*	Morocco	33°31´N	05°06´W	MH	X	X	X	X
8572	*D. kerryensis*	Ireland	53°01´N	09°17´W	MH	X	X	X	X
8172	*D. sphagnicola*	Denmark	57°07´N	08°54´E	MH	X	X	X	X

14323	*D. sphagnicola*	Belgium	50°14´N	05°46´E	MH	X	X	X	X
3581	*D. sphagnicola*	Norway	59°43´N	10°06´E	MH	X	X	X	X
3929	*D. sphagnicola*	Sweden	57°03´N	15°50´E	MH	X	X	X	X
8328	*D. sphagnicola*	Sweden	59°06´N	14°36´E	MH	X	X	X	X
2762	*D. urvilleana*	Turkey	40°37´N	35°32´E	KEW	X	X	X	X
3275	*D. urvilleana*	Georgia	42°02´N	44°42´E	MH	X	X	X	X
3309	*D. urvilleana*	Georgia	41°44´N	43°30´E	MH	X	X	X	X
572	*D. urvilleana*	Turkey	41°13´N	42°27´E	MH	X	X	X	X
589	*D. urvilleana*	Turkey	40°36´N	39°20´E	MH	X	X	X	X
773	*D. urvilleana*	Turkey	40°43´N	39°31´E	MH	X	X	X	X
6256	*D. cordigera*	Bulgaria	41°81´N	25°40´E	MH	X	X		
2127	*D. cordigera*	Greece	40°33´N	22°01´E	MH	X	X	X	X
2113	*D. cordigera*	Greece	41°16´N	23°34´E	MH	X	X	X	X
2118	*D. cordigera*	Greece	41°16´N	23°34´E	MH	X	X	X	X
2094	*D. cordigera*	Greece	40°10´N	20°57´E	MH	X	X	X	X
482	*D. cordigera*	Greece	40°10´N	20°57´E	MH	X	X	X	X
1000	*D. cordigera*	Greece	39°40´N	20°51´E	MH	X	X	X	X
2154	*D. cordigera*	Romania	45°22'	25°26'E	OP	X	X	X	X
946	*D. kalopissii*	Greece	39°46´N	21°11´E	MH	X	X	X	X
949	*D. kalopissii*	Greece	39°46´N	21°11´E	MH	X	X	X	X
2106	*D. kalopissii*	Greece	41°01´N	22°17´E	MH	X	X	X	X
2709	*D. pytagorae*	Greece	37°47´N	26°49´E	MH	X	X	X	X
10020	*D. brennensis*	France	46°51´N	01°17´E	MH	X	X	X	X
10425	*D. brennensis*	France	46°51´N	01°16´E	MH	X	X	X	X
4315	*D. nieschalkiorum*	Turkey	40°37´N	31°16´E	MH	X	X	X	X

TABLE S2. Comparison of previous studies of *Dactylorhiza* diploids with the present RADseq study. 1: Renz 1978, Tutin 1980, Luer 1975, Averyanov 1990; 2: Pridgeon et al. 1997, Bateman et al. 2003, Pillon et al. 2007; 3: Inda et al. 2012; 4: Inda et al. 2010; 5: Devos et al. 2005, 2006; 6: Pillon et al. 2007, Pedersen and Hedrén 2010, Baumann et al. 2006, Hedrén et al. 2012a, Tyteca et al. 1991, Bateman et al. 2003, Holzinger and Künkele 1988, Hedrén 2001, Tyteca and Gathoye 1999; 7: Pillon et al. 2007, Devos et al. 2006, Hedrén et al. 2011, 2012b; 8: Hedrén 1996a, 1996b, 1996c, 2001a, 2001b, Pedersen 1998, Hedrén et al. 2007, 2012a, Pedersen 2002; 9: Hedrén et al. 2001, 2007; 10: Hedrén et al. 2012a, 2012b Balao et al. 2016; 11: Devos et al. 2006, Pillon et al. 2006, 2007, Nordström and Hedrén 2009, Hedrén et al. 2007, 2011, 2012b, Shipunov and Bateman 2005, Devos et al. 2003, Hedrén 2003, Shipunov et al. 2005.

Diploids*	**Distribution (1)**	**ITS (2)**	**rpl16, ITS and coxI (3)**	**CHS region (4)**	**ITS, ETS (5)**	**RADseq (present study)**
D. viridis	Circumboreal	Embedded in Dac*tylorhiza*	Embedded in Dac*tylorhiza*		Sister group to *Dactylorhiza*	Embedded in Dac*tylorhiza*, sister to the rest of *Dactylorhiza*
D. iberica	Balkans, Cyprus, Crimea, Asia Minor, and the Caucasus	Basal position in *Dactylorhiza*	Embedded in *Dactylorhiza*, but position within unresolved	Sister group to *D. sambucina*		Sister to *Dactylorhiza* (minus *D. viridis*). Closest related to *D. sambucina*
D. sambucina	Mountain regions in S Europe, Asia Minor and the Caucasus + lowlands of C and N Europe	Embedded in *Dactylorhiza*	Embedded in *Dactylorhiza*	Most closely related to *D. iberica*		Sister to *Dactylorhiza* (minus *D. viridis and D. iberica*). Closest related to *D. iberica*
D. incarnata s.s.	Most of Europe (except Mediterranean climate)	Sister to rest of *Dactylorhiza* (incl. *D. viridis*)	Sister to rest of *Dactylorhiza* (incl. *D. viridis*) or sister to *D. viridis*	Sister to rest of *Dactylorhiza* (incl. *D. viridis*)		Not sister to rest of *Dactylorhiza*
D. umbrosa	Central Asia, Asia Minor, Caucasus					Closest related to *D. incarnata*, treated as own taxa
D. euxina	Endemic to montane NE Turkey and Caucasus	Sister to *D. incarnata*			Sister to *D. incarnata*	Sister to *D. incarnata*
D. aristata	E Russia, China, Japan and Alaska	Sister group to *D. fuchsii-maculata* clade	Sister group to *D. fuchsii-maculata* clade	Sister group to *D. fuchsii-maculata* clade		Sister group to *D. fuchsii-maculata* clade
D. fuchsii s.s.	Temperate-boreal Europe to W Asia	Same clade as *D. saccifera*, *D. foliosa* and *D. maculata*.	Same clade as *D. saccifera*, *D. foliosa* and *D. maculata*.	Same clade as *D. saccifera*, *D. foliosa* and *D. maculata*.	Sister to *D. saccifera*	Sister to *D. saccifera*
D. saccifera	Mediterranean region and Asia Minor				*D. fuchsii* and *D. saccifera* sister to *D. foliosa*	*D. fuchsii* and *D. saccifera* sister to *D. foliosa*
D. foliosa	Endemic to Madeira	Sister species to *D. maculata* s.s.	Sister species to *D. maculata* s.s.		Sister species to *D. maculata* s.s.	Most closely related to *D. maculata* (W)
D. maculata s.s.	Europe and W Asia, SE Europe or Asia Minor	Closely related to *D. fuchsii* s.l.			Closely related to *D. fuchsii* s.l.	*D. maculata* (W) arose from *D. foliosa* by autopolyploidy, *D. maculata* (C/N) show similarity to *D. foliosa* and *D. fuchsii*

Allopolyploids	**Distribution (6)**	**ITS (or ITS and ETS) (7)**	**Allozymes (8)**	**AFLP (9)**	**Microsatellites (nuclear or plastid) (10)**	**plastid (or plastid and nuclear markers) (11)**	**RADseq (the present study)**
D. majalis s.s.	Much of Central Europe (absent from oceanic areas such as the British Isles)	ITS-III shared with *D. maculata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Closely similar to *D. fuchsii* s.l. and *D. maculata* s.l. haplotypes, incl. extinct haplotypes	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.

D. traunsteineri	Temperate and N Europe (+CE Russia)	ITS-III and ITS-V shared with *D. maculata* s.l., ITS-X shared with *D. incarnata*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Closely similar to *D. fuchsii* s.l. and *D. maculata* s.l. haplotypes, including extinct haplotypes	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. praetermissa	NW Europe (continent) + S Britain	High frequency ITS allele shared with *D. saccifera*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. fuchsii* s.l. * *D. incarnata* s.l.	Closely similar to *D. fuchsii* s.l. haplotypes, including extinct haplotypes	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. purpurella	Oceanic; Ireland, N Britain-Scotland, Faroes, Jutland and W Norway	ITS-V shared with *D. maculata* s.l. and ITS-X shared with *D. incarnata*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l., closely related to western European population of *D. incarnata*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. fuchsii* s.l. * *D. incarnata* s.l.	Closely similar to common haplotypes in *D. fuchsii s.s.and D. incarnata s.l.*	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. baltica	Around Baltic Sea, incl. W Russia	ITS-V shared with *D. maculata* s.l. and ITS-X shared with *D. incarnata*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.		Allopolyploid: *D. fuchsii* s.l. * *D. incarnata* s.l.	Closely similar to common haplotypes in *D. fuchsii s.s.*and *D. incarnata s.l.*, maternal parent = similar/identical to *D. fuchsii*	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. ruthei	Southern sides of the Baltic		Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.		Allopolyploid: *D. fuchsii* s.l. * *D. incarnata* s.l.	Closely similar to common haplotypes in *D. fuchsii* s.s. and *D. incarnata* s.l.	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. elatior	Baltic islands of Gotland and Saaremaa	ITS-V shared with *D. maculata* s.l. and ITS-X shared with *D. incarnata*	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. fuchsii* s.l. * *D. incarnata* s.l.	Haplotype identical to haplotype in *D. fuchsii* s.s.	Allopolyploid: *D. fuchsii* s.s. * *D. incarnata* s.s.
D. elata	Wide distribution in SW Europe and NW Africa	ITS-I shared with western European *D. maculata* s.s.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Extinct *D. maculata* s.l. and *D. foliosa* plastid haplotypes	Allopolyploid: W *D. maculata* s.s. * *D. incarnata* s.l.
D. sphagnicola incl. *D. calcifugiens*	NW Europe (continent) N to mid Scandinavia and local; W coast of Jutland resp.	ITS-I shared with W *D. maculata* s. s.	Allopolyploid: *D. fuchsii-maculata* clade x *D. incarnata-euxina* clade	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	Maternal parent = *D. maculata*	Allopolyploid: W *D. maculata* s.s. * *D. incarnata* s.s.
D. kerryensis	Ireland	ITS-I shared with W *D. maculata* s. s.			Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.	W *D. maculata* plastid genome	Allopolyploid: W *D. maculata* s.s. * *D. incarnata* s.s.
D. cordigera	Balkans		Allopolyploid: *D. fuchsii-maculata* clade x *D. incarnata-euxina* clade	Allopolyploid: *D. maculata* s.l. * *D. incarnata* s.l.		Maternal parent = extinct member of the *D. fuchsii-maculata* clade	Allopolyploid: *D. saccifera* * *D. incarnata* s.s.
Dactylorhiza kalopissii and *D. macedonica*	NW and NE Greece, resp.		Allopolyploid: *D. fuchsii*/*D. saccifera* x *D. incarnata*	Allopolyploid: *D. fuchsii*/*D. saccifera* x *D. incarnata*		Identical *D. saccifera* or *D. incarnata* (*macedonica*) haplotypes	Allopolyploid: *D. saccifera* * *D. incarnata* s.s.

Dactylorhiza pythagoreae	Mediterra-nean region, Samos		Allopolyploid: *D. fuchsii*/*D. saccifera* x *D. incarnata*	Allopoly-ploid: *D. fuchsii*/*D. saccifera* x *D. incar-nata*		Similar to *D. fuchsii* s.l. haplotype	Allopoly-ploid: *D. saccifera* * *D. umbrosa*
Dactylorhiza armeniaca	NE Turkey, Caucasus		Allopolyploid: *D. incarnata* s.l. x *D. euxina*		Allopolyploid: *D. incarnata* s.l. x *D. euxina*	Identical to *D. euxina* and *D. incarnata* s.l. haplotypes	Allopoly-ploid: *D. euxina* * *D. umbrosa*
Dactylorhiza urvilleana	N-C Turkey, Caucasus		Allotetraploid: *D. fuchsii-maculata* clade x *D. euxina*		Allopolyploid: *D. fuchsii* s.l. x *D. euxina*	Similar to *D. fuchsii* s.l. haplotypes	Allopoly-ploid: *D. euxina* * *D. saccifera*
Dactylorhiza nieschal-kiorum	W Turkey		Allotetraploid: *D. fuchsii* s.l. x *D. incarnata* s.l.			Similar to *D. fuchsii* s.l. haplotypes	
Dactylorhiza brennensis	Small area in C France	Chimeric ITS between *D. fuchsii* s.l. and *D. maculata* s.l. or between *D. maculata* s.l. and *D. incarnata*				*D. elata* (S-C France) plastid genomes + *D. majalis-D.traunsteineri* and/or *D. praetermissa* plastid genomes	

TABLE S3. Information about input files used in downstream analyses of RADseq data.

Input file	ploidy	pecentage ind	individuals included	number of SNPs (loci)	program
phylip	di-ploid+autotetraploid	75 %	116	20,713	RAxML
phylip	di-ploid+autotetraploid	75 %	125	19,305	RAxML
xml	diploid	90 %	24	1,626	SNAPP
haplotype	di-ploid+autotetraploid	95 %	112	(385)	FINERADSTRUCTU-RE
nexus	polyploid	50 %	95	28,842	SPLITSTREE4
VCF	polyploid	90 %	95	26,988	R (dartR)
VCF	diploid+polyploid	50 %	185	47,22	VCF-TOOLS

FIGURE S1. Best-scoring maximum likelihood phylogenetic tree of a dataset of 19,305 SNPs, comprising 89 diploid *Dactylorhiza* individuals. Eighteen autotetraploid *D. maculata* accessions and 18 outgroup individuals from closely related genera are also included.

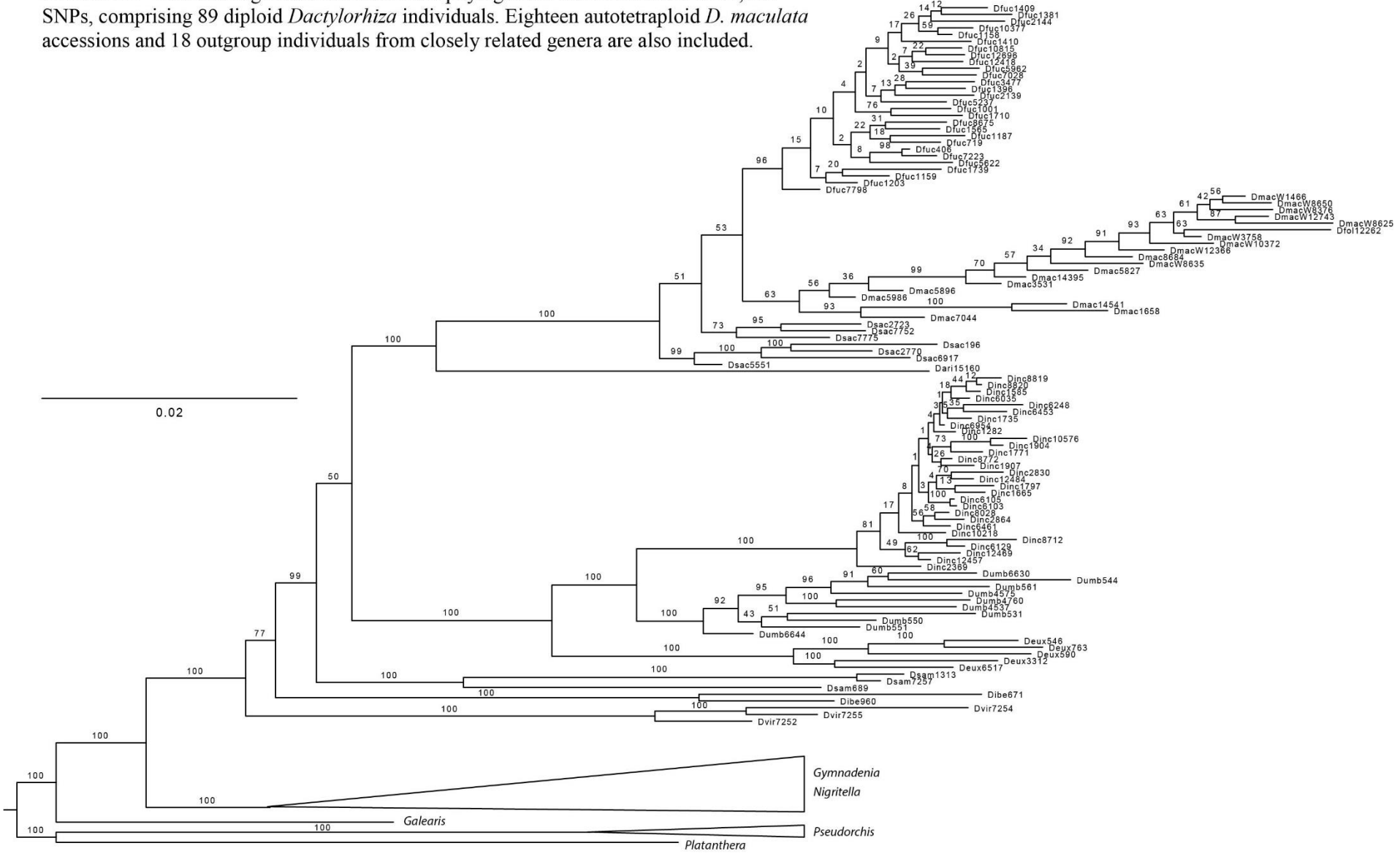


FIGURE S2. Heatmap produced by FINERADSTRUCTURE analysis performed on a dataset of 385 RADseq loci and 112 diploid and autopolyploid *Dactylorhiza* individuals. Ari = *D. aristata*, Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Ibe = *D. iberica*, Inc = *D. incarnata*, Mac = *D. maculata*, Mac W= western *D. maculata,* Sac = *D. saccifera*, Sam = *D. sambucina*, Umb = *D. umbrosa*, and Vir = *D. viridis*.

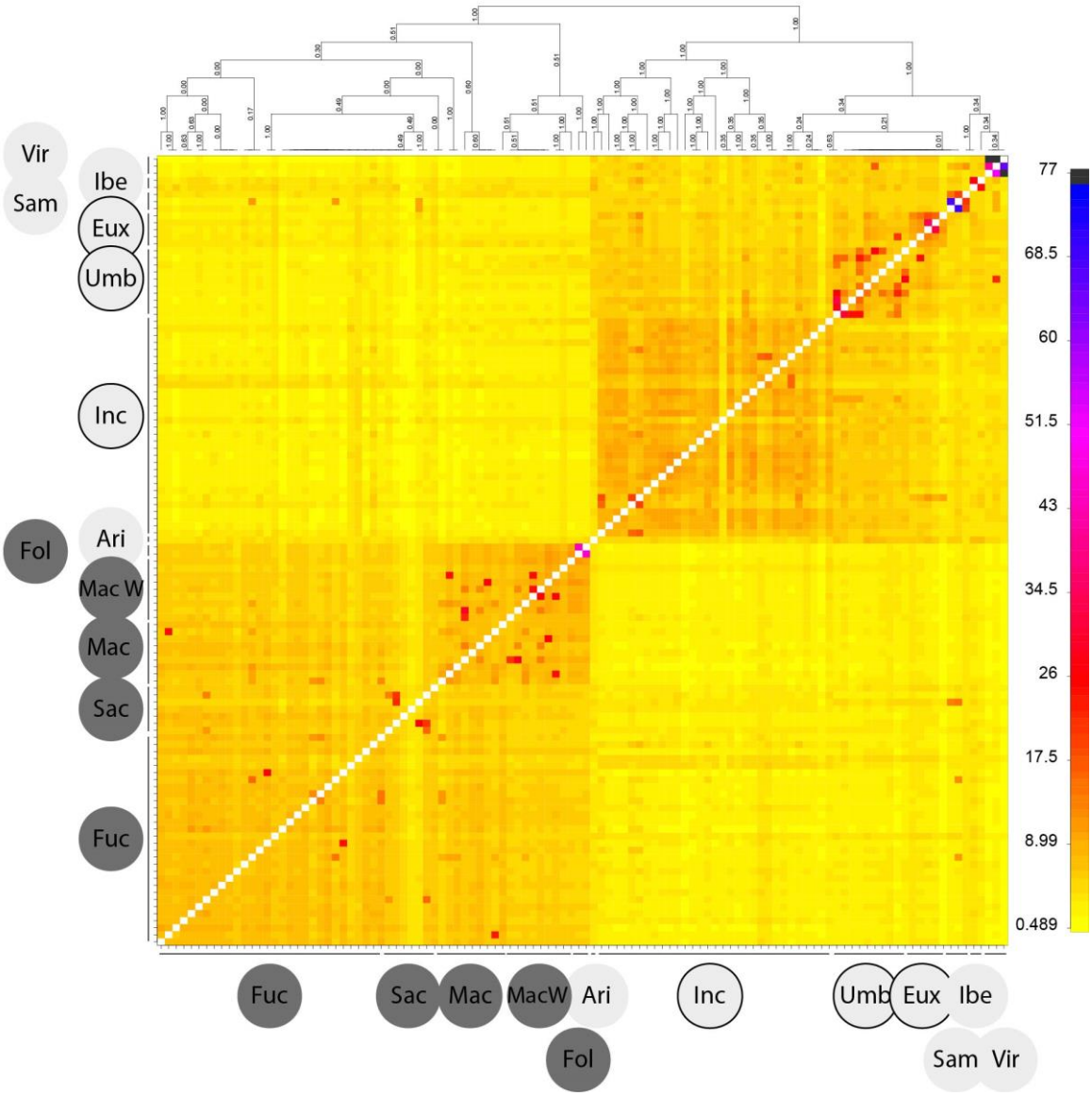


FIGURE S3. Vioplots of inbreeding coefficient (*F*) estimated by VCFTOOLS (Danecek et al. 2011) for the diploid and autotetraploid *Dactylorhiza* species included here (except *D. aristata*). Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Ibe = *D. iberica*, Inc = *D. incarnata*, Mac = *D. maculata,* Sac = *D. saccifera*, Sam = *D. sambucina*, Umb = *D. umbrosa*, and Vir = *D. viridis*.

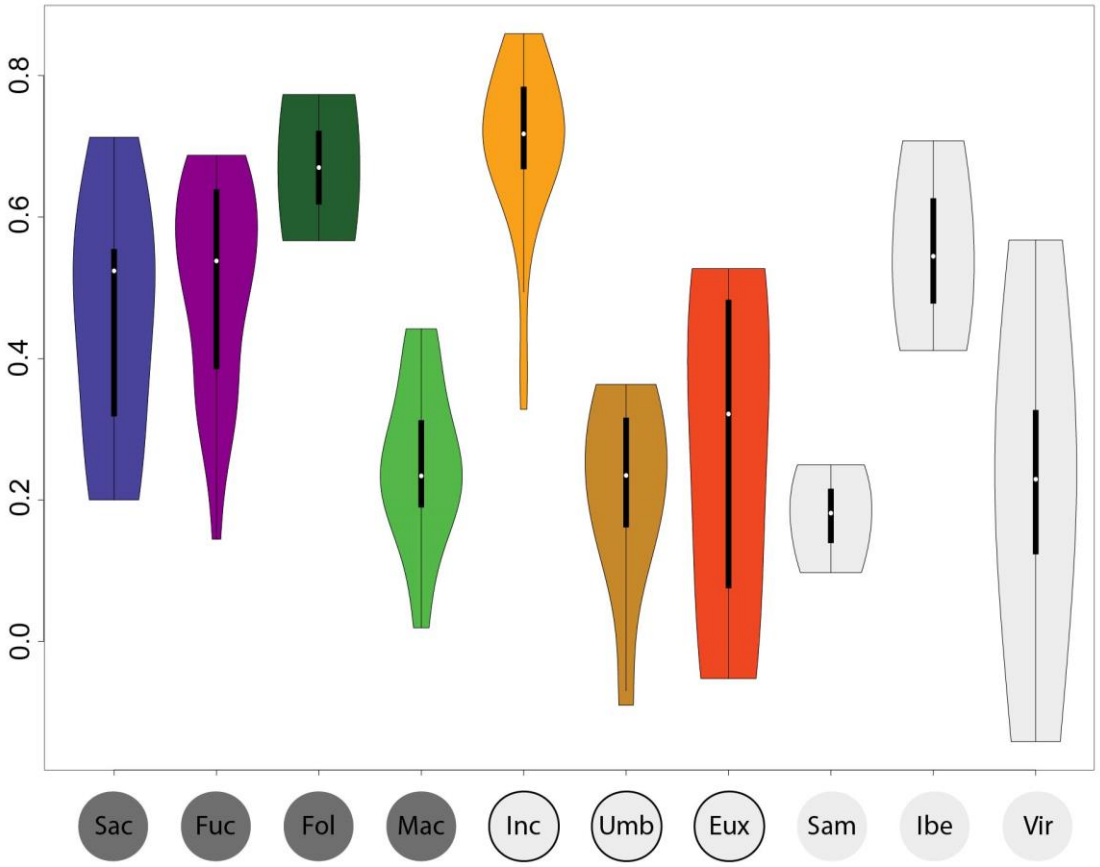


FIGURE S4. Plot resulting from a Neighbour Net analysis based on Jukes Cantor distance performed on 95 allopolyploid *Dactylorhiza* individuals using 28,842 SNPs from RADseq data. As the polyploids did not diverge from one another, this network should be taken as a phenetic representation of relationships. Arm = *D. armeniaca*, Bal = *D. baltica* (incl. *D. ruthei*), Bre = *D. brennensis,* Cor = *D. cordigera*, Ela = *D. elata*, Etr = *D. elatior*, Kal = *D. kalopissii* (incl. *D. macedonica*), Ker = *D. kerryensis*, Nie = *D. nieschalkiorum*, Maj = *D. majalis*, Pra = *D. praetermissa*, Pur = *D. purpurella*, Pyt = *D. pythagorae*, Sph = *D. sphagnicola* (incl. *D. calcifugiens*), Tra = *D. traunsteineri*, and Urv = *D. urvilleana*. Colors indicate different parental pairs for the allopolyploids. The symbols and colors follow Fig. 2. Lineages involved in origin of the allopolyploids are also indicated; Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata,* Sac = *D. saccifera*, Umb = *D. umbrosa*. Asterisks indicate *D. majalis-traunsteineri* intermediates found in sympatric areas. The analysis was performed with Splitstree4. ⟶

0.001
Sac
Eux
Urv
Eux
Umb
Arm
Mac
Inc
Ela
Sac
Umb
Pyt
Sph
Ker
Nie
Bre
Cor
Sac
Inc
Kal
Etr
Bal
Pur
Pra
Tra
(CE)
Tra
(NW)
Maj
Fuc
Inc

FIGURE S5. Violplots of relatedness of allopolyploids produced by *D. fuchsii* and *D. incarnata* to potential ancestral genomes (diploid or autotetraploid). Plots of relatedness of the allopolyploids to members of the *D. fuchsii-maculata* clade are shown to the left of dashed vertical bars, and to members of the *D. incarnata-euxina* clade to the right. Stars indicate significantly different distributions. Colors follow Fig. 1. Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata,* Sac = *D. saccifera*, Umb = *D. umbrosa.* Additional relatedness plots are given in Fig. 6 and Supplementary Fig. S6.

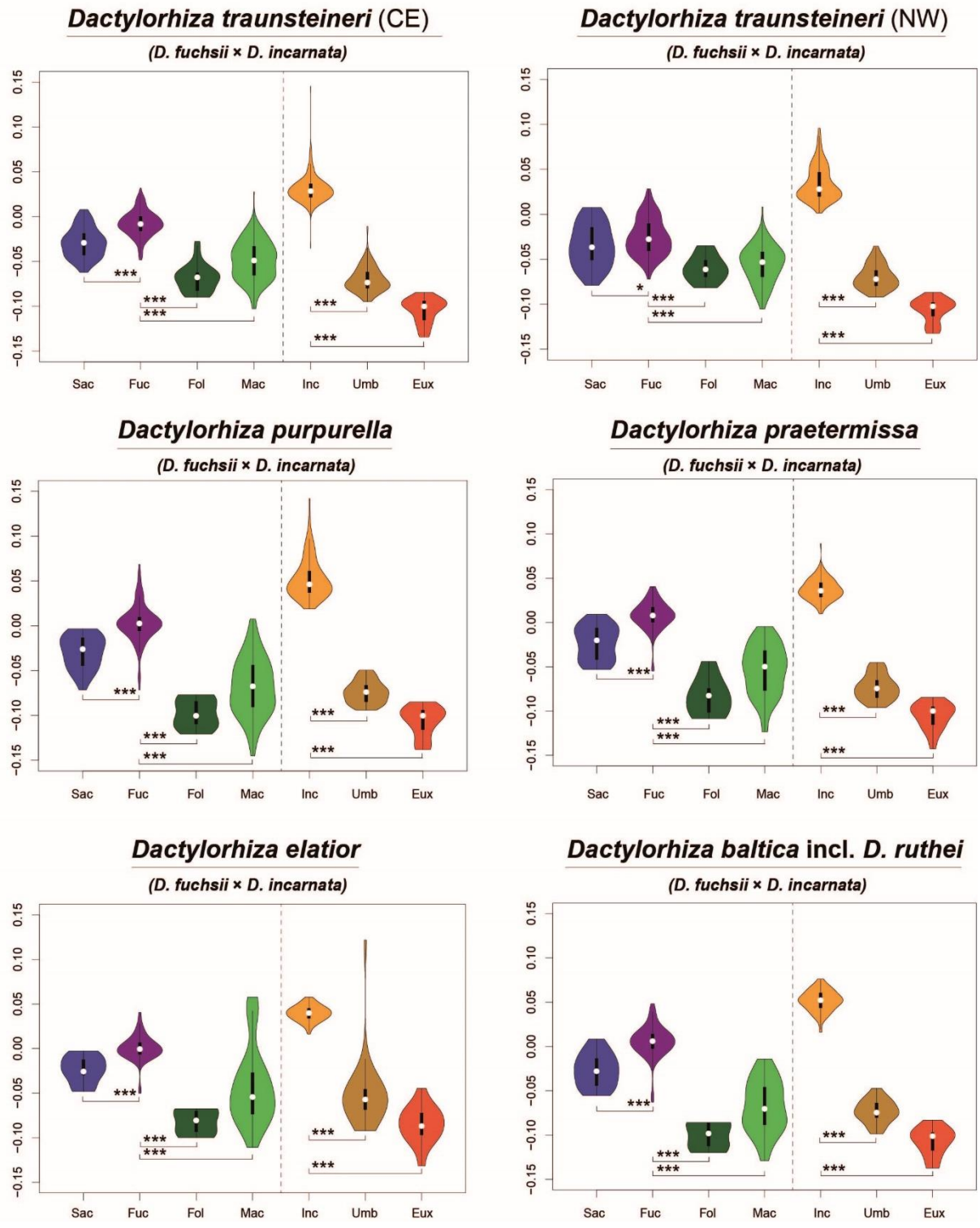
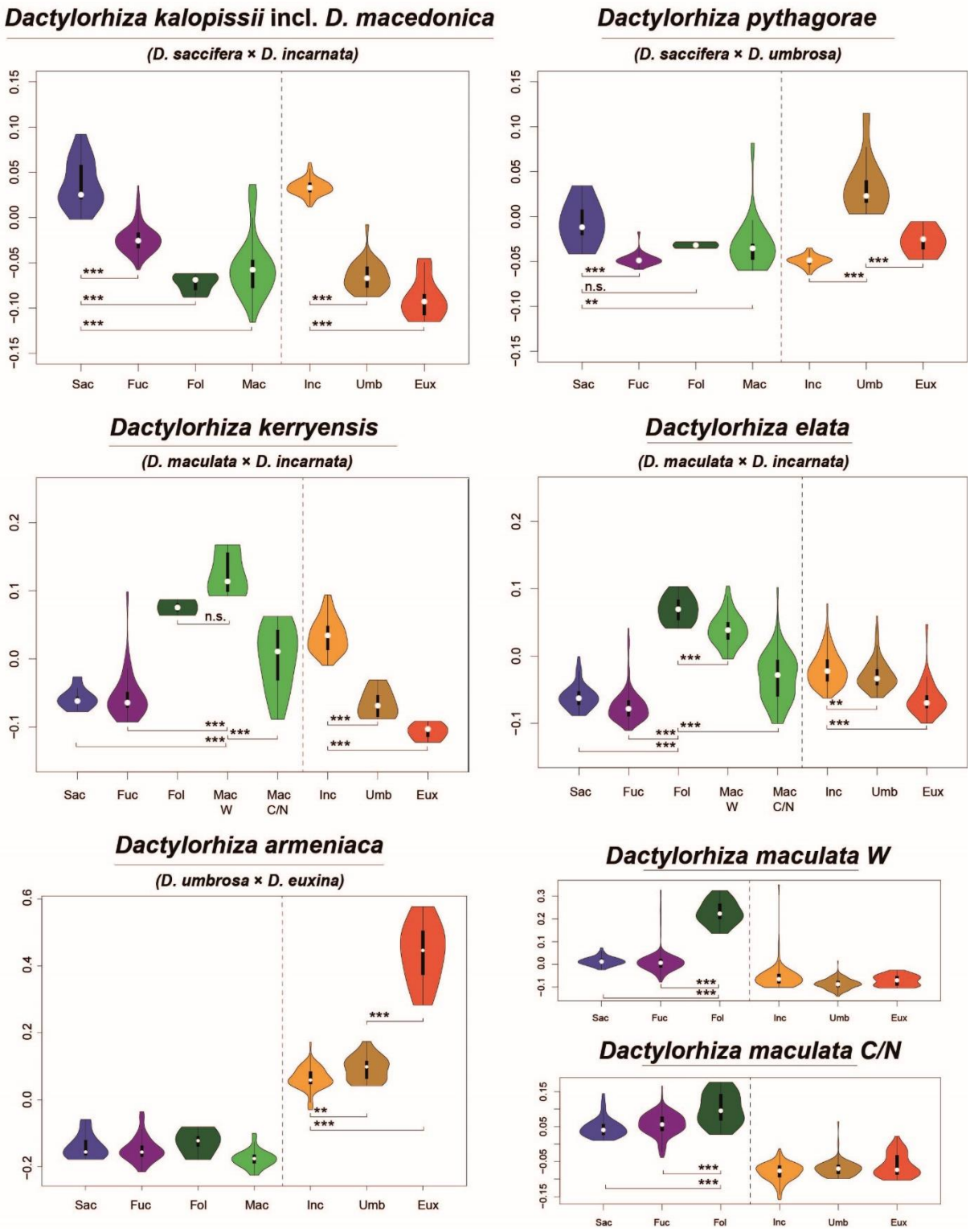


FIGURE S6. Violplots of relatedness of allopolyploids to potential ancestral genomes (diploid or autotetraploid). Plots of relatedness of the allopolyploids to members of the *D. fuchsii-maculata* clade are shown to the left of dashed vertical bars, and to members of the *D. incarnata-euxina* clade to the right. Stars indicate significantly different distributions. Colors follow Fig. 1. Eux = *D. euxina*, Fol = *D. foliosa*, Fuc = *D. fuchsii*, Inc = *D. incarnata*, Mac = *D. maculata,* Sac = *D. saccifera*, Umb = *D. umbrosa.* Additional relatedness plots are given in Fig. 6 and Supplementary Fig. S5.



ADDITIONAL REFERENCES FOR SUPPLEMENTARY MATERIAL

Baumann H., Künkele S., Lorenz R. 2006. Orchideen Europas: mit angrenzenden gebieten. Weltbild.

Hedrén M. 1996a. Genetic differentiation, polyploidization and hybridization in Northern European *Dactylorhiza* (Orchidaceae): evidence from allozyme markers. Plant Systematics and Evolution 201:31-55.

Hedrén M. 1996b. Electrophoretic evidence for allotetraploid origin of *Dactylorhiza purpurella* (Orchidaceae). Nordic Journal of Botany 16:127-134.

Hedrén M. 1996c. The allotetraploid nature of *Dactylorhiza praetermissa* (Druce) Soó (Orchidaceae) confirmed. Watsonia 21:113-118.

Hedrén M. 1996d. Notes on the esterase variation in Swedish *Dactylorhiza incarnata* sl (Orchidaceae). Nordic Journal of Botany 16:253-256.

Holzinger J., Kunkele S. 1988. *Dactylorhiza macedonica* sp. nova. Mitteilungsbl. Arbeitskr. Heim. Orch. Baden-Wurttemberg, 20:185-195.

Luer C. A. 1975. The native orchids of the United States and Canada. Botanical Garden.

Pedersen H. *Æ.* 2002. *Dactylorhiza majalis* sl (Orchidaceae) in acid habitats: variation patterns, taxonomy, and evolution. Nordic Journal of Botany 22:641-658.

Pridgeon A., Bateman R., Cox A. V., Hapeman J. R., Chase M. W. 1997. Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences: 1. Intergeneric relationships and polyphyly of Orchis sensu lato. Lindleyana 12:89-109

Renz J., Taubenheim G. 1984 *Dactylorhiza* Neck- er ex Nevski. In: Davis P. H. (ed.) Flora of Turkey and the East Aegean Islands, vol. 8. Edinburgh University Press, Edinburgh, 535-551

Shipunov A. B., Bateman R. M. 2005. Geometric morphometrics as a tool for understanding *Dactylorhiza* (Orchidaceae) diversity in European Russia. Biological Journal of the Linnean Society, 85:1-12.

Shipunov A. B., Fay M. F., Chase M. W. 2005. Evolution of *Dactylorhiza baltica* (Orchidaceae) in European Russia: evidence from molecular markers and morphology. Botanical Journal of the Linnean Society, 147:257-274.

Tutin T. G. 1980. Flora Europea; Vol. 5. Alismataceae to Orchidaceae.

Tyteca D., Gathoye J. L., Chas E. 1991. Le Dactylorhiza de Lesdiguiéres. The Lesdiguieres dactylorchid.) Orchidophile, 22:155-160.

Tyteca, D., Gathoye J. L. 1999. Morphometric analysis of *Dactylorhiza occitanica* and related populations from eastern France (Orchidaceae). Belgian Journal of Botany, 158-174.

— CHAPTER 2 —

The impact of recurrent origins, genetic drift and gene flow on the genetic structure of allopolyploid marsh orchids

---

Marie K. Brandrud, Thomas Wolfe, Juliane Baar,

Mikael Hedrén, Ovidiu Paun

Status: Prepared for submission to Molecular Biology and Evolution

Contribution: Data processing, analyses and visualization, writing manuscript

The impact of recurrent origins, genetic drift and gene flow on the genetic structure of allopolyploid marsh orchids

Marie K. Brandrud[1], Thomas Wolfe[1], Francisco Balao[3], Juliane Baar[1], Mikael Hedrén[2], Ovidiu Paun[1]

[1]*University of Vienna, Department of Botany and Biodiversity Research, Rennweg 14, Vienna, A-1030, Austria*

[2]*Department of Biology, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden*

[3]*Universidad de Sevilla, Dpto. Biología Vegetal y Ecología, Área de Botánica. Facultad de Biología, Avda. Reina Mercedes, 6, Sevilla, 41012, Spain*

ABSTRACT

Whole genome doubling has undoubtedly contributed to the present dominance and diversity of flowering plants. Early generation polyploids must quickly recover from the genetic bottleneck associated with their origin. Their genetic variation can be fueled either via multiple origins, out of distinct parental populations, or from subsequent introgression from relatives. Inferences based on reduced representations of genomes have become widely applied to answer evolutionary questions, mostly in diploid organisms. We focus here on the genetic structure of an allotetraploid complex of European marsh orchids (*Dactylorhiza*) which have originated from unidirectional crosses of the same diploid parental pair. Starting from tens of representatives of the diploid parents and using both genotype-based and genotype-free analytical methods, we propose an approach to analyse RADseq data for allopolyploids and investigate their evolutionary history. The consistent signal obtained from each subgenome of the allopolyploids uncovers a complex genetic structure shaped by two main independent origins of distinct, postglacial age. Each primary allopolyploid lineage expanded to occupy large European areas while segregating further as a result of genetic drift in allopatry; they secondary merged into a sympatric area around the Alps. Our results uncover main phases and contributors during the evolution of allopolyploid marsh orchids, whereas our analytic approach should prove useful for other studies featuring non-model allopolyploids.

INTRODUCTION

Polyploidy, i.e. whole genome duplication, and hybridization are frequently encountered within the plant kingdom and other groups (Cui et al. 2006, Rieseberg and Willis 2007, Barker et al. 2008, Soltis et al. 2009, Shi et al. 2010, Jiao et al. 2011, Arrigo and Barker 2012, Jiao et al. 2012, Vekemans et al. 2012, Cannon et al. 2014, Jiao et al. 2014, Kagale et al. 2014). Allopolyploids arise after hybridization between two progenitors and are normally the products of meiotic non-reduction followed by fertilization of unreduced gametes (Ramsey and Schemske 1998). Combining different parental genomes in the same nucleus normally have disastrous fitness effects, but can rarely boost functional possibilities allowing polyploid and homoploid hybrids to explore novel evolutionary paths (Rieseberg et al. 2003, Comai 2005, Mallet 2007). Surveying the literature of 47 plant genera (4,003 species) Barker et al. (2016) found that 24% of the species were of polyploid origin, also including autopolyploids that arise within the same species or population (Ramsey and Schemske 1998, Soltis et al. 2007, Wood et al. 2009).

After overcoming the critical formation, newly formed polyploids face a minority type disadvantage, i.e. being at low frequency and potentially swamped with parental gametes that will contribute to unfit/sterile odd-ploid offspring (Felber 1991, Ramsey and Schemske 1998, Husband 2000, Levin 2002). Apart from backcrossing with the parents, self-fertilization, asexual reproduction or crossing with other polyploid individuals are needed to reproduce. It is now recognized that recurrent origins are the rule for most polyploids (Soltis and Soltis 1999), and this can create an array of genetically, ecologically, morphologically, and physiologically distinct genotypes. Subsequent gene flow, independent assortment, and recombination may produce additional variation (Soltis and Soltis 1999, Paun et al. 2007). If recurrent formations are happening within the same population this can significantly aid allopolyploid establishment. On the other hand, recurrent origins out of different parental populations can give rise to different allopolyploids and ultimately result in species with different evolutionary fates.

In the origin of a polyploid only a small part of the variation is transferred from the diploids, leading to a bottleneck effect on the young polyploid lineage. The polyploid genetic pools may slowly be enriched via backcrossing with the parents or hybridizing with other relatives, but most frequently through recurrent formations, and these processes are likely to leave traces in their genomes. How recurrently formed allopolyploid species evolve further and if they

maintain their distinctiveness while sharing the same ploidy and genetic background is still a field with limited research. As multiple origins provide natural replicates, sibling allopolyploids are excellent models to uncover the different factors contributing to the evolutionary outcome of allopolyploidization.

Next generation sequencing, including methods delivering a reduced representation of the genome like restriction site-associated DNA sequencing, RADseq (Etter et al. 2012), is increasingly being applied to answer evolutionary questions (Twyford and Ennos 2012, Buggs et al. 2012, Abreu et al. 2018). The methods and programs available are mainly developed for diploid data and expect maximum two allele copies per locus with regular inheritance (Dufresne et al. 2014, Meirmans et al. 2018). For polyploid data where the inheritance can be ranging from disomic to polysomic, i.e. multiple allele copies should be considered at the methodological level. For autotetraploids that feature strict polysomic inheritance with non-preferential pairing, the copies from the two different subgenomes are expected to be genetically very similar and the alleles to share a common evolutionary history achieved through recombination. For this situation a methodology that allows for more than two alleles per locus would be required. For strict allopolyploids, disomic inheritance is expected and the genetic material inherited from the two parents are expected to be more different, because they evolve independently. The preferred methodology for this situation would be to split the data into the two subgenomes of the allopolyploids, which then each can be treated as diploid. In between these extremes there are cases of polyploids having a combination of polysomic and disomic inheritance for their chromosomes, but such constructions may have a general low fitness and be selected against (Soltis et al. 2010; Paun et al. 2011). From a technical point of view, depending on the time since the polyploid originated, there may be polyploids that will show a mixture of loci that can be distinguished to their original subgenomes and some loci that cannot.

To study how recurrent allopolyploid evolution unfolds we focused on allopolyploid marsh orchids in the core group of the *Dactylorhiza majalis* complex. Based on eco-geographical features and based on our results reported below, we will be using hereafter the formal names *Dactylorhiza majalis* (s. str.) and *D. traunsteineri* (s. lat.; including *D. lapponica*, *D. russowii*, *D. ebudensis* and *D. traunsteinerioides*, which are recognised names by some authors). We have investigated these allopolyploids as two evolutionary groups and do not assign to these names any taxonomic rank. These allopolyploids have originated, potentially multiple times, by unidirectional hybridization of diploid *D. fuchsii* (always the maternal parent) and *D. in-*

carnata (always as the paternal parent) (Heslop-Harrison 1953, Hedrén 1996a, Hedrén 1996b, Hedrén 1996c, Hedrén et al. 2001, Hedrén 2002, Pillon et al. 2007, Balao et al. 2016, Brandrud et al. submitted - Chapter 1), but the number of such origins is not yet clear. The *Dactylorhiza* genus exhibits high morphological variation, which is not always correlated with genetic variation, and it is considered one of the most taxonomically difficult groups of European orchids (Pillon et al. 2007, Hedrén et al. 2008, Paun et al. 2010, Hedrén et al. 2011).

The diploid parental species, *D. fuchsii* and *D. incarnata*, are widely distributed in Europe and extend their distribution to Siberia. Despite several different morphotypes recognized and a wide array of infraspecific taxa recognized, *D. incarnata* is considered remarkably genetically uniform (Hedrén and Nordström 2009, Balao et al. 2017). *Dactylorhiza fuchsii*, on the other hand, is genetically more variable and its genome is affected by hybridization to the autopolyploid *D. maculata* (Pillon et al. 2007, Ståhlberg and Hedrén 2008, Brandrud et al. submitted - Chapter 1).

The two focal allopolyploids for this study are not as widely distributed as their parents: *Dactylorhiza majalis* s. str. occurs mainly in C Europe with a northern limit around Denmark/S Sweden and western limit in the Pyrenees and the Netherlands (Pillon et al. 2007). *Dactylorhiza majalis* is sometimes used as a wider term to include all allopolyploids derived from *D. incarnata* and the *D. fuchsii*, but is here used in the strict sense following a recommended framework classification by Pillon et al. (2007) based on genetic evidence so far. *Dactylorhiza traunsteineri* s. lat. has a more boreal-montane distribution, extending to N Scandinavia and to western parts of the Britain and Ireland (Pillon et al. 2007). Both allopolyploids occur in moist, calcareous habitats, though *D. traunsteineri* more frequently grows in calcareous fens and mires, whereas *D. majalis* grows often in moist, grazed meadows (Pillon et al. 2007, Ståhlberg and Hedrén 2008, Hedrén et al. 2011).

Dactylorhiza traunsteineri is, based on the degree of gene conversion in ITS parental alleles, considered to be a young allopolyploid, likely with an origin after the last glaciation and *D. majalis* is considered to be older (Pillon et al. 2007). Previous genetic studies based on traditional markers have not been able to fully sort out the complicated history of this complex. With RADseq we analyse an extensive sampling including 32 populations of diploid (16 *D. fuchsii* and 16 *D. incarnata*) and 41 population of allopolyploid (26 *D. traunsteineri* and 15 *D. majalis*) *Dactylorhiza* throughout Europe (Fig. 1, Table S1) in order to investigate the his-

tory of this allopolyploid complex in detail and estimate the relative contributions of number of origins, genetic drift and gene flow to the observed genetic variation.

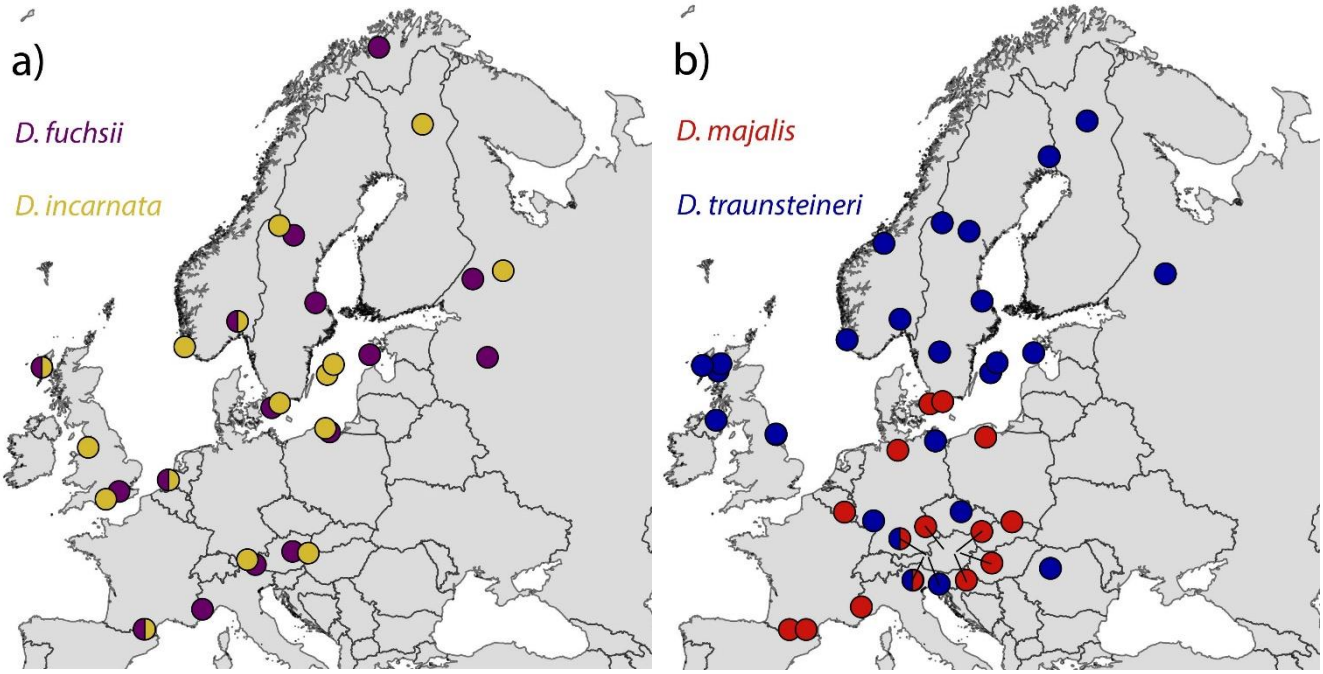


Figure 1: a) Sampling map for diploids, including 16 *D. fuchsii*, and 16 *D. incarnata* sampling localities. b) Sampling map of 15 sampling localities of *D. majalis* and 26 for *D. traunsteineri*. The maps were created with QGIS v. 2.4.071 (QGIS Development Team). GADM version 1.0 (available from www.gadm.org) was used to extract the map layer for both maps and exact coordinates of the populations are in Table S1.

RESULTS

On average ca. 1.5 million and ca 1.8 million RADseq reads were retained per diploid and polyploid accession, respectively, after demultiplexing and cleaning. A reference optimisation script by dDocent (Puritz et al. 2014a, Puritz et al. 2014b) was used to estimate the optimal number of loci that should be filtered from the data. For the *D. fuchsii* accessions the optimal number of loci was estimated to be ca 6,000 or ca 8,000 (i.e., two peaks) and ca 5,500 (i.e., one peak) for the *D. incarnata* accessions.

Two different pipelines were developed, one to maximize the number of unique loci retained, from which genotype likelihoods (Langmead and Salzberg 2012, Korneliussen et al. 2014) were extracted (Pipeline I), and a second, which was genotype-based and separated the loci in the two respective subgenomes (Pipeline II). First, the forward reads from all 153 diploid individuals (all *D. fuchsii* and *D. incarnata* individuals together) were used to build the synthet-

ic reference for Pipeline I. From the 6,737 loci assembled *de novo* with STACKS (Catchen et al. 2011, 2013), 4,355 were retained after removing redundant loci with BLAST. Both diploid and polyploid accessions were mapping on average with 50.2% of their reads to this synthetic reference.

Paired-end data, from 19 and 24 diploid individuals of *D. fuchsii* and *D. incarnata*, respectively, was used to build the synthetic references for pipeline II, which was optimized to separate the allopolyploid reads back to the original two subgenomes, with longer contigs that should aid in distinguishing *D. fuchsii* and *D. incarnata* loci. On average 21.4% of the paired-end reads could be combined to create longer contigs with FLASH (Magoč and Salzberg 2011). On average the 5,220 loci assembled with PyRAD (Eaton 2014) from *D. incarnata* were 190 bp long, and the 5,701 loci assembled from *D. fuchsii* were 185 bp long. Altogether 32% of the *D. incarnata* reads and 25% of the polyploid reads were mapping to the II subgenome. Only 29% of the *D. fuchsii* reads and 24% of the polyploid reads were mapping to the FF subgenome. Synthetic polyploids were created from pairing together ten *D. fuchsii* and ten *D. incarnata* and to test the accuracy of the pipeline. The pipeline had an accuracy of 77%. The samples had an average coverage of 120X to the reference when calling SNPs with REF_MAP.PL, which is part of STACKS.

Diploids

The PCoA created for *D. fuchsii* with the dartR package (Gruber et al. 2018) show significantly overlapping populations and no clear geographical structure (Fig. 2a). The first axis explains 5.1% of the variation and is mostly stretching out the variation within the populations. There is a slight geographical pattern to be seen from south to north on the second axis that explains 4.4% of the variation. The highest number of private alleles in the *D. fuchsii* populations were found in Austria, S England and W Russia. The lowest number of private alleles was found in S France, E Sweden and in The Netherlands. Isolation by distance was not found to be significant in *D. fuchsii* (p-value: 0.078; Fig. S1a). The PCoA created similarly for *D. incarnata* shows mostly distinct populations (Fig. 2b). The populations from CE Europe group together and are separated from the populations from NW Europe by the first axis that explains 14.3% of the variation. The second axis explains 7.3% of the variation and shows a (weak) geographical gradient from north to south among the CE populations. The highest number of private alleles in the *D. incarnata* populations was found in Scotland, Austria and S France, whereas the lowest number of private alleles are found in Hungary, N Poland and C

Sweden. Isolation by distance was found to be significant for *D. incarnata* (p-value: 0.048; Fig. S1b)

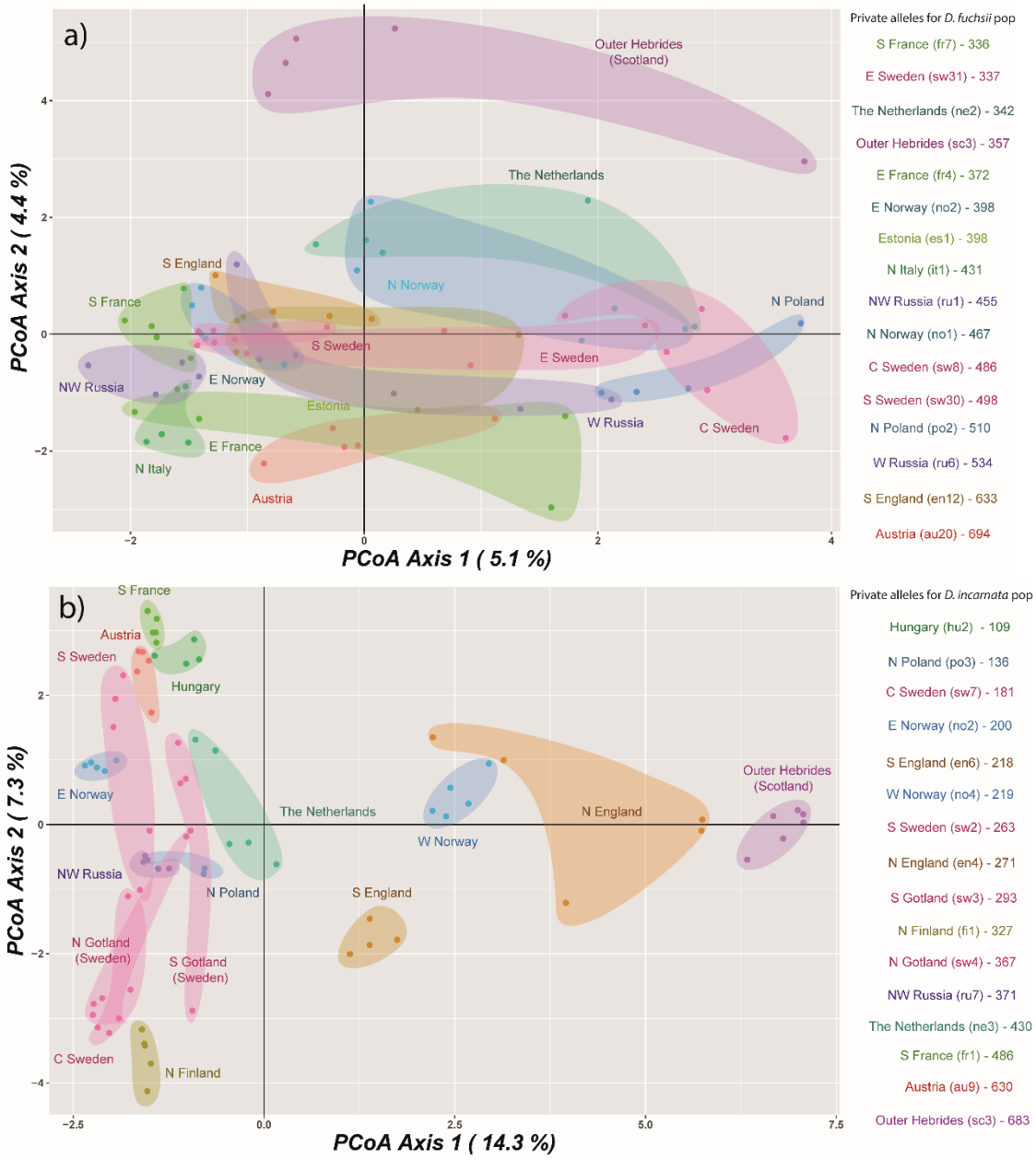


Figure 2: a) Ordination of 78 *D. fuchsii* individuals created with loci covered in minimum 90% of the individuals, comprising 3,115 SNPs filtered with Pipeline II. b) Ordination of 75 *D. incarnata* individuals created with loci covered in minimum 90% of the individuals, comprising 3,094 SNPs filtered with Pipeline II. The ordinations were created with the dartR package with R (Gruber et al. 2018). The private alleles in each population is stated on the left side.

Polyploids - Approach based on genotype likelihoods (pipeline I)

In the PCAngsd (Meisner et al. 2018) ordinations based on genotype-likelihoods extracted with pipeline I the allopolyploids are clearly placed in between the diploid parents (Fig. S2). Admixture analyses with NGSadmix (Skotte et al. 2013) were run for K=1-10, with ten independent runs with different seeds for each K. The admixture analyses were run for the allopolyploids together with *D. fuchsii* and, separately, with *D. incarnata*. The log probability values from the admixture runs were visualized with plots following Evanno et al. (2005). The separation in two groups corresponding to diploids versus allopolyploids had the highest DeltaK values both in the analysis with *D. incarnata* and in the analysis with *D. fuchsii* (Fig. S3-S4). The next highest value (K=4) was also visualized (Fig. 3a and 3c). Both CE *D. incarnata* and NW *D. incarnata* are largely affiliating to the same group. However, in K=6 (which is the next peak in the DeltaK plot), they are separated to their own groups. For K=4 *D. majalis* individuals are largely assigned to a single gene pool, whereas two main gene pools are observed within *D. traunsteineri*, corresponding to NW (i.e., British plus western Norwegian populations) and the rest (i.e., CE). Significant allele sharing between the allopolyploid gene pools is observed, especially between W *D. majalis* (Pyrenees) and NW *D. traunsteineri* and between the two CE groups. Gene flow between the diploids and the allopolyploids, in particular around CE Europe is also visible in the admixture plots.

Polyploids - Approach using genotypes for each subgenome (pipeline II)

PCoAs were created in the same way as for the diploids and were similar for the II subgenome and the FF subgenome. The first axis (explaining 11.1% of the variation in the FF subgenome PCoA and 10.9% in the II subgenome PCoA) is stretching out the geographical structure found in the allopolyploids (Fig. 3b, 3d). Within *D. traunsteineri* there is a NW group including populations from the British Isles as well as W Norway, and a CE group including all mainland Europe populations except W Norway. Within *D. majalis* there is less geographical structure, though a distinction can be seen between W *D. majalis* (from the Pyrenees) and CE *D. majalis* (which comprise the rest of the populations). The second axis (explaining 7.7% of the variation in the FF subgenome PCoA and 7.6% of the variation in the II subgenome PCoA) is stretching the distance between the two allopolyploids though in CE Europe, though there is observed mixture between *D. traunsteineri* and *D. majalis*. *Dactylorhiza majalis* show insignificant isolation by distance with a p-value: 0.051 (Fig. S1c) and *D. traunsteineri* show significant isolation by distance with a p-value of 0.001 (Fig. S1d).

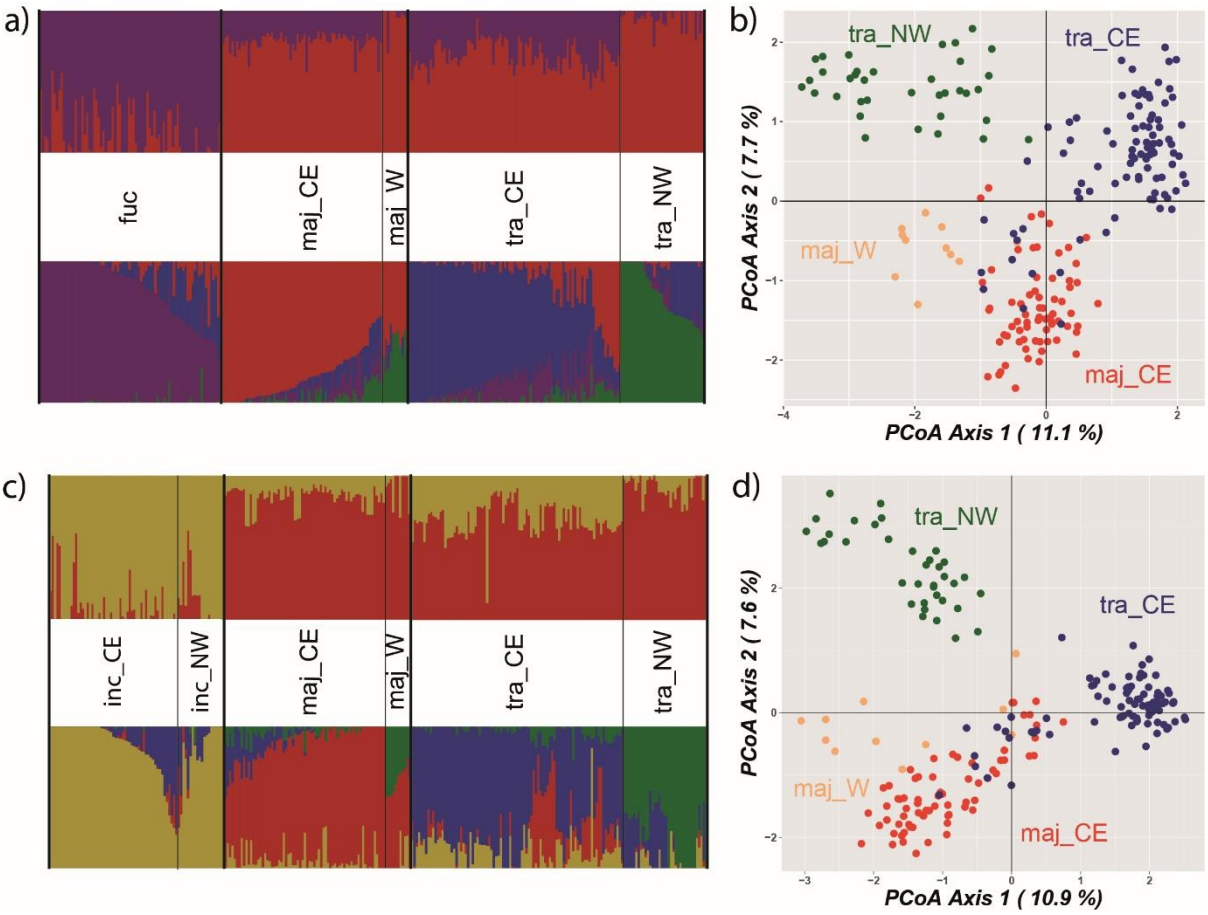


Figure 3: a) ADMIXTURE results based on 2,348 independent SNPs (pipeline I), for K=2 and K=4, for 207 polyploid *Dactylorhiza* individuals and 78 diploid *D. fuchsii* individuals. b) Ordination of 207 polyploid *Dactylorhiza* individuals created with 743 SNPs from the FF subgenome (pipeline II) covered in minimum 85% of the individuals. c) ADMIXTURE results based on 3,278 independent SNPs (pipeline I) for K=2 and K=4, for 207 polyploid *Dactylorhiza* individuals and 75 diploid *D. incarnata* individuals. d) Ordination of 207 polyploid *Dactylorhiza* individuals created with loci covered in minimum 85% of the individuals, comprising 1,087 SNPs from the II subgenome (pipeline II). The ADMIXTURE results were obtained with NGSadmix (Skotte et al. 2013) based on genotype likelihoods from ANGSD, and the genotype-based ordinations were created with the dartR package (Gruber et al. 2018) in R.

Several demographic scenarios regarding the number and order of origins of the allopolyploids were tested with DIY ABC (Cornuet et al. 2014), separately based on the data from each subgenome (Pipeline II). Three scenarios were tested for the history of *D. majalis* (Fig. 4). As both *D. majalis* occur only on mainland Europe, only *D. fuchsii* and *D. incarnata* from mainland Europe were used in these analyses. Altogether 3,000,000 datasets (1,000,000 for each scenario) were simulated and the posterior probabilities of the scenarios were compared

by using two approaches for calculating the probability of each scenario: a direct estimate and a logistic regression. The scenario indicating that *D. majalis* had a single origin consistently had the highest probability for both subgenomes and independent on the approach of estimating the probability (Fig. 4).

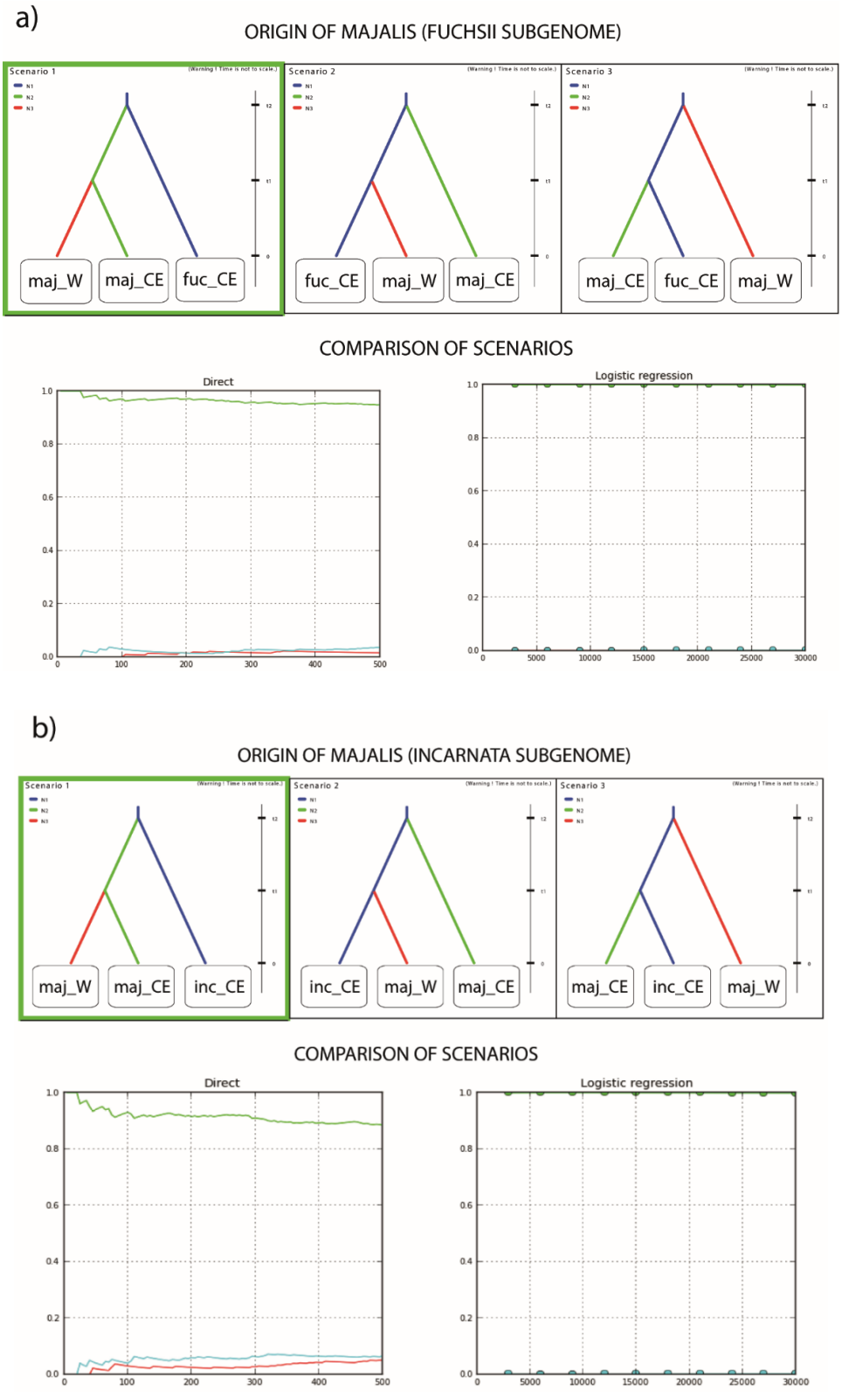
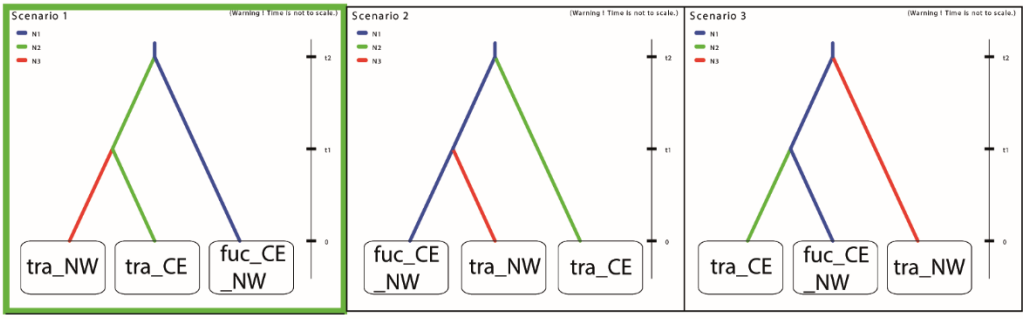


Figure 4: a) DIY ABC (Cornuet et al. 2014) results on three scenarios for the demographic history of *D. majalis* on the FF subgenome, including ten *D. fuchsii*, 21 CE *D. majalis* and six W *D. majalis* individuals on 628 SNPs. b) DIY ABC results on three scenarios for the demographic history of *D. majalis* on the II subgenome, including eight *D. incarnata*, 20 CE *D. majalis* and six W *D. majalis* individuals on 654 SNPs. The probability graphs for each scenario are indicated with a direct and a logistic approach. In both analyses scenario 1 (green box) obtained the highest probability and illustrates *D. majalis* originating once followed later by an allopatric split of CE *D. majalis* from W *D. majalis*.

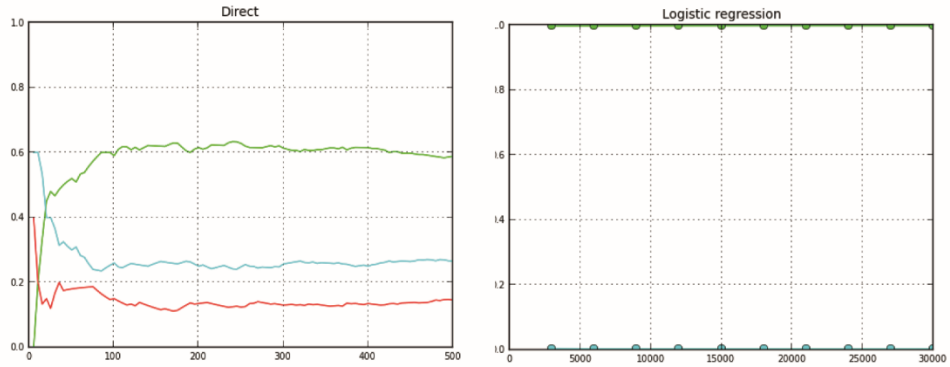
Three scenarios were also tested for the history of *D. traunsteineri* with the data from the FF subgenome (Fig. 5A). Since *D. traunsteineri* occurs both on mainland Europe and on the British Isles, *D. fuchsii* from both areas were included as one group because *D. fuchsii* does not show any significant geographical structure (Fig. 2a, Fig. 3a). For the *D. incarnata*, accessions were included from both mainland Europe and the British Isles, but were split in two groups following their geographical structure (Fig. 2b, Fig. 3c). Hence, more scenarios were needed for the test on the II subgenome (Fig. 5b). Altogether 5,000,000 datasets (1,000,000 for each scenario) and for both approaches and for both subgenomes *D. traunsteineri* had a single origin with the highest probability (Fig 5).

a)

ORIGIN OF TRAUNSTEINERI (FUCHSII SUBGENOME)

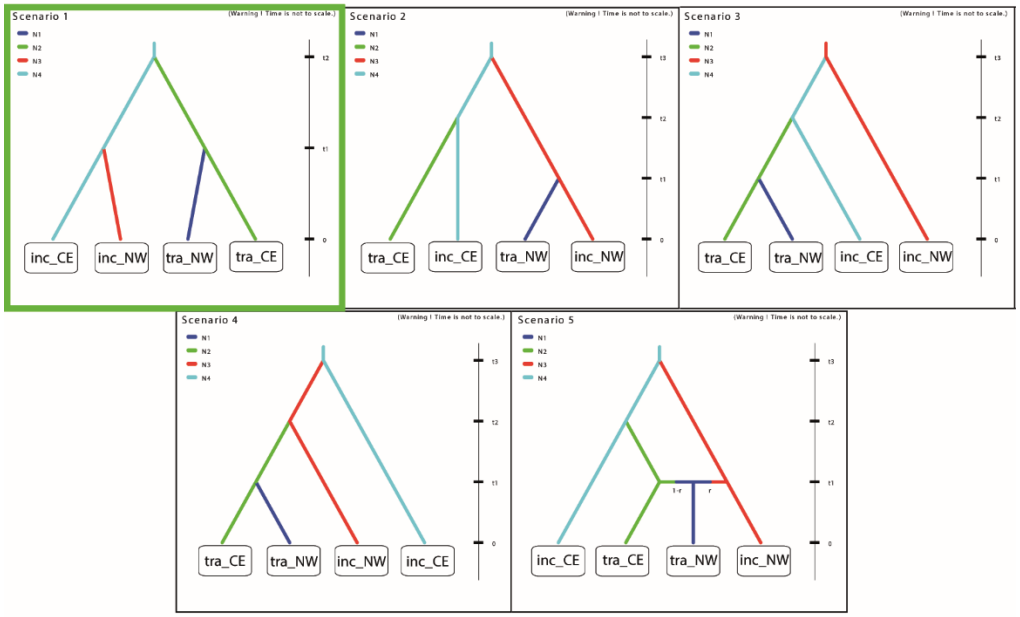


COMPARISON OF SCENARIOS



b)

ORIGIN OF TRAUNSTEINERI (INCARNATA SUBGENOME)



COMPARISON OF SCENARIOS

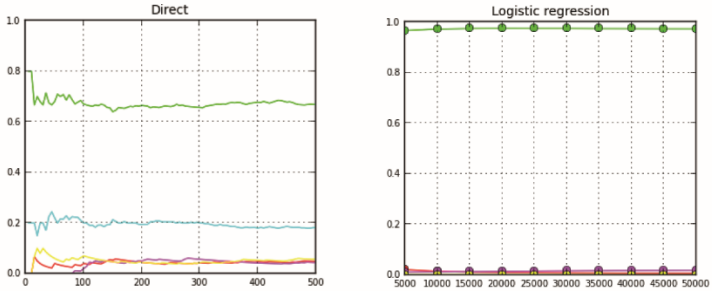
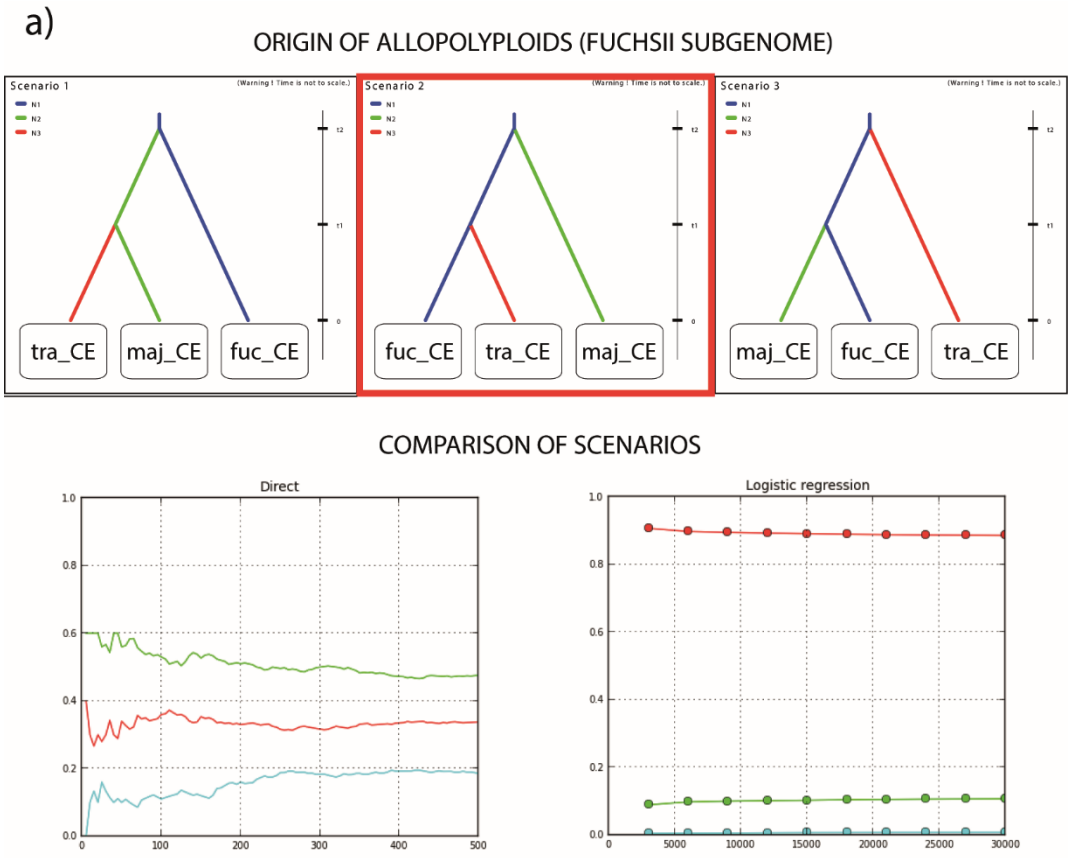


Figure 5: a) DIY ABC (Cornuet et al. 2014) results on three scenarios for the demographic history of *D. traunsteineri* on the FF subgenome, including 20 *D. fuchsii*, 23 CE *D. traunsteineri* and 15 NW *D. traunsteineri* individuals on 755 SNPs. The probability graphs of each scenario are indicated with a direct and a logistic approach. Scenario 1 (green) had the highest probability and illustrates *D. traunsteineri* having a single origin, followed by an allopatric split of CE *D. traunsteineri* from NW *D. traunsteineri*. b) DIY ABC results on five scenarios for the demographic history of *D. traunsteineri* on the II subgenome, including 8 CE *D. incarnata*, 12 NW *D. incarnata*, 24 CE *D. traunsteineri* and 20 NW *D. traunsteineri* individuals on 789 SNPs. Being the best, scenario 1 (green) illustrates *D. traunsteineri* originating only once, followed by allopatric splits within each *D. traunsteineri* and *D. incarnata*.

Finally, three scenarios were also tested for the two sibling allopolyploids, independently on the FF and on the II subgenomes (Fig. 6). Only material from the mainland was used in this analysis for both the diploids and their parents, and altogether 3,000,000 datasets (i.e., 1,000,000 for each scenario) were simulated. All posterior probabilities, except for the direct probability estimation approach for the FF subgenome, supported the scenario with independent origins for each allopolyploid, with *D. majalis* splitting from the diploids earlier than *D. traunsteineri* (Fig. 6).



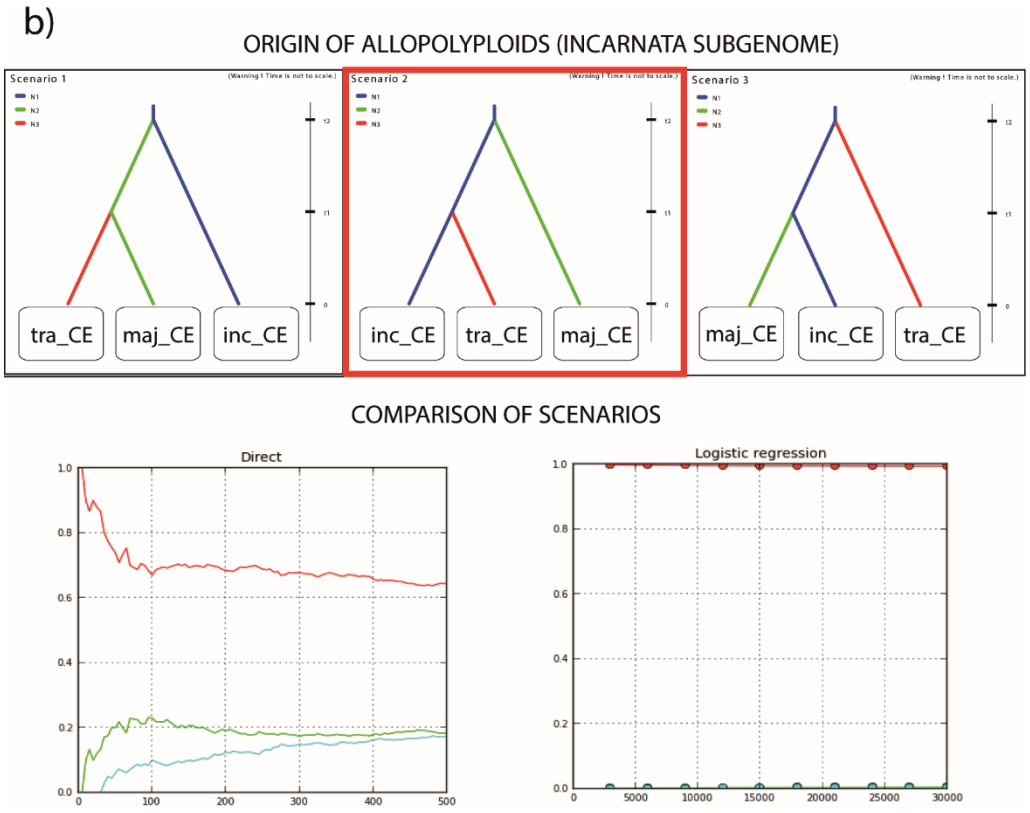


Figure 6: a) DIY ABC (Cornuet et al. 2014) results on three scenarios for the demographic history of *D. majalis* and *D. traunsteineri* on the FF subgenome, including ten *D. fuchsii*, 21 CE *D. majalis* and 23 CE *D. traunsteineri* individuals on 765 SNPs. b) DIY ABC results on three scenarios for the demographic history of *D. majalis* and *D. traunsteineri* on the II subgenome, including 8 individuals of *D. incarnata*, 20 individuals of CE *D. majalis* and 24 of CE *D. traunsteineri* on 758 SNPs. The probability graphs of each scenario are indicated with a direct and a logistic approach. The most likely hypothesis for both subgenomes is scenario 2 (green), which illustrates the allopolyploids having two independent origins for the allopolyploids, i.e., *D. majalis* originating first, and later *D. traunsteineri* originating independently.

Time estimates from the DIY ABC program is outputted in generation time. By using 5,8 years as generation time (Dahlgren et al. 2016), the 95% confidence interval for the time of the origin was estimated to be between 3,000-10,000 years ago (i.e., 540 to 1,730 generations ago) for *D. majalis* and between 2,000-5,350 years ago (i.e., 340 to 920 generations ago) for *D. traunsteineri* (Fig. S6).

DISCUSSION

Evaluation of methodology to process and treat allopolyploid RADseq data

Even though including only a fraction of the genome, the RADseq provides hundredfold or even thousandfold more information than traditionally single nucleotide markers and this way can provide higher resolution, it makes it suitable for entangling more complex situations such as reticulate systems or at the population level. Importantly the loci should provide a fairly equally distributed information from across the genome, providing confidence that the patterns observed reflect the true history experienced by lineages. RADseq is, still, a representation of the genome and RADseq and similar approaches have received criticism when being applied to genome scans for detecting sites under selection, where whole genome data would be desirable. However, compared to alternative methods like RNAseq or exome capture, RADseq offers a representation of the genome with, assumed to be mostly neutral, SNPs that are random and well suited for most aspects of population genetics and demographic inference (Lowry 2016, Catchen 2017, Mckinney 2017).

Bioinformatic processing of RADseq data can dramatically impact population genetic and demographic inferences, especially those making use of SFS, i.e. site frequency spectrum. Measures of population differentiation like F_{ST} (fixation index) and isolation by distance estimates seem to be more robust. Incorporating alignment information by applying a reference genome as well as applying multiple pipelines is advantageous to ensure robust evolutionary inferences (Paris 2017, Shafer 2017), which has also been followed through here. RADseq has been successfully applied to other polyploid plant systems (Chen et al. 2013; Wang et al. 2013).

When working with polyploids the dosage of the different alleles is the key to infer potential partial heterozygotes (Dufresne et al. 2014; Meirmans et al. 2018). Using the read depth of the loci to estimate dosage of alleles is, however, not advised for RADseq, because the read depth may be variable due to restriction fragment bias and GC content bias (Davey et al. 2013). Restriction fragment bias, though expected to be more of a problem for frequently cutting restriction enzymes, is the result of an incomplete shearing of shorter loci since the efficiency of shearing to a suitable length for Illumina sequencing decreases as restriction fragment length decreases. Unsheared or partially sheared fragments will be discarded in the size selection steps and decrease the read depth of those loci. Bias caused by GC content happens during the PCR cycle because RADloci or, even more problematic, alleles with a high GC content will

be less amplified compared to loci/alleles with a lower GC content. Lastly, restriction site polymorphism (i.e., mutations in the restriction site) can cause loci to be found missing in parts of the dataset. These downsides can be minimized by filtering out most loci affected by missing data (Davey 2013), which has been done in the present study.

Genetic structure of the diploids and the polyploids

The *D. fuchsii-maculata* clade and the *D. incarnata-euxina* clade split ~6 MYA (Brandrud et al. submitted - Chapter I) and the two diploid parents of the allopolyploids are genetically divergent (Balao et al. 2017) both when it comes to ecological and genetic makeup, even with visible differences in genome size, i.e. *D. fuchsii* with a 2C value of 5.78 pg and *D. incarnata* with 7.09 pg (Aagaard et al. 2005). *Dactylorhiza fuchsii* grows in semi-open forests on base-rich soils and has a widespread ecological environment that appears less patchy, i.e. more evenly distributed than *D. incarnata* over most of their sympatric distribution (Ståhlberg 2009). This is reflected in its population structure, showing populations with large variation within and little distinction between. *Dactylorhiza incarnata* grows in rich to extreme-rich fens and has more restricted and isolated populations (Hultén 1971; Hedrén 1996d). *D. incarnata* is genetically less variable than *D. fuchsii* (Balao et al. 2017). A low overall diversity may be due to a strong genetic bottleneck during postglacial recolonization (Hedrén and Nordström 2009), whereas low within population diversity may be due to higher inbreeding levels (Pedersen 2009). On a wide geographic scale, the NW European populations are genetically differentiated from the CE European populations, but local populations may be fairly inbred and largely invariant, depending on the marker system.

Despite the divergent evolutionary histories of the subgenomes incorporated in the allopolyploids, e.g. 2C=13.37 pg in *D. traunsteineri* (Aagaard et al. 2005), we find that consistent genetic signals are found in both subgenomes from the analyses performed (Pipeline II). It should be underlined that it is unclear what effect of the accuracy, of our pipeline when separating the loci to subgenomes, has on the consistency of the patterns observed in our demographic inference. Incomplete lineage sorting, which could retain ancestral polymorphism across evolutionary times would also be impairing the accuracy of our inference. Another factor is that depending on the level of divergence between the parents there might be parts of the genome that are more overlapping in the parents than others, called segmental polyploids (Stebbins Jr 1947; Stebbins 1950). The difficulty in our inference is most likely arising from the short reads available, that will be less likely finding diagnostic variants for one or the oth-

er genome. From this perspective approaches delivering longer reads, such as Pacific Biosciences, i.e. PacBio, would have been better to use and has already been applied by others to polyploid systems (Rothfels et al. 2017; Dauphin et al. 2018). Such approaches come, however, with significantly increased costs and error rates (Quail et al. 2012). The idea behind pipeline II was to analyse the loci that cannot be distinguished to parental subgenomes and, in that context, allowing for partial heterozygotes, by working with the genotype likelihoods directly. Interestingly, we find that the two pipelines show quite consistent results for the polyploids, which is also strengthening our findings (Shafer et al. 2017).

Demographic interpretations

D. majalis and *D. traunsteineri* are both members of the *D. majalis* core complex (Nordström and Hedrén 2009), which comprises more than half of all described allopolyploid taxa that have arisen from hybridization between the *D. fuchsii-maculata* clade and the *D. incarnata-euxina* clade in Europe (Landwehr 1977; Delforge 1995; Baumann et al. 2006; Brandrud et al. submitted - Chapter I). Members of the core complex are characterized by leaves that are usually spotted on the upper surface, and purplish flowers with a conical spur and a distinctly, but shallowly three-lobed lip with distinct bow-like streaks arranged concentrically between the spur entrance and the edge of the side lobes. Much-cited authorities on European orchid taxonomy subdivide the complex into a high number of species taxa of which e.g. *D. lapponica*, *D. russowii*, *D. curvifolia*, *D. alpestris*, *D. ebudensis*, *D. traunsteinerioides*, *D. angustata*, *D. parvimajalis* are often accepted in recent literature, in addition to *D. majalis* and *D. traunsteineri* (Delforge 1995; Tyteca 2001). Molecular studies have demonstrated considerable overlap between these taxa in plastid genomes, nuclear microsatellites, and ITS profiles (Pillon et al. 2007; Balao et al. 2016), and there is strong reason to believe that several of the above-mentioned segregates are indeed artificial assemblages of unrelated populations that happen to approach each other in superficial morphology. Much of the diversity of the *D. majalis* core complex is represented here in our sampling (Table S1). We take a conservative approach and use a subdivision of our material into two evolutionary groups that capture most of the morphological, ecological and genetic (see e.g. Fig. 3 and Brandrud et al. submitted - Chapter I) divergence of the complex: (1) *D. majalis*, a stout many-flowered species with a predominantly lowland and submontane distribution in calcareous meadows central Europe, and (2) *D. traunsteineri*, a more slender and few-flowered species with a scattered boreal-montane distribution in Europe, and mostly located in sloping calcareous fens with movable ground-water in the surface.

As evident from our ordinations (Fig. 3b, 3d), the divergence pattern is still somewhat incongruent with the two-group hypothesis. The analysed material diverges in three directions: NW *D. traunsteineri* distributed in Britain, Ireland and westernmost Scandinavia, CE *D. traunsteineri*, and *D. majalis*. Within the latter, there is still a tendency for the populations from the Pyrenees (W *D. majalis*) to form a separate cluster from the rest of *D. majalis*, and to share some alleles with the NW *D. traunsteineri*. This geographically correlated divergence pattern is sometimes reflected in taxonomic systems, where the British *D. traunsteineri* (i.e., but not the entire group of NW *D. traunsteineri*) has been separated as *D. traunsteinerioides* by some authors, and the W *D. majalis* somewhat vaguely corresponds to *D. alpestris*, which is used for low-grown montane forms of *D. majalis* from all over central Europe; the type is from Vorarlberg (W Austria) according to Delforge (1995). Admixture groups, however, should not be immediately treated as ancestral populations without further investigations as there can be different histories leading to similar patterns (Lawson et al. 2018). In our DIY ABC models we have tested these subgroups as independent entities (based on the genetic pattern observed) and we found no reason to test other segregates of the *D. majalis* core complex recognized in the taxonomic literature. In the ordinations (Figs. 3b, 3d), we also observed a fair degree of overlap between *D. traunsteineri* and *D. majalis* in the areas where their distributions overlap and especially in the two sympatric Alps populations. Likewise, in the Bayesian Admixture clustering graphs (Figs. 3a, 3c), many samples from sympatric areas showed similar probability profiles, disregarding their taxonomic belonging. The two factors responsible for the discrepancy between the subdivision into two groups and the observed patterns could be summarized as (1) a geographic subdivision of *D. traunsteineri* and to some extent also *D. majalis*, and (2) gene flow of the two species in the area of sympatry (Balao et al. 2016).

Origin of *D. majalis*: As the most likely scenario in the present analyses (Fig. 4) suggest one origin of *D. majalis*, the genetic clustering apparent in *D. majalis* is interpreted as geographical subdivision and might have appeared more continuous if *D. majalis* had been sampled denser between the Alps and the Pyrenees.

Origin of *D. traunsteineri*: Despite the quite clear division seen in the ordination and the admixture analyses between the NW and CE populations of *D. traunsteineri* (Fig. 3) the tests on *D. traunsteineri* also indicate one origin as the most likely scenario (Fig. 5). Secondary hybridization, with regional parental population leading to introgression, might be happening, but cannot have been the main factor leading to the differentiation found between the populations, as this was tested as scenario 5 on the data from the II subgenome (Fig. 5). The two

groups observed in the *D. traunsteineri* material is thus also interpreted as geographical subdivision. A single origin is the most parsimonious explanation, but to explain the current pattern observed, the populations from CE and NW could now be in the process of diverging to two groups. A model for such a process was proposed for polyploid Pteridophytes where a different pattern of gene silencing is observed in the geographically isolated populations (Werth and Windham 1991).

Origin of the allopolyploids: Independent origins for *D. majalis* and *D. traunsteineri* are found as the most likely scenario in the present study (Fig. 6). The distinction between the two allopolyploids is likely to come from having originated from slightly different parental populations. Further, the distinction is likely to be reinforced by their ecological differentiation, leading to partly different patterns of transcriptomic repatterning after each allopolyploidization event (Wolfe et al. in prep), and by geographical differentiation in areas where they are allopatric. The genetic overlap is found to be due to poor separation in areas of sympatry, i.e. secondary hybridization between the allopolyploids. One would expect that multiple polyploidizations from overlapping sets of parents would result in shared genetic patterns throughout the material, rather than just in sympatric areas. The similarity between NW *D. traunsteineri* and W *D. majalis* could result for example from regional introgression from relatives, such as *D. pratermissa* or *D. maculata*. Alternatively, this pattern might be explained by ancient hybridization between NW *D. traunsteineri* and W *D. majalis.*

Evaluation of demographic inference

DIY ABC is an applied Approximate Bayesian Computations to infer population history (Cornuet et al. 2014). This is done by first defining scenarios, including historical models for how the populations are connected to each other and their ancestor, and generate simulated datasets and then select those closest to the observed data set. Lastly, posterior distributions of parameters are estimated, which is how different scenarios can be compared. DIY ABC provides two measures of posterior probabilities, i.e. directly as the relative proportion of each scenario in the simulated data set closest to the observed data (Miller et al. 2005; Pascual et al. 2007), and as a logistic regression for each scenario's probability calculated from the deviations between the simulated and observed summary statistics (Fagundes et al. 2007; Beaumont 2008).

As it follows one should be aware when applying these kind of ABC analyses, that if the right scenario is not present this will still deliver one of the tested scenarios as the most likely. As

the scenarios are operating with merging and splitting of groups, which data is included in the groups will of course affect the result as well. For this reason, individuals from sympatric populations were not included to avoid, as much as possible, recent admixture to affect the results. Major hybridization events were added to the scenarios when deemed likely; it is, however, difficult to model all contact that has been in the past. Smaller events of gene flow can result in overestimation of effective population sizes and underestimation of divergence times (Leaché et al. 2013) and we suggest that the max 95% estimate is likely closer to the actual age, i.e. ~5.000 years ago for *D. traunsteineri* and ~10.000 years ago for *D. majalis*.

The obtained timepoints are actually an estimate for the time when the groups defined are starting to lose contact with each other, which may not necessarily pinpoint to the exact age of the polyploidization event, rather colonization of one of the groups to a new area as interpreted in (Stenøien et al. 2011). Indeed, due to a lack of endosperm in *Dactylorhiza*, triploids could be frequent and have a degree of fertility, having the potential to ensure genetic cohesion between diploids and allopolyploid derivatives in close sympatry. Hybridization across ploidal levels are found in other *Dactylorhiza* studies (Lord et al. 1977; Lammi et al. 2003; De hert et al. 2011a, 2011b). That *D. majalis* is older than *D. traunsteineri* has been suggested in several studies (e.g. Pillon et al. 2007, Brandrud et al. submitted - Chapter I) and this study can further postulate that it has a likely glacial origin (i.e., most likely it originated towards the end of the last glacial period), perhaps sharing refugia for a short while with its parents. *Dactylorhiza traunsteineri*, on the other hand, is likely to have originated in the stress of rapidly changing environments, which is found to often trigger the production of unreduced gametes (Mason and Pires 2015), after the last glacial age. It is likely that *D. traunsteineri* originated as the ice sheet was withdrawing; its establishment may have been favoured given the newly available habitats and niches unveiled after the ice retreated, and it is now only occupying previously glaciated areas (Bateman 2011). Other papers have looked in more detail on the dispersal patterns of *D. traunsteineri* and *D. majalis* pattern to the north (Hedrén and Olofsson 2018; Nordström and Hedrén 2008; Hedrén et al. 2018), and found multiple routes of colonization as well as multiple colonization events leading to genetically diverse populations of *D. traunsteineri* and *D. majalis*.

MATERIALS AND METHODS

Sampling

Altogether leaf/petal samples from 78 sampling localities of at least five individuals if available per locality were sampled (see Table 1S). Diploid parental individuals of *D. fuchsii* (78 ind, 16 localities) and *D. incarnata* s.l. (75 ind, 16 localities) were sampled throughout their European range (see Table 1S, Fig. 1). Polyploid individuals of D. *traunsteineri* (127 ind, 26 localities) and *D. majalis* (80 ind, 15 localities) were sampled to cover their diversity as well as areas were the two allopolyploids occur sympatrically, i.e. in the Alps, as well as allopatric, i.e. in the Pyrenees (only *D. majalis*) and e.g. in Britain (only *D. traunsteineri*). QGIS v. 2.4.0 was used to created maps of the sampling (QGIS 2015), using a map layer extracted from GADM version 1.0 (available from www.gadm.org).

Laboratory

For molecular analyses DNA was isolated from dried leaves/petals of the samples following a cetyl trimethylammonium bromide (CTAB) procedure or using the DNeasy Plant Mini Kit (Qiagen). Further the DNA was cleaned with the Nucleospin gDNA clean-up Kit (Machery-Nagel). In general, 40 accessions including repeats when necessary, were pooled together per library. For each sample 100-400 ng DNA was used and the amount of each sample was normalized within libraries as well as taking ploidy into account, i.e., double DNA amount was used for tetraploids compared to diploids. The libraries were prepared following the protocol detailed in Brandrud et al. (submitted - Chapter I) using using SbfI-HF (NEB). Eight of the RADseq libraries were sequenced as paired-end 125 bp reads with inline-indexing and the remaining libraries were sequenced as single-end 100 bp reads and inline-index-indexing on an Illumina HiSeq platform at VBCF NGS Unit (www.vbcf.ac.at/ngs), Vienna, Austria.

The retrieved reads were demultiplexed and cleaned with BamIndexDecoder v. 1.03 (included in Picard Illumina2Bam package, available from http://gq1.github.io/illumina2bam/) and PROCESS_RADTAGS with STACKS v. 1.47 (Catchen et al. 2011, 2013). The quality filtered was performed with settings to rescue barcodes and cut sites with a maximum of one mismatch relative to expectation, as well as remove any read with an uncalled base or low scores on average.

Pipeline I

All diploid individuals were used for assembly of loci *de novo* with minimum number of reads for each stack set to five, maximum mismatches allowed between alleles when creating the loci for each individual, as well as number of mismatches allowed when merging loci from different individuals to a common catalog, set to one, and was allowing indels with STACKS. EXPORT_SQL.PL was used to extract the loci present in minimum a 40% of the individuals allowing monomorphic loci as they might still be polymorphic in the polyploids. The loci were blasted (Altschul et al. 1990) against each other by using a python script (https://bitbucket.org/mccannj/ngs_analysis) and redundant loci (i.e., with identity score of 100) were removed. The remaining loci were used as synthetic reference I. Diploid and polyploid reads were mapped with default settings with BOWTIE2 v. 2.2.6 (Langmead and Salzberg 2012) and were sorted and locally realigned with SAMtools v. 1.3.1 (Li et al. 2009, Li 2011), picard-tools v. 2.1.0 (http://broadinstitute.github.io/picard/) and the Genome Analysis Toolkit, i.e. GATK (McKenna et al. 2010).

Genotype likelihoods were calculated with ANGSD v. 0.921 after the GATK model (-GL 2) with a minimum nucleotide and mapping quality of 20, appearing in at least 90% of the individuals and with alleles present in minimum two individuals to generate a beagle variant file (McKenna et al. 2010, Korneliussen et al. 2014). Ordinations were produced with PCAngsd v. 0.95 (Meisner and Albrechtsen 2018) and plotted in R. Admixture analyses were run using only one variant from each locus with NGSadmix v.32 (Skotte et al. 2013). Ten independent runs starting from different seeds were run for K ranging from one to ten. The optimal value of K was selected by calculating deltaK after Evanno et al. (2005) and the final admixture plots were visualized by using a python script (https://github.com/rajanil/fastStructure/blob/master/distruct.py).

Pipeline II

Paired-end diploid individuals were used to extend overlapping reads into contigs for each individual with FLASH v. 1.2.11 (Magoč and Salzberg 2011). The extended contigs were used in pyRAD v. 3.0.36 to assemble loci for *D. fuchsii* and, separately, for *D. incarnata* (Eaton 2014). For the assemblies a minimum depth of ten and a clustering threshold of 95% was used, only allowing for loci present in minimum 20% of the individuals. The two sets of loci obtained were concatenated into a synthetic reference and then the diploid and polyploid data were mapped to it with BOWTIE 2 with default setting. Only uniquely mapping reads with the highest mapping score (42) were further used; the mapped reads were separated to

the FF or the II subgenomes depending on the reference locus they mapped to. The reads were sorted and then variants were called in the form of a vcf-file with REF_MAP.PL and POPULATIONS with STACKS. The accuracy of the pipeline was tested by creating synthetic allopolyploids by matching ten randomly-chosen *D. fuchsii* and ten randomly-chosen *D. incarnata* individuals and running them through the pipeline.

Isolation by distance was tested for each diploid and allopolyploid lineage (i.e., for the latter with data from subgenomes concatenated) by using the function mantel.randtest to perform a Mantel test with 999 permutations with the package ADE4 v1.6-2 (Dray and Dufour 2007). Private alleles were calculated with STACKS for the diploid populations. Ordinations for each diploid species and, separately, for each subgenome in the polyploids were created with the dartR package allowing for maximum 10% missing data with R v. 3.2.3 in RSTUDIO v. 1.0.44 (RStudio Team 2015). Finally, different demographic scenarios were tested in particular with regard to the number of origins of the allopolyploids and their estimated age with DIY ABC v. 2.1.0 (Cornuet et al. 2014). To this aim, vcf-files specific for each ABC-modelling test were created allowing for only one SNP per locus and maximum 10% missingness with populations in STACKS and further converted with a DIY ABC specific python script (https://github.com/loire/vcf2DIYABC.py). To increase feasibility of the analyses, a subset of individuals was selected from those available for each group, aiming to still cover the breadth of genetic variation in each group (Table S1).

ACKNOWLEDGEMENTS

We thank Maria T. Lorenzo, Mark Chase, Mike Fay, Richard M. Bateman, Tobias Kaiser and Jamie McCann for contributing with samples, photos or helpful suggestions or comments. Financial support for this study came from the Austrian Science Fund (FWF, Project Y661-B16 to OP).

REFERENCES

Aagaard S., Såstad S. M., Greuhuber J., Moen A. 2005. A secondary hybrid zone between diploid *Dactylorhiza incarnata* ssp. *cruenta* and allotetraploid *D. lapponica* (Orchidaceae). Heredity 94:488.

Abreu N. L., Alves R. J. V., Cardoso S. R. S., Bertrand Y. J. K., Sousa F., Hall C. F., Pfeil B. E., Antonelli A. 2018. The use of next generation sequencing to solve phylogenetic incongruences in Orchidaceae Juss. PeerJ Preprints 6: e26530v26531.

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. 1990. Basic local alignment search tool. Journal of molecular biology 215:403-410.

Arrigo N., Barker M. S. 2012. Rarely successful polyploids and their legacy in plant genomes. Current opinion in plant biology 15:140-146.
Balao F., Tannhäuser M., Lorenzo M. T., Hedrén M., Paun O. 2016. Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. Heredity 116:351.
Balao F., Trucchi E., Wolfe T., Hao B., Lorenzo M. T., Baar J., Sedman L., Kosiol C., Amman F., Chase M. W., Hedrén M., Paun O. 2017. Adaptive sequence evolution is driven by biotic stress in a pair of orchid species *Dactylorhiza* with distinct ecological optima. Molecular ecology 26:3649-3662.
Barker M. S., Arrigo N., Baniaga A., Li Z., Levin D. A. 2016. On the relative abundance of autopolyploids and allopolyploids. New Phytologist 210:391-398.
Barker M. S., Kane N. C., Matvienko N. C., Kozik A., Michelmore R. W., Knapp S. J., Riesenberg L. H. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Molecular biology and evolution 25:2445-2455.
Baumann H., Künkele S., Lorenz R. 2006. Orchideen Europas: mit angrenzenden gebieten. Weltbild.
Bateman R. M. 2011. Glacial progress: do we finally understand the narrow-leaved marsh orchids? New Journal of Botany 1:2-15.
Beaumont M. A. 2008. Joint determination of topology, divergence time, and immigration in population trees.
Brandrud, M. K., Baar, J., Lorenzo, M. T., Bateman, R. M., Chase, M. W., Paun, O., Hedrén, M. submitted – Chapter 1. Phylogenomic relationships of diploids and the origin of allotetraploids in *Dactylorhiza* (Orchidaceae): RADseq data track reticulate evolution.
Buggs R. J., Renny-Byfield S., Chester M., Jordon-Thaden I. E., Viccini L. F. Chamala S., Leith A. R., Schnable P. S., Barbazuk W. B., Soltis P. S., Soltis D. E. 2012. Next generation sequencing and genome evolution in allopolyploids. American Journal of Botany 99:372-382.
Cannon S. B., McKain M. R., Harkness A., Nelson M. N., Dash S., Deyholos M. K., Peng Y., Joyce B., Stewart C. N., Rolf M., Kutchan T., Tan X., Chen C., Zhang Y., Carpenter E., Wong G. K.-S., Doyle J. J., Leebens-Mack J. 2014. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Molecular biology and evolution 32:193-210.
Catchen J., Hohenlohe P. A., Bassham S., Amores A., Cresko W. A. 2013. Stacks: an analysis tool set for population genomics. Molecular ecology 22:3124-3140.
Catchen J. M., Amores A., Hohenlohe P., Cresko W., Postlethwait J. H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, genomes, genetics 1:171-182.
Catchen J. M., Hohenlohe P. A., Bernatchez L., Funk W. C., Andrews K. R., Allendorf F. W. 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular Ecology Resources, 17:362-365.
Chen X., Li X., Zhang B., Xu J., Wu Z., Wang B., Li H., Younas M., Huang L., Luo Y., Wu J., Hu S., Liu K. 2013. Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in allotetraploid Brassica napus. BMC genomics 14:346.
Comai L. 2005. The advantages and disadvantages of being polyploid. Nature reviews genetics 6:836.
Cornuet, J.-M., Pudlo P., Veyssier J., Dehne-Garcia A., Gautier M., Leblois R., Marin J.-M., Estoup A. 2014. DIYABC v2. 0: a software to make approximate Bayesian

computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30:1187-1189.
Cui L., Wall P. K., Leebens-Mack J. H., Lindsay B. G., Soltis D. E., Doyle J. J., Soltis P., S., Carlson J. E., Arumugabathan K., Barakat A., Albert V. A., Ma H., dePamphilis C. W. 2006. Widespread genome duplications throughout the history of flowering plants. Genome research 16:738-749.
Dahlgren J. P., Colchero F., Jones F., Øien D. I., Moen A., Sletvold N. 2016. Actuarial senescence in a long-lived orchid challenges our current understanding of ageing. Proc. R. Soc. B 283:20161217.
Dauphin B., Grant J. R., Farrar D. R., Rothfels C. J. 2018. Rapid allopolyploid radiation of moonwort ferns (Botrychium; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions. Molecular Phylogenetics and Evolution, 120:342-353.
Davey J. W., Cezard T., Fuentes-Utrilla P., Eland C., Gharbi K., Blaxter M. L. 2013. Special features of RAD Sequencing data: implications for genotyping. Molecular ecology 22:3151-3164.
De Hert K., Jacquemyn H., Van Glabeke S., Roldán-Rui I., Vandepitte K., Leus L., Honnay O. 2011a. Patterns of hybridization between diploid and derived allotetraploid species of *Dactylorhiza* (Orchidaceae) co-occurring in Belgium. American Journal of Botany, 98:946-955.
De Hert K., Jacquemyn H., Van Glabeke S., Roldán-Ruiz I., Vandepitte K., Leus L., Honnay O. 2011b. Reproductive isolation and hybridization in sympatric populations of three *Dactylorhiza* species (Orchidaceae) with different ploidy levels. Annals of Botany, 109:709-720.
Delforge P. 1995. Orchids of Britain and Europe (Vol. 2200244). London: Harper Collins
Dray S., Dufour A.-B. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software 22:1-20.
Dufresne F., Stift M., Vergilino R., Marble B. K. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular ecology 23:40-69.
Eaton D. A. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844-1849.
Etter P. D., Bassham S., Hohenlohe P. A., Johnson E. A., Cresko W. A. 2012. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Molecular methods for evolutionary genetics, Springer: 157-178.
Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14:2611-2620.
Fagundes N. J., Ray N., Beaumont M., Neuenschwander S., Salzano F. M., Bonatto S. L., Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. Proceedings of the National Academy of Sciences 104(45): 17614-17619.
Felber F. 1991. Establishment of a tetraploid cytotype in a diploid population: effect of relative fitness of the cytotypes. Journal of Evolutionary Biology 4:195-207.
Gruber B., Unmack P.J., Berry O.F., Georges A. 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molec. Ecol. Resour. 18:691–699.
Hedrén M. 1996a. Genetic differentiation, polyploidization and hybridization in Northern European *Dactylorhiza* (Orchidaceae): evidence from allozyme markers. Plant Systematics and Evolution 201:31-55.
Hedrén M. 1996b. Electrophoretic evidence for allotetraploid origin of *Dactylorhiza purpurella* (Orchidaceae). Nordic Journal of Botany 16:127-134.

Hedrén M. 1996c. The allotetraploid nature of *Dactylorhiza praetermissa* (Druce) Soó (Orchidaceae) confirmed. Watsonia 21:113-118.
Hedrén M. 1996d. Notes on the esterase variation in Swedish *Dactylorhiza incarnata* sl (Orchidaceae). Nordic Journal of Botany 16:253-256.
Hedrén M. 2002. Speciation patterns in the *Dactylorhiza incarnata/maculata* polyploid complex (Orchidaceae): evidence from molecular markers. J. Eur. Orchideen 34:707-731.
Hedrén M., Fay M. F., Chase M. W. 2001. Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidaceae). American Journal of Botany 88:1868-1880.
Hedrén M., Nordström S. 2009. Polymorphic populations of *Dactylorhiza incarnata* sl (Orchidaceae) on the Baltic island of Gotland: morphology, habitat preference and genetic differentiation. Annals of botany 104:527-542.
Hedrén M., Nordström S., Bateman R. M. 2011. Plastid and nuclear DNA marker data support the recognition of four tetraploid marsh orchids (*Dactylorhiza majalis* sl, Orchidaceae) in Britain and Ireland, but require their recircumscription. Biological Journal of the Linnean Society 104:107-128.
Hedrén M., Nordström S., Ståhlberg D. 2008. Polyploid evolution and plastid DNA variation in the *Dactylorhiza incarnata/maculata* complex (Orchidaceae) in Scandinavia. Molecular ecology 17:5075-5091.
Hedrén M., Olofsson S. N. 2018. High levels of genetic diversity in marginal populations of the marsh orchid *Dactylorhiza majalis* ssp. *majalis*. Nordic Journal of Botany.
Hedrén M., Olofsson S. N., Paun O. 2018. Orchid colonization: multiple parallel dispersal events and mosaic genetic structure in *Dactylorhiza majalis* ssp. *lapponica* on the Baltic island of Gotland. Annals of botany 122:1019-1032.
Heslop-Harrison, J. 1953. Microsporogenesis in some triploid dactylorchid hybrids. Annals of botany 17:539-550.
Hultén E. 1971: Atlas of the distribution of vascular plants in North West Europe. 2nd ed. Generalstabens litografiska anstalts förlag, Stockholm
Husband B. C. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. Proceedings of the Royal Society of London B: Biological Sciences 267:217-223.
Jiao Y., Leebens-Mack J., Ayyampalayam S., Bowers J. E., McKain M. R., McNeal J., Rolf M., Ruzicka D. R., Wafula E., Wickett N. J., Wu X., Zhang Y., Wang J., Zhang Y., Carpenter E. J., Deyholos M. K., Kutchan T. M., Chanderbali A. S., Soltis P. S., Steveson D. W., McCombie R., Pires J. C., Wong G. K.-S., Soltis D. E., dePamphilis C. W. 2012. A genome triplication associated with early diversification of the core eudicots. Genome biology 13:R3.
Jiao Y., Li J., Tang H., Paterson A. H. 2014. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. The Plant Cell: tpc. 114.127597.
Jiao Y., Wickett N. J., Ayyampalayam S., Chanderbali A. S., Landherr L., Ralph P. E., Tomsho L. P., Hu Y., Liang H., Soltis P. S., Soltis D. E., Clifton S. W., Schlarbaum S. E., Schuster S. C., Ma H., Leebens-Mack J., dePamphilis C. W. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97.
Kagale S., Robinson S. J., Nixon J., Xiao R., Huebert T., Condie J., Kessler D., Clarke W. E., Edger P. P., Links M. G., Sharpe A. G., Parkin I. A. P. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell: tpc. 114.126391.
Korneliussen T. S., Albrechtsen A., Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC bioinformatics 15:356.
Lammi A., Vallius E., Vauhkonen T., Kuitunen M. 2003. Outcrossing, hybridization, pollen quantity, and the evolution of deceptive pollination in Dactylorhiza incarnata.

In Annales Botanici Fennici (Vol. 40, No. 5, pp. 331-338). Helsinki: Societas Biologica Fennica Vanamo, 1964-.
Landwhr J. 1977. Wilde orchideeën van Europa I. – Vereniging tot Behoud van Natuurmonumenten in Nederland
Langmead B., Salzberg S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357.
Lawson D. J., Van Dorp L., Falush D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature communications, 9:3258.
Leaché A. D., Harris R. B., Rannala B., Yang Z. 2013. The influence of gene flow on species tree estimation: a simulation study. Systematic biology 63:17-30.
Levin D. A. 2002. The role of chromosomal change in plant evolution, Oxford University Press.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987-2993.
Li, H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.
Lord R. M., Richards A. J. 1977. A hybrid swarm between the diploid *Dactylorhiza fuchsii* (Druce) Soó and the tetraploid *D. purpurella* (T. & TA Steph.) Soó in Durham. Watsonia, 11:205-210.
Lowry D. B., Hoban S., Kelley J. L., Lotterhos K. E., Reed L. K., Antolin M. F., Storfer A. 2017. Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular ecology resources, 17:142-152.
Magoč T., Salzberg S. L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963.
Mallet J. 2007. Hybrid speciation. Nature 446:279.
Mason A. S., Pires J. C. 2015. Unreduced gametes: meiotic mishap or evolutionary mechanism*?* Trends in Genetics, 31:5-10.
McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research.
McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. 2017. RAD seq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al.(2016). Molecular ecology resources, 17:356-361.
Meirmans P. G., Liu S., van Tienderen P. H. 2018. The analysis of polyploid genetic data. Journal of Heredity 109:283-296.
Meisner J., Albrechtsen A. 2018. Inferring Population Structure and Admixture Proportions in Low Depth NGS Data. bioRxiv: 302463.
Miller N., Estoup A., Toepfer S., Bourguet D., Lapchin L., Derridj S., Kim K. S., Reynaud P., Furlan L., Guillemaud T. 2005. Multiple transatlantic introductions of the western corn rootworm. science 310:992-992.
Nordström S., Hedrén M. 2008. Genetic differentiation and postglacial migration of the *Dactylorhiza majalis* ssp. *traunsteineri/lapponica* complex into Fennoscandia. Plant Systematics and Evolution 276:73-87.
Nordström S., Hedrén M. 2009. Genetic diversity and differentiation of allopolyploid *Dactylorhiza* (Orchidaceae) with particular focus on the *Dactylorhiza majalis* ssp. *traunsteineri/lapponica* complex. Biological Journal of the Linnean Society, 97:52-67.

Pascual M., Chapuis M. P., Mestres F., Balanya J., Huey R. B., Gilchrist G. W., Serra L., Estoup A. 2007. Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. Molecular ecology 16:3069-3083.

Paun O., Bateman R. M., Fay M. F., Hedrén M., Civeyrel L., Chase M. W. 2010. Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorhiza*: Orchidaceae). Molecular biology and evolution 27:2465-2473.

Paun O., Fay M. F., Soltis D. E., Chase M. W. 2007. Genetic and epigenetic alterations after hybridization and genome doubling. Taxon 56:649-656.

Paun O., Forest F., Fay M. F., Chase M. W. 2011 Parental divergence and hybrid speciation in angiosperms revisited. Taxon 60:1241-1244

Paris J. R., Stevens J. R., Catchen J. M. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution, 8:1360-1373.

Pedersen H. Æ. 2009. Apochromic populations of *Dactylorhiza incarnata* s.l. (Orchidaceae): diversity and systematic significance as revealed by allozyme markers and morphology. Bot. J. Linnean Soc. 159:396-407.

Pillon Y., Fay M. F., Hedrén M., Bateman R. M., Devey D. S., Shipunov A. B., van der Bank M., Chase M. W. 2007. Evolution and temporal diversification of western European polyploid species complexes in *Dactylorhiza* (Orchidaceae). Taxon 56:1185-1185.

Puritz J. B., Hollenbeck C. M., Gold J. R. 2014a. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2: e431.

Puritz J. B., Matz M. V., Toonen R. J., Weber J. N., Bolnick D., Bird C. E. 2014b. Demystifying the RAD fad. Molecular ecology 23:5937-5942.

Quail M. A., Smith M., Coupland P., Otto T. D., Harris S. R., Connor T. R., Bertoni A., Swerdlow H. P., Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics, 13:341.

QGIS. 2015. QGIS Geographic Information System. Open Source Geospatial Foundation Project http://www.qgis.org/.

Ramsey J., Schemske D. W. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual review of ecology and systematics 29:467-501.

Rieseberg L. H., Raymond O., Rosenthal D. M., Lai Z., Livingstone K., Nakazato T., Durphy J. L., Schwarzbach A. E., Donovan L. A., Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211-1216.

Rieseberg L. H., Willis J. H. 2007. Plant speciation. Science 317:910-914.

Rothfels C. J., Pryer K. M., Li F. W. 2017. Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. New Phytologist, 213:413-429.

RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com.

Shafer A., Peart C. R., Tussi S., Maayan I., Brelsford A., Wheat C. W., Wolf J. B. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution 8:907-917.

Shi T., Huang H., Barker M. S. 2010. Ancient genome duplications during the evolution of kiwifruit (Actinidia) and related Ericales. Annals of Botany 106:497-504.

Skotte L., Korneliussen T. S., Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics: genetics. 113.154138.

Soltis D. E., Albert V. A., Leebens-Mack J., Bell C. D., Paterson A. H., Zheng C., Sankoff D., de Pamphilis C. W., Wall P. K., Soltis P. S. 2009. Polyploidy and angiosperm diversification. American Journal of Botany 96:336-348.

Soltis D. E., Buggs R. J. A., Doyle J.J., Soltis P.S. 2010 What we still don't know about polyploidy. Taxon 59:1387-1403.
Soltis D. E., Soltis P. S. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution 14:348-352.
Soltis D. E., Soltis P. S., Schemske D. W., Hancock J. F., Thompson J. N., Husband B. C., Judd W. S. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56:13-30.
Stebbins G. L. 1950. Variation and evolution in plants, Geoffrey Cumberlege.; London.
Stebbins Jr G. L. 1947. Types of polyploids: their classification and significance. Advances in genetics, Elsevier. 1:403-429.
Stenøien H., Shaw A., Stengrundet K., Flatberg K. I. 2011. The narrow endemic Norwegian peat moss *Sphagnum troendelagicum* originated before the last glacial maximum. Heredity 106:370.
Ståhlberg D. 2009. Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid *Dactylorhiza maculata* sl (Orchidaceae). Evolutionary Ecology 23:295.
Ståhlberg D., Hedrén M. 2008. Systematics and phylogeography of the *Dactylorhiza maculata* complex (Orchidaceae) in Scandinavia: insights from cytological, morphological and molecular data. Plant Systematics and Evolution 273:107.
Tyteca D. 2001. Systematics and biostatistics of *Dactylorhiza* in western Europe: some recent contributions. J Eur Orchid 33:179–199
Twyford A., Ennos R. 2012. Next-generation hybridization and introgression. Heredity 108:179.
Vekemans D., Proost S., Vanneste K., Coenen H., Viaene T., Ruelens P., Maere S., Van de Peer Y., Geuten K. 2012. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. Molecular biology and evolution 29:3793-3806.
Wang N., Thomson M., Bodles W. J. A., Crawford R. M. M., Hunt H. V., Featherstone A. W., Pellicer J., Buggs R. J. A. 2013. Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. Molecular ecology 22:3098-3111.
Werth C. R.,Windham M. D. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. The American Naturalist 137:515-526.
Wood T. E., Takebayashi N., Barker M. S., Mayrose I., Greenspoon P. B., Riesenberg L. H. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences 106:13875-13879.

SUPPLEMENTARY MATERIAL

Table S1: Sampling details for each locality of each taxon. The individuals which were included in different analyses are also indicated. The coordinates are approximate.

Taxon	Country	Loc code	Latitude	Longitude	number ind	genetic structure	DIY ABC
D. fuchsii	Austria	au20	47°46'42.0"N	15°13'48.0"E	5	X	
D. fuchsii	England	en12	51°15'18.0"N	00°18'30.0"W	5	X	X
D. fuchsii	Estonia	es1	58°17'00.0"N	22°08'00.0"E	5	X	
D. fuchsii	France	fr4	44°11'06.0"N	07°09'42.0"E	5	X	
D. fuchsii	France	fr7	42°53'42.0"N	01°55'48.0"E	5	X	
D. fuchsii	Italy	it1	46°58'48.0"N	11°54'12.0"E	4	X	
D. fuchsii	Norway	no1	70°03'00.0"N	22°57'00.0"E	5	X	

D. fuchsii	Norway	no2	59°49'00.0"N	10°18'00.0"E	5	X	
D. fuchsii	Poland	po2	54°28'00.0"N	18°33'00.0"E	5	X	X
D. fuchsii	Russia	ru1	61°40'00.0"N	31°22'00.0"E	4	X	
D. fuchsii	Russia	ru6	58°09'36.0"N	32°40'42.0"E	5	X	
D. fuchsii	Scotland	sc3	57°41'06.0"N	07°12'18.0"W	5	X	X
D. fuchsii	Sweden	sw30	55°42'06.0"N	13°22'06.0"E	5	X	
D. fuchsii	Sweden	sw31	60°38'18.0"N	17°17'42.0"E	5	X	
D. fuchsii	Sweden	sw8	63°28'00.0"N	15°20'00.0"E	5	X	X
D. fuchsii	The Netherlands	ne2	51°53'48.0"N	04°05'12.0"E	5	X	
D. incarnata	Austria	au9	47°17'00.0"N	11°11'06.0"E	5	X	
D. incarnata	England	en4	53°39'12.0"N	03°04'00.0"W	5	X	X
D. incarnata	England	en6	50°48'59.8"N	01°28'59.9"W	4	X	X
D. incarnata	Finland	fi1	67°34'00.0"N	26°52'00.0"E	5	X	
D. incarnata	France	fr1	42°51'42.0"N	01°58'48.0"E	5	X	
D. incarnata	Hungary	hu2	47°40'36.0"N	16°38'24.0"E	4	X	
D. incarnata	Norway	no2	59°49'00.0"N	10°18'00.0"E	5	X	
D. incarnata	Norway	no4	58°39'00.0"N	05°34'00.0"E	5	X	X
D. incarnata	Poland	po3	54°39'13.5"N	18°10'12.8"E	3	X	X
D. incarnata	Russia	ru7	62°00'56.9"N	34°04'55.3"E	4	X	
D. incarnata	Scotland	sc3	57°41'06.0"N	07°12'18.0"W	6	X	X
D. incarnata	Sweden	sw2	55°55'51.6"N	14°04'05.9"E	5	X	
D. incarnata	Sweden	sw3	57°20'24.7"N	18°19'16.3"E	5	X	
D. incarnata	Sweden	sw4	57°49'01.5"N	18°53'43.3"E	5	X	
D. incarnata	Sweden	sw7	63°51'00.0"N	14°01'60.0"E	4	X	
D. incarnata	The Netherlands	ne3	51°53'48.0"N	04°05'12.0"E	5	X	X
D.traunsteineri	Austria	au8	47°27'39.5"N	12°21'56.3"E	6	X	
D.traunsteineri	Austria	au7	47°20'28.7"N	12°48'18.9"E	6	X	
D.traunsteineri	Austria	au3	47°31'45.4"N	12°34'42.7"E	6	X	
D.traunsteineri	Czech	cz2	50°03'24.5"N	15°42'42.3"E	5	X	X
D.traunsteineri	England	en1	54°15'10.0"N	00°41'06.3"W	5	X	X
D.traunsteineri	England	en2	54°16'53.7"N	00°41'23.2"W	5	X	X
D.traunsteineri	Estonia	es1	58°16'59.9"N	22°07'59.9"E	5	X	
D.traunsteineri	Finland	fi1	67°33'59.8"N	26°51'59.8"E	5	X	
D.traunsteineri	Germany	ge3	49°31'60.0"N	07°57'00.0"E	4	X	X
D.traunsteineri	Germany	ge4	53°55'14.0"N	13°25'59.9"E	3	X	X
D.traunsteineri	Ireland	ir1	54°59'26.9"N	05°59'36.1"W	5	X	X
D.traunsteineri	Norway	no6	63°01'60.0"N	08°52'00.0"E	4	X	X
D.traunsteineri	Norway	no10	58°53'02.6"N	05°36'14.5"E	5	X	X
D.traunsteineri	Norway	no2	59°49'00.0"N	10°18'00.0"E	5	X	
D.traunsteineri	Romania	ro5	46°42'53.8"N	23°37'18.0"E	5	X	X
D.traunsteineri	Russia	ru4	61°46'60.0"N	33°48'00.0"E	4	X	X
D.traunsteineri	Scotland	sc1	57°25'19.3"N	05°49'09.6"W	5	X	X
D.traunsteineri	Scotland	sc2	57°46'01.2"N	05°34'08.4"W	5	X	X
D.traunsteineri	Scotland	sc3	57°41'07.4"N	07°12'20.4"W	6	X	
D.traunsteineri	Sweden	sw13	63°31'00.0"N	16°25'00.0"E	3	X	X
D.traunsteineri	Sweden	sw15	58°19'07.4"N	13°48'10.7"E	4	X	X
D.traunsteineri	Sweden	sw5	60°36'53.0"N	17°33'30.0"E	5	X	X
D.traunsteineri	Sweden	sw6	66°19'21.9"N	23°30'30.6"E	5	X	X

D.traunsteineri	Sweden	sw7	63°51'00.0"N	14°01'60.0"E	5	X	
D.traunsteineri	Sweden	sw3	57°20'24.7"N	18°19'16.3"E	6	X	
D.traunsteineri	Sweden	sw4	57°49'01.5"N	18°53'43.3"E	5	X	
D. majalis	Austria	au1	47°41'07.9"N	15°39'20.9"E	6	X	X
D. majalis	Austria	au2	47°44'41.5"N	15°21'03.0"E	5	X	X
D. majalis	Austria	au5	47°35'08.4"N	15°05'58.2"E	6	X	X
D. majalis	Austria	au6	47°54'19.5"N	14°09'55.8"E	5	X	X
D. majalis	Austria	au8	47°27'39.5"N	12°21'56.3"E	5	X	
D. majalis	Austria	au3	47°31'45.4"N	12°34'42.7"E	5	X	
D. majalis	Belgium	be3	50°01'55.0"N	05°22'25.1"E	5	X	X
D. majalis	France	fr27	44°19'10.7"N	06°52'17.5"E	5	X	
D. majalis	France	fr2	42°51'31.4"N	00°29'43.2"E	5	X	X
D. majalis	France	fr1	42°51'41.9"N	01°58'50.7"E	6	X	X
D. majalis	Germany	ge1	53°25'60.0"N	10°06'00.0"E	5	X	X
D. majalis	Poland	po1	54°06'59.8"N	17°52'59.9"E	5	X	X
D. majalis	Poland	po5	49°28'26.0"N	20°12'44.2"E	5	X	X
D. majalis	Sweden	sw1	55°49'04.1"N	12°56'44.2"E	5	X	X
D. majalis	Sweden	sw2	55°55'51.6"N	14°04'05.9"E	7	X	

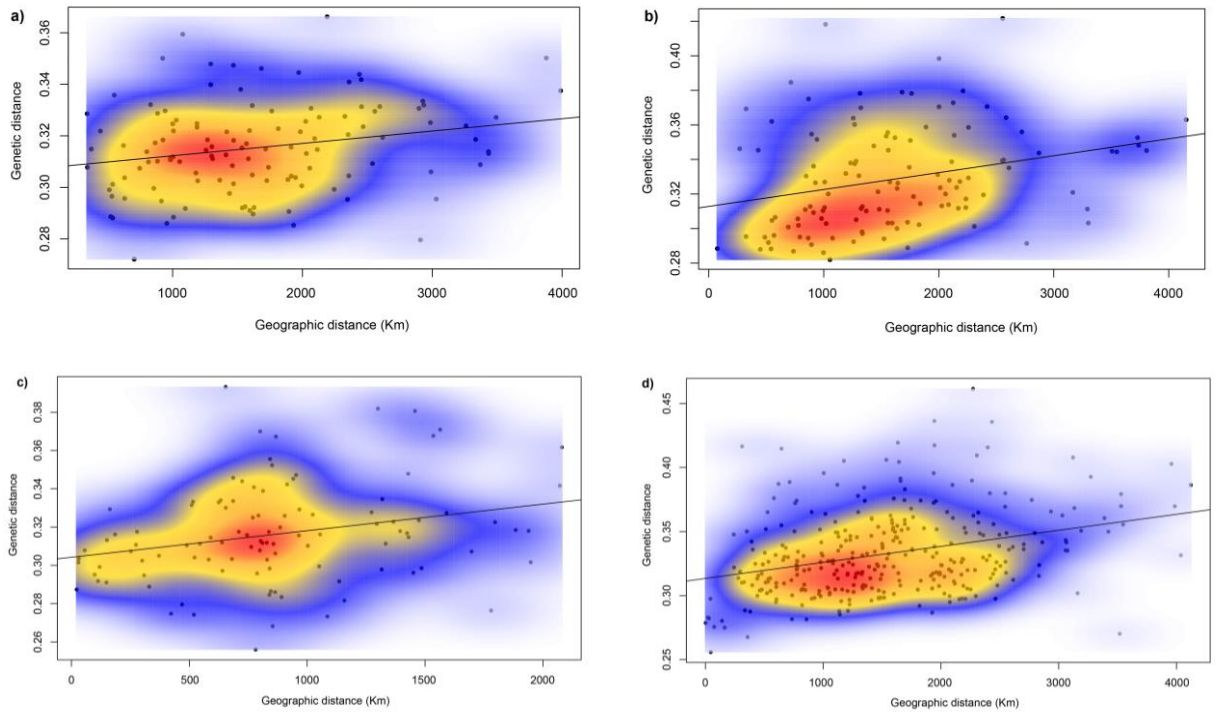


Figure S1: Relationship between genetic distance and geographical distance, i.e. Euclidean distances, performed with a mantel.randtest for a) the *D. fuchsii* populations on 13,242 SNPs (rM = 0.23, n.s.); b) the *D. incarnata* populations on 8,580 SNPs (rM = 0.28, p = 0.048); c) *D. majalis* populations on 21,535 SNPs (rM = 0.27, n.s.); and d) *D. traunsteineri* populations on 24,289 SNPs (rM = 0.32, p = 0.001). All genetic data was extracted with Pipeline II. The col-

our shading is reflecting the density of points modelled from a two-dimensional kernel density estimation.

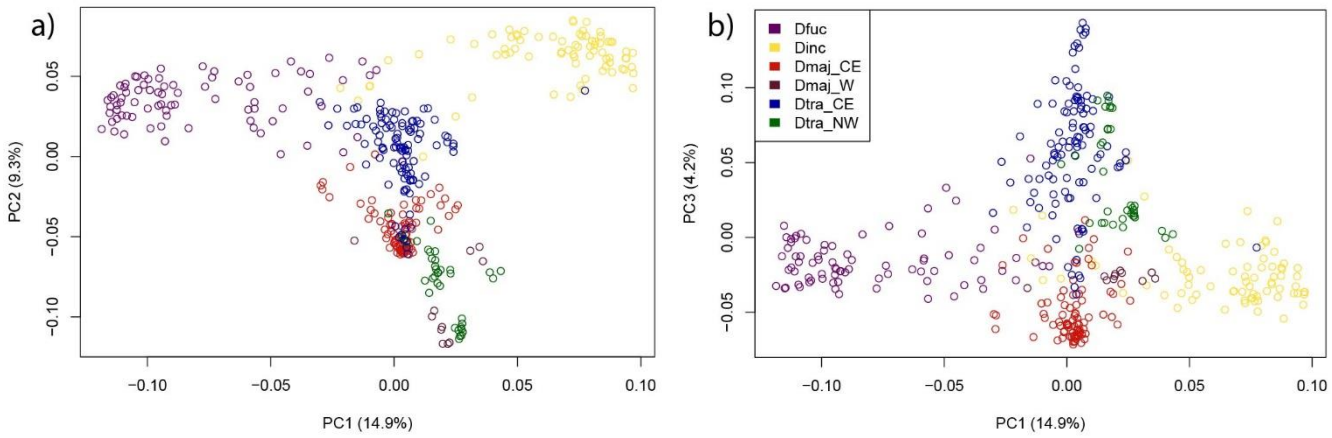


Figure S2: Ordinations performed on genotype likelihoods (Pipeline I) of 12,474 variants by using PCAngsd (Meisner and Albrechtsen 2018) on 153 diploid and 207 polyploid *Dactylorhiza* individuals. a) is showing axis 1 and axis 2. b) is showing axis 1 and axis 3.

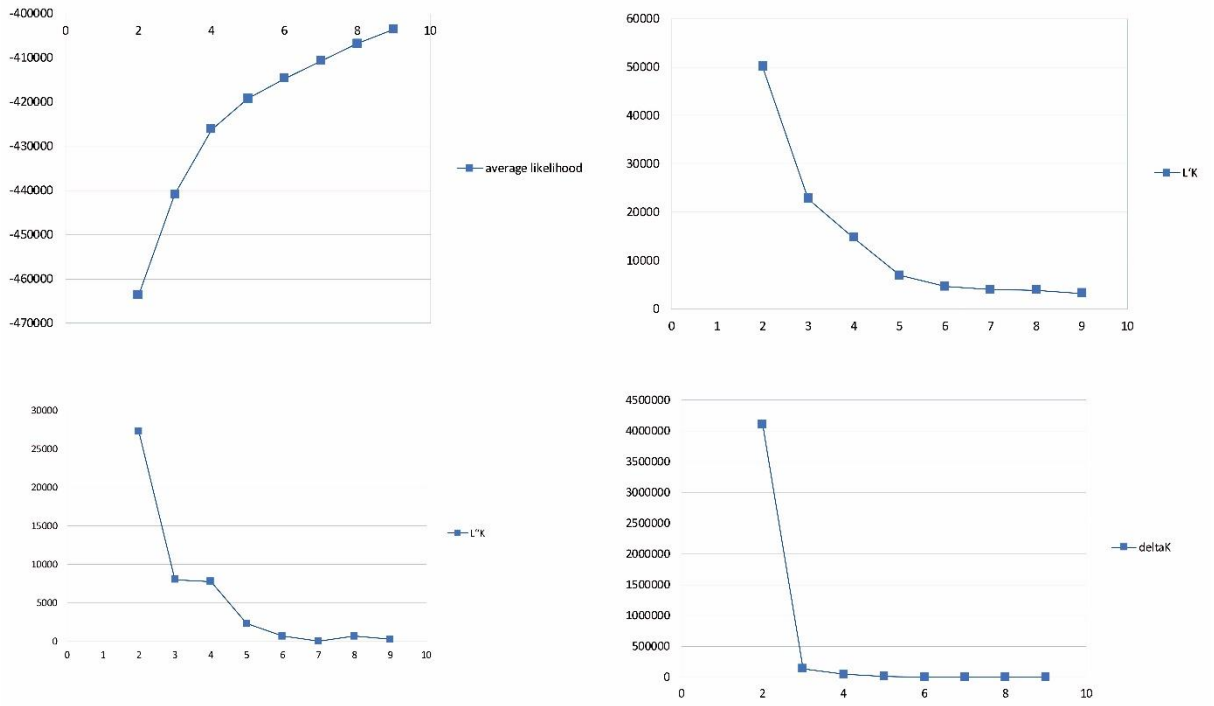


Fig. S3: Plots calculated from the log probability from ADMIXTURE runs for K=1-10 after Evanno (2005) for 207 polyploids and 78 diploids on 2,348 SNPs from the FF subgenome (Pipeline I).

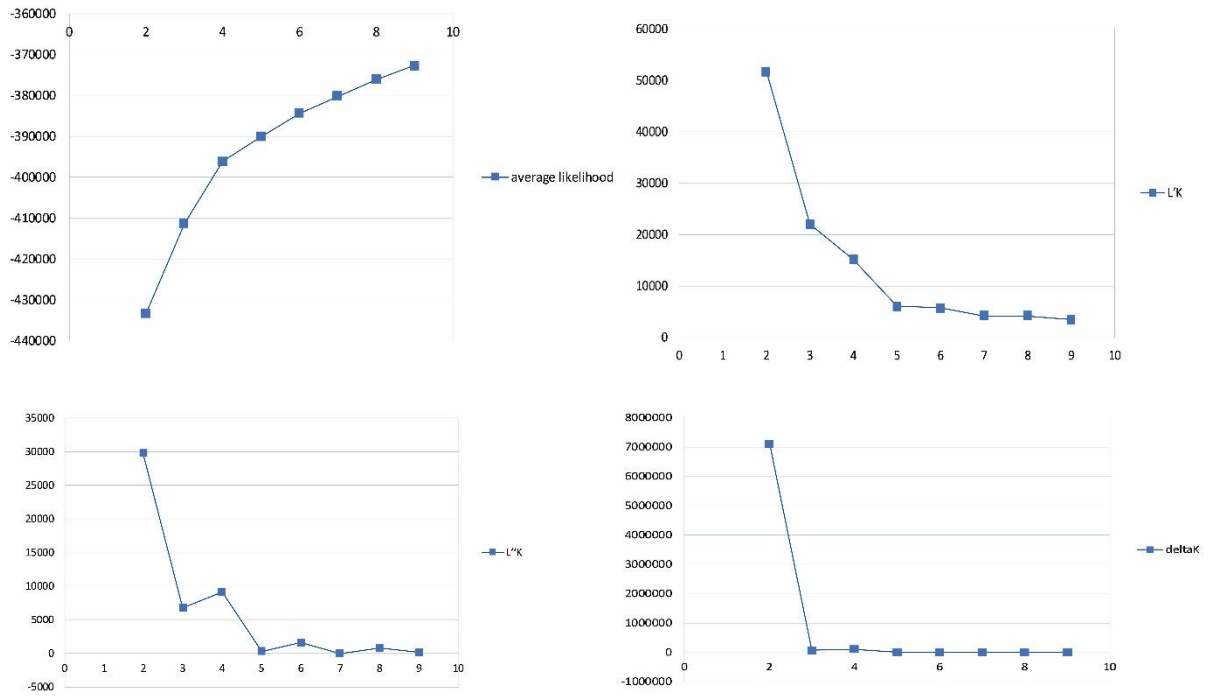


Figure S4: Plots calculated from the log probability from ADMIXTURE runs for K=1-10 after Evanno (2005) for 207 polyploids and 75 diploids on 3,278 SNPs from the II subgenome (Pipeline I).

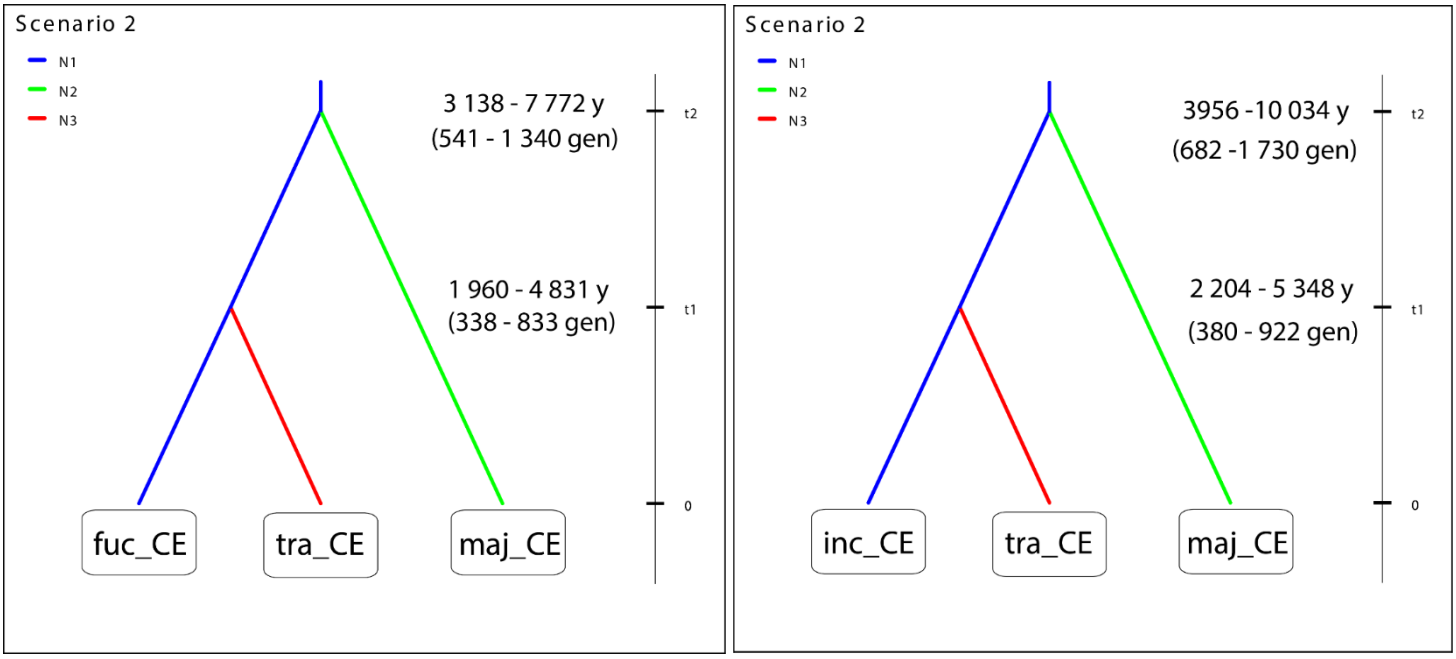


Figure S5: a) Time estimates indicated for the most likely scenario from the DIY ABC (Cornuet et al. 2014) on *D. traunsteineri* and *D. majalis* on the FF subgenome. b) Time estimates indicated for the most likely scenario from the DIY ABC on *D. traunsteineri* and *D. majalis* on the II subgenome.

— CHAPTER 3 —

Restriction site-associated DNA sequencing supports a sister group relationship of *Nigritella* and *Gymnadenia* (Orchidaceae)

Marie K. Brandrud, Ovidiu Paun, Richard Lorenz,

Juliane Baar, Mikael Hedrén

Status: Submitted to Molecular Phylogenetics and Evolution

Contribution: Data processing, analyses and visualization, writing manuscript

Restriction-site associated DNA sequencing supports a sister group relationship of *Nigritella* and *Gymnadenia* (Orchidaceae)

Marie K. Brandrud[1], Ovidiu Paun[1], Richard Lorenz[2], Juliane Baar[1] and Mikael Hedrén[3,4]

1 *Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria*

2 AHO Baden-Württemberg, Leibnizstrasse 1, D-69459 Weinheim, Germany

3 *Department of Biology, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden*

4 Author for correspondence

ABSTRACT

The orchid genus *Nigritella* is closely related to *Gymnadenia* and has from time to time been merged with the latter. Although *Nigritella* is morphologically distinct, it has been suggested that the separating characters are easily modifiable and subject to rapid evolutionary change. So far, molecular phylogenetic studies have either given support for the inclusion of *Nigritella* in *Gymnadenia*, or for their separation as different genera. To resolve this issue, we analysed data obtained from Restriction-site associated DNA sequencing, RADseq, which provides a large number of SNPs distributed across the entire genome. To analyse samples of different ploidies, we take an analytical approach of building a reduced genomic reference based on *de novo* RADseq loci reconstructed from diploid accessions only, which we further use to map and call variants across both diploid and polyploid accessions. We found that *Nigritella* is distinct from *Gymnadenia* forming a well-supported separate clade, and that genetic diversity within *Gymnadenia* is high. Within *Gymnadenia*, taxa characterized by an ITS-E ribotype (*G. conopsea* s.str. (early flowering) and *G. odoratissima*), are divergent from taxa characterized by ITS-L ribotype (*G. frivaldii, G. densiflora* and late flowering *G. conopsea*). *Gymnigritella runei* is confirmed to have an allopolyploid origin from diploid *Gymnadenia conopsea* and tetraploid *N. nigra* ssp. *nigra* on the basis of RADseq data. Within *Nigritella* the aggregation of polyploid members into three clear-cut groups as suggested by allozyme and nuclear microsatellite data was further supported.

Keywords: *Gymnadenia*, *Nigritella*, Orchidaceae, phylogenomics, polyploidy, RADseq

INTRODUCTION

Orchids are often subject to intense taxonomic debate regarding generic delimitation and species circumscription (Bateman et al. 1997). On one hand their immense diversity summing up to 10% of angiosperm species may be challenging to grasp and catalogue and on the other hand their often complex pollination syndromes, opening opportunities for gene flow may blurr phenotypic and genetic patterns as inconclusive for taxon delimitation. Lastly, the flagship conservation efforts and their ornamental qualities make orchids particularly prone for subjective taxonomic opinions.

The two orchid taxa *Gymnadenia* R.Br. and *Nigritella* Rich. are easily separated in the field and are regarded by many authors as different genera based on morphological evidence (e.g., Moore 1980, Baumann et al. 2006). Members of *Gymnadenia* are characterized by an elongate, inflorescence with resupinate flowers carrying a thin, medium-sized to long spur, whereas *Nigritella* species have a short and dense head-like inflorescence with non-resupinate flowers carrying a minute sac-like spur (Moore 1980, Baumann et al. 2006). Nevertheless, the two genera are obviously related, with common features such as a deeply divided palmate-digitate tuber, narrow unspotted leaves, and similar morphologies of the column (Pridgeon et al. 2001; Claessens and Kleynen 2011). Since the columnar structure is emphasized in orchid classification, some orchid systematists have argued that the genera should be collapsed under *Gymnadenia* on the basis of this apomorphic morphological feature (Løjtnant 1977; Sundermann 1980).

Gymnadenia s. str. has a wide distribution in temperate-boreal Eurasia, where it is found in open grasslands in mountain regions, as well as lowland areas including semi-open woodlands. *Nigritella* is confined to mountain regions of Europe and is considered as one of the few orchid genera endemic to Europe (Gjærevoll 1992). They both produce nectar and are mainly pollinated by *Lepidoptera* (Vöth 2000; Claessens and Kleynen 2011), however, largely by different species (Vöth 2000; Claessens and Kleynen 2011).

Population genetic studies performed in *Gymnadenia* have previously found relatively high levels of diversity within populations, low levels of differentiation between populations, and usually moderate levels of inbreeding (Scacchi and De Angelis 1989; Soliva and Widmer 1999; Gustafsson 2000; Gustafsson and Sjögren-Gulve 2002; Gustafsson and Lönn 2003). Both genera include diploids with $2n$=40 and polyploids with multiples of the base number x=20 (Moore 1980). Most species and populations of *Gymnadenia* are typically diploid, but in the mountains of Central Europe, populations of *G. conopsea* can be tetraploid or mixed dip-

loid/tetraploid (Trávníček et al. 2012). Small numbers of individuals with odd ploidies can also be found. As far as known, polyploid *Gymnadenia* are sexual and outcrossing, as are the diploids.

Diploid members of *Nigritella* are sexual and predominantly outcrossing, with a high degree of genetic variation within and between populations (Teppner and Klein 1990; Hedrén et al. 2000, 2018). Polyploid members of *Nigritella* are all reproducing asexually by agamospermy (i.e., nucellar embryony; Teppner 1996, 2004) and they have been suggested to be allopolyploids (Hedrén et al. 2000). These taxa are less variable, but populations are still often multiclonal, and some clones can be regionally widespread and shared between populations (Hedrén et al. 2018). Most polyploid *Nigritella* are tetraploid, but there is also one triploid, and one pentaploid species (Teppner 2004).

Despite the significant difference in spur length and the resulting differential deposition of pollinaria on visiting insects, hybrids between members of *Gymnadenia* and *Nigritella* are rarely encountered. It is doubtful whether hybridization ever goes beyond primary hybrid formation (Gerbaud and Schmid 1999). However, one of the hybrids has inherited the capacity to reproduce by agamospermy from its *Nigritella* parent, and is recognized as a separate species, *Gymnigritella runei* (Teppner and Klein 1989).

Molecular studies have so far shown conflicting results or have been inconclusive on the exact relationships between *Gymnadenia* and *Nigritella.* Phylogenetic analyses of nuclear ITS sequences, which are highly variable spacer regions separating ribosomal genes in the tandemly repeated rDNA regions of the nuclear genome (Jorgensen and Cluster 1988), have given support to the argument that *Nigritella* should be included in *Gymnadenia* (Pridgeon et al. 1997, Bateman et al. 1997, Bateman et al. 2003, Bateman et al. 2006, Stark et al. 2011, Bateman et al. 2018). According to ITS phylogenies, the primary subdivision of the group results in one small clade comprising *G. conopsea* s.str. and *G. odoratissma*, and a second more species-rich and diverse clade containing other members of typical *Gymnadenia* including *G. densiflora*, *G. borealis*, *G. frivaldii*, the Asian *G. orchidis* and *G. crassinervis*, as well as a distinct subclade comprising all members of *Nigritella*. Thus, *Nigritella* appears as a monophyletic subgroup, but it is fully embedded within *Gymnadenia*. Emphasizing the monophyly criterion when circumscribing genera (Bateman 2009), all members of *Nigritella* should accordingly be recognized as *Gymnadenia*, when based on evidence from ITS sequences alone.

However, phylogenies presented in Inda et al. (2012), based on plastid rpl16 and mitochondrial coxI, suggest a sister group relationship between *Nigritella* and *Gymnadenia* s.str.

Similarly, Bateman (2001) demonstrated a complete separation between the genera in the plastid trnL intron, and this difference was also confirmed in a more comprehensive study (Hedrén et al. 2018), which included several additional plastid marker loci. Moreover, the two genera were shown to be divergent in phenetic analyses of AFLPs (Ståhlberg 1999), and nuclear allozymes (Hedrén et al. 2000).

Still, the molecular studies performed so far have been based on restricted numbers of molecular markers. Given the apparent conflicts between phylogenies based on different data sets, phylogenies based on single genes will often be discordant to each other and to the species phylogeny. We therefore conducted the present study in which we obtained thousands of genome-wide SNPs derived from restriction site associated DNA sequencing (RADseq; Hohenlohe et al. 2012). Data sets obtained by RADseq should provide an accurate representation of the phylogenetic relationships across the genome and, accordingly, they should be less biased than other data sets that only describe specific portions of a genome. We use our data to analyze the phylogenetic position of *Nigritella* relative to *Gymnadenia*, and we analyze species relationships within each genus. To enable comparison with previous analyses based on ITS data, we also compiled information on major ITS sequence type in the studied material and added these data onto the trees derived from RADseq.

MATERIALS AND METHODS

Plant Material and DNA extraction. Forty-two samples were included in our analysis, seventeen samples of *Nigritella*, eighteen samples of *Gymnadenia*, two samples of *Gymnigritella runei*, and five samples of *Dactylorhiza viridis* (syn. *Coeloglossum viride*) which were used as outgroup. The selected samples of *Nigritella* covered all taxa treated as species in Hedrén et al. 2018 except for the rare *N. carpatica*. The samples also included four recently described segregates of *Nigritella miniata* (Lorenz and Perazza 2012; Table 1), although these were poorly separated according to nuclear microsatellites (Hedrén et al. 2018). Since *Nigritella* is restricted to Europe, the sampling of *Gymnadenia* was also focused on the European species (Table 2). The species used as outgroup, *Dactylorhiza viridis*, represents the earliest branch to split from the rest of the genus *Dactylorhiza* (Brandrud et al. submitted - Chapter 1), which is the sister group to *Gymnadenia/Nigritella*. Like the latter, it also produces nectar. Most of the samples have been used previously in population-based molecular studies (Hedrén and Pedersen 2016, Hedrén et al. 2018) and agree with the taxa they have been affiliated with here.

Sampling was performed carefully to allow for continued survival of the plants, and only portions of above-ground parts were collected. Total DNA was isolated from silica dried flowers with bracts (ca. three to four flowers from *Gymnadenia* accessions, part of a flowering head from *Nigritella* accessions; Chase and Hills 1991) by following a cetyl trimethylammonium bromide (CTAB) procedure (Doyle 1990) or using a DNeasy Plant Mini Kit (Qiagen,Venlo, Netherlands). DNA extracts are stored in the DNA bank of MH at Department of Biology, Lund University or in the DNA bank of RMB at Jodrell Laboratory, RBG Kew, as indicated in Table 1. Vouchers in the form of dried flowers are deposited in the Lund University botanical museum (LD) or Royal Botanic Gardens Kew (K), Table 1. Maps of the sampling locations (Fig. 1) were generated using QGIS v. 2.4.071 (QGIS 2015), with a map layer that was extracted from GADM version 1.0 (available from www.gadm.org).

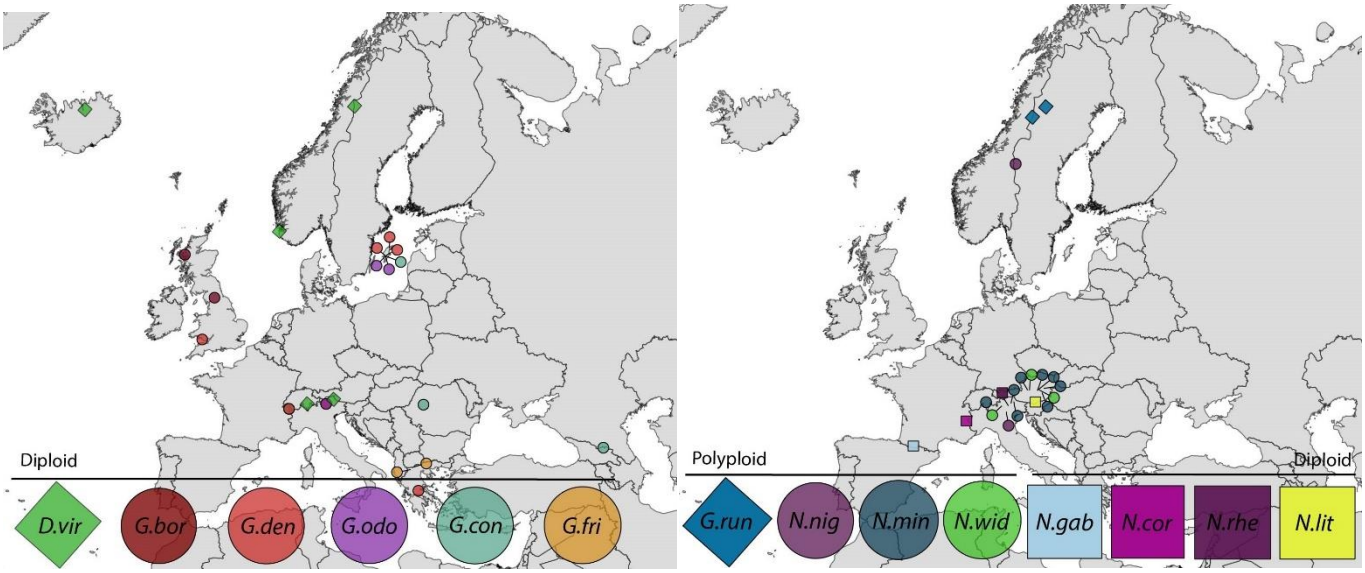


Figure 1. Sampling maps of five diploid *Dactylorhiza*, 18 diploid *Gymnadenia*, 15 polyploid *Nigritella* and *Gymnigritella* as well as four diploid *Nigritella* individuals included in the present study. The map layer was extracted from GADM version 1.0 (available from www.gadm.org). The *N. nigra*, *N. widderi* and *N. miniata* group each comprise multiple taxa. *D.vir=D. viridis*, *G.bor=G. borealis*, *G.den=G. densiflora*, *G.odo=G. odoratissima*, *G.con=G. conopsea*, *G.fri=G. frivaldii*, *G.run=G. runei*, *N.nig=N. nigra*, *N.min=N. miniata*, *N.wid=N. widderi*, *N.gab=N. gabasiana*, *N.cor=N. corneliana*, *N.rhe=N. rhellicani*, *N.lit=N. lithopolitanica*

Library Preparation and Sequencing. Part of the material used in Brandrud et al. submitted - Chapter 1, were also used for the present study. DNA was purified with the Nucleospin gDNA clean-up kit (Macherey-Nagel, Düren, Germany) following the manufacturer's

protocol. RADseq libraries of 30-72 individuals per library, were prepared following the protocol detailed in Paun et al. (2016), with the following modifications. Depending on the library, for each sample 100–400 ng DNA was used. The DNA was sheared with a Bioruptor Pico using 0.65 ml tubes (Diagenode) and three cycles of 30 sec ON and 60 sec OFF. The inline and index barcodes used differed from each other by at least three sequence positions. The DNA amount of each sample was normalized within libraries. All RADseq libraries were sequenced as single-end 100 bp reads on an Illumina HiSeq platform at VBCF NGS Unit (www.vbcf.ac.at/ngs), Vienna, Austria.

Filtering SNPs from RADseq Data. A similar bioinformatics pipeline to that used in Brandrud et al. submitted - Chapter 1 was also applied in the present study, but the loci for the synthetic reference were built *de novo* only from diploid *Gymnadenia* and *Nigritella* (i.e., in the absence of a reference genome), followed by mapping both diploid and polyploid accessions to this reference. Finally, variants were called and filtered across all samples.

The raw reads were demultiplexed based on index reads using BamIndexDecoder v. 1.03 (included in Picard Illumina2Bam package, available from http://gq1.github.io/illumina2bam/) and on inline barcodes using PROCESS_RADTAGS from STACKS v. 1.44 (Catchen et al. 2011, 2013). Together with demultiplexing a quality filtering was performed that removed uncalled bases, discarded reads with low quality scores and rescued barcodes and cut sites with maximum one mismatch. RADseq loci were built *de novo* for the set of diploid individuals with DENOVO_MAP.PL. The final settings used were requiring at least three reads to create a stack (m), allowing maximum one mismatch when merging the loci within individuals (M) as well as among individuals when building the catalog (n). The loci present in at least 50% of individuals and containing between one and 15 SNPs were extracted with EXPORT_SQL.PL in STACKS and a consensus for each of these loci was retained to produce the reference for further analysis.

The reads of diploids and polyploids were mapped back to this synthetic reference by using BOWTIE2 v. 2.2.6 (Langmead and Salzberg 2012) with default settings. Variants were then called with REF_MAP.PL and POPULATIONS in STACKS using default settings to produce a vcf file. PGDSpider v. 2.0.8.2 (Lischer and Excoffier 2011) was used to convert the vcf file to the phylip file used in the maximum likelihood phylogenetic tree and VCFTOOLS v. 0.1.14 (Danecek, Auton et al. 2011) was used to filter the vcf file used for the ordinations.

RADseq Data Analyses. A maximum likelihood phylogeny of diploid accessions was obtained by running RAxML v. 8.2.9. (Stamatakis 2014). The analysis was performed on the phylip file of concatenated SNPs with the following settings: 1,000 rapid Bootstrap replicates;

searching for the best-scoring ML tree with a general time reversible model of nucleotide substitutions; disabled rate heterogeneity among sites model (i.e., the GTRCAT model). As the dataset contained only concatenated SNPs, the ascertainment bias correction of likelihoods (Lewis 2001) was applied, and 1,000 alternative runs on distinct starting trees was used. *Dactylorhiza viridis* was used as outgroup.

As an alternative approach to assess the phylogenetic relationships among the diploids, a TREEMIX analysis was run, including only accessions of *Gymnadenia* and *Nigritella*. For this analysis only one SNP per locus was used to approximate unlinked markers - the input file for TREEMIX was obtained by using VCF-tools, PLINK v. 1.07 (Purcell et al. 2007) and the python script (plink2treemix.py) available at the TREEMIX website (https://bitbucket.org/nygcresearch/treemix/downloads/). TREEMIX was run by successively allowing for migration events and comparing the increase of total variation explained to infer how many migrations to allow. TREEMIX graphs were visualised with RSTUDIO v. 1.0.44 (Rstudio Team 2015) using Rscripts provided with TREEMIX.

To visually illustrate the similarity between diploids and polyploids in the entire dataset (i.e., except *Dactylorhiza*), Principal Coordinates Analyses, PCoAs based on Euclidian distance, were conducted using the dartR package (Gruber and Georges 2018), based on a dataset including loci with maximum 10% missing data and with each allele found in at least two individuals (--maf). For a better visualization, separate PCoAs were performed for *Nigritella* diploids and polyploids. Finally, a coancestry analysis was run with FINERADSTRUCTURE v. 0.2. for the *Nigritella* polyploids only to assess their relationships. The program is using the SNPs on each locus to calculate a nearest neighbour haplotype coancestry between the individuals. The haplotype input file for FINERADSTRUCTURE was created by using the python script (Stacks2fineRAD.py) available from http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html.

To infer the most likely parental species of *Gymnigritella runei,* VCFTOOLS was used to calculate the unadjusted Ajk statistic (--relatedness option; Yang, Benyamin et al. 2010) between *Gymnigritella* and each potential parental diploid, as well as between *Gymnigritella* and each polyploid group of *Nigritella*, and the results were presented as vioplots in R following Brandrud et al. submitted - Chapter 1. Relatedness coefficients, such as Ajk, have been shown to be unbiased with respect to ploidy (Meirmans et al. 2018). Ajk can take values of one for an individual compared to itself, be around zero if individuals from the same population are compared, and take negative values up to -1 otherwise. Finally, for each polyploid lineage a relative measure of inbreeding *F* was calculated with VCFTOOLs with the option --

het. The *F* results were plotted as vioplots in RSTUDIO. Similar to the population index F_{IS}, the per-individual *F* estimate is bound from -1 (maximum outcrossing) to 1 (maximum inbreeding), but as it is derived from a vcf file containing only variable sites, it should be regarded as a relative measure of inbreeding. The difference between *F* distributions between *Gymnigritella*, on one hand, and each of the three *Nigritella* polyploids, on the other hand, was tested for statistical significance with Mann-Whitney-Wilcoxon tests in R.

Determination of major ITS sequence types. Two major ITS sequence variants (ribotypes) have been identified in *Gymnadenia* sensu lato (Pridgeon et al. 1997, Gustafsson and Lönn 2003). These variants differ consistently at five positions in ITS1 and at five positions in ITS2, but do not differ in length (Gustafsson and Lönn 2003, Stark et al. 2011). In the following, the two variants are denoted ITS-E (found in early-flowering *G. conopsea/G. odoratissima*), and ITS-L (found in late-flowering *G. conopsea/G.densiflora*), respectively (cf. Gustafsson and Lönn 2003). For samples that were not already sequenced for ITS, we took advantage of a simplified protocol to rapidly screen the accessions for major ITS type by means of tetra-primer ARMS-PCR (Amplification-Refractory Mutation System), which is a protocol designed for SNP identification (Chiapparino et al. 2004). In ARMS-PCR fragments differing at single positions are selectively amplified by alternative primer pairs giving rise to fragments of different lengths, which enables simple screening by gel electrophoresis. Details of our protocol have been given in a previous paper (Hedrén et al. 2018).

RESULTS

The demultiplexed data contained on average 1.2 million (+/- 1.1 SD) quality reads per individual. The data has been deposited in the NCBI Short Reads Archive (BioProject ID XXX, SRA Study XXX). From a *de novo* diploid catalog building with STACKS we retained 3,793 polymorphic loci of 94 bp for the synthetic reference according to the criteria mentioned above. The average mapping percent on this reference was ~30%.

For the phylogenetic analysis on a dataset of 18,007 RADseq-derived SNPs including all samples of *Gymnadenia*, as well as four diploid members of *Nigritella* (Fig. 2), *Gymnadenia* and *Nigritella* came out as fully supported monophyletic sister groups. Within *Gymnadenia*, the following clades were distinguished with 100% support: *G. conopsea* s. str. (early flowering), *G. odoratissima*, *G. borealis*, *G. densiflora/G. conopsea* (late flowering) and *G. frivaldii*. The latter came out as sister to the rest of the *Gymnadenia* species. Except for the *G.*

densiflora/*G. conopsea* late flowering complex, all RADseq clades were well in correspondence with the classification based on morphology.

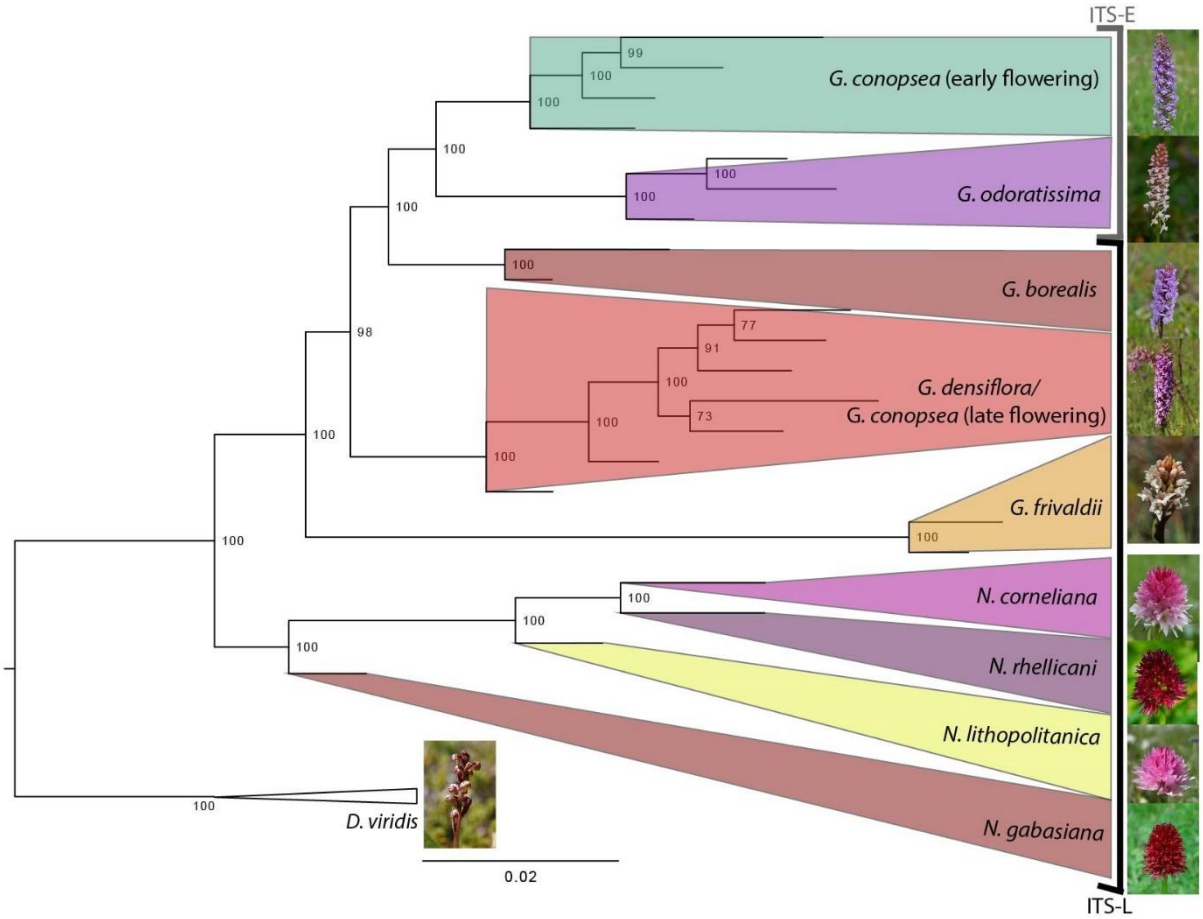


Figure 2. The best-scoring maximum likelihood phylogenetic tree of a concatenated dataset of 18,007 SNPs from 22 accessions of diploid *Gymnadenia* and *Nigritella*. Five samples of the diploid *Dactylorhiza viridis* were included as outgroup. Photos by Sven Birkeland, Richard Lorenz and Heinz-Werner Zaiss.

We also mapped the major ITS types recorded in the samples onto this tree. The clade composed of *G. odoratissima* and *G. conopsea* s.str. all had ITS-E, while the remaining samples of *Gymnadenia*, i.e. *G. frivaldii*, *G. densiflora*, late flowering *G. conopsea*, and all *Nigritella* had ITS-L. The general topology of the TREEMIX analysis (Fig. S1) was consistent with the phylogenetic relationships found in the RAxML analysis. One migration event between *Gymnadenia* and *Nigritella*, i.e. between the common ancestor of *G. conopsea* s.str. and *G. odoratissima* and *N. gabasiana*, improved the variation explained from 95.8% (without any migration event) to 98.1% (one migration allowed). An additional migration event (summing up two altogether) improved the variation contained by only 0.1%, which was considered insignificant and the scenario with one migration event was finally chosen.

In the resulting plot given by the first two axes of the PCoA of all material (Fig. 3a), *Gymnadenia* and *Nigritella* are separated along the first axis that explains as much as 32.2% of the genetic variation. *Gymnigritella runei* and *N. gabasiana* take a somewhat intermediate position, although still closer to the reminder of *Nigritella*. The second axis separates *G. densiflora* from the other *Gymnadenia*, and the third axes separates *G. frivaldii* from the rest (result not shown). In the PCoA of *Nigritella* only, the polyploids are separated along axis 2, with *N. miniata* in one end and *N. widderi* and *N. nigra* in the other. The diploid *N. rhellicani* and *N. corneliana* appear close to the *N. nigra* group, whereas *N. lihopolitanica* and, in particular, *N. gabasiana* differ from the rest. The first three axes of this PCoA describe 47.7 % of the total variation.

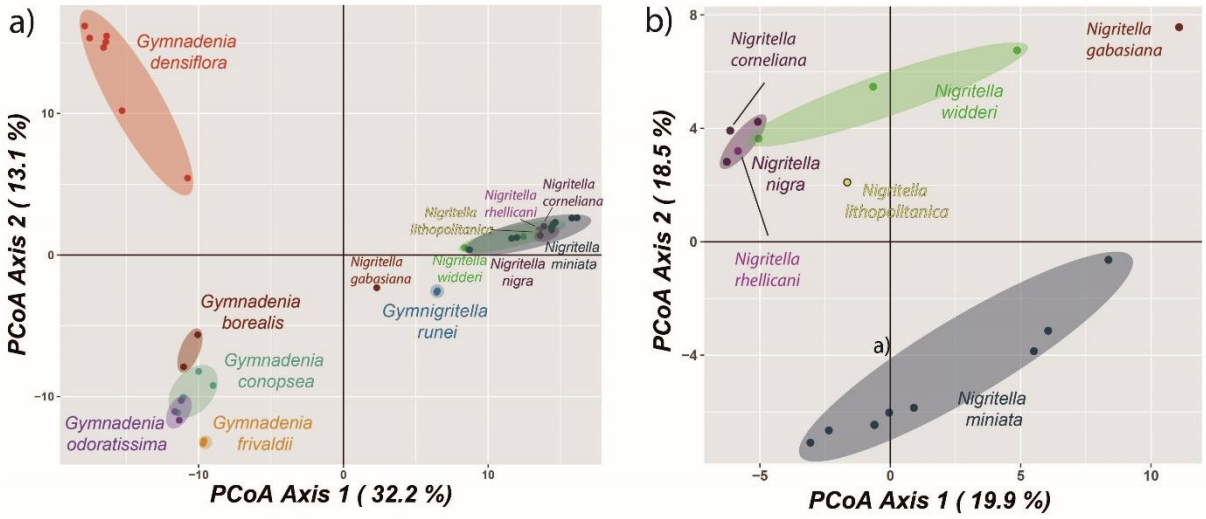


Figure 3. PCoA analysis performed with the dartR package (Gruber and Georges 2018). (a) PCoA on 6,487 SNPs from 18 diploid *Gymnadenia*, four diploid *Nigritella*, 13 polyploid *Nigritella*, and two *Gymnigritella* individuals. (b) PCoA on 2,038 SNPs from with 13 polyploid *Nigritella* and four diploid *Nigritella* individuals. The *N. nigra*, *N. widderi* and *N. miniata* group each comprise multiple taxa.

The FINERADSTRUCTURE result shows three groups within the *Nigritella* polyploids; the *N. nigra* group, the *N. widderi* group and the *N. miniata* group. The amount of coancestry shared within each group differed, with *N. nigra* accessions sharing comparatively many haplotypes with each other, followed by *N. widderi* and *N. miniata*. Futher, *N. nigra* and *N. widderi* seem to share more with each other that either of them does with *N. miniata*. The relatedness plots of the polyploids to the putative parents was analysed. The relatedness violplots show the highest relatedness of *Gymnigritella* to *G. conopsea* among the *Gymnadenia* diploids (Fig. S2a) and to polyploid *N. nigra* s.l. among the *Nigritella* taxa (Fig. S2b), confirming *Gymnigritella* as an allopolyploid. Investigating the degree of heterozygosity in

the polyploid lineages, the relative *F* values of *Gymnigritella* took positive but low values, not different from *N. widderi* and *N. miniata* suggesting an allopolyploid origin for these lineages (Fig. S2c). *Nigritella nigra* showed higher *F* values than *Gymnigritella,* and even though the distributions were not significantly different, this most likely indicate an origin starting from more closely related parentals.

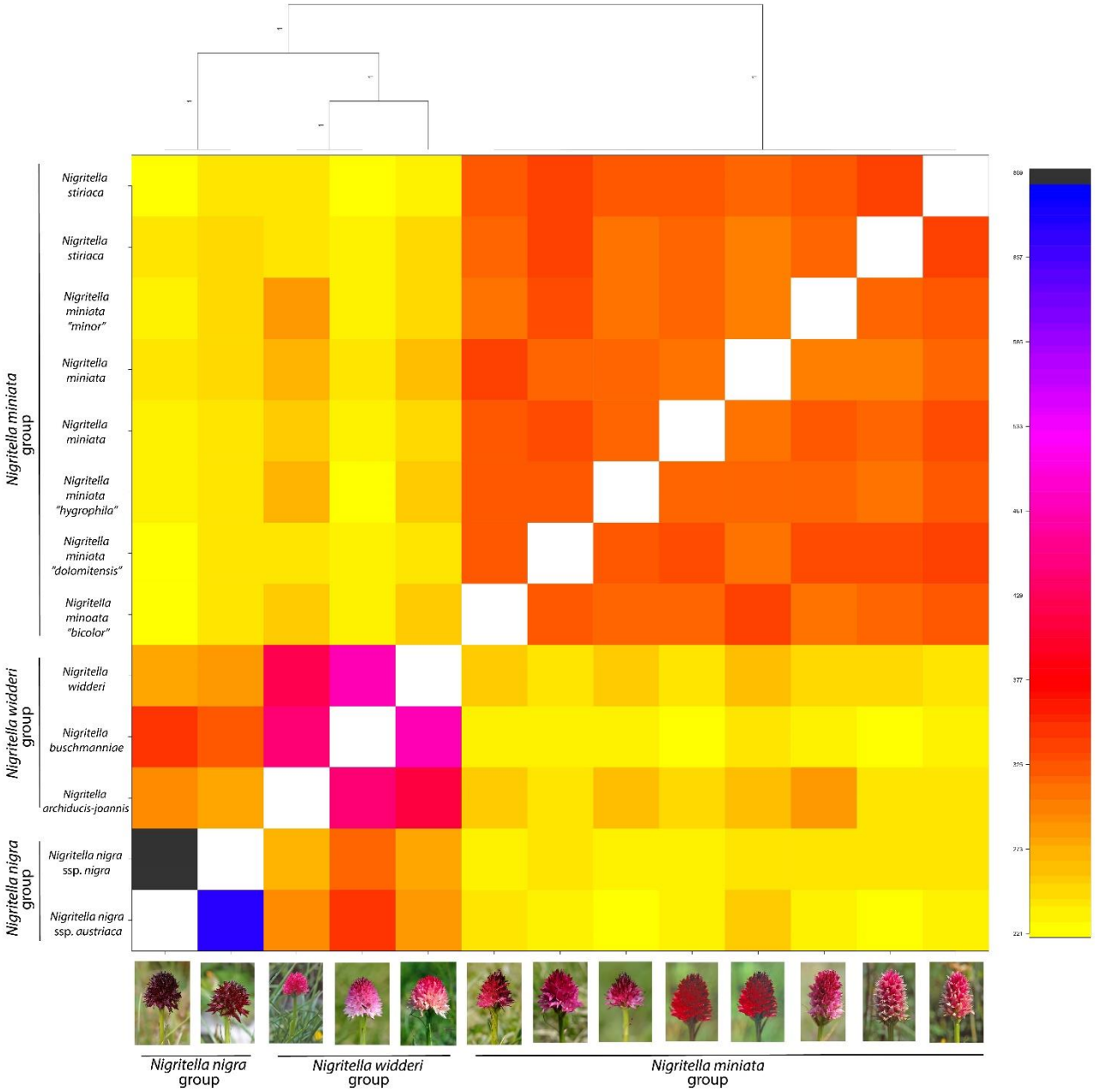


Figure 4. Heatmap obtained with FINESTRUCTURE on 3,447 RADseq loci for 13 polyploid *Nigritella* individuals. Photos by Sven Birkedal and Richard Lorenz.

DISCUSSION

The relationship between *Gymnadenia* and *Nigritella*. All ITS-based analyses published so far (Pridgeon et al. 1997, Bateman et al. 2003, Bateman et al. 2006, Stark et al. 2011), have shown *Nigritella* to be embedded within *Gymnadenia*. Additionally, phylogenetic analyses combining ITS with other sequence data have also shown such a pattern (Inda et al. 2012, Sun et al. 2015). Specifically, it has been found that *G. conopsea* s.str. (early-flowering) together with *G. odoratissima* form one monophyletic clade characterized by ITS-E, and all other members of *Gymnadenia* (e.g. late-flowering *G. conopsea*, *G. densiflora*, *G. borealis*, *G. orchidis*, *G. crassinervis*, *G. frivaldii*) together with members of *Nigritella* form another monophyletic clade characterized by ITS-L (Gustafsson and Lönn 2003). Based on ITS sequences, it could thus be argued that *Nigritella* should be part of *Gymnadenia* (Bateman et al. 1997).

However, all our analyses based on tens of thousands of genome-wide SNPs derived from RADseq indicate that the genus *Nigritella* is separated from the genus *Gymnadenia*. The phylogenetic analyses using RAxML and TREEMIX arranged all *Gymnadenia* and all *Nigritella*, respectively, in two separate and well-supported monophyletic clades. In the PCoA plots (Fig. 3b), members of *Nigritella* formed a distinct cluster clearly separated from members of *Gymnadenia*.

The split between *Nigritella* and *Gymnadenia* observed here also agree with patterns obtained from other genomic regions than ITS, including AFLPs (Ståhlberg 1999), allozymes (Hedrén et al. 2000), nuclear microsatellites (Hedrén et al. 2018), plastid tRNA-Leu intron (Bateman 2001) and plastid VNTRs (Hedrén et al. 2018). Also phylogenetic analyses of plastid rpl16 intron and mitochondrial cox1 sequence data agree with this basal split (Inda et al. 2012), although the number of samples in these analyses were restricted. When all evidence is taken together, we conclude that *Nigritella* and *Gymnadenia* can be validly treated as separate genera, also when a phylogenetic criterion is applied (Bateman 2009).

Because of its high variability, the ITS region has been used extensively to analyse species delimitations and species relationships in many groups of flowering plants (Baldwin et al. 1995, Álvarez and Wendel 2003). However, several types of genetic mechanisms are known to contribute to molecular evolution of the ITS regions. These mechanisms result in sequence divergence, incomplete lineage sorting, as well as homogenization, and for these reasons analysis of ITS sequence data may not necessarily reflect organismal phylogeny (Álvarez and Wendel 2003). First, the nuclear DNA regions harboring ITS are subject to a biased concerted evolution (Baldwin et al. 1995), by which divergent copies of the rDNA repeats are homogenized against each other, often in a repeated way. Such homogenization

may go in the direction towards one or the other of the divergent copies, or may result in repeats of intermediate appearance (Wendel et al. 1995). Similarly, whenever divergent rDNA repeats occur together within a genome, chimeric repeat types combining portions of the parental sequences may be produced in addition to the parental types (Devos et al. 2006), and may eventually become fixed within the genome. Yet a possibility is that multiple repeat types are retained, and maintained within a single genome without much of integration. In the phylogenies obtained from analysis of RADseq data, it appears that ITS differentiation does not fully reflect the phylogeny of the *Nigritella*/*Gymnadenia* clade within the Orchidinae, and that ITS artifactually places *Nigritella* as a subclade within *Gymnadenia*. Specifically, it appears that the ITS-E sequence type evolved in the common ancestor to *G. conopsea* s.s and *G. odoratissima* from a plant with ITS-L type of ITS. Alternatively, the two major ITS types may both have been present in the common ancestor to the *Nigritella/Gymnadenia* clade. The present-day distribution of major ITS types may then be the result of fixation of alternative types in branches that have diverged later on, one of which corresponds to present day *Nigritella*. Gene flow (Fig. S1), perhaps via allopolyploids similar to the extant *Gymnigritella*, followed by concerted evolution may have also played a role in shaping a similar ITS type between the two genera.

Relationships within *Gymnadenia*. *Gymnadenia frivaldii* was the sister to the rest of *Gymnadenia* in our SNPs-based analyses. This position differs from that given by ITS-based phylogenies, according to which *G. frivaldii* is embedded in the clade of *Gymnadenia/Nigritella* characterised by ITS-L (Bateman et al. 2006, Stark et al. 2011). However, *Gymnadenia frivaldii* differs from other species of *Gymnadenia* in, e.g., a relatively short spur and in structure of the column (Bateman et al. 2006), and its basal position in the genus could be seen in light of its somewhat deviating morphology.

The finding that *G. conopsea* s.str. and *G. odoratissima* are sister species is in agreement with ITS-based phylogenies, as they are both characterized by the ITS-E type. *Gymnadenia borealis* was found to be the successive sister species to this group. This placement may seem surprising as it is characterized by ITS-L (Pridgeon et al. 1997, Bateman et al. 2003, 2006, Stark et al. 2011). However, it resembles *G. conopsea* s.str. as well as *G. densiflora* in general morphology (Rich 2012).

Finally, as in Gustafsson and Lönn (2003), we found that late-flowering *G. conopsea* grouped together with *G. densiflora*. Both are characterized by having ITS-L. These results suggest that late-flowering plants similar to *G. conopsea* in size and overall morphology, should be included in *G. densiflora*, which requires an emended circumscription of the latter.

Relationships within *Nigritella*. The diploid members of *Nigritella* constitute a monophyletic sister group to *Gymnadenia* in the phylogenetic tree as well as in the TREEMIX analysis. The placement of *N. gabasiana* as the basalmost species in *Nigritella* is in agreement with plastid VNTR data (Hedrén et al. 2018), according to which it is clearly separated from other diploid members of *Nigritella*. Furthermore, all members of *Nigritella*, including polyploid representatives could be separated as relatively distinct clusters separated from the *Gymnadenia* samples in the PCoA ordination (Fig. 3a). These observations support the hypothesis that polyploid members of *Nigritella* have originated from diploid members of the same genus without contribution from *Gymnadenia* (Hedrén et al. 2000).

Except for *N. nigra*, which seems to be related to the extant *N. corneliana* and *N. rhellicani*, we were not able to match polyploid *Nigritella* to any specific diploid member of the genus included in the analysis. The polyploids have previously been hypothesized to at least partly have originated from now extinct diploid ancestors (Hedrén et al. 2000, 2018). Because they are highly heterozygous at nuclear codominant loci including allozymes and microsatellites, polyploid members of *Nigritella* are believed to be of hybrid origins, which is confirmed here by results of *F* coefficients that take similar values as *Gymnigritella*. On basis of AFLPs (Ståhlberg 1999), allozymes (Hedrén et al. 2000) and microsatellites (Hedrén et al. 2018), the polyploids have been found to aggregate in three groups, the *nigra* group (*N. nigra* ssp. *nigra*, *N. nigra* ssp. *austriaca* and, less strongly attached, *Gymnigritella runei*, see below), the *widderi* group (*N. widderi*, *N. buschmanniae* and *N. archiducis-joannis*) and the *miniata* group (*N. miniata* and *N. stiriaca*). These groups were also identified in our FINERADSTRUCTURE analyses based on RADseq (Fig. 4), but the numbers of samples were insufficient to examine the exact subdivision of each group. Given that different taxa in *Nigritella* are much less differentiated from each other than taxa in *Gymnadenia*, the number of variable SNPs within *Nigritella* should be relatively few. We would therefore need an extended data set including additional samples per taxon, additional populations of the diploids and perhaps also deeper sequencing with higher coverage to obtain detailed knowledge on fine-scale relationships in *Nigritella*.

The RADseq was informative on the position and origin of *Gymnigritella runei*, which is a polyploid with a restricted distribution in the Scandinavian mountains (Teppner and Klein 1989, Rune 1993). According to our data, this polyploid has originated from a recent hybridization between *G. conopsea* s.str. and *N. nigra* ssp. *nigra*, which is evident from nuclear codominant markers as well as ITS at which it combines ITS-L from *Nigritella* with ITS-E from *G. conopsea* s.str. (Hedrén et al. 2018). We included two samples of *Gymnigritella runei*

in our analyses. In the PCoA, they take a position intermediate between the *Nigritella* cluster and the portion of the *Gymnadenia* cluster in which samples with ITS-E are located (Fig, 4a). The relatedness plot identifies *G. conopsea* s. str. and *N. nigra* as the most likely parental species (Fig. S2), which is in correspondence with the conclusion of former studies (see e.g. Hedrén et al. 2018). The localities for *Gymnigritella runei* are situated relatively close to the main distribution area of *N. nigra* ssp. *nigra* in mid Scandinavia, suggesting that the former has a post-glacial origin (Rune 1993).

ACKNOWLEDGEMENTS

We thank Richard Bateman, Sven Birkedal, Mark Chase, Wolfram Foelsche, Olivier Gerbaud, Cesario Giotta, Norbert Griebl, Sven Hansson, Erich Klein, Marc Lewin, Henrik Ærenlund Pedersen, Giorgio Perazza, David Ståhlberg, Eva Waldemarson and Heinz-Werner Zaiss for providing samples, photos, and/or valuable comments on this study.

Financial support was given by the Royal Physiographic Society Lund, Nilsson-Ehle foundation to MH and Austrian Science Fund (FWF, Project Y661-B16) to OP.

REFERENCES

Álvarez I., Wendel J. F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29: 417-434.

Baldwin B. G., Sanderson M. J., Porter M. J., Wojciechowski M. F., Campbell C. S., Donoghue M. J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of Missouri Botanical Garden 82: 247-277.

Bateman R.M. 2001. Evolution and classification of European orchids: insights from molecular and morphological characters. J. Eur. Orch. 33: 33-119.

Bateman R.M. 2009. Evolutionary classification of European orchids: the crucial importance of maximizing explicit evidence and minimizing authoritarian speculation. J. Eur. Orch. 41: 243-318.

Bateman R. M., Murphy A. R., Hollingsworth P. M., Hart M. L., Denholm I., Rudall P. J. 2018. Molecular and morphological phylogenetics of the digitate-tubered clade within subtribe Orchidinae ss (Orchidaceae: Orchideae). Kew Bulletin, *73*:54.

Bateman R. M., Pridgeon A.M., Chase M.W. 1997. Phylogenetics of subtribe Orchidineae, Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 2. Infrageneric relationships and classification to achieve monophyly of *Orchis* sensu stricto. Lindleyana 12: 113–141.

Bateman R.M., Hollingsworth P.M., Preston J., Yi-Bo L., Pridgeon A.M., Chase M.W. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). – Bot. J. Linn. Soc. 142:1–40.

Bateman R. M., Rudall P. J., James K. E. 2006. Phylogenetic contexts, genetic affinities and evolutionary origin of the enigmatic Balkan orchid *Gymnadenia frivaldii* Hampe ex Griseb. Taxon 55:107-118.

Baumann H., Künkele S., Lorenz R. 2006. Orchideen Europas. – Ulmer Naturführer.

Brandrud, M. K., Baar, J., Lorenzo, M. T., Bateman, R. M., Chase, M. W., Paun, O., Hedrén, M. submitted – Chapter I. Phylogenomic relationships of diploids and the origin of allotetraploids in *Dactylorhiza* (Orchidaceae): RADseq data track reticulate evolution.

Catchen J. M., Amores A., Hohenlohe P., Cresko W., Postlethwait J. H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics 1:171-182.

Catchen J. M., Hohenlohe P. A., Bassham S., Amores A., Cresko W. A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124-3140.

Chase M. W., Hills H. H. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon: 215-220.

Chiapparino E., Lee D., Donini P. 2004. Genotyping single nucleotide polymorphisms in barley by tetraprimer ARMS-PCR. – Genome 47:414–420.

Claessens J., Kleynen J. 2011. The flower of the European orchid. Form and function. 1-440. Privately published.

Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T., Sherry S. T., McVean G., Durbin R. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Devos N., Raspé O., Oh S.-H., Tyteca D., Jacquemart A.-L. 2006. The evolution of *Dactylorhiza* (Orchidaceae) allotetraploid complex: Insights from nrDNA sequences and cpDNA PCR-RFLP data. Molecular Phylogenetics and Evolution, 38:767–778.

Doyle J. J. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Gerbaud O., Schmid W. 1999: Die Hybriden der Gattungen *Nigritella* und/oder *Pseudorchis* [Les hybrids des genres *Nigritella* et/ou *Pseudorchis*]. – Cahiers de la Société Française d´Orchidophilie 5: 1–128

Gjærevoll O. 1992. Plantegeografi. – Tapir.

Gruber B., Unmack P. J., Berry O. F., Georges A. 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources, 18:691-699.

Gustafsson S. 2000. Patterns of genetic variation in *Gymnadenia conopsea*. Molecular Ecology 9: 1863-1872.

Gustafsson S., Sjögren-Gulve P. 2002. Genetic diversity in the rare orchid, *Gymnadenia odoratissima* and a comparison with the more common congener, *G. conopsea*. Conservation Genetics 3:225-234.

Gustafsson S, Lönn M. 2003. Genetic differentiation and habitat preference of flowering-time variants within *Gymnadenia conopsea*. Heredity 91:284–292.

Hedrén M., Pedersen H. Æ. 2016. Plastid DNA microsatellite data do not support recognition of subspecies in *Coeloglossum viride* (L.) Hartm.(Orchidaceae) in northern Europe. Symbolae Botanicae Upsalienses, Uppsala University**:** 81-91.

Hedrén M., Klein E., Teppner H. 2000. Polyploid evolution in the European orchid genus *Nigritella*: evidence from allozyme data. – Phyton 40:239–275.

Hedrén M., Lorenz R., Teppner H., Dolinar B., Giotta C., Griebl N., Hansson S., Heidtke U., Klein E., Perazza G., Ståhlberg D., Surina B. 2018. Evolution and systematics of polyploid *Nigritella* (Orchidaceae). Nordic Journal of Botany, in press.

Hohenlohe P. A., Catchen J., Cresko W. A. 2012. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Methods in Molecular Biology 888:235-260.

Inda, L.A., Pimentel M., Chase M. W. 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. – Ann. Bot. 110:71–90.

Jorgensen R. A., Cluster P. D. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Annals of Missouri Botanical Garden 75:1238-1247.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913-925.

Langmead B., Salzberg S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359.

Lischer H. E., Excoffier L. 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.Lorenz R., Perazza G. 2012. Beiträge zur Kenntnis der Blütenmorphologie der Artengruppe *Nigritella miniata* s.l. (Orchidaceae) in den Ostalpen. – Gredleriana 12: 67–146.

Løjtnant B. 1977: – I: Mossberg, B., Nilsson, S., *Nordens orkidéer (Danish edition.)* Gyldendals, Köpenhamn. 76

Meirmans P.G., Liu S., van Tienderen P.H. 2018. The analysis of polyploid genetic data. J. Heredity 109:283-296.

Moore D.M. 1980. *Gymnadenia* R.Br., *Nigritella* L.C.M. Richard. – In: Tutin, T.G., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A. Flora Europea, 5. – Cambridge University Press, 332–333.Paun O., Turner B., Trucchi E.,

Munzinger J., Chase M. W., Samuel R. 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. Systematic Biology 65:212-227.

Pridgeon A. M., Cribb P. J., Chase M.W., Rasmussen F. N. 2001. Genera orchidacearum 2, 1. New York: Oxford University Press.

Pridgeon A. M., Bateman R. M., Cox A.V., Hapeman J. R., Chase M.W. 1997. Phylogenetics of subtribe Orchidineae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 1. Intergeneric relationships and polyphyly of *Orchi*s sensu lato. Lindleyana 12: 89-109.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D., Maller J., Sklar P., de Bakker P. I. W., Daly M. J., Sham P. C. 2007. PLINK: a tool set for whole genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.

QGIS. 2015. QGIS Geographic Information System. Open Source Geospatial Foundation Project http://www.qgis.org/.

RStudio Team. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com.

Rich T.C.G. 2012 Plant Crib 3, *Gymnadenia conopsea*, 2nd ed. Botanical Society of the British Isles.

Rune O. 1993. Distribution and ecology of *Gymnigritella runei*: a new orchid in the Scandinavian mountain flora. – Opera Bot. 121:29–34.

Scacchi R., de Angelis G. 1989. Isoenzyme polymorphisms in *Gymnadenia conopsea* and its inferences for systematics within this species. Biochemical Systematics and Ecology 17: 25-33.

Soliva M., Widmer A. 1999. Genetic and floral divergence among sympatric populations of *Gymnadenia conopsea* s.l. Orchideaceae) with different flowering phenology. Int. J. Plant Sci. 160: 897–905.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysisof large phylogenies. Bioinformatics 30:1312-1313.

Stark C., Michalski S.G., Babik W., Winterfeld G., Durka W. 2011. Strong genetic differentiation between *Gymnadenia conopsea* and *G. densiflora* despite morphological similarity. Plant Syst Evol 293:213–226.

Ståhlberg D. 1999. Polyploid evolution in the European orchid genus *Nigritella*: evidence fromDNA fingerprinting. – Unpublished Master´s thesis. Institutionen för Systematisk Botanik, Lunds Universitet. <https://lup.lub.lu.se/record/2026859>. Accessed 2016-11-14.

Sun M., Schlüter P.M., Gross K., Schiestl F.P. 2015. Floral isolation is the major reproductive barrier between a pair of rewarding orchid sister species. Journal of Evolutionary Biology 28:117-129.

Sundermann H. 1980. Europäische und Mediterrane Orchideen, ed. 3. Schmersow, Hildesheim.

Teppner H. 1996. Adventitious embryony in *Nigritella* (Orchidaceae). – Folia Geobot. Phytotax. 31:323–331.

Teppner H. 2004. A review of new results in *Nigritella* (Orchidaceae). – Sprawozdania z posiedzeń komisji naukowych [reports of the scientific committee], Polska Akademia Nauk, 46.2:111–116.

Teppner H. and Klein, E. 1989. *Gymnigritella runei* spec. nova (Orchidaceae-Orchideae) aus Schweden. – Phyton 29:161–173.

Teppner H., Klein, E. 1990. *Nigritella rhellicani* spec. nova und *N. nigra* (L.) Rchb. f. s.str. (Orchidaceae-Orchideae). – Phyton 31:5–26.

Trávníček P., Jersáková J., Kubátová B., Krejčíková J., Bateman R.M., Lučanová M. Krajníková E., Těšitelová T., Štípková Z., Amardeilh J.-P-, Brzosko E., Jermakowicz E., Cabanne O., Durka W., Efimov P., Hedrén M., Hermosilla C.E., Kreutz K., Kull T., Tali K., Marchand O., Rey M., Schiestl F.P., Čurn V., Suda J. 2012: Minority cytotypes in European populations of the *Gymnadenia conopsea* complex (Orchidaceae) greatly increase intraspecific and intrapopulation diversity. Ann. Bot. 110:977–986.

Vöth W. 2000. *Gymnadenia, Nigritella* und ihre Bestäuber. – J. Eur. Orch. 32:547–573.

Wendel J. F., Schnabel A., Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences 92: 280-284.

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., Nyholt D. R., Madden P. A., Heath A. C., Martin N. G., Montgomery G. W., Goddard M. E., Visscher P. M. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42:565.

SUPPLEMENTARY MATERIAL

Table 1. Origins of samples analysed by RADseq in the present study. DNA bank numbers and voucher information are also provided.

Taxon	Segregate	Country	Locality	Latitude	Longitude	Major ITS type	Comments	DNA bank accession number	Voucher location
Dactylorhiza viridis		Sweden	Lycksele Lappmark, Rödingsnäset, NW Tängvattnet	65°52´N	14°43´E		syn. *Coelo-glossum viride*	**6785**	LD (MH)
Dactylorhiza viridis		Norway	Rogaland, Orre, Haugen	58°44´N	05°31´E		syn. *Coelo-glossum viride*	**7252**	LD (MH)
Dactylorhiza viridis		Switzerland	Ticino, Blenio, Lukmanier	46°36´N	08°48´E		syn. *Coelo-glossum viride*	**7254**	LD (MH)
Dactylorhiza viridis		Italy	Südtirol, Lungiarü	46°38´N	11°49´E		syn. *Coelo-glossum viride*	**7255**	LD (MH)
Dactylorhiza viridis		Iceland	Norur-Ísland, Ljosavatn	65°42´N	17°42´W		syn. *Coelo-glossum viride*	**8424**	LD (MH)
Gymnadenia borealis		Scotland	Wester Ross, Highlands, Depression S Applecross	57°26´N	05°49´W			**1843**	K (RMB)
Gymnadenia borealis		England	Cumbria, Brough, Stainmore, Rampson's Farm	54°31´N	02°13´W			**3328**	K (RMB)
Gymnadenia conopsea		Romania	Valea Morii	46°42´N	23°36´E	**E**		**7215**	LD (MH)
Gymnadenia conopsea		Greece	Fthiótida, Itis Oros, Pavliani	38°44´N	22°20´E	**L**		**10452**	LD (MH)
Gymnadenia conopsea		France	Savoie, Col de l'Iseran	45°25´N	07°02´E	**E**		**10459**	LD (MH)
Gymnadenia conopsea		Italy	Südtitol, Val Gardena	46°32´N	11°47´E	**L**		**10465**	LD (MH)
Gymnadenia conopsea		Georgia	Kazbegi, Kazbegi	42°39´N	44°39´E	**E**		**10496**	LD (MH)

Gymnadenia conopsea		Sweden	Gotland, Gerum, Botes käll-myr	57°21´N	18°19´E	**L**	late-flowering plant	**15273**	LD (MH)
Gymnadenia conopsea		Sweden	Gotland, Gerum, Botes käll-myr	57°21´N	18°19´E	**E**	early-flowering plant	**15316**	LD (MH)
Gymnadenia densiflora	"*friesica*"	Wales	Glamorgan, Kenfig Dunes NNR	51°31´N	03°44´W	**L**		**2370**	K (RMB)
Gymnadenia densiflora		France	Savoie, Col de l'Iseran	45°25´N	07°02´E	**L**		**10460**	LD (MH)
Gymnadenia densiflora		Sweden	Gotland, Rute, Kauparve	57°49´N	18°54´E	**L**		**10484**	LD (MH)
Gymnadenia densiflora		Sweden	Gotland, Gerum, Botes käll-myr	57°21´N	18°19´E	**L**	late-flowering plant	**15295**	LD (MH)
Gymnadenia frivaldii		Greece	Flórina, Vitsi Mts, 17 km from Drossopigi	40°42´N	21°27´E	**L**		**2769**	LD (MH)
Gymnadenia frivaldii		Greece	Flórina, Kajmakcalan, Vori ski center	40°55´N	21°48´E	**L**		**2774**	LD (MH)
Gymnadenia odoratissima		Sweden	Gotland, Follingbo, Klinte	57°35´N	18°21´E	**E**		**10493**	LD (MH)
Gymnadenia odoratissima		Italy	Südtitol, Val Gardena	46°32´N	11°47´E	**E**		**10472**	LD (MH)
Gymnadenia odoratissima		Sweden	Gotland, Gerum, Botes käll-myr	57°21´N	18°19´E	**E**		**15249**	LD (MH)
Gymnigritella runei		Sweden	Åsele Lappmark, Ransaren	65°15´N	15°03´E	**E+L**		**15537**	LD (MH)
Gymnigritella runei		Sweden	Lycksele Lappmark, Rödings-näset	65°52´N	14°43´E	**E+L**	locus classicus for *G. runei*	**15546**	LD (MH)
Nigritella archiducis-joannis		Austria	Steiermark, Totes Gebirge, Tauplitz/Lawinenstein, 1960 m asl	47°36´N	13°58´E			**9796**	LD (MH)
Nigritella buschmanniae		Italy	Trentino, Brenta, Le Crosette - Monte Turrion, 2360 m asl	46°12´N	10°55´E			**15052**	LD (MH)
Nigritella corneliana		France	Hautes-Alpes, Col du Lautaret	45°02´N	06°24´E		locus classicus for *N. corneliana*	**15398**	LD (MH)

Nigritella gabasiana		France	Pyrénées-Orientales, Fontrabiouse, Val de Galbe, 2000 m asl	42°35´N	02°04´E			**15002**	LD (MH)
Nigritella lithopolitanica		Austria	Kärnten, Karawanken, Petzen, Kniepsattel	46°30´N	14°46´E			**9346**	LD (MH)
Nigritella miniata		Switzerland	Ticino, Blenio, Lukmanier, 1870 m asl	46°33´N	08°28´E			**9867**	LD (MH)
Nigritella miniata	"*N. hygrophila*"	Italy	Trentino, Sellagruppe, Passo Pordoi, 2190 m asl	46°29´N	11°48´E		locus classicus for for *N. hygrophila*	**9936**	LD (MH)
Nigritella miniata	"*N. minor*"	Austria	Steiermark, Hochschwab-Gruppe, Trenchtling, 1810 m asl	47°32´N	15°02´E		locus classicus for *N. minor*	**15041**	LD (MH)
Nigritella miniata	"*N. bicolor*"	Austria	Steiermark, Hochschwab-Gruppe, Trenchtling	47°32´N	15°03´E		locus classicus for *N. bicolor*	**15368**	LD (MH)
Nigritella miniata		Austria	Steiermark/Niederösterreich, Rax	47°21´N	15°41´E			**15569**	LD (MH)
Nigritella miniata	"*N. dolomitensis*"	Italy	Südtirol, Fanes, Col Bechei, 2307 m asl	46°36´N	12°02´E		locus classicus for *N. dolomitensis*	**15704**	LD (MH)
Nigritella nigra ssp. *austriaca*		Italy	Trentino, Monti Lessini	45°43´N	11°00´E			**15070**	LD (MH)
Nigritella nigra ssp. *nigra*		Sweden	Härjedalen, Ljusnedal, Klinken	62°43´N	12°17´E	**L**		**9779**	LD (MH)
Nigritella rhellicani		Italy	Südtirol, Lungiarü, Medalgesalm, 2125 m asl	46°38´N	11°49´E			**9696**	LD (MH)
Nigritella stiriaca		Austria	Steiermark, Salzkammergut, Ausseer Zinken	47°35´N	13°46´E			**9341**	LD (MH)
Nigritella stiriaca		Austria	Steiermark, Teichalm	47°22´N	15°26´E			**15584**	LD (MH)
Nigritella widderi		Austria	Steiermark, Hochschwab-Gruppe, Aflenzer Bürgeralm, 1450 m asl	47°34´N	15°14´E			**9974**	LD (MH)

Table 2. Information on distribution, habitat, breeding system, ploidal level, and major ITS type for the taxa analysed by RADseq in the present study.

Taxon	Distribution	Ecology	Sexual system and ploidy	Major ITS type
Gymnadenia frivaldii	southeastern Europe,	Moist to wet mountain grasslands	sexual; 2x (?)	ITS-L; Bateman et al. 2003
Gymnadenia conopsea	European continent, Britain and Ireland, Scandinavia, Asia	Dry to wet grasslands and pastures, often on calcareous soils	sexual; 2x, 4x	ITS-E; Pridgeon et al. 1997
Gymnadenia odoratissima	mountain regions of Central and Eastern Europe, Baltic region incl. southern Scandinavia	Moist to wet calcareous grasslands	sexual; 2x	ITS-E; Bateman et al. 2003
Gymnadenia densiflora	European continent, Britain and Ireland, Scandinavia, Asia	Moist to wet calcareous grasslands	sexual 2x	ITS-L; Gustafsson and Lönn 2003
Gymnadenia borealis	Britain and Ireland	Acidic Molinia grassland	sexual; 2x (?)	ITS-L; Pridgeon et al. 1997
Nigritella gabasiana	Cantabrian Mountains and the Pyrenees	Mountain grasslands	sexual 2x	
Nigritella rhellicani	mountain regions of central Europe and the Balkans	Meadows and pastures, mountain grasslands, on calcareous and acidic soils	sexual 2x	
Nigritella corneliana	western Alps	Calcareous mountain grasslands	sexual 2x	
Nigritella lithopolitanica	southeastern Alps	Calcareous mountain grasslands	sexual 2x	
Nigritella nigra ssp. *nigra*	Scandinavian mountain range	Meadows and pastures, mountain grasslands, often on calcareous soils	asexual; 3x	ITS-L; Pridgeon et al. 1997
Nigritella nigra ssp. *austriaca*	Pyrenees, Jura, Alps	Calcareous mountain grasslands	asexual; 4x	ITS-L; Bateman et al. 2003
Nigritella widderi	northern limestone Alps, Grazer Bergland, Moiazza in the southern limestone Alps, central Apennines	Calcareous mountain grasslands	asexual; 4x	
Nigritella archiducis-joannis	northeastern and southeastern limestone Alps	Calcareous mountain grasslands	asexual; 4x	
Nigritella buschmanniae	Brenta massif in southern limestone Alps	Calcareous mountain grasslands	asexual; 5x	
Nigritella miniata	Eastern Alps and the Carpathians	Montane meadows and pastures, mountain grasslands, on calcareous soils	asexual; 4x	ITS-L; Bateman et al. 2003
Nigritella stiriaca	northeastern limestone Alps (Austria)	Calcareous mountain grasslands	asexual; 4x	
Gymnigritella runei	southern Lapland	Calcareous mountain grasslands	asexual; 4x	ITS-E/ITS-L; Hedrén et al. 2018

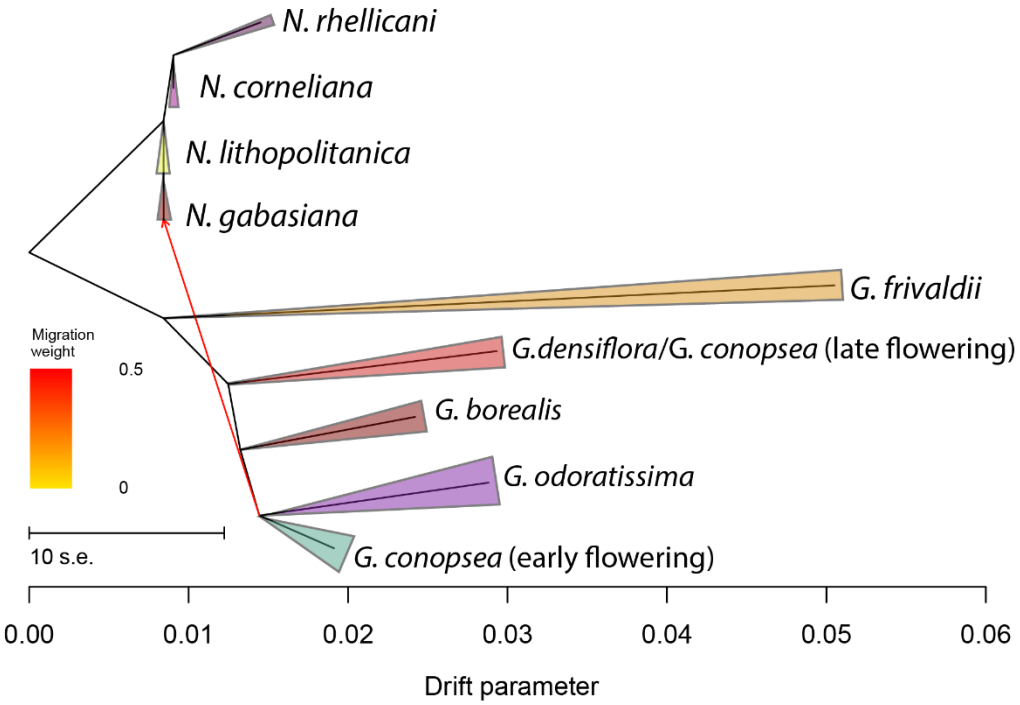


Figure S1. TREEMIX results based on 1,371 SNPs (one SNP per locus) for 21 diploid *Gymnadenia* and *Nigritella* individuals.

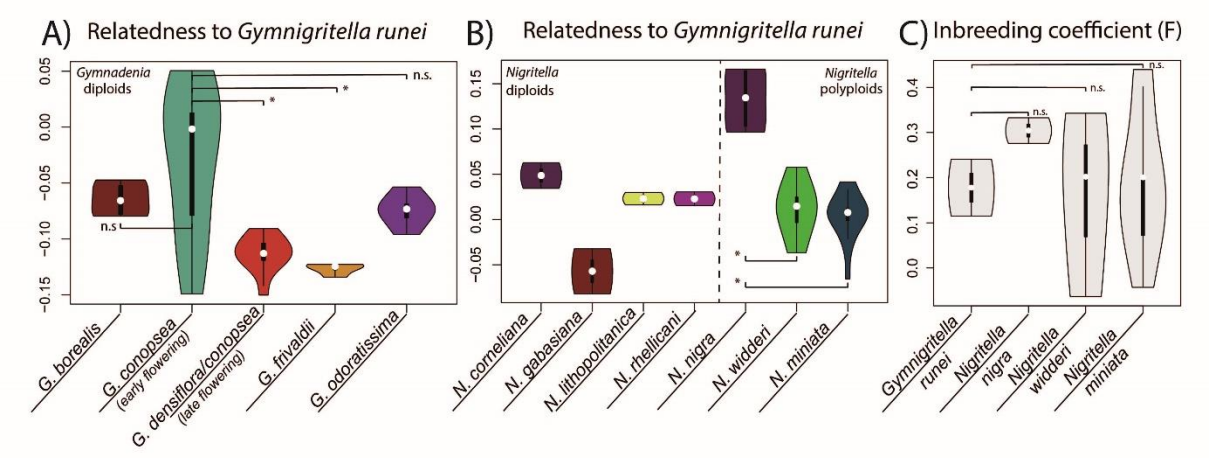


Figure S2. Vioplots showing levels of relatedness between *Gymnigritella runei* and potential ancestral genomes and inbreeding coefficients (F) estimated by VCFTOOLS (Danecek et al. 2011). (a) Plots of relatedness of the allopolyploids to members of the *Gymnadenia* diploids. (b) Plots of relatedness of the allopolyploid to members of *Nigritella* (diploid or polyploid). (c) Vioplots of the relative inbreeding coefficient (*F*) calculated for each polyploid lineage included in this study. Stars indicate significantly different distributions (* $p < 0.05$).

— CHAPTER 4 —

Iterative allogamy-autogamy transitions drive actual and incipient speciation during the ongoing radiation within the orchid genus *Epipactis* (Orchidaceae)

Gábor Sramkó, Ovidiu Paun, Marie K. Brandrud,

Levente Laczkó, Attila Molnár V., Richard M. Bateman

Status: Submitted to Annals of Botany

Contribution: Data processing and analyses

Iterative allogamy–autogamy transitions drive actual and incipient speciation during the ongoing evolutionary radiation within the orchid genus *Epipactis* (Orchidaceae)

Gábor Sramkó[1,2,*], Ovidiu Paun[3], Marie K. Brandrud[3], Levente Laczkó[1], Attila Molnár V.[1] and Richard M. Bateman[4]

[1] *Department of Botany, University of Debrecen, Egyetem tér 1., H4032 Debrecen, Hungary,* [2] *MTA-DE 'Lendület' Evolutionary Phylogenomics Research Group, H4032 Debrecen, Hungary,* [3] *Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, Vienna 1030, Austria,* [4] *Jodrell Laboratory, Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3DS, UK*

* For correspondence. *E-mail:* sramko.gabor@science.unideb.hu (G. Sramkó)

Running title:
Sramkó et al. — *Phylogenomics reveals speciation process in* Epipactis

Abstract

• **Background and Aims** The terrestrial orchid genus *Epipactis* has become a model system for the study of speciation via transitions from allogamy to autogamy, but close phylogenetic relationships have proven difficult to resolve through Sanger sequencing.

• **Methods** We analysed with RAD-seq 108 plants representing 29 named taxa that together span the genus, focusing on section *Euepipactis*. Our filtered matrix of 12,543 single nucleotide polymorphisms was used to generate an unrooted network and a rooted, well-supported likelihood tree. We further inferred genetic structure through a co-ancestry heat-map and admixture analysis, and estimated inbreeding coefficients per sample.

• **Key Results** The 27 named taxa of the ingroup were resolved as 11 genuine, geographically widespread species: four dominantly allogamous and seven dominantly autogamous. A single allogamous species, *E. helleborine*, is the direct ancestor of most of the remaining species, though one of the derived autogams has generated one further autogamous species. An assessment of shared ancestry suggested only sporadic hybridisation between the recircumscribed species. Taxa with greatest inclination towards autogamy show less if any admixture, whereas the gene pools of more allogamous species contain a mixture alleles found in the autogams.

• Conclusions This group is presently undergoing an evolutionary radiation likely driven by a wide spectrum of genotypic, phenotypic and environmental factors. *Epipactis helleborine* has also frequently generated many local variants showing inclinations toward autogamy (and occasionally cleistogamy) and thus becomes an example of a paraphyletic species. Autogams are often as widespread and ecologically successful as allogams.

Key words: Admixture, autogamy, evolutionary dead-end, evolutionary radiation, phylogeography, paraphyly, phylogenomics, RAD-seq, speciation

INTRODUCTION

Self-fertilisation ('selfing') is an important evolutionary mechanism that can provide selective leverage under suboptimal ecological conditions via transmission advantage and reproductive assurance (Stebbins, 1957; Takebayashi and Morrell, 2001; Igic and Busch, 2013; Wright *et al.*, 2013). The Eurasian genus *Epipactis* Zinn. (tribe Neottieae, family Orchidaceae) provides an important example of a clade widely regarded as comparatively rich in selfing species (Claessens and Kleynen, 2011), its hypothesised elevated speciation rate having been attributed to autogamy-triggered isolation (Richards, 1982; Robatsch, 1995; Pridgeon and Light, 2005).

Genera Orchidacearum considered the genus to contain approximately 15 species (Wood, 2005), where – according to a molecular phylogenetic study based on nrITS and four plastid regions (Bateman *et al.*, 2005) – half of these species constitute a paraphyletic group (the so-called 'Section *Arthrochilum*') relative to the monophyletic Section *Euepipactis* (more commonly known as the *Epipactis helleborine* alliance/aggregate). Section *Euepipactis* has a natural distribution that is confined to Eurasia, though the most widespread species, *E. helleborine* sensu stricto (s.s.), has become an enthusiastic occupier of anthropogenic habitats (Rewicz *et al*., 2018) and an invasive adventive in North America (Squirrell *et al.*, 2001; Light and MacConaill, 2006; Kolanowska, 2013). Within the *Euepipactis* clade modest levels of morphological variation exists among sometimes geographically restricted 'swarms' (Fig. 1) – variation that some taxonomists have converted into as many as 65 species (Delforge, 2016) in Europe and Asia Minor, at least partly by invoking the selfing inclination of the genus.

Members of Section *Euepipactis* collectively occupy a wide range of soils and habitat types, and have become a model system for the study of tritrophic mycorrhizal interactions, wherein the orchid parasitises adjacent trees via plumbing provided by fungal intermediaries (Bidartondo *et al.*, 2004). The alliance has also enabled studies of the active transition from autotrophic to facultative mycoheterotrophic nutrition (Selosse *et al.*, 2004; Julou *et al.*, 2005; Schiebold *et al.*, 2017). More importantly in the context of the present study, compared with members of 'Section *Arthrochilus*', those species attributed to Section *Euepipactis* differ among each other less markedly in morphology (Fig. 1) and show negligible divergence in both nrITS and plastid regions (Bateman *et al.*, 2005; Hollingsworth *et al.*, 2006; Drouzas *et al.*, 2017; Zhou and Jin, 2018). Nonetheless, collectively they exhibit the full range of reproductive modes from entomophilous allogamy approximating Hardy-Weinberg equilibrium

through enhanced geitonogamy and facultative autogamy to near-obligate autogamy by means of cleistogamy (Richards, 1986; Robatsch, 1995; Ehlers and Pedersen, 2000; Squirrell *et al.*, 2002; Pridgeon and Light, 2005; Hollingsworth *et al.*, 2006; Tranchida-Lombardo *et al.*, 2011). Thus, the Section may provide an example of the evolution of a bidirectional mating system in which species evolve towards either complete selfing or complete outcrossing, depending on the balance between automatic selection promoting self-fertilisation and the costs of inbreeding via inbreeding depression (Charlesworth and Charlesworth, 1987; Goodwillie *et al.*, 2005; Brys and Jacquemyn, 2016).

Further studies involving population-level allozyme and sequencing analyses (Harris and Abbott, 1997; Ehlers and Pedersen, 2000; Pedersen and Ehlers, 2000; Squirrell *et al.*, 2002; Bateman *et al.*, 2005; Hollingsworth *et al.*, 2006; Tranchida-Lombardo *et al.*, 2011) strongly suggested that the near-obligate autogams had originated iteratively across Europe from within the more widespread, dominantly allogamous *E. helleborine* s.s., leading Section *Euepipactis* to become a textbook case of speciation via the transition from allogamy to autogamy (Proctor and Yeo, 1973; Richards, 1986). Some authors further suggested that this transition in breeding system had a reliable polarity, allogams never arising from autogams (Richards, 1982), and that autogamous *Epipactis* should therefore be viewed as evolutionary dead-ends (Hollingsworth *et al.*, 2006; Bateman, 2009; Bateman, 2012a; Bateman, 2012b; Claessens and Kleynen, 2016). Some authors also argued that a clade containing substantial numbers of putative species, yet possessing so little molecular divergence in otherwise polymorphic regions that an origin more than 1 Myr ago is unlikely, offers a particularly high probability of constituting an active evolutionary radiation (Bateman, 1999).

There exists interest well beyond the realm of orchid studies in resolving the general questions of (a) whether transitions from allogamy to autogamy can be iterative (e.g., Squirrell *et al.*, 2002), (b) whether autogams can speciate to form further autogams (Stebbins, 1970; Takebayashi and Morrell, 2001; Bateman, 2009; Igic and Busch, 2013; Wright *et al.*, 2013), and (c) whether they might even undergo 'reverse speciation' to form novel allogams. Capturing such transitions in the midst of a genuine evolutionary diversification would be an additional bonus (e.g., Bateman *et al.*, 1998; Bateman, 1999).

Recently developed molecular techniques collectively termed next-generation sequencing (NGS; Schloss, 2008; Olson *et al.*, 2016) offer the opportunity to bring to bear on such questions far larger numbers of genome-wide, phylogenetically informative characters and thereby improve resolution of phylogenetic relationships among such controversial taxa (Kircher and Kelso, 2010; Harrison and Kidner, 2011). Among these techniques, restriction site-associated

sequencing (RAD-seq) utilises the Illumina high-throughput sequencing technique (Baird *et al.*, 2008) to screen hundreds of thousands of base-pairs of genomic DNA for informative sites (SNPs) sampled throughout the whole genome, yielding incomparably more data than conventional candidate-gene sequencing (Rubin *et al.*, 2012; Pante *et al.*, 2015). RAD-seq is increasingly being employed in studies of the plant kingdom (e.g., Eaton and Ree, 2013; Paun *et al.*, 2016; Ren *et al.*, 2017; Trucchi *et al.*, 2017; Heckenhauer *et al.*, 2018). Having applied the RAD-seq technique to the other orchid genera (Bateman *et al*., 2018; Brandrud *et al*. submitted – Chapter I) we have now turned our attention to *Epipactis* section *Euepipactis*. We sampled the genus extensively across Europe and (to a lesser degree) Asia Minor, initially within a framework provided by the divisive taxonomy of Delforge (2016). Here, we selected a subset of samples for detailed analysis that gave greater emphasis than previous studies to material from the less intensively researched eastern European taxa. Our main objectives in pursuing the present study were:

(1) To use well-founded molecular estimates of monophyly and disparity as a guide to determine which of the many named taxa have the strongest support to be recognised as *bona fide* species;

(2) To determine with greater confidence the relationships among those species;

(3) To explore any phylogenetic patterns that can be discerned from the putative universal ancestor of this taxonomic diversity, *E. helleborine*;

(4) To combine those phylogenetic insights with knowledge of the reproductive biology of taxa in order to test previous hypotheses relevant to the broader discipline of evolutionary biology. More specifically, we explore whether (a) few or many autogamous taxa have arisen iteratively across Europe from within the widespread allogam *E. helleborine* s.s., (b) the transition from allogamous to autogamous lineage is irreversible, the converse transition never taking place, and (c) autogams are evolutionary dead-ends, no autogam ever giving rise to further autogamous species.

We conclude by speculating on the geographic regions of origin of the derived species and whether section *Euepipactis* has indeed acquired the rare scientific credentials necessary to be viewed as undergoing an active evolutionary radiation.

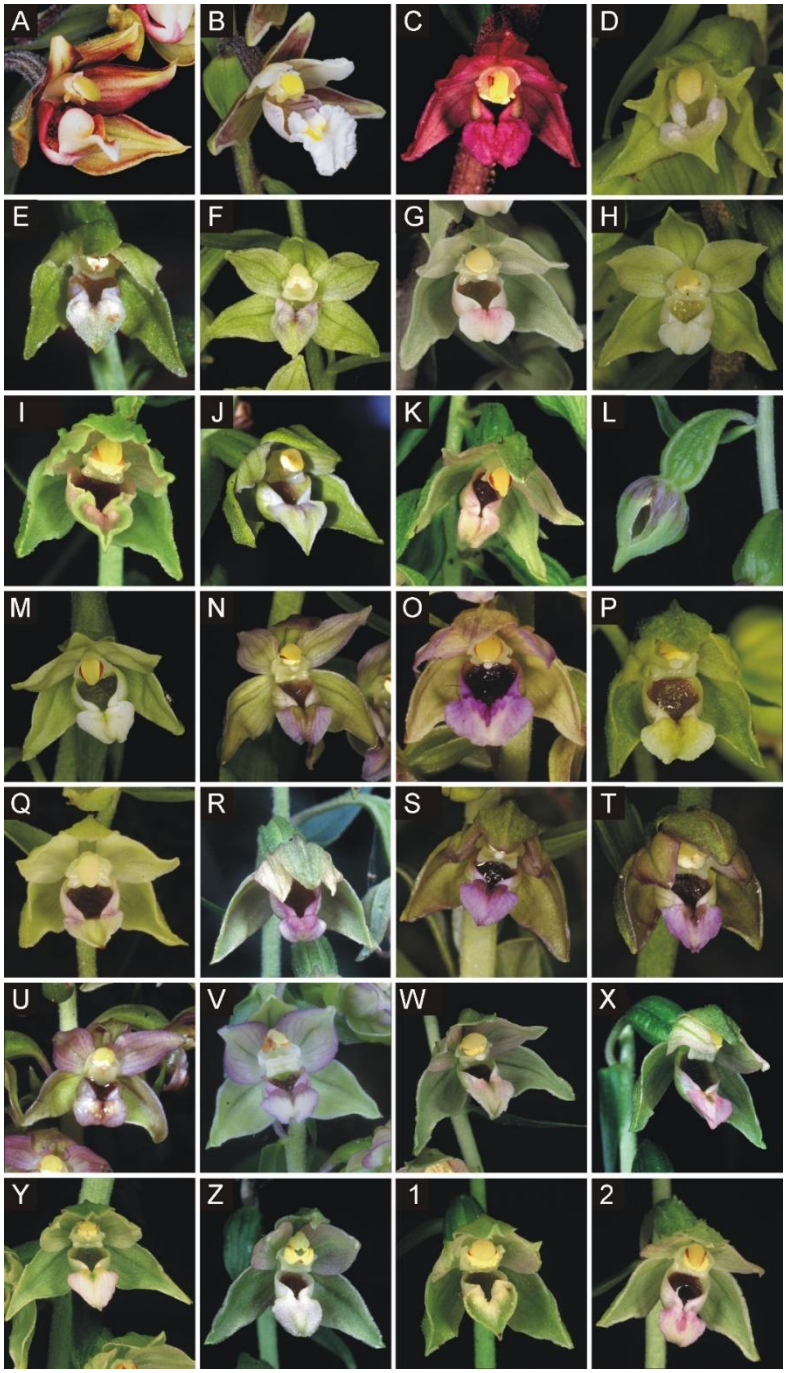


FIGURE 1. Flowers of 11 *Epipactis* taxa representing the morphological variation displayed by the studied genus. Taxa with a verified specific rank in our analyses are preceded by 'E.'; the remaining taxa are best viewed as infraspecific, but further research using modern methods is needed before the most appropriate infraspecific rank for each of these taxa can be determined. Taxa: A – *E. veratrifolia*, Cyprus; B – *E. palustris*, Hungary; C – *E. atrorubens*, Hungary; D – *E. phyllanthes*, Belgium; E –*persica*, Turkey; F –*exilis*, Bulgaria; G – *E. purpurata*, Hungary; H –*pseudopurpurata*, Hungary; I – *E. leptochila*, Hungary; J –*peitzii*, Hungary; K –*neglecta*, Hungary; L –*futakii*, Hungary; M – *E. greuteri*, Greece; N –*nauosaensis*, Greece; O –*densifolia*, Turkey; P – *E. pontica*, Hungary; Q – *E. muelleri*, Hungary; R – *voethii*, Hungary; S–*neerlandica*, Denmark; T –*renzii*, Denmark; U –*distans*, Serbia; V – *E.*

helleborine, Hungary; W –*minor*, Hungary; X –*mecsekensis*, Hungary; Y –*tallosii*, Hungary; Z –*nordeniorum*, Hungary; AA –*albensis*, Hungary; AB –*bugacensis*, Hungary. Flowers are shown at the same scale. Images: A–C, E, G, I–L, O, Q, R, U–AB by A. Molnár V.; D, F, H, M, N, P, S, T by M. Óvári.

MATERIALS AND METHODS

Sampling

Previous molecular phylogenetic analyses (notably Bateman *et al.*, 2005) were used to select as outgroups from within the apparently paraphyletic section *Arthrochilus* two successively diverging species, *E. veratrifolia* and *E. palustris*. Within the ingroup (i.e., section *Euepipactis*), the 29 samples selected of *E. helleborine* s.s. – the putative progenitor of the autogamous species – extended west–east from the coast of Wales to the eastern shore of the Black Sea (Fig. 2). Within this area, we selected a further 77 ingroup samples, together encompassing 26 named taxa. Each taxon was represented by between one and six samples (mean 2.9 ± 1.4), with a concentration of sampling in Eastern Europe. Taxa were chosen to encompass the full range of breeding systems from putative obligate allogams to putative obligate autogams (Table 1). Field-collected samples of leaf tissue were immediately immersed in sachets of fine-grained silica gel, and an open flower from each plant was immediately placed in 96% ethanol as a voucher to be deposited in the herbarium of the University of Debrecen (DE).

RAD-seq library preparation and SNP filtering

The silica gel-stored tissue was used as a template for DNA extraction following the protocol detailed by Sramkó *et al*. (2014). A modified CTAB protocol was used with RNase treatment to isolate high molecular weight genomic DNA. Checks on 1.8% agarose gel enabled removal of any samples showing signs of fragmentation and/or RNA contamination, and identified a positive correlation between period of storage and degree of genomic DNA degradation. Double-stranded DNA (dsDNA) contents of all acceptable samples were assessed using a Qubit v.3.0 fluorimeter (Thermo-Fisher Scientific Inc., USA).

Three single-digest RAD libraries were prepared from between 40 and 60 individuals (including repeats where judged necessary). For each accession, 210 ng of dsDNA was digested with *Sbf*I-HF enzyme (New England Biolabs Inc., USA), reflecting the comparatively large genome size of *E. helleborine* (1C = ca 14 pg: Leitch *et al.*, 2009). The protocol of library preparation largely followed Paun *et al*. (2016) but with the minor modifications described by Trucchi *et al*. (2017). The only deviation from these past protocols was applying a different regime of sonication using Bioruptor Pico (Diagenode, Belgium), which in this case involved

three cycles of 45 s 'on' and 45 s 'off' at 6 °C. After library control, the libraries were submitted to the VBCF NGS Unit (www.vbcf.ac.at/ngs) for sequencing on an Illumina HiSeq as 100 bp single-end reads.

Raw Illumina reads were first demultiplexed based on index reads with BamIndexDecoder v.1.03 (included in Picard Illumina2Bam package, available from http://gq1.github.io/illumina2bam/). Stacks v.1.44 (Catchen *et al.*, 2011) was used to further process the RAD-seq data. The reads were de-multiplexed and quality checked with the process_radtags program under the following settings: remove any uncalled base, discard both reads with low quality scores and rescue barcodes, and cut sites with a maximum of one mismatch. Loci were produced *de novo*, only allowing 'sequence stacks' to be formed with a minimum depth of five reads. One mismatch was allowed when merging loci within individuals (i.e., setting "M" in denovo_map.pl of Stacks) but two mismatches were allowed when merging loci among individuals (i.e., setting "n"). With export_sql.pl of Stacks we further extracted from this catalogue a set of loci that occurred in at least 50% of the individuals and contained between one and 15 single nucleotide polymorphism sites (SNPs), yielding 3,927 RAD-seq loci. A consensus haplotype of each of these loci was further used to build a synthetic reference, including each locus as a "chromosome". To better accommodate the biological and technical variation in the data (i.e., in coverage per individual and locus), the raw reads of all individuals were then mapped back to this reference using BOWTIE2 v.2.2.6 (Langmead and Salzberg, 2012). We then followed the bioinformatics pipeline used by Heckenhauer *et al.* (2018), sorting the aligned sam files by reference coordinate and adding read groups using Picard tools v.2.9.2 (available from http://broadinstitute.github.io/picard), followed by local realigning around indels using the Genome Analysis Toolkit v.3.7.0 (GATK; McKenna *et al*., 2010). Finally, SNPs were called from the realigned bam files using default settings for ref_map.pl of Stacks (Catchen *et al.*, 2011). A VCF file was produced using Stacks' populations and the file was further filtered using VCFtools v.0.1.14 (Danecek *et al.*, 2011) to only contain SNPs present in at least 90% of the individuals. The vcf file was finally converted to phylip format with PGDSpider v.2.1.1.3 (Lischer and Excoffier, 2012) using IUPAC symbols to represent heterozygosity.

A measure of inbreeding was estimated for each individual using a method of moments with VCFtools (option --het). The results were drawn as beanplots (Kampstra, 2008) for each of four putative breeding groups (i.e., obligate allogams, predominantly allogams, predominantly autogams and obligate autogams) in R v.3.1.2, run under Rstudio v.0.98.1102. The

statistical significance of differences in distributions of the inbreeding coefficient *F* between the groups were estimated with Mann-Whitney tests in R (command wilcox.test()).

To run admixture analyses, we used the genotype-free method of calling variants implemented in ANGSD v.0.910 (Korneliussen *et al*., 2014) on the indel-realigned mapping files. Here, we excluded the two outgroup accessions, retained only reads with base and mapping qualities of at least 20 (-minMapQ 20 -minQ 20), and implemented a SAMtools-derived algorithm (-GL 1) when inferring alleles from genotype likelihoods (-doMajorMinor 1). We kept only variable sites at $p < 0.000001$ (-SNP_pval 1e-06) that had data for at least 50% of individuals and showed a minor allele frequency of 0.02. Finally, we estimated allele frequencies (-doMaf 2) and output three possible genotypes in a beagle genotype likelihood format (-doGlf 2).

Phylogenetic analyses of SNP data

For phylogenetic tree reconstruction we used Maximum Likelihood (ML), as implemented in the web-based version of PhyML v.3.0 (Guindon *et al.*, 2010) with automatic model selection (Lefort *et al.*, 2017). This found the GTR+G model of sequence evolution to be most appropriate for the present dataset. Branch robustness was tested using the Approximate Likelihood-Ratio Test (aLRT) approach (Anisimova and Gascuel, 2006), where branch support was categorised as 'strong' (aLRT ≥ 0.95), 'moderate' (0.95 > aLRT ≥ 0.90), 'weak' (0.90 > aLRT ≥ 0.81) or 'none' (0.80 > aLRT). The resulting ML phylogenetic tree was visualised using PRESTO (available at: www.atgc-montpellier.fr/presto/).

To explore possible reticulations within the phylogeny, we also built a phylogenetic network using the uncorrected-p distance in a NeighbourNet analysis, as implemented in SplitsTree v.4.14.4. This approach provides a better representation of reticulate evolution, which is likely to be present when outcrossing species are considered (Huson and Bryant, 2006).

Lastly, shared ancestry of *Epipactis* taxa was explored on a subsampled dataset of unlinked SNPs (i.e., selecting one SNP per locus) with fineRADstructure v.0.2 (Malinsky *et al.*, 2017) using default settings.

To assess the genetic structure of the ingroup, we performed an admixture analyses with NGSadmix v.32 based on the genotype likelihoods dataset obtained from ANGSD. Ten independent runs from different starting seeds were carried out for values of K ranging from one to 22. The optimal number of clusters (K) was determined by implementing the Evanno method (Evanno *et al*., 2005), whereas the graphical representations of the results were generated as barplots using R.

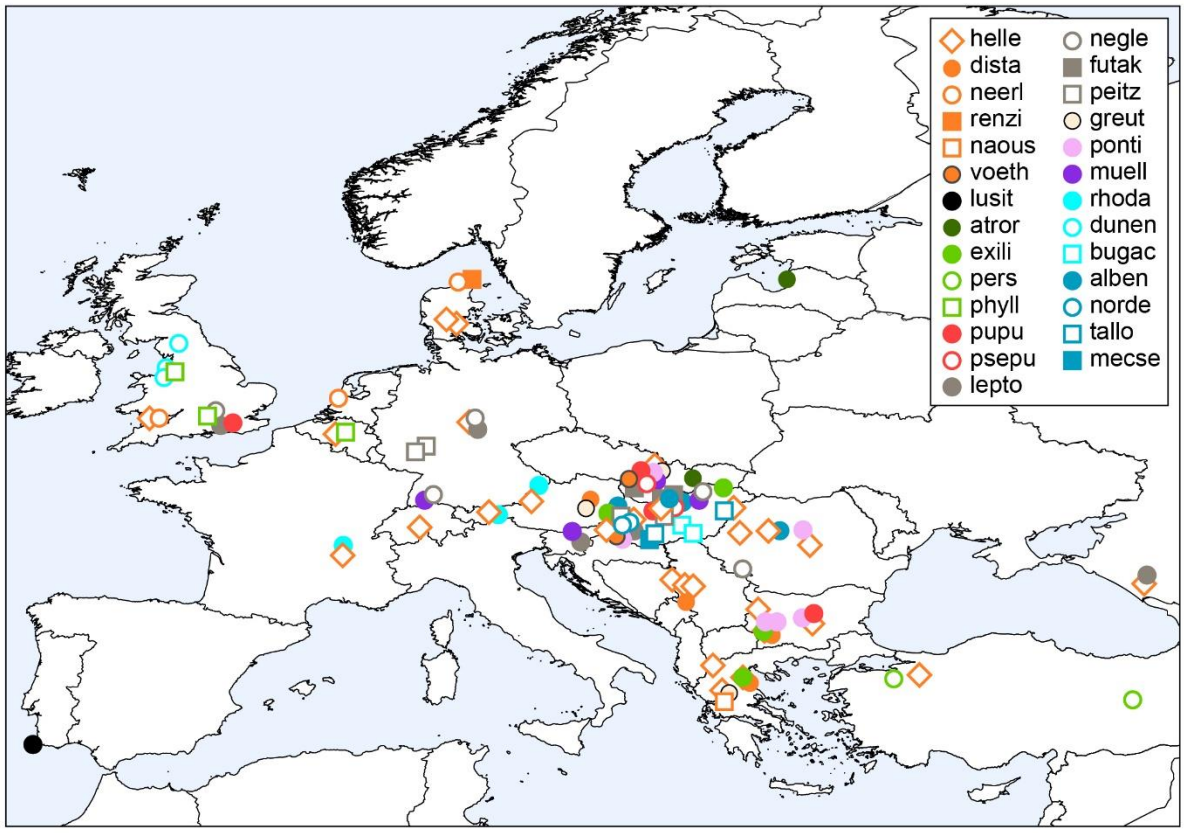


FIGURE 2. Collecting locations of ingroup samples. Taxa codes follow Supplementary Table 1. The map layer was extracted from GADM version 1.0 (available from www.gadm.org).

RESULTS AND DISCUSSION

Phylogenetic overview

The average number of Illumina reads per accession retained after demultiplexing and quality filtering was 1.9 (SD = 0.8) million. When allowing for the presence of 10% missing values, the 108 *Epipactis* accessions representing 29 species collectively produced 12,543 filtered SNPs for use in downstream phylogenetic analyses. The average coverage across variants and individuals was 125× (SD 60×). The variant calling method of ANGSD inferred genotype likelihoods at 9,415 sites that passed the criteria mentioned above.

As would be anticipated for a phylogeny based on such a large volume of informative characters, most of the branches on the resulting ML tree (Fig. 3) receive either strong or moderate statistical support (aLRT ≥ 90). In summary, as expected, branches separating the outgroups from the ingroup are much longer than those within the ingroup. The tree (described in greater detail below) reveals several monophyletic groups subtended by comparatively long branches that are nested with a paraphyletic (arguably polyphyletic) *E. hellebor-*

ine. The deeper branches are considerably shorter; three receive only weak statistical support and a further three collapse.

The weak confidence in the relationships recovered between some of these early-divergent lineages is even more evident in the star-shaped unrooted network (Fig. 4). Nonetheless, major groups of samples identified as monophyletic in the rooted tree remain evident, once again subtended by comparatively long branches. Relationships among the individual samples are strikingly similar, and only two samples occupy placements significantly different from those evident in the rooted tree. The Bulgarian sample attributed to *E. 'exilis'* (223) is placed well below the other *'exilis'* samples, and the English plant attributed by some workers to the supposedly entirely Continental *E. 'neglecta'* (348) is no longer placed within the *E. leptochila* clade that includes the remaining four samples attributed to *'neglecta'*; instead, it is placed close to the origin of the network.

One further sample merits explanation because of its unexpected phylogenetic placement. A Russian sample (165) that when collected was attributed to *E. helleborine* is actually nested phylogenetically well within *E. leptochila* s.l., adjacent to a Russian sample of *E. leptochila* s.s. (166). This proximity leads us to believe that sample 165 was misidentified when collected, actually being a plant of *E. leptochila* that possesses an unusually well-developed viscidium. Admittedly, *E. leptochila* has not previously been formally reported from Russia (Efimov, 2004).

Having attempted to account for these 'rogue' samples, we can now proceed to use the remaining samples to attempt an optimal circumscription of species within section *Euepipactis*. Named taxa regarded by us as *bona fide* species are preceded by *'E.'*; taxa regarded by us as infraspecific are simply given as single epithets (e.g. *'exilis'*).

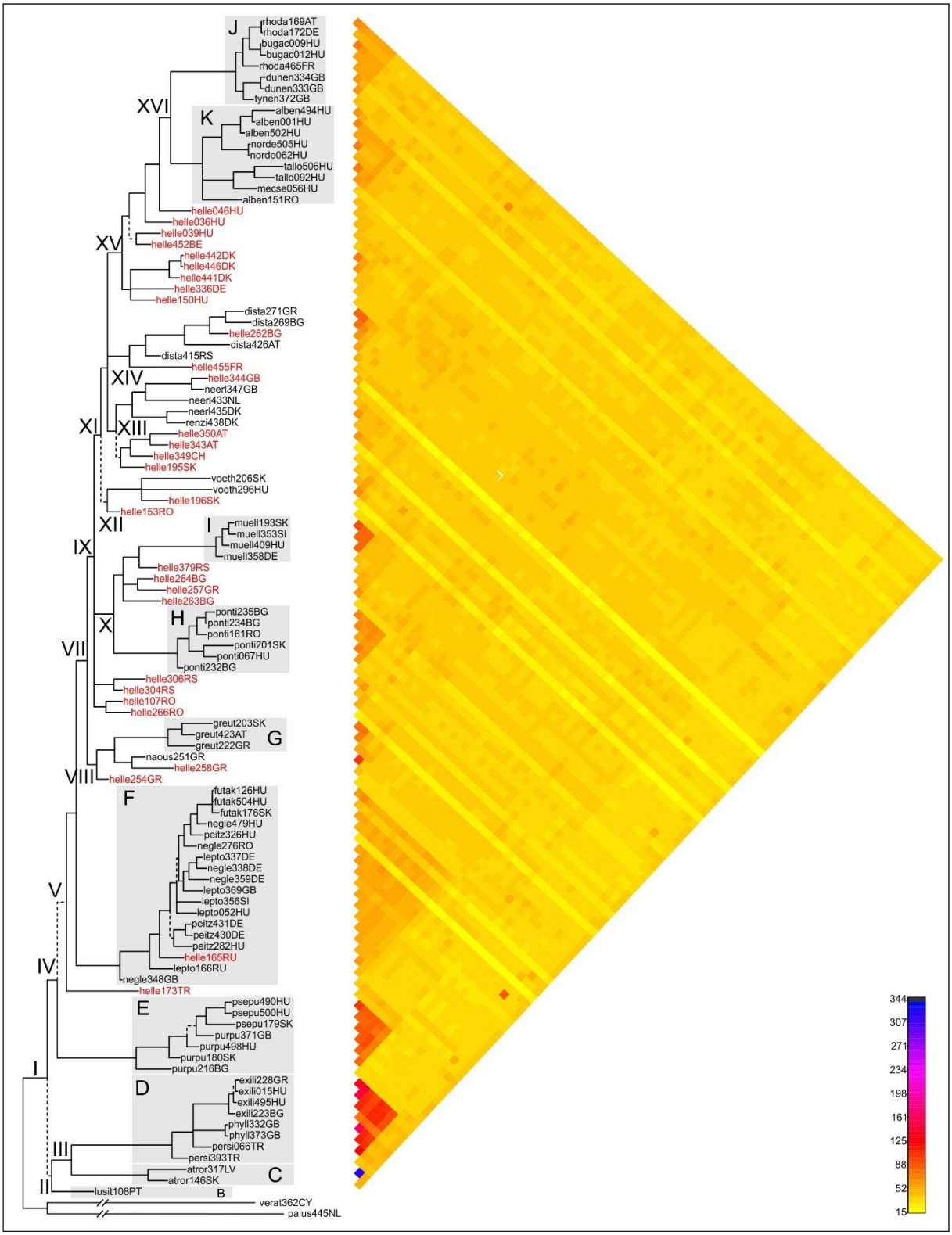


FIGURE 3. Maximum Likelihood phylogram depicting the evolutionary relationship of the studied *Epipactis* species (left) and 'heat map' of shared ancestry (right) based on RADseq data. On the tree unsupported branches (aLRT < 80) are collapsed, weakly supported branches (80 ≤ aLRT <90) are dashed, whereas moderately to strongly supported branches (aLRT ≥90) are presented as a continuous line. Sample codes follow Supplementary Table 1 suffixed

by a two-letter country code. Major groups of samples identified as monophyletic entities are grouped by grey shading and identified by capital letters (B–K), whereas main branches referenced in the text are indicated by roman numerals. Samples of the supposed progenitor species, forming the paraphyletic group 'A', are highlighted in red. Branches leading to the two outgroup samples are severely truncated to facilitate viewing. The 'heat map' indicates the level of shared ancestry between samples according to the attached legend, where yellow is the lowest and dark blue is the highest co-ancestry.

Re-circumscription of species

Our exploration of the more evolutionary aspects of this study will be made simpler and more comprehensible if we begin the more detailed discussion with our conclusions regarding species delimitation. In doing so, it is important to bear in mind that the genetic disparities between any of the ingroup plants shown in Fig. 3 are much smaller than those detected between most other closely related species of European orchids – even the major lineages within *Ophrys* offer greater collective disparity (Bateman *et al.*, 2003).

The more relevant of Bateman's (2012a) rules for converting a tree into species are, briefly:

(1) Recognise only monophyletic groups;

(2) Preferentially divide the tree at branches that are relatively robust (and usually comparatively long within the context of the tree in question);

(3) Preferentially divide the tree in a way that minimises the need to (a) create new names (a low risk in this case, given the huge number of epithets already available within section *Euepipactis*) and (b) create new combinations of existing names.

Based on applying these rules to the topology and comparative branch lengths of Figs. 3 and 4, we conclude that optimally 11 ingroup species can be recognised, ten of them ostensibly monophyletic and seven of them dominantly autogamous. These groups of samples that comprise these species are labelled A–K on Fig. 3; additional internal branches of potential interest are labelled in roman numerals (I–XVI). These 11 ingroup genetic pools are also evident in the admixture results (Fig. 5), which find optimal peaks in clustering the accessions in four, seven and 15 groups (Supplementary Fig. S1) according to the deltaK method of Evanno *et al*. (2005).

It is immediately clear from Fig. 3 that most of the autogamous lineages are embedded within the ancestral plexus that is *E. helleborine* (species A). The samples attributed by us to *E. helleborine* are distributed throughout much of the tree, though concentrated in the derived

portion of the tree above branch VII. In the network (Fig. 4) they are placed around the centre of the graph on shorter branches. The admixture results (Fig. 5) show this lineage to be a mixture of alleles from all other genetic pools. Imposing rigorous monophyly on *E. helleborine* s.s. would require us to allocate the 29 samples of *E. helleborine* among no less than 16 'species'. Given that all of these 'species' that would likely be indistinguishable using morphological criteria (or even standard candidate-gene sequencing), we have little option but to continue to accept *E. helleborine* as a most likely paraphyletic species. Ongoing uncertainty rests over two samples located at important nodes located low in the tree: Might the single analysed sample (108) of the Iberian allogam *E. lusitanica* (species B) be more justifiably attributed to *E. helleborine*? Or conversely, could the single Turkish plant attached above branch V, diverging immediately below species F and initially attributed by us to *E. helleborine* (173), actually represent a new and as yet unrecognised allogamous species?

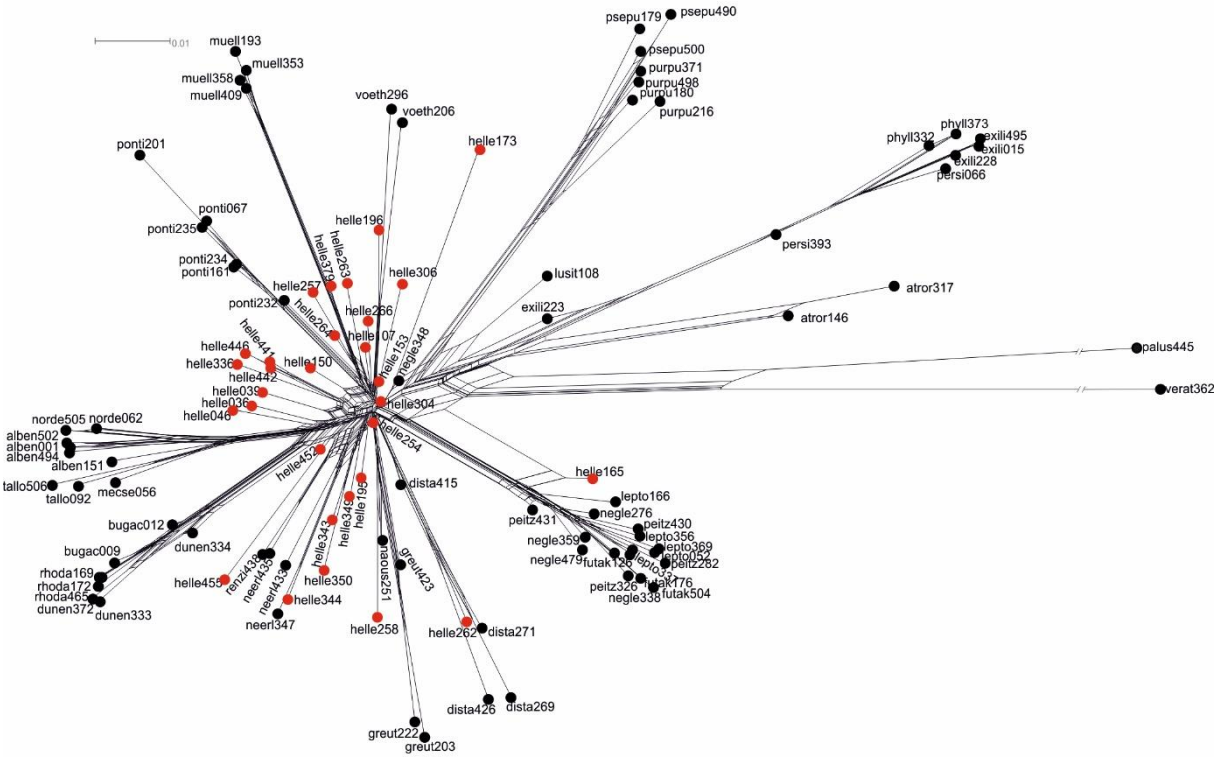


FIGURE 4. Phylogenetic network depicting the genetic relationship of the studied *Epipactis* samples presented as a NeighbourNet network based on uncorrected-p distance. The dots represent an edge with a sample (codes are given according to Supplementary Table 1); red dots represent samples of the supposed progenitor species. The branches leading to the two outgroup samples are severely truncated to facilitate viewing.

The nine remaining derived species are distributed throughout the tree. All are nucleated around at least one well-known and comparatively widespread taxon (when selecting the pre-

ferred epithet for each species, we adopted the apparently earliest name but did not perform an in-depth nomenclatural analysis). Within our spectrum of analysed taxa that were treated as 'species' by Delforge (2016), four of the nine *bona fide* species contain samples attributed to only one Delforgean species: *E. atrorubens* (C), *E. greuteri* (G), *E. pontica* (H) and *E. muelleri* (I). Between two and four Delforgean species are encompassed by the five remaining species: *E. phyllanthes* (D), *E. purpurata* (E), *E. leptochila* (F), *E. albensis* (J) and *E. dunensis* (K). All of these groups are monophyletic (Fig. 3) and genetically relatively homogeneous (Fig. 5) (Rule 1); it is mainly the criterion of comparatively long molecular branches subtending species (Rule 2) that has dictated these amalgamations of putative species.

For example, in species K (*E. dunensis*), samples attributed to *rhodanenisis* and *bugacensis* are intermingled, thus failing the criterion of monophyly, but could in theory be combined into a single monophyletic taxon, which would be sister to *E. dunensis* (apparently monophyletic) and its localised segregate *tynensis* (monophyly not tested, as only one sample was included in our analysis). However, the very short branch lengths within the group (Fig. 3) and the unique genetic cluster identified in the admixture analyses (Fig. 5) clearly show that these samples constitute a single cohesive biological entity, particularly when considering that together these eight samples of species K extend from Britain eastward as far as Hungary and share a clear habitat preference for disturbed sandy soils. In the case of species E (*E. purpurata*), four widely geographically separate samples attributed to *E. purpurata* form a paraphyletic group subtending a monophyletic (Fig. 3) group of three samples of the eastern European 'local endemic' obligate autogam *pseudopurpurata*, but comparative branch lengths indicate these two taxa should be treated as a single species. They also form a single gene pool (Fig. 5). Species F (*E. leptochila*) encompasses three additional named taxa, *neglecta*, *peitzii* and *futakii*, that have often been recognised as species but were grouped within *E. leptochila* as subspecies by Kreutz (2004) and as varieties by Delforge (2016).

Other Delforgean species clearly cannot be separated at species level from *E. helleborine* s.s., either because they are phylogenetically intermingled with samples of *E. helleborine*, thus failing Rule 1 (*distans*), or are placed close to unequivocal samples of *E. helleborine* s.s., thus failing Rule 2 (e.g., *neerlandica, naousaensis, voethii*). Although absent from our analysis, *E. microphylla* has already proven its qualifications as an autogamous species in previous molecular studies (Bateman *et al.*, 2005; Hollingsworth *et al.*, 2006; Tranchida-Lombardo *et al.*, 2011; Zhou and Jin, 2018). With that exception, we suspect that most, but probably not all, of the 39 putative species of Section *Euepipactis* that were listed by Delforge (2016) but not included in our analysis would similarly fail to qualify as *bona fide* species if analysed via

RAD-seq, each instead either being placed phylogenetically within one of the derived species B–K or in some cases within the ancestral plexus that is *E. helleborine*. As is the case in several other genera of Eurasian orchids, severe taxonomic inflation has occurred in recent years, driven largely by intensity of interest within the orchidological community. Of course, the challenge remaining to systematic biology is to generate morphometric matrices of equal rigour and taxonomic breadth, in order to assess the degree of genus-wide congruence between genotype and phenotype. Thus far, with very few geographically restricted exceptions (e.g., Tyteca and Dufrene, 1994), the morphology of *Epipactis* has been seriously under-researched using genuinely scientific methods. Such studies could be particularly effective in apportioning named taxa that do not merit species rank among the ranks of subspecies (e.g., *E. helleborine* subsp. *neerlandica*) or varieties (e.g., *E. helleborine* subsp. *neerlandica* var. *renzii*) – a topic to which we return in the next section (Pedersen and Ehlers, 2000; Brys and Jacquemyn, 2016; Jacquemyn *et al.*, 2018).

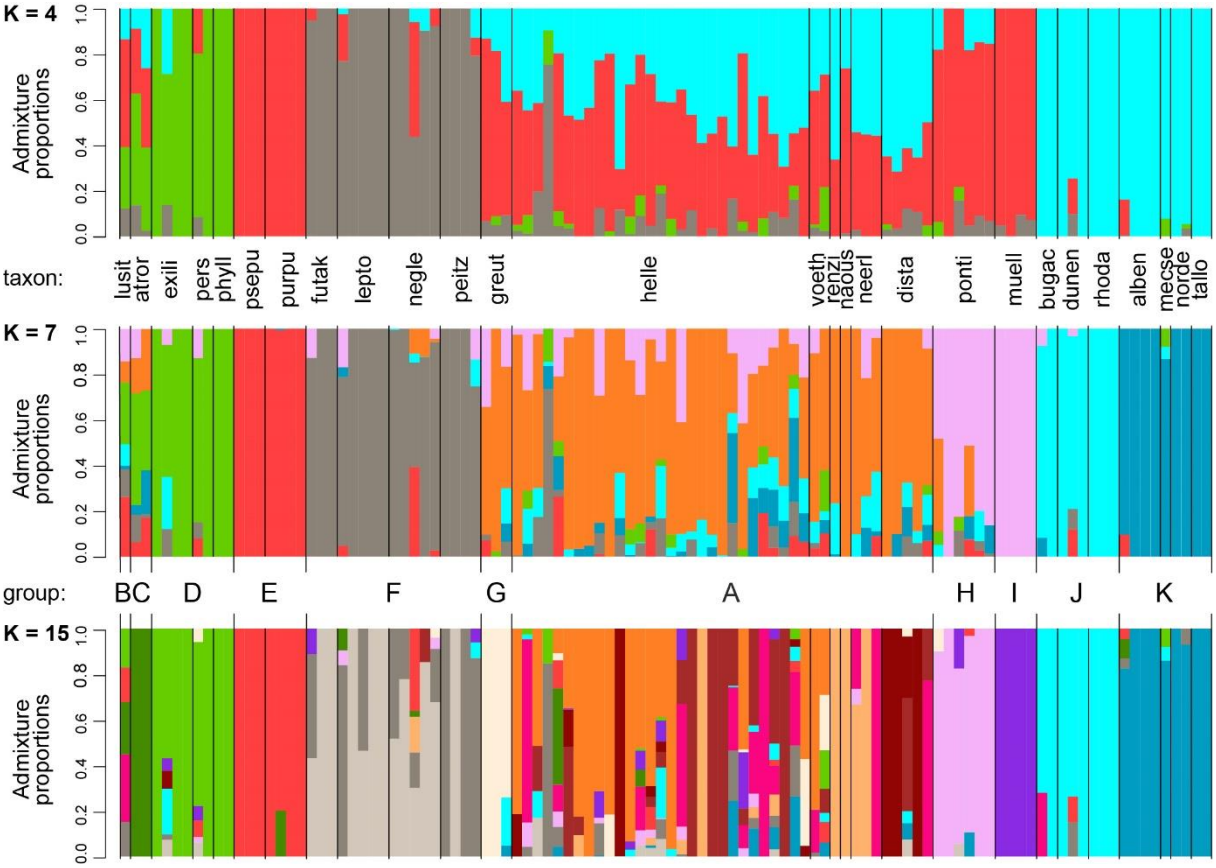


FIGURE 5. Barplots of results of admixture analyses based on 9,415 polymorphic sites inferred with ANGSD. The method by Evanno *et al.* (2005) identified four, seven and 15 groups as best clustering patterns across the ingroup samples. Taxa codes follow Supplementary Table 1. Major groups of samples identified in Fig. 3 are indicated by capital letters (B–K) and are compared with the supposed progenitor paraphyletic group A.

Ancestry and breeding system during speciation

Branches subtending these ten derived monophyletic species are reliably considerably longer than those that form the backbone of the phylogeny, suggesting that delimitation of these species can be discussed with greater confidence than relationships among them. This result is unsurprising, as previous sequencing studies have given meaningful support to only one major branch: Branch V, which separates the early-divergent species *E. phyllanthes*, *E. purpurata* and *E. leptochila* (also *E. microphylla*) from the more derived clade containing the remaining species, was evident previously in both nrITS sequences (Bateman *et al.*, 2005; Hollingsworth *et al.*, 2006; Zhou and Jin, 2018) and a marker in the plastid *rbc*L–*acc*D spacer (Tranchida-Lombardo et al., 2011). Interestingly, this is the branch above which incontrovertible specimens of *E. helleborine* are confined (Fig. 3). Moreover, of the three species that diverge below this point on Fig. 3, two (*E. atrorubens*, *E. purpurata*) are dominantly allogamous; only *E. phyllanthes* is autogamous (though previous candidate gene studies show that the autogam *E. microphylla* also diverges below this point). In contrast, all five species that diverged subsequently from within the *E. helleborine* alliance are at least facultative autogams (*E. leptochila*, *E. greuteri*, *E. muelleri*, *E. albensis*, *E. dunensis*). Although in most cases the autogamous species have as their closest relatives allogamous species there is one exception – the eastern European *E. albensis* is sister to the western European *E. dunensis*. This relationship implies that one of these autogamous species most likely gave rise to the other, though it does not indicate which is the more probable ancestor. In summary, our initial hypotheses that (a) autogams evolved from allogams and (b) autogams are evolutionary dead-ends do constitute good guidelines, as was previously believed; however, we have also shown that they fall short of being unbreakable rules.

In addition to *bona fide* species derived from *E. helleborine*, there are signs of incipient speciation within the clade subtended by branch XI (Fig. 3), yielding taxa most appropriately treated as subspecies that have not (yet) gained clear molecular monophyly or become clearly distinct morphologically from *E. helleborine*. One of these examples, located above branch XII, is the localised eastern European facultative autogam *voethii*, intermingled with a co-occurring sample of typical *E. helleborine* from western Slovakia. The other example, located above branch XIV, is the localised eastern European, facultative autogam *distans*, again intermingled with a co-occurring sample of typical *E. helleborine* from a lower elevation on the same Bulgarian mountain. Also of interest is *neerlandica*, located above branch XIII and here, as expected, shown to be a coastal ecotype of *E. helleborine* (Brys and Jacquemyn, 2016).

Within this clade, geography dominates over both taxonomy and breeding system. The Danish plant of the allogam *neerlandica* is sister to, and molecularly very similar to, the Danish plant attributed to the rare and reputedly obligate autogam *renzii*. The Welsh duneland plant attributed to *neerlandica* has a similarly close relationship with a plant of typical *E. helleborine* sampled from nearby back-dune woodland. These results endorse previous, more detailed investigations of this duneland ecotype (Jacquemyn *et al.*, 2016; Jacquemyn *et al.*, 2018), but are the first to demonstrate a close relationship between the Welsh and Dutch/Danish populations in a molecular phylogenetic context. Further down the tree, above branch VIII, a Greek sample of *E. helleborine* s.s. is shown to be sister to the molecularly similar facultative autogam *nauosaensis*, reputedly endemic to Greece and Bulgaria and said to "often form hybrid swarms with *E. helleborine*" (Delforge, 2016: 72) – an observation readily explained as these few localised populations are, in fact, a minor variant within *E. helleborine*.

All of the *bona fide* species present in our analysis of section *Euepipactis* are geographically widespread, extending across at least half of Europe. The most extensive are *E. helleborine* and *E. atrorubens* – the only species to reach as far northward as Sweden and Norway. In contrast, only *E. helleborine* and *E. phyllanthes* (together with *E. microphylla*) appear comfortable growing around the Mediterranean. Central Europe has marginally the highest species diversity, where species with a western bias (*E. leptochila*, *E. muelleri*, *E. dunensis*) overlap with species with an eastern bias (notably *E. pontica*) and others have the centres of their distributions (*E. greuteri*, *E. albensis*). As would be expected (e.g., Burns-Balogh *et al.*, 1987), taxa apparently undergoing incipient speciation (*distans* and especially *voethii*) consist of fewer, more localised populations.

With regard to seeking likely centres of origin of these derived species, there is obvious interest in exploring the 'background' genetic structure within the ancestor of most, *E. helleborine*. This task was made rather more difficult by the realisation that the two easternmost samples attributed by their collectors to *E. helleborine* yielded substantially deviant RAD-seq profiles. As already noted, we suspect that sample 165 from the Caucasus Mountains was a morphologically misidentified plant actually attributable to *E. leptochila* – a species not previously confirmed from Russia (Efimov, 2004). The western Turkish sample 173, isolated in the phylogeny on a comparatively poorly-supported branch V immediately below *E. leptochila* (Fig. 3), presents a more ambiguous challenge. We could continue to view this plant as the earliest-divergent of our *E. helleborine* samples, but alternatively it could be treated as a geographically disparate, as-yet unnamed allogam – analogous with, and geographically mirror-

ing, the Iberian endemic *E. lusitanica* located near the base of the tree (species B). Allogamous populations of Asia Minor evidently merit further investigation.

As expected from mainstream evolutionary theory (Wright *et al.*, 2013), lineages of all autogamous species (as recircumscribed here) form monophyletic entities on our phylogenetic tree (Fig. 3), suggesting that gene-flow with their progenitor, *E. helleborine* s.s., is limited or absent (Brys and Jacquemyn, 2016). Inbreeding coefficients calculated from RAD-seq data indicate a tendency for frequent selfing in all hypothesised obligately and predominantly autogamous species (Fig. 6). However, all inbreeding coefficients recovered were positive within a theoretical scale of –1 to +1 for *F* values, most likely a sign of often small effective population sizes in *Epipactis*. Previous studies reported much narrower genetic variability in self-pollinating *Epipactis* taxa compared with the outcrossing progenitor species *E. helleborine* s.s. (Ehlers and Pedersen, 2000; Pedersen and Ehlers, 2000; Squirrell *et al.*, 2002; Hollingsworth *et al.*, 2006). A similar pattern is evident in our study, as *E. helleborine* s.s. occupies a much greater genetic space on the phylogenetic network than any other species (Fig. 4). Moreover, all but one of the autogamous species arise from within this genetic variation, thus arguing for the universality of *E. helleborine* s.s. as the progenitor for *Epipactis* species originating above node IV. In some cases, the geographic origin of the allogamous sister sample of a monophyletic autogamous species allows inference of their approximate geographical origin. For example, the predominantly central European clade XVI has as sister samples *E. helleborine* s.s. accessions collected in Hungary, hinting at a central European origin of this clade that has since greatly expanded its geographic distribution as far as the British Isles. Similarly, the predominantly Balkanic species *E. greuteri* (species G) has sister lineages from exclusively Greek accessions of *E. helleborine* s.s. and the Greek infraspecific endemic *naousoensis* making up the monophyletic clade VIII (Fig. 3). We can conclude a southern Balkanic origin of this species, which then successfully colonised suitable habitats northward to central Germany.

Because in several cases the samples of a given autogamous species were collected from a geographically distinct locations but nonetheless formed a monophyletic group, we can also argue for a single origin of these species and for their subsequent spread. This ecological success is partly the result of being able to colonise environments where the outcrossing progenitor has a disadvantage because of the limited availability of pollinators (cf. 'reproductive assurance hypothesis': Wright *et al.*, 2013; Barrett *et al.*, 2014). For example, *E. pontica* is a typical plant of shaded beech-forests from northern Turkey to eastern Austria, where the most typical pollinating insects of *E. helleborine* s.s., wasps, are usually less frequent (Claessens

and Kleynen, 2011). Improved colonisation ability of self-pollinating plants has been reported in the literature (Baker, 1955; Wright *et al.*, 2013) and might have contributed significantly to the large geographic area that autogamous *Epipactis* species can occupy.

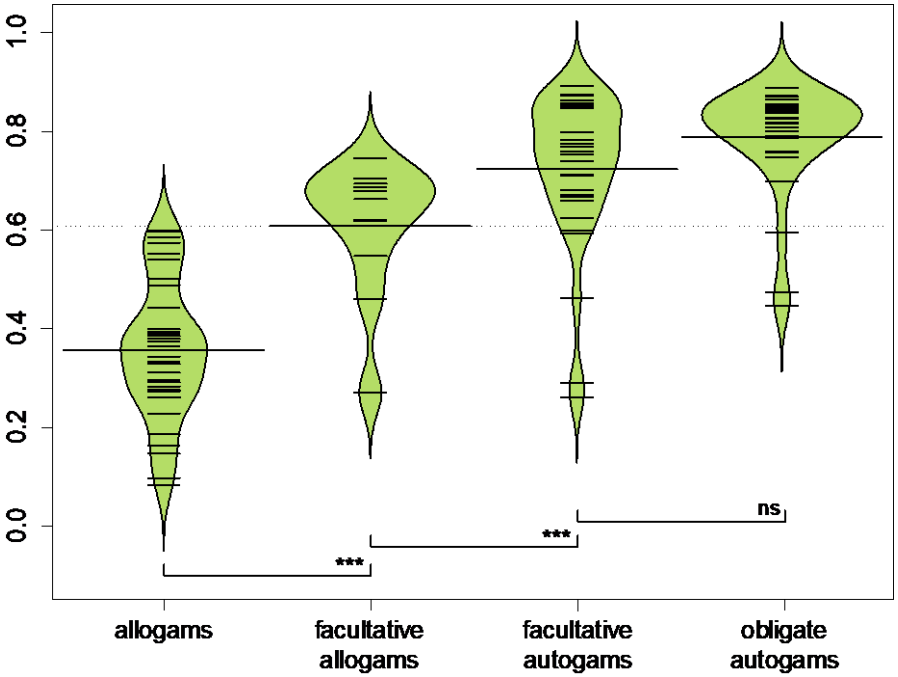


FIGURE 6. Beanplots of estimated inbreeding coefficient *F* for four hypothetical categories of breeding system in *Epipactis* section *Euepipactis.* The index *F* reported here is estimated per individual using a method of moments implemented in VCFtools. The significance of differences in distribution between categories calculated with a Mann-Whitney test is indicated as *** for $p < 0.001$ and as "ns" if $p > 0.05$.

Allogamy versus autogamy: transition or gradation?

We initially rather boldly divided the taxa in our analysis among four categories: obligate and facultative allogamy, and facultative and obligate autogamy. The inbreeding coefficients calculated by us yielded statistically significant differences between some prior categories but not between facultative and obligate autogams; also, all four categories registered notable outliers (Fig. 6). We are increasingly inclined to view these categories as gradational, though previous population genetic studies that encompassed numerous putative species (Squirrell *et al.*, 2002; Hollingsworth *et al.*, 2006) have suggested a profound contrast between species that are dominantly allogamous and those that are dominantly autogamous – the latter reliably show fixed homozygosity in allozyme profiles. Squirrell *et al.* (2002: 1963) further noted that "With each generation of complete selfing, homozygosity increases by 50%. In this fashion, a large genetic distance arises rapidly between progenitor and derivative species." Additionally,

Brys and Jacquemyn (2016) found that rapid evolution of selfing can be achieved by outbreeding depression acting within *Epipactis* populations, thus effectively isolating autogamous forms from outbreeders. We believe that this phenomenon is evident in the comparative branch lengths of our phylogeny; a much shorter molecular distance separates the root of the tree from individuals of the allogams *E. helleborine*, *E. lusitanica* and *E. atrorubens* than from the various autogams. Admittedly, it is also a longer distance to the morphologically allogamous *E. purpurata*, but allozyme data have shown this species to be more vulnerable to geitonogamy (Hollingsworth *et al.*, 2006).

Although some authors have argued that some species (including *E. helleborine*) are obligately allogamous, both crossing experiments (Tałałaj and Brzosko, 2008; Jacquemyn *et al.*, 2018) and some pollinator exclusion experiments (Brantjes, 1981; Ehlers *et al.*, 2002; Tałałaj and Brzosko, 2008) demonstrated that intrinsic sterility barriers facilitating self-incompatibility are absent across the genus. Given the elongate inflorescences bearing numerous large flowers, effective self-pollination via geitonogamy inevitably occurs frequently (Ehlers and Olesen, 1997; Kropf and Renner, 2008; Tałałaj and Brzosko, 2008; Claessens and Kleynen, 2016). Fruit-set percentages are predictably high in both the autogams and the allogams, as all species secrete considerable quantities of nectar and insect-attracting volatiles (Brodmann *et al.*, 2008).

Whether wholly obligate autogamy occurs is also questionable (Fig. 6). On the spectrum of allogamy to autogamy transition some of the supposed 'species' of section *Euepipactis*, particularly those based on a small number of geographically localised populations, are prone to cleistogamy; some have a proportion of cleistogamous populations (e.g., *E. greuteri*, *E. phyllanthes*), whereas *futakii* (placed within *E. leptochila* in Fig. 3) is predominantly cleistogamous and shows some morphological adaptations to cleistogamy (Fig. 1L). These taxa, showing the extreme end of the spectrum as near-obligately autogamous (Richards, 1982, 1986), appear on our tree as monophyletic entities terminating comparatively long branches. This apparently accelerated retention of mutations may reflect rapid isolation of autogamous species, as their separation from their genetic background (e.g., *E. leptochila* in case of *futakii*) is fairly recent in evolutionary terms. We doubt, however, that cleistogamy alone is adequate grounds for separation at the species level, as the appearance of cleistogamous individuals within dominantly autogamous species makes it difficult to accept cleistogamy (which is in any case generally imperfectly expressed) as a pre-eminent criterion for species delimitation. Multiple samples of most of the dominantly cleistogamous taxa such as *futakii* (within *E. leptochila*: Fig. 3) and *pseudopurpurata* (within *E. purpurata*) not only formed monophyletic

groups but also differed little if at all in genotype, despite the vast number of SNPs included in our analysis; these are expected properties of a lineage.

On the above reproductive spectrum, the primary morphological indictors of autogamy are said to be a reduced viscidium (a sphere of milky adhesive liquid surrounded by a viscidial membrane connected to the pollinarium with the function of adhering it to the pollinator) separating the pollinaria from the broad, adhesive stigmatic surface below and more friable pollinia capable of disaggregation (Richards, 1986; Claessens and Kleynen, 2011). Nonetheless, the relevance of environmental factors should not be under-estimated. The actual pollination syndrome can be influenced by floral ontogeny, environment, or both (Bateman, 2012b; Claessens and Kleynen, 2016). In particular, the quantity and effectiveness of the viscidial glue decreases during anthesis, as does the physical integrity of the pollinia. Thus, *Epipactis* flowers are well-adapted for a mixed breeding system that emphasises allogamy (or geitonogamy, providing *de facto* autogamy) early in anthesis but incurs an increased risk of autogamy with time. Warm temperatures, drying winds and/or low humidity can all conspire to accelerate this process (e.g., Claessens *et al.*, 1998; Pedersen and Ehlers, 2000), thereby causing contrasting frequencies of autogamy in successive flowering seasons. Consequently, some taxa that possess the suite of floral features regarded as being indicative of autogamy have in practice been shown to possess population genetic profiles typical of allogams: an example is the British *youngiana* (Hollingsworth *et al*., 2006).

Close morphological similarities make the confident identification of hybrids difficult in the absence of diagnostic molecular markers. Also, the difficulty of cultivating species of section *Euepipactis* (in contrast with Section *Arthrochila*) discourages attempts at artificial crossing. Nonetheless, hybrids between *bona fide* species within section *Euepipactis* are often reported in nature (Vlčko *et al.*, 2003; Batoušek, 2010). The limited genetic evidence of gene flow between species reliably indicates flow from allogams (typically *E. helleborine*) into autogams, including *E. dunensis tynensis* and *E. phyllanthes* (Hollingsworth *et al.*, 2006), but such hybridisation has also been inferred through morphological means in the case of *E. purpurata* × *albensis* (Jakubska-Busse *et al.*, 2017).

Our shared ancestry analysis carried out using fineRADstructure (Fig. 3) offers a direct test of possible hybridisation events; SNPs shared between samples at the tips of the phylogenetic tree are indicated on the ‘heat map’. Higher levels of shared SNPs located close to the left-hand margin of this heat map indicate higher levels of shared ancestry between phylogenetically closely related entities – an expectation if lineages are phylogenetically well-isolated

from the rest of the samples. In contrast, shared ancestry deeper in the heat map should represent recent gene flow between phylogenetically distant entities.

Although we deliberately excluded from our sample set any suspected hybrids, it is nonetheless reassuring to see little evidence of gene-flow in the heat map (Fig. 3). The highest levels of shared SNPs are indicated for the autogamous species *E. phyllanthes* (species D), *E. greuteri* (G), *E. pontica* (H) and *E. muelleri* (I). Less pronounced elevated levels of shared ancestry are evident in other autogamous lineages, including *E. dunensis* (species J), *E. albensis* (K) and *E. leptochila* (F). These findings, plus the relatively homogeneous gene pools revealed by the admixture results (Fig. 5), indicate a comparatively cohesive genetic background of the autogamous lineages, a conclusion that accords with previous population genetic studies (Ehlers and Pedersen, 2000; Pedersen and Ehlers, 2000; Squirrell *et al.*, 2002; Hollingsworth *et al.*, 2006). The lack of significant gene flow from *E. helleborine* towards the predominantly autogamous species does not support the 'evolutionary detour' hypothesis in the genus formulated by Hollingsworth et al. (2006).

There are nonetheless some examples of possible gene-flow evident in the fineRADstructure analysis (Fig. 3). The most significant is an indication of higher coancestry than expected between *nauosaensis* (sample 251) and a typical *E. helleborine* (sample 258) from the same mountain in Greece. The second-most localised sign of shared coancestry involves a German sample of *E. leptochila* s.s. (sample 334) and a Turkish sample of *E. phyllanthes* (sample 393), followed by sharing of SNPs between a British *E. helleborine* (sample 344) and French *E. dunensis* (sample 465).

Indications of gene-flow between phylogenetically distantly related samples are otherwise absent from our dataset. This comes as a surprise, as sample 348 of the present study represents the controversial Risborough population in England that has variously been suggested as being *E. leptochila leptochila*, *E. leptochila neglecta*, or *E. leptochila* introgressed by *E. helleborine*. The strongly contrasting phylogenetic placement of this plant in Figs. 3 and 4, and its location well below the remaining 17 samples of *E. leptochila* in Fig. 3, may indicate that introgression hypothesis is the most likely, but there is little evidence of gene flow in the shared ancestry analysis. A possible explanation of this finding is that a genotype closely similar to the Risborough population of *E. helleborine* is not represented in our dataset, so the method cannot possibly find traces of gene flow. Whatever the explanation for the actual gene flow patterns reported here for given samples, the shared ancestry analysis detected only limited and localised gene flow between phylogenetically distant samples in section *Euepipactis*.

However, local-scale studies indicated that even a modest amount of gene-flow between dominantly allogamous and autogamous species should be sufficient to maintain a degree of heterozygosity within the autogams (Durka, 2002; Hollingsworth *et al.*, 2006). An additional point made by Claessens and Kleynen (2011) is that airborne pollinators (most commonly wasps in most of the allogams) are by no means the only insects that routinely inhabit *Epipactis* plants, and that the friable nature of the pollinia can in practice permit transfer of pollen tetrads by much smaller insects. Indeed, friability could also lead to mixed matings, pollen tubes from different source plants competing to reach the several thousand ovules held within a particular *Epipactis* ovary.

Clearly, the transition from allogamy to autogamy is far more complex and labile than a simple binary switch. However, our data support the view that species with a tendency to be autogamous are – despite occasional allogamous pollination or hybridisation – isolated from the genetic background provided by *E. helleborine*. If the switch to autogamy happened relatively long in the evolutionary history of the genus, the autogamous lineages might have gained severe genetic isolation from their genetically rather polymorphic progenitor species.

Does speciation in the* E. helleborine *alliance constitute an active evolutionary radiation?

After reviewing in detail prior definitions and underlying concepts, Bateman (1999: 441) defined an evolutionary radiation as "a large surplus in the rate of natality over the rate of mortality for species and/or character states within a specified clade through a specified time interval." Given the extreme paucity of fossil orchids – none are definitively known from tribe Neottieae, though the amber-entombed Eocene pollinarium *Succinanthera* bears some resemblance (Poinar and Rasmussen, 2017) – we are poorly placed to comment on the rate of species mortality within *Epipactis* section *Euepipactis*. This means that we cannot conclusively differentiate between a genuine radiation and a lineage that simply indulges in an unusually high rate of species turnover.

However, we do now possess conclusive evidence that section *Euepipactis* constitutes a clade, that all of the speciation events within the section have occurred comparatively recently, that each such event is underpinned by a molecular branch considerably longer than those within the group, and that several unequivocal species have emerged from within a single ancestral species, *E. helleborine* s.s. This impression is further enhanced by the NeighbourNet analysis (Fig. 4), which clearly depicts comparatively long-branch clusters of samples, each cluster constituting a single re-circumscribed species, radiating outward from a central core of comparatively short-branch allogams. Thus, we have in place most of the features that we would hope to see in a genuine radiation. And the fact that we can identify examples of incip-

ient speciation that are employing the same evolutionary mechanisms as the lineages that have already achieved speciation strongly suggests that we could legitimately brand this radiation as ongoing and thus active.

When used in its evolutionary context, the term radiation is rarely widely separated from the term adaptive (Bateman, 1999). Referring to the drivers that allow the transition from allogamy to autogamy in the genus – reduced size of rostellum and increased friability of pollinium – these represent a combination of developmental genetic, epigenetic and ecophenotypic influences. The highly iterative trend toward autogamy could be taken as evidence of prolonged selection pressure, but equally it could be viewed simply as weak control of rostellum development leading to genetic drift (Hollingsworth *et al*., 2006) combined with environmentally-driven desiccation of the pollinarium (R. Bateman, unpublished obs.).

In the case of the sister-pairing of re-circumscribed autogamous species, all of the named taxa now encompassed by *E. albensis* flower later than those encompassed by *E. dunensis*, and most also prefer less alkaline soils. Given that many Neottieae species have been shown to maintain contrasting yet broadly predictable mycorrhizal communities (e.g., Bidartondo *et al.*, 2004; Selosse *et al.*, 2004; Jacquemyn *et al.*, 2016; Schiebold *et al.*, 2017; Jacquemyn *et al.*, 2018), it seems likely to one of us (RMB) that mycorrhizal switching may have enabled colonisation of contrasting soils and habitats, thereby at least assisting – if not driving – speciation (Bateman, 2012b). Certainly, mycorrhizally mediated nutrient sequestration is greater in species that preferentially inhabit shadier woodlands (Schiebold *et al.*, 2017).

Shady woodlands with comparatively impoverished ground floras are the preferred habitats of most of the autogams. This observation has led some authors to conclude that the autogams are adapted to these habitats because insect faunas are likely to be similarly impoverished and hence pollinator visits are likely to be less frequent (e.g., Claessens and Kleynen, 2011, 2016). In light of our data presented above, the frequency of successful pollination events must be of little if any evolutionary or ecological significance, given the huge number of seeds produced by each fertilised *Epipactis* capsule; only if pollination fails completely through a considerable period of time will the population be under serious threat of extirpation through non-renewal.

CONCLUSIONS

It is notoriously difficult to identify with any confidence ancestor–descendant relationships (rather than the less informative sister-species relationships) and to identify with adequate rigour an evolutionary radiation. Nonetheless, it is highly probable that *E. helleborine* is

the ancestor of at least ten recently derived species, the majority of them near-obligate autogams. Genuine speciation events can derive one autogamous species from another, though the majority of autogamous species are likely to prove to be evolutionary dead-ends. As noted by Claessens & Kleynen (2011), comparison of the dominantly allogamous species versus the dominantly autogamous species within section *Euepipactis* does not reveal any great disadvantage incurred by the latter, despite the theoretical ravages of inbreeding depression. Excepting the ancestral plexus that is *E. helleborine*, the average distributional areas or population sizes do not differ greatly between the allogams and the autogams. However, *E. helleborine* does occupy a greater range of phylogenetic space (Figs. 3–5), suggesting that the combination of its greater generic diversity and its predominance of allogamous populations allows it to function far more successfully as a source of further novel species.

The vast majority of the (usually more geographically localised) taxa awarded species status by some systematists actually merit the rank of subspecies, variety or forma. The great frequency and rapidity with which populations that are increasingly oriented toward autogamy can emerge from within dominantly autogamous species, and the difficulty of using floral morphology alone to determine within species the comparative frequency of allogamy versus autogamy, together dissuade us from placing too much taxonomic emphasis on perceived breeding system. One of the many advantages of acquiring genetic data is their ability to summarise the recent breeding history of the populations under scrutiny.

Although *Epipactis* section *Euepipactis* has become an increasingly well-understood model system, several disparate factors contribute to speciation within the section. The comparative significance of genetic versus epigenetic or ecophenotypic influences on phenotype, preadaptation to adaptation, selection versus drift or saltation, and of partnerships with pollinators versus those with mycorrhizae, all remain open for further informed discussion.

FUNDING

This work was supported by the Hungarian Scientific Research Fund (OTKA PD109686 to GS), the Austrian Science Fund (FWF Y661-B16 to OP).

ACKNOWLEDGEMENTS

We are grateful to many colleagues who helped us with field sampling: Kath Barrett, Philip Cribb, Miklós Csábi, Ian Denholm, Matthias Fiedler, Olivier Gerbaud, Gergely Gulyás, Ernst Gügel, Stefan and Kerstin Hertel, Dirk W. Kapteyn den Boumeester, Balázs András Lukács, Gábor Magos, Richard Manuel, Pavel Mered'a junior and senior, Miklós Óvári, Ian

Phillips, Paula Rudall, András Schmotzer, Jess Stone, József Sulyok, Polina A. Volkova and Wolfgang Wucherpfennig.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

Anisimova M, Gascuel O. 2006. Approximate Likelihood-Ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology* 55: 539–552.

Baird NA, Etter PD, Atwood TS, *et al.* 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.

Baker HG. 1955. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* 9: 347–349.

Barrett SCH, Arunkumar R, Wright SI. 2014. The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.

Bateman RM. 1999. Integrating molecular and morphological evidence for evolutionary radiations. In: Hollingsworth PM, Bateman RM, Gornall RJ, eds. *Molecular systematics and plant evolution*. London: Taylor & Francis.

Bateman RM. 2009. Evolutionary classification of European orchids: the crucial importance of maximising explicit evidence and minimising authoritarian speculation. *Journal Europäischer Orchideen* 41: 243–318.

Bateman RM. 2012a. Circumscribing genera in the European orchid flora: a subjective critique of recent contributions. *Berichte aus den Arbeitskreisen heimische Orchideen* Beiheft 8: 94–126.

Bateman RM. 2012b. Circumscribing species in the European orchid flora: multiple datasets interpreted in the context of speciation mechanisms. *Berichte aus den Arbeitskreisen Heimische Orchideen* Beiheft 8: 160–212.

Bateman RM, Crane PR, DiMichele WA, *et al.* 1998. Early evolution of land plants: Phylogeny, physiology, and ecology of the primary terrestrial radiation. *Annual Review of Ecology and Systematics* 29: 263-292.

Bateman RM, Hollingsworth PM, Preston J, Yi-Bo L, Pridgeon AM, Chase MW. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society* 142: 1–40.

Bateman RM, Hollingsworth PM, Squirrell J, Hollingsworth ML. 2005. Tribe Neottieae. Phylogenetics. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, eds. *Genera Orchidacearum. Volume 4. Epidendroideae (Part One)*. Oxford: Oxford University Press.

Bateman RM, Sramkó G, Paun O. 2018. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. *Annals of Botany* 121: 85–105.

Batoušek P. 2010. Drei neue Bastarde der Gattung *Epipactis* Zinn. (Orchidaceae). *Journal Europäischer Orchideen* 41: 507-518.

Bidartondo MI, Burghardt B, Gebauer G, Bruns TD, Read DJ. 2004. Changing partners in the dark: isotopic and molecular evidence of ectomycorrhizal liaisons between forest orchids and trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271: 1799-1806.

Brantjes NBM. 1981. Ant, bee and fly pollination in *Epipactis palustris* (L.) Crantz (Orchidaceae). *Acta Botanica Neerlandica* 30: 59-68.
Brodmann J, Twele R, Francke W, Hölzler G, Zhang Q-H, Ayasse M. 2008. Orchids mimic green-leaf volatiles to attract prey-hunting wasps for pollination. *Current Biology* 18: 740-744.
Brys R, Jacquemyn H. 2016. Severe outbreeding and inbreeding depression maintain mating system differentiation in *Epipactis* (Orchidaceae). *Journal of Evolutionary Biology* 29: 352-359.
Burns-Balogh P, Szlachetko DL, Dafni A. 1987. Evolution, pollination, and systematics of the tribe *Neottieae* (Orchidaceae). *Plant Systematics and Evolution* 156: 91-115.
Brandrud, M. K., Baar, J., Lorenzo, M. T., Bateman, R. M., Chase, M. W., Paun, O., Hedrén, M. submitted – Chapter I. Phylogenomic relationships of diploids and the origin of allotetraploids in *Dactylorhiza* (Orchidaceae): RADseq data track reticulate evolution.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics* 1: 171–182.
Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* 18: 237-268.
Claessens J, Kleynen J. 2011. *The flower of the European orchid. Form and function*: Jean Claessens & Jacques Kleynen.
Claessens J, Kleynen J. 2016. Many ways to get happy: pollination modes of European *Epipactis* species. *Orchid Digest* 80: 142–150.
Claessens J, Kleynen J, Wiellnga R. 1998. Some notes on *Epipactis helleborine* (L.) Crantz ssp. *neerlandica* (Vermeulen) Buttler and *Epipactis renzii* K. Robatsch. *Eurorchis* 10: 55-64.
Danecek P, Auton A, Abecasis G, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
Delforge P. 2016. *Orchidées d'Europe, d'Afrique du Nord et du Proche-Orient. 4e édition revue et augmentée.* Paris: Delachaux et Niestlé.
Drouzas AD, Charitonidou M, Tsiftsis S. 2017. Chloroplast DNA variation in *Epipactis atrorubens* populations from northern Greece. *Botany Letters* 164: 55-62.
Durka W. 2002. Blüten- und Reproduktionsbiologie [Floral and reproductive biology]. *Schriftenreihe für Vegetationskunde* 38: 133–175.
Eaton DAR, Ree RH. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.
Efimov P. 2004. Genus *Epipactis* Zinn. (Orchidaceae) in Russia [in Russian]. *Turczaninowia* 7: 8-42.
Ehlers BK, Olesen JM. 1997. The fruit-wasp route to toxic nectar in *Epipactis* orchids? *Flora* 192: 223-229.
Ehlers BK, Olesen JM, Agren J. 2002. Floral morphology and reproductive success in the orchid *Epipactis helleborine*: regional and local across-habitat variation. *Plant Systematics and Evolution* 236: 19-32.
Ehlers BK, Pedersen HAE. 2000. Genetic variation in three species of *Epipactis* (Orchidaceae): geographic scale and evolutionary inferences. *Biological Journal of the Linnean Society* 69: 411-430.
Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology* 14: 2611–2620.

Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* 36: 47-79.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.
Harris SA, Abbott RJ. 1997. Isozyme analysis of the reported origin of a new hybrid orchid species, *Epipactis youngiana* (Young's helleborine), in the British Isles. *Heredity* 79: 402-407.
Harrison N, Kidner CA. 2011. Next-generation sequencing and systematics: What can a billion base pairs of DNA sequence data do for you? *Taxon* 60: 1552-1566.
Heckenhauer J, Samuel R, Ashton PS, Abu Salim K, Paun O. 2018. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe Shoreeae (Dipterocarpaceae). *Molecular Phylogenetics and Evolution* 127: 1–13.
Hollingsworth PM, Squirrell J, Hollingsworth ML, Richards AJ, Bateman RM. 2006. Taxonomic complexity, conservation and recurrent origins of self-pollination in *Epipactis* (Orchidaceae). In: Bailey J, Ellis RG, eds. *Current taxonomic research on the British & European flora*. London: BSBI.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254-267.
Igic B, Busch JW. 2013. Is self-fertilization an evolutionary dead end? *New Phytologist* 198: 386–397.
Jacquemyn H, Kort HD, Broeck AV, Brys R. 2018. Immigrant and extrinsic hybrid seed inviability contribute to reproductive isolation between forest and dune ecotypes of *Epipactis helleborine* (Orchidaceae). *Oikos* 127: 73–84.
Jacquemyn H, Waud M, Lievens B, Brys R. 2016. Differences in mycorrhizal communities between *Epipactis palustris*, *E. helleborine* and its presumed sister species *E. neerlandica*. *Annals of Botany* 118: 105–114.
Jakubska-Busse A, Żołubak E, Łobas Z. 2017. *Epipactis* × *subtilis* (Orchidaceae), a new hybrid between *E. albensis* and *E. purpurata*. *Annales Botanici Fennici* 54: 139-144.
Julou T, Burghardt B, Gebauer G, Berveiller D, Damesin C, Selosse M-A. 2005. Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytologist* 166: 639-653.
Kampstra P. 2008. Beanplot: a boxplot alternative for visual comparison of distributions. *Journal of Statistical Software* 28: 1–9.
Kircher M, Kelso J. 2010. High-throughput DNA sequencing–Concepts and limitations. *BioEssays* 32: 524–536.
Kolanowska M. 2013. Niche conservatism and the future potential range of *Epipactis helleborine* (Orchidaceae). *PLOS ONE* 8: e77352.
Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* 15: 356.
Kreutz CAJ. 2004. *Catalogue of European orchids.* Landgraaf: Kreutz Publisher.
Kropf M, Renner S. 2008. Pollinator-mediated selfing in two deceptive orchids and a review of pollinium tracking studies addressing geitonogamy. *Oecologia* 155: 497–508.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357.
Lefort V, Longueville J-E, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* 34: 2422–2424.

Leitch IJ, Kahandawala I, Suda J, *et al.* 2009. Genome size diversity in orchids: consequences and evolution. *Annals of Botany* 104: 469-481.
Light MHS, MacConaill M. 2006. Appearance and disappearance of a weedy orchid,Epipactis helleborine. *Folia Geobotanica* 41: 77-93.
Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298-299.
Malinsky M, Trucchi E, Lawson D, Falush D. 2017. RADpainter and fineRADstructure: population inference from RADseq data. *bioRxiv*.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.
Olson PD, Hughes J, Cotton JA. 2016. Next Generation Systematics. *Systematics Association Special Volume Series*. Cambridge: Cambridge University Press.
Pante E, Abdelkrim J, Viricel A, *et al.* 2015. Use of RAD sequencing for delimiting species. *Heredity* 114: 450–459.
Paun O, Turner B, Trucchi E, Munzinger J, Chase MW, Samuel R. 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros* , Ebenaceae) in New Caledonia, a biodiversity hotspot. *Systematic Biology* 65: 212-227.
Pedersen HAE, Ehlers BK. 2000. Local evolution of obligate autogamy in *Epipactis helleborine* subsp. *neerlandica* (Orchidaceae). *Plant Systematics and Evolution* 223: 173–183.
Poinar JRG, Rasmussen FN. 2017. Orchids from the past, with a new species in Baltic amber. *Botanical Journal of the Linnean Society* 183: 327–333.
Pridgeon AM, Light MHS. 2005. Tribe Neottieae. Pollination. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, eds. *Genera Orchidacearum. Volume 4. Epidendroideae (Part One)*. Oxford: Oxford University Press.
Proctor MCF, Yeo P. 1973. *The pollination of flowers.* London: Collins.
Ren G, Mateo RG, Liu J, *et al.* 2017. Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing. *New Phytologist* 213: 1500–1512.
Rewicz A, Rewers M, Jedrzejczyk I, Rewicz T, Kolodziejek J, Jajubska-Busse A. 2018. Morphology and genome size of *Epipactis helleborine* (L.) Crantz (Orchidaceae) growing in anthropogenic and natural habitats. *PeerJ* 6: e5992.
Richards AJ. 1982. The influence of minor structural changes in the flower on breeding systems and speciation in *Epipactis* Zinn. (Orchidaceae). In: Armstrong JA, Powell JM, Richards AJ, eds. *Pollination and evolution*. Sydney: Royal Botanical Gardens.
Richards AJ. 1986. *Plant breeding systems.* London: Chapman and Hall.
Robatsch K. 1995. Beiträge zur Kenntnis des europäischen *Epipactis*-Arten (Orchidaceae) und zur Evolution der Autogamie bei europäischen und asiatischen Gattungen der Neottioideae. *Journal Europäischen Orchideen* 27: 125–177.
Rubin BER, Ree RH, Moreau CS. 2012. Inferring phylogenies from RAD sequence data. *PLoS ONE* 7: e33394.
Schiebold JMI, Bidartondo MI, Karasch P, Gravendeel B, Gebauer G. 2017. You are what you get from your fungi: nitrogen stable isotope patterns in *Epipactis* species. *Annals of Botany* 119: 1085-1095.
Schloss JA. 2008. How to get genomes at one ten-thousandth the cost. *Nature Biotechnology* 26: 1113.
Selosse M-A, Faccio A, Scappaticci G, Bonfante P. 2004. Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are

associated with ectomycorrhizal septomycetes, including truffles. *Microbial Ecology* 47: 416-426.
Squirrell J, Hollingsworth PM, Bateman RM, *et al.* 2001. Partitioning and diversity of nuclear and organelle markers in native and introduced populations of *Epipactis helleborine* (Orchidaceae). *American Journal of Botany* 88: 1409–1418.
Squirrell J, Hollingsworth PM, Bateman RM, Tebbitt MC, Hollingsworth ML. 2002. Taxonomic complexity and breeding system transitions: conservation genetics of the *Epipactis leptochila* complex (Orchidaceae). *Molecular Ecology* 11: 1957–1964.
Sramkó G, Molnár V. A, Hawkins JA, Bateman RM. 2014. Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus *Himantoglossum* s.l. (Orchidaceae). *Annals of Botany* 114: 1609–1626.
Stebbins GL. 1957. Self fertilization and population variability in the higher plants. *American Naturalist* 91: 337–354.
Stebbins GL. 1970. Adaptive radiation of reproductive characteristics in Angiosperms. I: pollination mechanisms. *Annual Review of Ecology and Systematics* 1: 307–326.
Takebayashi N, Morrell PL. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany* 88: 1143–1150.
Tałałaj I, Brzosko E. 2008. Selfing potential in *Epipactis palustris*, *E. helleborine* and *E. atrorubens* (Orchidaceae). *Plant Systematics and Evolution* 276: 21-29.
Tranchida-Lombardo V, Cafasso D, Cristaudo A, Cozzolino S. 2011. Phylogeographic patterns, genetic affinities and morphological differentiation between *Epipactis helleborine* and related lineages in a Mediterranean glacial refugium. *Annals of Botany* 107: 427-436.
Trucchi E, Frajman B, Haverkamp THA, Schönswetter P, Paun O. 2017. Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). *New Phytologist* 216: 267-278.
Tyteca D, Dufrene M. 1994. Biostatistical studies of western European allogamous populations of the *Epipactis helleborine* (L.) Crantz species group (Orchidaceae). *Systematic Botany* 19: 424-442.
Vlčko J, Dítě D, Kolník M. 2003. *Orchids of Slovakia.* Zvolen: ZO SZOPOK Orchidea.
Wood JJ. 2005. 387. *Epipactis*. Distribution. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN, eds. *Genera Orchidacearum. Volume 4. Epidendroideae (Part One)*. Oxford: Oxford University Press.
Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B: Biological Sciences* 280.
Zhou T, Jin X-h. 2018. Molecular systematics and the evolution of mycoheterotrophy of tribe *Neottieae* (Orchidaceae, Epidendroideae). *PhytoKeys* 94: 39-49.

SUPPLEMENTARY MATERIAL

TABLE S1

Species of Epipactis sampled for this study. The general distribution, regions and populations sampled, pollination mode and abbreviated name in analyses are given for each taxon. Distribution data and pollination mode information are from the comprehensive online monograph of the association "Arbeitskreis Heimische Orchideen Bayern e.V.", which treats all named taxa as species and bases pollination mode on personal observations without making judgements regarding their biological validity. Available at: http://www.aho-bayern.de/epipactis/fs_epipactis_1.html

Species	Distribution (countries*)	Sampled countries*	No. of studied populations	Pollination mode	Abbreviation in analyses
Epipactis albensis	AT,CZ,DE,HU,PL,RO,SK	HU,RO	4	obligate autogam	alben
Epipactis atrorubens	throughout Europe & TR	SK,LT	2	facultative allogam	atror
Epipactis bugacensis	HU	HU	2	obligate autogam	bugac
Epipactis distans	AT,CZ,DE,FR,HR,HU,IT,LT,PL,SI,SK	AT,BG,GR,RS	4	facultative allogam	dista
Epipactis dunensis	GB	GB	2	facultative autogam	dunen
Epipactis exilis	BG,ES,FR,GR,HR,HU,IT,RO,SK	BG,GR,HU	4	facultative autogam	exili
Epipactis futakii	HU,SK	HU,SK	3	obligate autogam	futak
Epipactis greuteri	AT,BG,CZ,DE,GR,HR,HU,RO,SI,SK	AT,GR,SK	3	obligate autogam	greut
Epipactis helleborine s.s.	throughout Europe & TR	AT,BE,BG,CH,DE,DK,FR,GB,GR,HU,RO,RS,RU,SK,TR	29	allogam	helle
Epipactis leptochila	throughout Europe & TR	DE,GB,HU,RU,SI	5	facultative autogam	lepto
Epipactis lusitanica	ES,PT	PT	1	allogam	lusit
Epipactis mecsekensis	HU	HU	1	facultative autogam	mecse
Epipactis muelleri	AT,CZ,DE,ES,FR,HR,HU,IT,PL,SI,SK	DE,HU,SI,SK	4	obligate autogam	muell
Epipactis naousoensis	IT,GR	GR	1	facultative autogam	naous
Epipactis neerlandica	BE,DE,DK,FR,GB,NL	DK,GB,NL	3	allogam	neerl
Epipactis neglecta	AT,BE,CH,CZ,DE,FR,GB,GR,HR,HU,IT,ME,RS,SI,SK	DE,GB,HU,RO	5	facultative autogam	negle
Epipactis nordeniorum	AT,HR,HU,SI	HU	2	facultative autogam	norde
Epipactis palustris	throughout Europe & TR	NL	1	allogam	palus
Epipactis peitzii	DE,HU	DE,HU	4	obligate autogam	peitz
Epipactis persica	RO,TR	TR	2	facultative autogam	pers

Epipactis phyllanthes	BE,DK,ES,FR,GB,NL	BE,GB	3	obligate autogam	phyll
Epipactis pontica	AT,BG,CZ,GE,GR,HR,HU, IT,RO,SI,SK,TR	BG,HU,RO,SK	6	facultative autogam	ponti
Epipactis pseudopurpurata	HU,SK	HU,SK	3	obligate autogam	psepu
Epipactis purpurata	whole Europe	BG,GB,HU,SK	4	allogam	purpu
Epipactis renzii	DK	DK	1	obligate autogam	renzi
Epipactis rhodanensis	AT,CH,DE,FR,IT	AT,DE,FR	3	obligate autogam	rhoda
Epipactis tallosii	CZ,HU,IT,RO,SK	HU	2	facultative autogam	tallo
Epipactis tynensis	GB	GB	1	facultative autogam	tynen
Epipactis veratrifolia	Middle East	CY	1	allogam	verat
Epipactis voethii	AT,CZ,HU,SK	HU,SK	2	facultative autogam	voeth

* Country codes are given according to their ISO standard

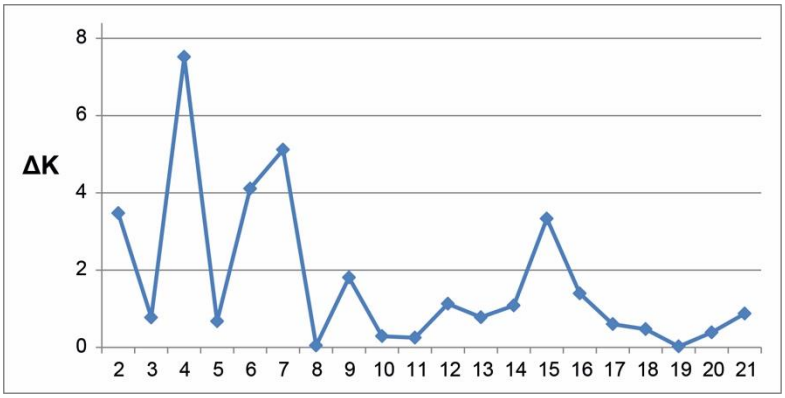


FIGURE S1: Plot of the *ad hoc* statistic deltaK proposed by Evanno *et al.* (2005) to identify the optimal number of gene pools within genetic data.

THESIS CONCLUSION

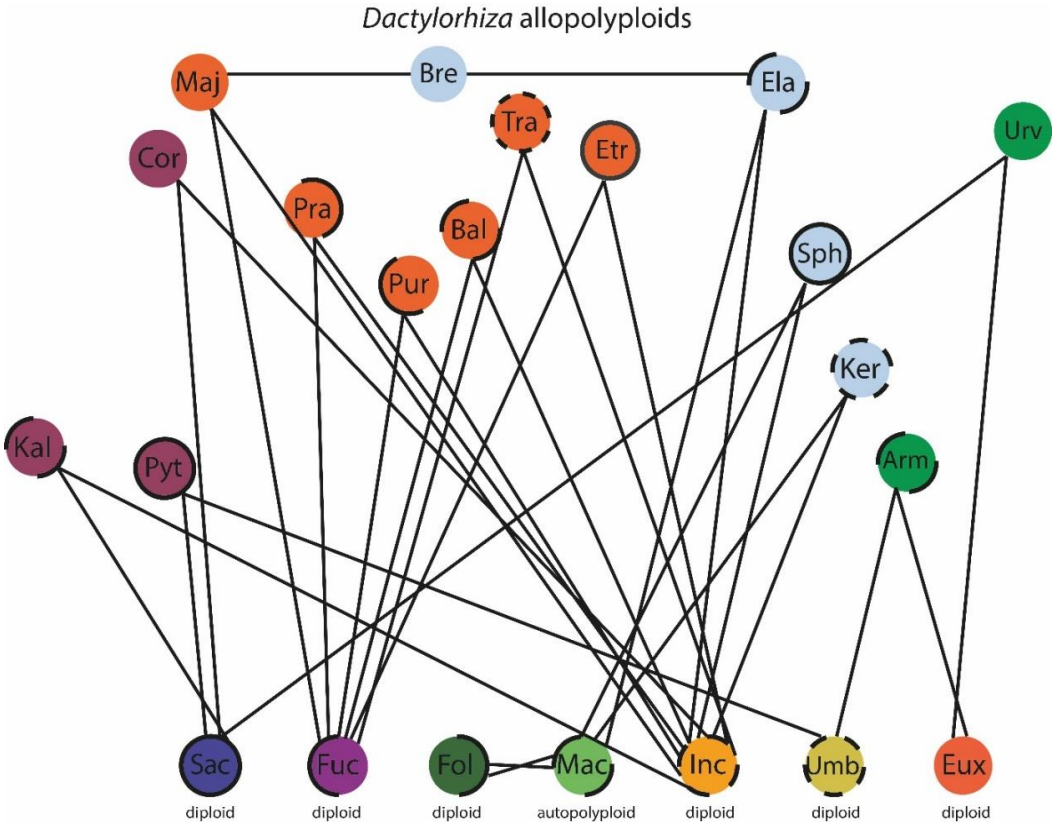
RADseq has successfully been applied to four orchid genera to investigate different aspects of their evolutionary history. Diploid *Dactylorhiza* formed a well-supported clade with ten species recognized in Europe and adjacent Asia. *Dactylorhiza euxina*, *D. umbrosa* and *D. incarnata* belong to the *D. incarnata-euxina* clade and *D. saccifera*, *D. fuchsii*, *D. foliosa* as well as *D. maculata* (the autopolyploid of *D. foliosa*) belong to the *D.fuchsii-maculata.* These two clades coalesce between six and five million years ago, further colalesce with *D. sambucina* and *D. iberica* between seven and six million years ago and lastly coalesce with *D. viridis* between nine and eight million years ago. *Dactylorhiza viridis* is supported to be included in *Dactylorhiza*, where it earlier has been assigned to its own genus *Coeloglossum viride*. Of the outgroups used *Gymnadenia* and *Nigritella* came out as the closest sister group to *Dactylorhiza*, followed by *Galearis*, *Pseudorchis* and *Platanthera*.

The members in both the *D. fuchsii-maculata* clade and the *D. incarnata-euxina* clade have different ages and distributions. Crosses between members from the two clades have recurrently produced allopolyploids with resulting different ages and distributions as well. The member from the *D.fuchsii-maculata* clade always acted as the maternal and the member from the *D.incarnata-euxina* clade acted as the paternal to produce the allopolyploids. Even though it is expected that certain features of the genome need to follow a specific path, there are also much of the genome that can evolve stochastically to produce this array of allopolyploids differing in morphological and ecology. The number of private alleles was used to infer the relative age of the polyploids. The oldest allopolyploids are found in southern and eastern Europe, whereas the youngest allopolyploids are found in northern and western Europe.

Two of the allopolyploid originating from *D. fuchsii* and *D. incarnata* were studied in further detail. The two genomes combined in the allopolyploids are very different in the way that *D. fuchsii* have much variation within the populations, but little variation between, which is probably a reflection of its comparatively wide ecology and distribution. On the other hand, *D. incarnata* has populations with little variation, but more variation found between, which is also reflecting its narrow ecology and fragmented/patchy distribution. Because of all the variation within the population it proved more difficult to make a good (synthetic) reference of *D. fuchsii* than *D. incarnata*, but the final references were able to distinguish the reads back to the original subgenome with an accuracy of 77%. Two bioinformatic pipelines were applied

in parallel to analyse the allopolyploids with RADseq data. One pipeline was optimized to distinguish the loci originating from the two subgenomes and the other pipeline was optimized to capture the loci that could not be distinguished to the specific subgenomes and allowing for partial heterozygotes, which is a result of having more allele copies that two. The application of two bioinformatic pipelines and application of a (synthetic) genome reference should produce robust data if consistent. We found the results from the two pipelines to be well in agreement.

Overall less structure is found in the diploids compared to the polyploids and local origins from regional diploid parents could not be inferred. One origin each for *D. majalis* and *D. traunsteineri* are postulated. *Dactylorhiza majalis* is homogenous and the distance seen between material from the western Europe and the material from central and eastern Europe is likely originated from isolation by distance. *Dactylorhiza majalis* is estimated to be at least ~10.000 years old. *D. traunsteineri* is a lot more heterogenous and the large distance seen between material from northwestern Europe and material from central and eastern Europe is also likely originated from isolation by distance. *Dactylorhiza traunsteineri* is younger; estimated to be at least ~5.000 years old and it is likely to have originated after the last Pleistocene ice age. In the areas where the allopolyploids are now occurring in sympatry they hybridize and exchange genetic material.



The closest European sister taxa to *Dactylorhiza* is found to be a clade containing *Gymnadenia* and *Nigritella.* Analysed more in detail it is supported phylogenetically to treat *Gymnadenia* and *Nigritella* as separate genera, as they both form a monophyletic clade each. A similar approach used to infer the parental species of the allopolyploids in *Dactylorhiza* was applied for the intergeneric *Gymnigritella runei* and was found most likely to have originated from diploid, early flowering, *Gymnadenia conopsea* s.str. and polyploid *Nigritella nigra.* In the *Nigritella* polyploids, as might be expected in apomixe groups, they come out as clearly defined, being in an almost stagnating situation evolutionary. On the other hand, high levels of heterozygosity are observed (at least for *N. widderi* and *N. miniata*), which support that they arose through a hybrid origin. One would predict that apomoxis might limit the evolutionary potential and that they could be suffering from inbreeding depression in the long run. The polyploids are, however, not obligate apomixe, and might still exchange genetic material.

Eleven *Epipactis* species were recognized as *bon fide* species and the rest was regarded as infraspecific variation. Altogether seven of these species had dominantly autogamy and four had dominantly allogamy. All were found to be monophyletic, except allogamous *E. helleborine* which was found distributed throughout the tree and at the base of most autogamous lineages. From this it is concluded that the transition from allogamy to autogamy is found repeatedly in *Epipactis*, whereas no examples of autogame species forming new autogame or allogame species are observed. Autogamy might, if obligatory, be interpreted as an evolutionary dead-end. However, just one rare event of cross-fertilization might be enough to get away from stagnation and inbreeding depression. Inbreeding coefficient seems to be a good indicator of reproductive mode in *Epipactis*. As observed in the inbreeding coefficients there is a gradual increase in the values from allogame, facultative allogame, facultative autogame to obligate autogame, though non-significant distributions between facultative autogame to obligate autogame. It is found that autogams are as widespread and ecologically successful as allogams.

APPENDIX

Conference Contributions

International Conference on Polyploidy, Hybridization and Biodiversity, Rovinj, Croatia, 11-14th May 2016

Using RAD-seq as a genome-wide approach to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

Marie K. Brandrud[1], Juliane Baar[1], Maria T. Lorenzo[1], Mikael Hedrén[2], Emiliano Trucchi[1], Ovidiu Paun[1]

[1]Department of Systematic Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria [2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden

Poster Presentation

Genomic inferences based on RAD-seq (i.e. restriction-site associated DNA sequencing) data have recently started to be widely applied to answer evolutionary questions in diploid organisms. However, difficulties in distinguishing between homoeologs and paralogs in duplicated non-model genomes have made it challenging to use RAD-seq to answer questions related to polyploid evolution. We describe here a reliable RAD-seq bioinformatics pipeline that integrates information obtained from parental diploid genomes in order to analyse polyploid genomic data. We further apply it to a dataset containing 184 representatives of several allopolyploid *Dactylorhiza* species that have originated independently from the same parents and overlap in distribution, but have different ecological preferences and morphology. We first assemble an artificial genomic reference based on non-overlapping RAD loci from the diploid parental species, which are expected to present a disomic inheritance in the allopolyploid genomes. We further implement and compare two pipelines (i.e., STACKS vs. GATK) for aligning the polyploid reads to this reference and calling/filtering SNPs. Based on the 5,000+ RAD loci obtained, we finally investigate the patterns of genetic divergence and gene flow dynamics, and formulate hypotheses regarding the polyploid origins across 35 populations sampled in Europe, over the distribution range of the *D. majalis* complex. Our results show a phylogeographic signal, but only a minimal genetic differentiation between the allopolyploids, giving evidence for frequent and extensive gene flow between the sympatric sibling allopolyploids in Central Europe. The data further supports two possible origins of *D. traunsteineri*; one in Great Britain and one in Continental Europe, whereas *D. majalis* and *D. purpurella* seem most likely to be of a single origin. We conclude that, in the face of gene flow, the observed phenotypic divergence between the sibling *Dactylorhiza* polyploids is maintained by a strong divergent selection, potentially related to their ecological specialization.

Using RAD-seq as a genome-wide approach to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

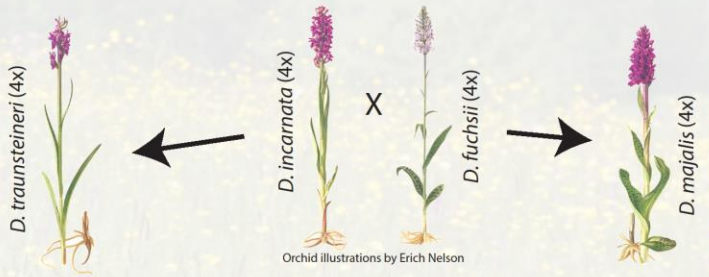
Marie Kristine Brandrud[1], Juliane Baar[1], Maria Teresa Lorenzo[1], Mikael Hedrén[2], Emiliano Trucchi[1], Ovidiu Paun[1]

[1]Department of Botany and Biodiversity Research, University of Vienna, Austria
(marie.kristine.brandrud@univie.ac.at, juliane.baar@univie.ac.at,
maylorenzo@gmail.com, emi.trucchi@gmail.com, ovidiu.paun@univie.ac.at)

[2]Department of Biology, Biodiversity, University of Lund, Sweden (mikael.hedren@biol.lu.se)

How to treat polyploid data ?

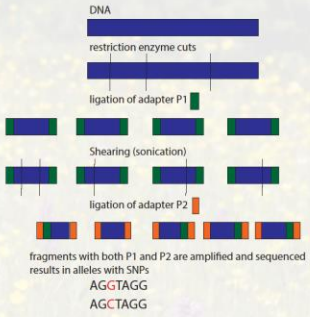
Next generation sequencing is increasingly being applied to answer evolutionary questions, namely in population genetics and phylogeny. The methods and programs available are, however, mainly developed for diploid data, i.e. maximum two alleles with a regular inheritance. How to best treat polyploid data, i.e. multiple alleles and loci as well as inheritance ranging from disomic to polysomic, is still challenging. Here we describe two methods on how to treat polyploids in a system containing sibling allopolyploid species originated independently from replicated crosses of two divergent diploid species.



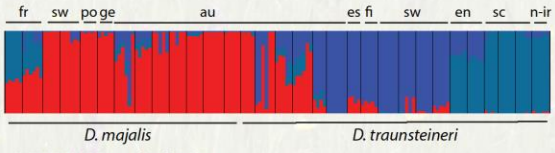
Dactylorhiza polyploid complex

Hybridization between *Dactylorhiza fuchsii* (2*n*=40) and *D. incarnata* (2*n*=40) has resulted in several independently formed several allopolyploids, here focusing on two of these; *D. traunsteineri* (2*n*=80) and *D. majalis* (2*n*=80).

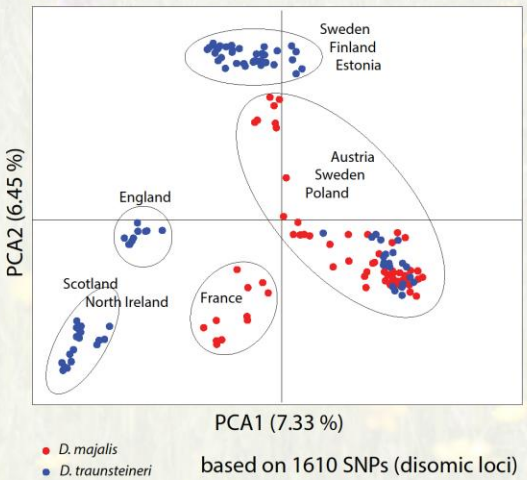
Restriction site associated DNA-sequencing (RAD-seq)
A reduced representation of the genome
(single digest protocol)



Results

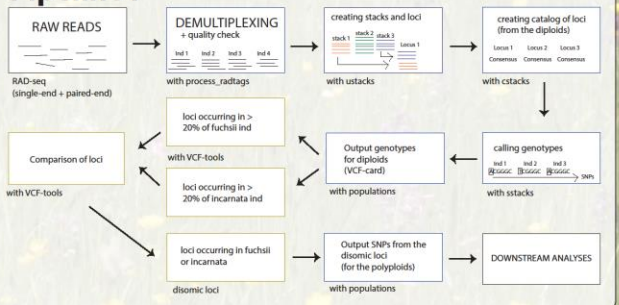


STRUCTURE results for K=3

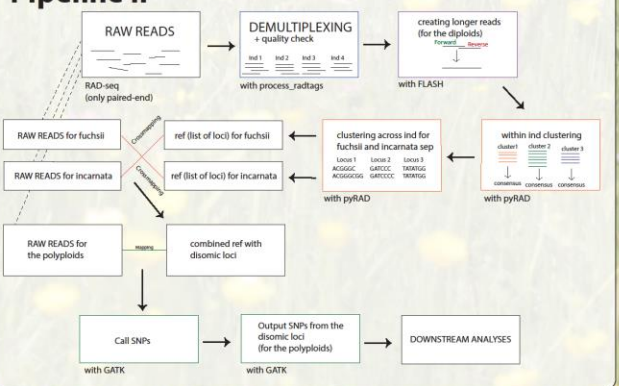


based on 1610 SNPs (disomic loci)

Pipeline I



Pipeline II



Conclusions

Geography explains genome-wide genetic structure due to independent origins and/or isolation by distance patterns
Gene flow between the two species in sympatry
The results indicate at least two independent origins of *D. traunsteineri*

Acknowledgements http://plantgenomics.univie.ac.at
Mikael Hedren
Mark Chase
Thomas Wolfe
Richard Bateman
FWF Der Wissenschaftsfonds.
Project nr: Y661-B16
universität wien

Symposium: Origins and Natural History of the Scandinavian Biota, Stockholm, Sweden, 7-8th November 2016

Origin and Phylogeography of Polyploid *Dactylorhiza* in Europe

Marie K. Brandrud[1], Juliane Baar[1], Maria T. Lorenzo[1], Mikael Hedrén[2], Emiliano Trucchi[1], Ovidiu Paun[1]

[1]Department of Botany and Biodiversity Research, University of Vienna, Austria
[2]Department of Biology, Biodiversity, University of Lund, Sweden

Poster Presentation

While Fennoscandia and Britain was covered in ice during the last glaciation (c. 22 000 to 17 000 years ago), several plants are believed to have survived in ice free refugia further south and east in Europe. During the retreat of the ice these plants have gone through a stepwise migration towards the north. Previous microsatellite and ITS studies on polyploid *Dactylorhiza* (Orchidaceae) have hypothesized at least three immigration routes into Scandinavia (see fig. 1). In this RAD-seq based study we aim to elucidate the origins and colonization of the two allopolyploids *D. traunsteineri* and *D. majalis*.

Origin and Phylogeography of Polyploid *Dactylorhiza* (Orchidaceae) in Europe

Marie Kristine Brandrud[1], Juliane Baar[1], Maria Teresa Lorenzo[1], Mikael Hedrén[2], Emiliano Trucchi[1], Ovidiu Paun[1]

[1]Department of Botany and Biodiversity Research, University of Vienna, Austria
(marie.kristine.brandrud@univie.ac.at, juliane.baar@univie.ac.at,
maylorenzo@gmail.com, emi.trucchi@gmail.com, ovidiu.paun@univie.ac.at)

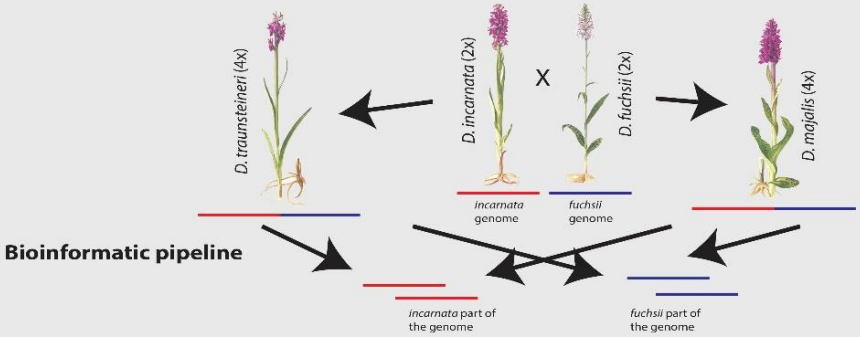
[2]Department of Biology, Biodiversity, University of Lund, Sweden (mikael.hedren@biol.lu.se)

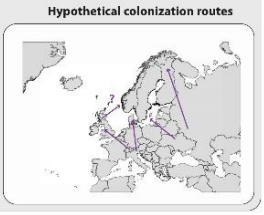
Introduction

While Fennoscandia and Britain was covered in ice during the last glaciation (c. 22 000 to 17 000 years ago), several plants are believed to have survived in ice free refugia further south and east in Europe. During the retreat of the ice these plants have gone through a stepwise migration towards the north. Previous microsatellite and ITS studies on polyploid *Dactylorhiza* (Orchidaceae) have hypothesized at least three immigration routes into Scandinavia. In this NGS-based study using RAD-seq (restriction site associated DNA-sequencing) we aim to elucidate the origins and colonization as well as investigate the genetic structure of the two sibling allopolyploids *D. traunsteineri* and *D. majalis*.

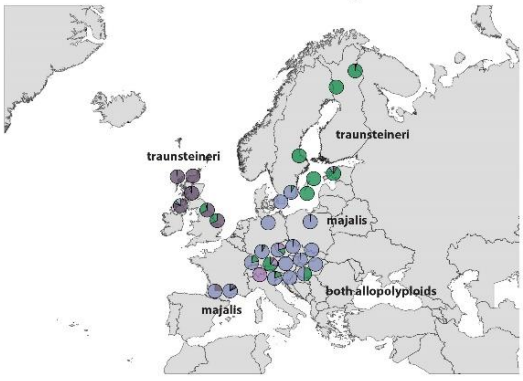
Dactylorhiza polyploid complex

Hybridization between *Dactylorhiza fuchsii* (2*n*=40) and *D. incarnata* (2*n*=40) has resulted in several independently formed allopolyploids, here focusing on two of these; *D. traunsteineri* (2*n*=80) and *D. majalis* (2*n*=80).



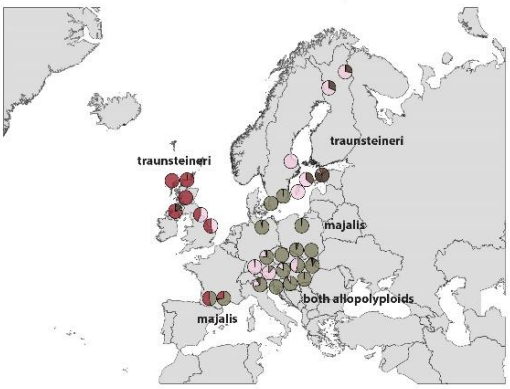


Results - *incarnata* part



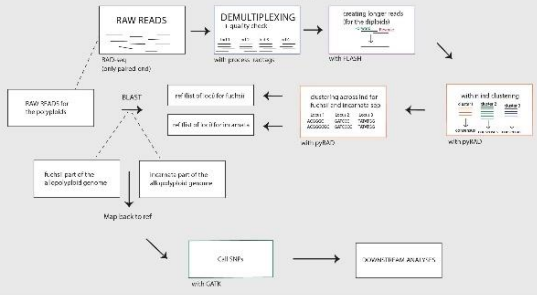
STRUCTURE analysis K=4 for 156 allopolyploid Dactylorhiza ind (5 ind in each of the 31 populations) based on 1083 SNPs from RAD-seq visualized on a map of Europe with QGIS. Each population is represented by a pie chart, which show the allocation to the four structure groups.

Results - *fuchsii* part



STRUCTURE analysis K=4 for 156 allopolyploid Dactylorhiza ind (5 ind in each of the 31 populations) based on 761 SNPs from RAD-seq visualized on a map of Europe with QGIS. Each population is represented by a pie chart, which show the allocation to the four structure groups.

Bioinformatic pipeline



Conclusions

- Geography explains genome-wide genetic structure due to independent origins or isolation by distance patterns
- Gene flow between the two species in sympatry
- The results indicate three allopolyploid origins and at least two colonization routes into Fennoscandia

Acknowlegdements http://plantgenomics.univie.ac.at
Mark Chase
Thomas Wolfe
Richard Bateman

FWF Der Wissenschaftsfonds.
Project nr: Y661-B16

universität wien

RAD-seq as a genome-wide approach to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

Marie K. Brandrud[1], Juliane Baar[1], Maria T. Lorenzo[1], Mikael Hedrén[2], Emiliano Trucchi[1], Ovidiu Paun[1]

[1]Department of Systematic Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria [2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden

Oral Presentation

Genomic inferences based on RAD-seq (i.e. restriction-site associated DNA sequencing) data are currently widely applied to answer evolutionary questions in diploid organisms, but their use in polyploids is hindered by their complicated genomes and inheritance patterns. We focus here on several *Dactylorhiza* allopolyploids that have been suggested to have originated independently from hybridization of two diploid lineages; *D. fuchsii* and *D. incarnata*. To investigate the evolutionary history of the polyploid group, representatives of both parental diploids were first made use of. Based on the diploid data, sampled throughout their European distribution, we build a separate synthetic reference for each of the parental subgenomes in the polyploids. The RAD-seq loci in these two references are expected to present disomic inheritance in the allopolyploid genomes. We further implemented a pipeline including BLAST, BOWTIE2 and GATK for aligning the polyploid reads to these references and calling/filtering variants. Based on the 5,000+ RAD loci obtained, we finally investigated the patterns of genetic divergence, gene flow dynamics and polyploid origins of 53 populations sampled across Europe, over the distribution range of the allopolyploid complex. Our results show a phylogeographic signal, but document a genome-wide absence of genetic differentiation between allopolyploids, giving evidence for frequent and extensive gene flow between the sympatric sibling allopolyploids in Central Europe. The data further supports two possible origins of *D. traunsteineri*; one in Great Britain and one in Continental Europe, whereas the widely distributed *D. majalis* and the western *D. purpurella* seem most likely to be of a single origin each. We conclude that, in the face of gene flow, the observed phenotypic divergence between the sibling *Dactylorhiza* polyploids is maintained by a strong divergent selection, potentially related to their ecological specialization.

Symposium: Plant epigenetics, from mechanisms to ecological relevance, Vienna, Austria, 12-15th September 2017

RADseq and bsRADseq as genome-wide approaches to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

Marie K. Brandrud[1], Emiliano Trucchi[1], Juliane Baar[1], Maria T. Lorenzo[1], Mikael Hedrén[2], Ovidiu Paun[1]

[1]Department of Systematic Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria [2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden

Poster Presentation

Genomic inferences based on RADseq (i.e. restriction-site associated DNA sequencing) are currently widely applied to answer evolutionary questions in diploid organisms, but their use in polyploids is hindered by their complicated genomes and inheritance patterns. We focus here on *Dactylorhiza* allopolyploids that have originated independently from hybridization of two diploid lineages; *D. fuchsii* and *D. incarnata*. Based on thousands of RAD loci, we investigate the patterns of genetic divergence, gene flow dynamics and polyploid origins of 53 populations sampled across Europe, over the distribution range of the allopolyploid complex. We complement this with results obtained in selected individuals with bisulfiteconverted RADseq, an approach to quantify the level of DNA methylation differentiation across multiple individuals. Our genetic results show a phylogeographic signal, but document a genome-wide absence of genetic differentiation between allopolyploids, giving evidence for frequent and extensive gene flow between the sympatric sibling allopolyploids in Central Europe. In contrast, significant methylation differences are uncovered that are fully methylated in one polyploid and not methylated in the other. We conclude that, in the face of gene flow, the observed phenotypic divergence between the sibling *Dactylorhiza* polyploids is maintained by a strong divergent selection, potentially related to their ecological specialization.

RADseq and bsRADseq as genome-wide approaches to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

M. K. Brandrud[1], E. Trucchi[1], J. Baar[1], M. T. Lorenzo[1], M. Hedrén[2], O. Paun[1]

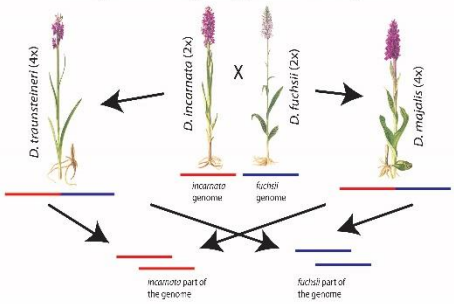
[1]Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria
[2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden

Introduction

Restriction-site associated DNA sequencing (i.e. RADseq) is being widely used to make genomic inferences for answering evolutionary questions. This approach has been mostly applied to diploid organisms, because of the complicated genomes and inheritance patterns in polyploids.
In this study we apply it to allopolyploid *Dactylorhiza* that have originated independently from hybridization of diploid *D. fuchsii* and *D. incarnata*.

Based on thousands of RAD loci, we investigate the patterns of genetic divergence, gene flow dynamics and polyploid origins of 53 populations sampled across Europe, over the distribution range of the allopolyploid complex. Epigenetics have been shown to play an important part in the establishment of polyploids. To quantify the level of DNA methylation differentiation in the polyploids bisulfite-converted RADseq (Trucchi *et al. MolEcol* 25(8) 2016) were also obtained for two selected individuals of each allopolyploid.

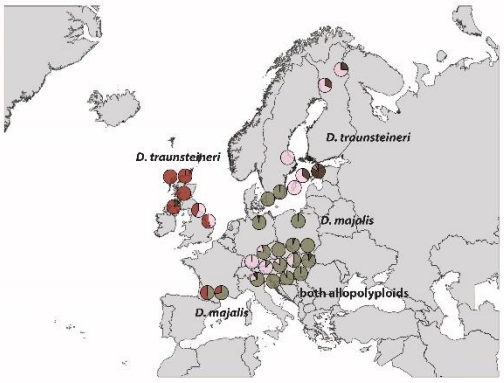
Dactylorhiza polyploid complex



Hybridization between *Dactylorhiza fuchsii* (2*n*=40) and *D. incarnata* (2*n*=40) has resulted in several independently formed allopolyploids, here focusing on two of these; *D. traunsteineri* (2*n*=80) and *D. majalis* (2*n*=80).

RADseq shows gene flow in sympatry

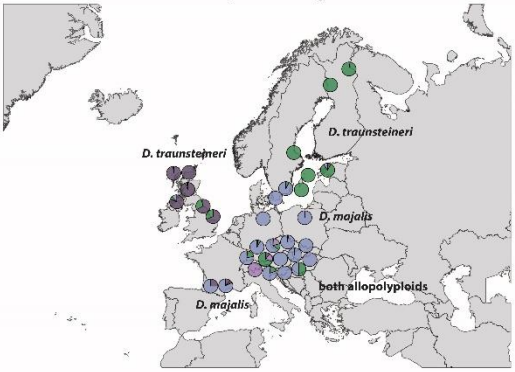
D. fuchsii part of the genome



STRUCTURE analysis K=4 for 156 allopolyploid *Dactylorhiza* ind (5 ind in each of the 31 populations) based on 761 SNPs extracted from the *D. fuchsii* part of the genome (RAD-seq) visualized on a map of Europe with QGIS. Each population is represented by a pie chart, which show the allocation to the four structure groups.

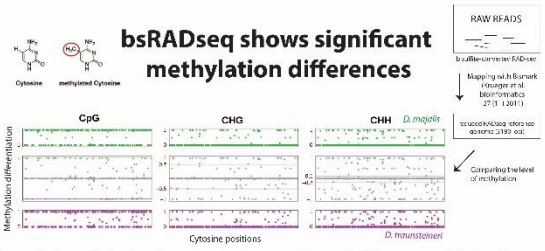
RADseq shows a phylogeographic signal

D. incarnata part of the genome



STRUCTURE analysis K=4 for 156 allopolyploid *Dactylorhiza* ind (5 ind in each of the 31 populations) based on 1083 SNPs extracted from the *D. incarnata* part of the genome (RAD-seq) visualized on a map of Europe with QGIS. Each population isrepresented by a pie chart, which show the allocation to the four structure groups.

bsRADseq shows significant methylation differences



Differential methylation between the two *Dactylorhiza* tetraploids, given separately for each methylation context. 705, 1153 and 3533 cytosine positions were assessed in the CpG, CHG and CHH context respectively, where H can be A, C or T. The top and bottom panels show the average methylation of the two species separately, wheras the middle panels show the methylation difference between the two species. A value of 1 indicates complete methylation in *D. majalis* and complete lack of methylation in *D. traunsteineri*, while a value of −1 indicates the opposite.
The lines in the middle panels indicate 95% quantiles.

Trucchi *et al. MolEcol* 25(8) 2016

Conclusions

We conclude that, in the face of gene flow, the observed phenotypic divergence between the sibling *Dactylorhiza* polyploids is maintained by a strong divergent selection, potentially related to their ecological specialization and connected to high epigenetic differentiation between the two allopolyploid species.

Acknowlegdements

Mark Chase, Thomas Wolfe, Richard Bateman
http://plantgenomics.univie.ac.at
Project nr: Y661-B16

universität wien
FWF Der Wissenschaftsfonds.

Young systematists' forum 2017, London, England, 1st December, 2017

Origin of Polyploid *Dactylorhiza* (Orchidaceae) in Europe

Marie K. Brandrud[1], Mikael Hedrén[2,], Mark W. Chase[3], Richard M. Bateman[3], Juliane Baar[1], Maria T. Lorenzo[1], Ovidiu Paun[1]

[1]Department of Systematic Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria [2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden [3]Royal Botanic Gardens Kew, UK
marie.kristine.brandrud@univie.ac.at

Oral Presentation

Polyploidy, i.e. whole genome doubling, is a predominant and powerful evolutionary force, especially in plants. Allopolyploidization can trigger speciation, through immediate isolation from diploid relatives. In addition, independently originated sibling allopolyploids tend to hybridize adding onto this complexity. One genus harbouring many such allopolyploids is *Dactylorhiza* in the incredibly diverse orchid family. This study makes use of RAD-seq by first unravelling the diploid species tree in the genus and then develops a bioinformatics approach to investigate the history of different allopolyploid lineages. A coalescent species delimitation analysis was run for the putative diploid parents to get the best defined groups that could have given rise to allopolyploids. *D. saccifera*, *D. fuchsii*, *D. foliosa* and *D. maculata* come out as most likely species delimitation, the same for European *D. incarnata*, Asian *D. incarnata* and *D. euxina*. Further fineRADstructure was run for the different allopolyploid and diploid groups to identify most likely parents.

RAD-seq as a genome-wide approach to understand allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

Marie Kristine Brandrud[1], Juliane Baar[1], Mikael Hedrén[2], Thomas Wolfe[1], Ovidiu Paun[1]

[1]Department of Systematic Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria [2]Department of Biology, Biodiversity, University of Lund, Sölvegatan 37, SE-223 62 Lund, Sweden

Oral Presentation

Genomic inferences based on RAD-seq (i.e. restriction-site associated DNA sequencing) are currently widely applied to answer evolutionary questions in diploid organisms, but their use in polyploids is hindered by complicated genomes and inheritance patterns. We focus here on several *Dactylorhiza* allopolyploids that have been suggested to have originated independently from hybridization of the diploids *D. fuchsii* and *D. incarnata*. We propose here a new analysis pipeline for polyploid groups, in which representatives of parental diploids, sampled throughout their European distribution, are used to create a reference by building loci *de novo.* The alleles at the RAD loci were further classified into sub-genomes with the program HyLiTE and the corresponding reads assigned to either parent. With these data we further investigated the patterns of genetic divergence, gene flow dynamics and polyploid origins of 48 populations sampled across Europe, over the distribution range of the allopolyploid complex. Our results show a phylogeographic signal, but document a genome-wide absence of genetic differentiation between allopolyploids, giving evidence for frequent and extensive gene flow between the sympatric sibling allopolyploids in Central Europe, in particular around the sympatric areas of the Alps. The data further give evidence of two genetic clusters for *D. traunsteineri*: one in Great Britain and one in Continental Europe, whereas the widely distributed *D. majalis* and the western *D. purpurella* form a single genetic pool each. We conclude that, in the face of gene flow, the observed phenotypic divergence between the sibling *Dactylorhiza* polyploids is maintained by a strong divergent selection, potentially related to their ecological specialization.

The macro- and microevolutionary processes driving allopolyploid evolution in *Dactylorhiza* (Orchidaceae)

Marie K. Brandrud[1], Thomas Wolfe[1], Emiliano Trucchi[2], Francisco Balao[3], Mikael Hedren[4], Ovidiu Paun[1]

[1]University of Vienna, Department of Botany and Biodiversity Research, Rennweg 14, Vienna, A-1030, Austria [2]Università degli Studi di Ferrara, Dipartimento di Scienze della vita e biotecnologie, Via Luigi Borsari 46, Ferrara, 44121, Italy [3]Universidad de Sevilla, Dpto. Biología Vegetal y Ecología, Área de Botánica. Facultad de Biología, Avda. Reina Mercedes, 6, Sevilla, 41012, Spain [4]Lund University, Department of Biology, Sölvegatan 37, Lund, Sweden

Oral Presentation

Whole genome doubling (WGD) is a central force shaping plant evolution. Early-generation allopolyploids need to quickly accommodate divergent genomes into one nucleus by adjusting organization and function, with potential ecological consequences. The orchid genus *Dactylorhiza*, with its diverse array of polyploid lineages that are often ecologically distinct, constitutes an excellent system to investigate allopolyploid evolution. With different phylogenetic and coalescent methods applied to RADseq, we first construct the species tree for the diploid species in this genus and then document the parentage of 18 independent *Dactylorhiza* allopolyploids. We bring evidence for frequent gene flow in contact zones between related polyploids of different ages. This process enriches the genetic pool of the involved lineages, but the patterns observed point to the existence of porous genomes, with few genomic regions resilient to admixture. Such a pattern suggests a strong divergent selection acting at specific loci in order to maintain the observed phenotypic divergence. We further exemplify with RNAseq and smRNAseq the molecular basis of phenotypic and ecological divergence between two of these independent, sibling allopolyploids (*D. majalis* and *D. traunsteineri*). These two polyploids are distributed across large European areas, but occupy distinct habitats in particular with regard to soil chemistry (i.e., available nitrate, but also K and P) and moisture. Based on Ks estimates, their diploid parental species

(*D. fuchsii* and *D. incarnata*) have diverged 10 MYA; today their transcriptomes are highly divergent, indicating a major transcriptomic shock associated with the origin of the allopolyploids. We find significant expression differences between both polyploids that affect several ecologically-relevant genes. For example, genes in the photosynthesis metabolic pathway have been significantly upregulated in *D. traunsteineri*, which adapted to nitrate-poor environments. Alternative parental dominance is confirmed by differential homoeolog expression in each of the two polyploids, and does not support a recently formulated hypothesis of a generally dominant genome that is retained over different WGD events. Finally, we conclude that the major transcriptomic divergence observed among the diploid parents became reconciled in different ways in the sibling *Dactylorhiza* polyploids, as a result of an interplay between stochastic genomic alterations and distinct selection pressures, resulting in specific adaptation to their respective divergent environments.