Genomic study of hybridization and introgression between two annual plant species

Khaled Mirzaei
September 2021

Thesis submitted in fulfilment of
the degree of Doctor in Sciences

Jury members

President:	Prof. Marie-Laurence De Keersmaecker (UCL, Belgium)
Supervisor:	Prof. Renate A. Wesselingh (UCL, Belgium)
Readers:	Prof. Pierre Bertin (UCL, Belgium)
	Prof. Thierry Hance (UCL, Belgium)
	Prof. Patrick Mardulyn (ULB, Belgium)
	Prof. Alok Arun (Inter-American University of Puerto Rico, USA)

Dedicated to all my teachers, from elementary school until my PhD

Acknowledgments

I would like to first express my deepest gratitude to my supportive supervisor, Professor Renate A. Wesselingh for her continuous support, patience and encouragement during my Ph.D. study. Under her supervision, I have learned many lessons and experiences to progress not only towards my research goals but also towards turning into an independent researcher.

Special thanks to the member of my thesis jury, Professor Pierre Bertin, Professor Thierry Hance, Professor Patrick Mardulyn, Professor Alok Arun and Professor Marie-Laurence De Keersmaecker for accepting to read my dissertation and providing valuable advice.

I would like to express my gratitude and appreciation for ELIV (ELIB) members for their helps and providing a good atmosphere for students.

Finally, I am indebted to my family and friends in Belgium and Iran for their endless love, support and encouragement during these years. Undoubtedly, the achievement of this success would not have been possible without their support. I would like thanks to all of them for being with me in the past, present and future. I want to thank all of those people who, once upon a time, were my teachers and who helped me begin this long journey. Their names are too numerous to mention, but many of them inspired me to continue learning and sharing with others. I thank you all!

Khaled Mirzaei

Thesis Abstract

The general framework of this work is study of hybridization between two *Rhinanthus* species (Orobanchaceae), *Rhinanthus minor* L. and *Rhinanthus major* using advanced genomics tools. These species are annual hemi-parasitic plants in which hybridization and introgression have been documented before. In naturally mixed populations of both species, hybridization leads predominantly to introgression from *R. minor* into *R. major*. Furthermore, hybrids with *R. major* as the maternal parent ($F1_a$) have often been reported to have low germination rates in comparison with hybrids with *R. minor* as maternal parent ($F1_m$). Therefore, in order to better estimate the degree of introgression between both species, a large panel of SNP markers was developed in this thesis project and used on samples collected from several mixed populations. In order to understand the mechanisms behind the low germination rate of $F1_a$ seeds, a differential gene expression study was conducted during dormancy and germination.

Using ddRAD sequencing nearly 1100 species-specific SNP markers were obtained for both species and tracked in nine mixed populations from Belgium and the Netherlands. Introgression was much more frequent than could be detected with previously used ISSR and RAPD makers. Further investigation of our species-specific markers in *R. major* individuals with a low admixture rate with *R. minor* showed that only a few of the loci tended to introgress frequently, while the others had low introgression rates. Although with ddRAD sequencing we only covered a small part of the genome, we were still able to find similarity and dissimilarity of introgressed loci among the populations.

Using RNA-seq in $F1_a$ and $F1_m$ seeds during cold stratification and after germination revealed that there were large transcriptome differences between seeds with different maternal plants, but we could not detect any differentially expressed genes in the plastid genome, what would be expected in case of a nuclear-cytoplasmic incompatibility. In $F1_m$ seeds, genes active in the development of seeds and seedlings were upregulated, but in $F1_a$ seeds, upregulation of genes related to stress and enduring dormancy were observed. The main cause of this maternal effect is still unclear and further investigation is needed to find out how these differences in gene expression are caused.

The combined action of this thesis has opened the way towards a full understanding of hybridization and introgression between *R. major* and *R. minor*.

Contents

General introduction

1. Speciation and hybridization 1
2. Reproductive barriers 2
3. Frequency of hybridization 3
4. Influence of human activity on the occurrence of hybridization 4
5. Consequences of hybridization 6
   5.1. Homoploid hybrid speciation 6
   5.2. Polyploid hybrid speciation 7
   5.3. Heterosis and transgressive segregation 8
   5.4. Adaptive introgression 9
6. Genetic techniques for studying hybridization 12
   6.1 Understanding the mechanisms behind barriers to hybridization ... 14
7. *Rhinanthus minor* and *Rhinanthus major* as model species 16
   7.1. General biology 16
   7.2. Patterns of hybridization 20
   7.3. Reproductive barriers 20
   7.4. Chronology of the work on introgression and reproductive barriers between *R. minor* and *R. major* at UCLouvain 21
8. Objectives of this thesis 22

Development of a large set of diagnostic SNP markers

Abstract 25

Introduction 26

Materials, methods and results 26

Introgression patterns in mixed *Rhinanthus* populations: increasing detectability with 1100 species-specific SNP markers

Abstract 31

Introduction 32

Materials and methods 34

Results....36
Discussion....41
Conclusion....44
Introgression from *Rhinanthus minor* into *R. major* in various habitats
Abstract....45
Introduction....46
Material and methods....47
Results....51
Discussion....59
Transcriptome profiling and comparison of *Rhinanthus major* and *R. minor* reciprocal F_1 hybrids during seed stratification and germination
Abstract....63
Introduction....65
Material and methods....67
Results....69
Discussion....77
General discussion
6.1 Similarity and dissimilarity of introgression in studied populations ..99
6.2 Was it a good idea to use ddRAD seq for studying introgression?...100
6.3 Conclusions and perspectives....103
References....105

Chapter 1

General introduction

1. Speciation and hybridization

From the very beginning of human history, we have had a penchant for classification and labeling of other organisms to establish a sense of order in the world around us. However, for biologists from different branches, classifying organisms in one category known as “species” has always been a controversial subject (Ruse 1969; Mallet 1995, 2010; Aldhebiani 2018). This challenge has driven us to have nearly 27 definitions for species, beside the fact that genetic exchanges among species generally have been ignored in these definitions (Bridle 2002; de Queiroz 2007; Aldhebiani 2018). The root of the problem probably lies in the acceptance of species as a static concept and not as a dynamic entity, given that speciation is a long and complex process (de Queiroz 2007; Genovart 2008). The process of speciation happens throughout a period of evolutionary time when a new species arises in the form of an isolated population (Gavrilets 2000; Mallet 2008). During this span of time, there it can be moments in which diverged individuals may eventually come into contact again and start to mate because of incomplete reproductive barriers (Abbott et al. 2013). This process, known as natural hybridization, is a rare process in terms of the total number of individuals that hybridize within populations, but it is not so rare when we compare it with the number of species that are actually known to be involved in hybridization (Mallet 1995; Arnold et al. 1999; Mallet 2005a). The average frequency of hybridization among animals species is close to 10% mean, while it goes to nearly 25% in plant species (Mallet 2005a). However, the occurrence of hybridization is not uniformly distributed among family and genera (Ellstrand et al. 1996). In the animal world, hybridization between different phyla, classes and orders has been reported recently, but in some taxonomic groups, such

as birds, it is considered to be more common (Grant and Grant 1992; Genovart 2008; Zhang et al. 2014). Indeed, in some cases, for instance in British ducks, the rate of animal hybridization is higher than the average in the plant kingdom, but animal hybridization is often considered as harmful (Mallet 2005a). Botanists generally agree that hybridization is important for plant evolution, while zoologists have called hybridization a harmful and unnatural process (Seehausen et al. 2008). While the short-term effect of interspecific hybridization is a great cause for concern for conservation biologists due to retarding or reversing population diversification through mixing or swamping of gene pools, in the long term it is a fascinating topic for evolutionary biologists (Ishida et al. 2011). If hybrids have a very low fitness in comparison to their parents due to outbreeding depression, they represent a waste of resources invested in reproduction. If reproduction falls below the replacement rate, this can lead to a decline of one of the parental lines and ultimately drive its population towards extinction (Wolf et al. 2001; Todesco et al. 2016). This process is called demographic swamping. However, if hybrid fitness is equal to the parental lines, and population growth rates are more than replacement rates, one of the parental linages may be replaced by hybrids. This is called genetic swamping (Wolf et al. 2001; Todesco et al. 2016). Both swamping processes can be a menace to rare or endemic species that come into contact with related taxa (Fitzpatrick et al. 2015).

2. Reproductive barriers

Reproductive isolation is essential for the process of speciation in order to prevent species from being able to successfully mate with other species. These reproductive barriers fall into two categories: prezygotic and postzygotic barriers, which are by-products of other evolutionary changes (Coyne and Orr 2004). They can guarantee a first barrier for formation and maintenance of species diversity, and the final isolation mechanism usually comes with a combination of both kinds of barriers (Widmer et al. 2009a). In animals, according to research on *Drosophila* (Coyne and Orr 1997) and the fish genus *Etheostoma* (Mendelson 2003), prezygotic isolation evolves faster than postzygotic reproductive isolation in species living in sympatry (Turissini et al. 2018), but no concrete evidence for this has been found in plants (Widmer et al. 2009a). Indeed, intrinsic and extrinsic factors have effects on pre- and postzygotic barriers, and levels of genomic distance between parents can affect this as well (Scopece et al. 2008; Widmer et al. 2009a). Prezygotic

barriers can be favored by selection, because they avoid the potentially negative fitness effects on hybrids, whereas postzygotic barriers cannot be favored by selection, since they still represent a loss of parental fitness. Prezygotic barriers, which are favored by selection, can cause reinforcement (Widmer et al. 2009a). Research on the contributions of pre- and postzygotic barriers to total reproductive isolation in plants has revealed that prezygotic isolation is usually stronger than postzygotic isolation (Ramsey et al. 2003; Lowry et al. 2008; Baack et al. 2015). However, some plant lineages show a wide variation in patterns of reproductive isolation (Baack et al. 2015), and in some cases, like *Phlox* (Suni and Hopkins 2018), sunflower (Ostevik et al. 2016) and jewel flowers (Christie and Strauss 2019), postzygotic barriers can be quite strong (Widmer et al. 2009a). Hybrid seed inviability is a postzygotic barrier and it is a common barrier found in diploid and polyploid hybrids (Widmer et al. 2009a). Failure of endosperm development is the main cause of this barrier in flowering plants and it can happen even between different populations of the same species (Barnard-Kubow et al. 2016; Hämälä et al. 2017). In this nourishing tissue, which is responsible for many events from seed development to germination, maternal and paternal genome ratios and interactions between epistatic alleles of different loci can disrupt embryo development and cause hybrid seed mortality (Lafon-Placette et al. 2017). The endosperm develops after fertilization of the homodiploid central cell to support embryo growth and it has major effects on hybrid seed survival. Genomic imprinting, an epigenetic phenomenon causing differential expression of genes depending on their parent-of-origin, plays a predominant role in the endosperm (Lafon-Placette and Köhler 2016).

3. Frequency of hybridization

The evolution of reproductive barriers requires a long time and many generations, during which population size and spatial distribution will undergo many changes (Hewitt 2011). Factors that are involved in strengthening or removing the gene exchange barriers, such as transposable elements or epigenetic changes (Serrato-Capuchina and Matute 2018) will probably be activated during this long period (Arnold 1997; Hewitt 2011; Abbott et al. 2013). According to the estimation of natural hybridization in plant species, 30–70 % of all flowering plant species have gone through hybridization events in their phylogenetic histories, which shows the importance of hybridization in plant evolution (Soltis and Soltis 2009),

considering that many hybrids remain undiscovered and recent or distant past hybridization events can be difficult to detect. The latest survey of hybridization in 37,000 species of vascular plants from 282 families showed that interspecific hybridization occurs in 40% of families and 16% of genera (Whitney et al. 2010), and in some genera like *Eucalyptus* more than 50% of species are involved in hybridization (Field et al. 2011). In the Whitney et al (2010) survey, flowering plant families such as *Orchidaceae*, *Lamiaceae*, *Asparagaceae* and *Asteraceae* contain high numbers of hybridizing species, while other families such as *Caryophyllaceae*, *Cyperaceae* and *Apiaceae* appear to have few hybrids. They also detected that the tendency for plants to hybridize varies across the planet (Whitney et al. 2010), although hybridization within a group is determined more by intrinsic characters than by environmental conditions. A recent study of plant hybridization frequency in the United States that showed hybridization clearly is not uniform among taxonomic groups: *Potamogetonaceae*, an aquatic family, had the highest occurrence among the studied families (Guo 2014). This study of spatial variation in hybrid richness at the county level across the United States not only revealed variation in the number species involved in hybridization, but also that there were three hotspots for hybridization, namely the north-east corner (New England), the Great Lakes region and the coast of California (Guo 2014). Modern analyses of hybridization using molecular tools have revealed that rates of hybridization differ across ferns and their relatives (Mitchell et al. 2019a). Rothfels et al. (2015) showed that one fern species (×*Cystocarpium roskamianum*) has recently been formed by natural hybridization between parental lineages that diverged from each other approximately 60 million years ago. Why do some groups hybridize more than others? According to a study by Mitchell *et al* (2019a), characters such as perenniality and woodiness across all taxonomic levels, higher outcrossing rates and abiotic pollination syndromes at the family level, and less variable genome sizes at the genus level were associated with increased hybridization rates and may be acting by increasing the formation of hybrids.

4. Influence of human activity on the occurrence of hybridization

Reports of new hybrids are increasing nowadays, not only because of the rise in accurate methods with which hybrids can be studied or easy access to genetic information for large numbers of samples, but also because of the increase in human activity over the centuries (Genovart 2008; Harrison and

Larson 2014; Ortego et al. 2017). Nearly 60 years ago, Anderson & Stebbins (1954) predicted that human activity would create new habitats for newly formed hybrids, allowing them to both establish and persist. Human activity, which always comes with disturbance of natural habitats, can be a catalyst that stimulates the formation and persistence of hybrid plants (Guo 2014; Todesco et al. 2016). While these activities have been known to increase the likelihood of contact between previously isolated species or populations, it is unclear how these newly formed hybrids will persist in some habitats (Ortego et al. 2017). Not only human activities and related biological invasions increase the diversity of exotic species, but they may also produce hybrids that would not form naturally (Todesco et al. 2016). With the continued expansion of the human population through growth and migration, these combined events can be expected to increase hybridization over time (Genovart 2008; Chunco 2014; Todesco et al. 2016). Human infrastructure like roads have facilitated the dispersal of many organisms and created new opportunities for allopatric species to come into contact again (Chunco 2014; Todesco et al. 2016). However, human disturbance alone cannot explain the high frequency of hybrid formation in some places, because it also requires source (parental) species and suitable habitats (Todesco et al. 2016). An example of hybridization due to road construction has been observed in the genus *Pericallis* (*Asteraceae*) on Tenerife, Canary Islands. In this case, hybridization was promoted by secondary contact caused by human activities and this process may eventually threaten the existence of the three species and counteract the process of evolutionary radiation (van Hengstum et al. 2012). In addition, humans have changed much of the planet and have brought many disturbances to the planet both directly and indirectly. Wildfire also has been confirmed as an important source of environmental disturbance that promotes hybridization between two ecologically well-differentiated native species of Californian oak (Ortego et al. 2017). Climate change is significantly influencing the evolutionary trajectory of species and making different ecological communities through new species assemblages and consequently providing new reproductive interactions (Chunco 2014). The occurrence of hybridization could also increase due to climate change because of changes to the time of flowering in many plant species (Baack et al. 2015; Todesco et al. 2016). In this situation, species that were previously isolated through different flowering phenologies may flower at the same time and thus provide opportunities for cross-pollination and hybrid production (Chunco 2014; Baack et al. 2015; Todesco et al. 2016).

5. Consequences of hybridization

Hybridization can provide several outcomes for the evolution of natural populations in plant species. It can quickly introduce adaptive variation into a population or it can generate new lineages (Goulet et al. 2017). If we define hybridization as reproduction between genetically distinct individuals, then it occurs in every process of speciation (Barton and Hewitt 1985). Through this process, a wide range of genetic interactions, from precluding divergence to providing opportunities for stronger reproductive barriers, can be facilitated between the parental species (Arnold 1997; Abbott et al. 2013; Goulet et al. 2017). The new phenotypes provided by genomic interactions will go through neutral divergence, local adaptation and coevolution (Abbott et al. 2013). This process is mainly promoted by heterosis or transgressive segregation, and it will fail due to intrinsic or environmentally mediated incompatibilities. Therefore, hybridization can act in two opposite ways, either retarding or promoting speciation (Rieseberg and Carney 1998; Abbott et al. 2013; Goulet et al. 2017; Kagawa and Takimoto 2018). However, the consequences of hybridization can be more complex if we consider temporal and spatial contexts (Abbott et al. 2013).

5.1. Homoploid hybrid speciation

Hybridization that produces a new taxon by reassembling the genomes of two parental species, without changes in ploidy level and distinct from both parents, is called homoploid hybrid speciation or recombinational speciation (Abbott et al. 2010, 2013; Nieto Feliner et al. 2017). It has been stated that homoploid hybrid speciation is a rare and difficult phenomenon, due to the homogenizing effects of the parental lineages. The crucial stage for hybrid descendants in this kind of speciation is escaping from the dissolving effect of gene flow from backcrosses with the parental species to become independent species under the influence of natural selection (Schumer et al. 2014; Lukhtanov et al. 2015). Among existing reproductive isolation mechanisms, prezygotic reproductive barriers have been considered the main tool of isolation between hybrid and parental lineages through formation of novel recombinant phenotypes to avoid crossing between parents and hybrids (Lukhtanov et al. 2015). Rearrangement of genetic blocks related to reproductive barriers in their parents is amongst the main mechanisms (Abbott et al. 2013; Schumer et al. 2014; Sun et al. 2020). A recent study of chestnut

trees showed that putative barrier segments have been accumulated in the hybrids of *Castanea henryi var. henryi* and *C. mollissima* in genomic regions with less recombination (Sun et al. 2020). The newly formed hybrid phenotype might allow them to occupy ecological niches that are unreachable for their parental species, alongside rejecting them as potential mates (Lukhtanov et al. 2015). Other effective tools, which are intraspecific factors, may influence the genomics of divergence. Factors like the age of the hybrid and the level of divergence between the parental taxa influence the probability of homoploid hybrid speciation (Buggs et al. 2009; Paun et al. 2009; Abbott et al. 2013). With a high level of divergence between parents, the chance of having major novelties in the phenotype of the offspring is high, but intrinsic incompatibilities may prevent successful hybridization. However, whether there is an optimal genetic distance for homoploid hybrid speciation is unclear (Arnold et al. 1999; Gross 2012; Abbott et al. 2013).

5.2. Polyploid hybrid speciation

With polyploidization involved, the hybrid will receive three or more homologous chromosome sets from its parents rather than the two resulting from a diploid combination of genomes (Abbott et al. 2013; Alix et al. 2017). Polyploidy has been a common and important player in plant evolution, since it has been estimated that nearly all of the extant flowering plants have gone through some polyploidization event. It is estimated that 40–70% of plant species are polyploids (Otto and Whitton 2000), and that 31% of ferns and 15% of angiosperms have been involved in polyploidy (Wood et al. 2009; Abbott et al. 2013). Through polyploidy, a strong postzygotic reproductive barrier will be created to prevent mating with the parental species and this will provide opportunities for divergence from the ancestral species (Alix et al. 2017). Although this is very common among plants, in the animal kingdom it is not very common except in fish and amphibians (Otto and Whitton 2000; Mable et al. 2011). In general, in organisms with sex chromosomes and mechanisms for balancing the gene products between the sexes the polyploidy rate is low (Mable 2004). There are two types of polyploidy: autopolyploidy and allopolyploidy. In autopolyploidy, chromosome sets originate from a single species, while in allopolyploids the chromosome sets are derived from different species due to interspecific hybridization (Abbott et al. 2013). Allopolyploidy has been considered to be more common in nature than autopolyploidy, and there are multiple examples of allopolyploidy in

domesticated crop species, such as wheat, potato and strawberry (Lyu 2016). Known cases of allopolyploid speciation involve hybridization with invasive species (Vallejo-Marin 2012), but it is also known from nature, such as in the *Dactylorhiza* species complex (Devos et al. 2003).

5.3. Heterosis and transgressive segregation

Even though hybrids are often inviable or sterile, such that genetic flow between parental species is not possible, hybridization can still affect the evolutionary trajectory of lineages. Hybridization resulting in low hybrid fitness and consequently low gene flow between offspring and parents has few evolutionary consequences (Rieseberg and Carney 1998; Abbott et al. 2013). Nevertheless, hybridization that results in at least some viable offspring can affect the genetic and phenotypic composition of populations by providing a bridge for gene flow (Harrison and Larson 2014).

After crossing two different plant species, hybrids may show superior phenotypes with a higher growth rate and biomass production than their parents. In this event, known as hybrid vigor or heterosis, offspring that have alleles from both parent perform much better than either of the parental sources (East 1936). There are two hypotheses to explain this: dominance and over-dominance (East 1936; Goulet et al. 2017). According to the dominance model, deleterious alleles received from one parent are mostly suppressed by beneficial alleles from the other parent. In the over-dominance hypothesis, heterozygotes have more extreme phenotypes than both homozygotes (Rieseberg and Carney 1998; Goulet et al. 2017). A study of heterosis for yield in tomato has shown that heterozygosity in a nonfunctional allele at the *SINGLE FLOWER TRUSS* (*SFT*) gene was giving 60% more yield in the hybrid compared to its parents (Krieger et al. 2010). In rice, several heterozygous loci have been detected showing heterosis of yield-related traits (Huang et al. 2016). However, recently it has been stated that, besides dominance and over-dominance, other mechanisms are involved in hybrid vigor, such as epistasis and epigenetics (Goulet et al. 2017; Mackay et al. 2021).

Whereas heterosis has been described for F_1 hybrids, a similar phenomenon called transgressive segregation is found when the phenotypic trait values of F_2 hybrids are strongly different from their parents (Rieseberg et al. 1999). Transgressive segregation can provide opportunities for the hybrids to occupy more diverse niches than their parents (Goulet et al. 2017).

Transgressive segregation is complex and depends on many genetic factors, but it is most often found in hybrids between inbred plants and intraspecific crosses (Rieseberg et al. 1999; Goulet et al. 2017; Mackay et al. 2021).

5.4. Adaptive introgression

When reproductive barriers are incomplete, the process of hybridization between linages is inevitable. With the occurrence of viable hybrids among parents from diverged populations, the door is open for other evolutionary events (Rieseberg and Carney 1998). When hybrids can mate again with their ancestors, after a few generations some parts of the genome of one species can flow into the other population and vice versa (Harrison and Larson 2014). This process, which is known as introgression, can provide opportunities for alleles to reveal themselves alongside other alleles already present in a parental population. This sudden gene flow from a different source into a receiver population is an important source of novel genetic variation, alongside standing variation existing within populations and ongoing mutation and recombination (Hedrick 2013; Suarez-Gonzalez et al. 2018a). Introgression, which starts with backcrosses between hybrids and pure parental individuals (Figure 1.1), was first defined by Lotsy (Hedrick 2013; Suarez-Gonzalez et al. 2018a). It was later named “introgressive hybridization” by Edgar Anderson (Anderson and Hubricht 1938; Goulet et al. 2017). Figure 1.1. shows the percentage of foreign genome remaining after each backcross.

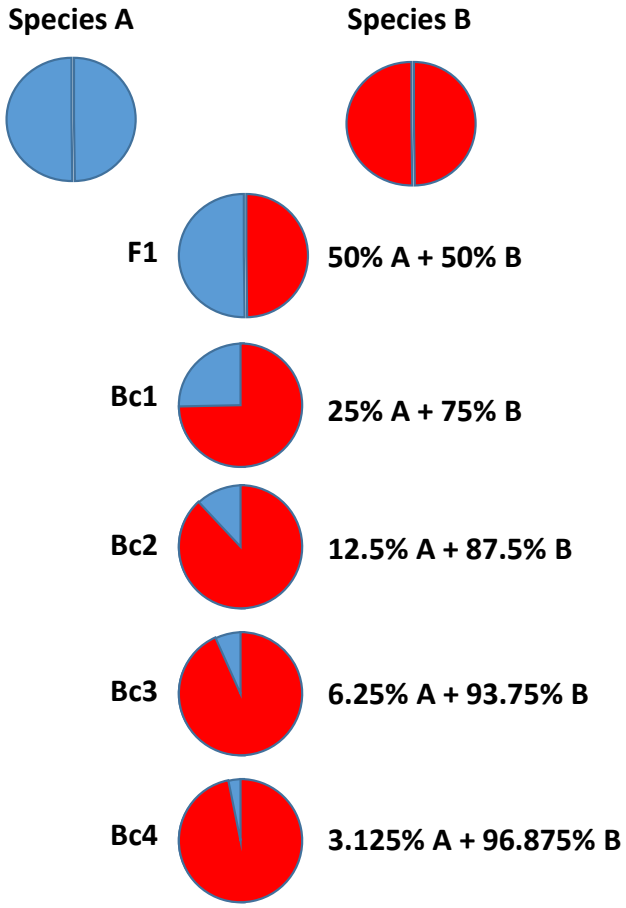


Figure 1.1. Percentage of the genome of species A remaining in species B after each generation of backcrossing. F1= First generation hybrid, Bc=backcross with species B

Introgression can bring new combinations of genetic variation, and can potentially move an adaptive trait into another species. Adaptive introgression happens when the introgressed alleles are maintained through natural selection. Not all alleles have the same probability of introgression. Alleles that are involved in a reproductive barrier mainly cluster in a region of the genome with less recombination, and tends to introgress less, while loci with advantages have more chances to introgress. Neutral loci will introgress to varying extents depending on their linkage to genes with contributions to local adaptation or reproductive isolation (Harrison and Larson 2014; Suarez-Gonzalez et al. 2018a).

Thanks to the improvements in molecular techniques (see section 6), several studies have documented adaptive introgression among species. In sunflowers, herbivore resistance traits have transferred from *Helianthus debilis to H. annuus* (D. Whitney et al. 2015). Other studies have shown adaptive introgression between *Iris* species for flood resistance (Martin et al. 2006), and for RAY genes that promote flower asymmetry and increase the outcrossing rate in groundsel species (Martin et al. 2006; Kim et al. 2008a; Hedrick 2013). Even though hybridization is not as common in the animal world, there are some examples of adaptive introgression like pesticide resistance in mice (Song et al. 2011), mimicry genes in *Heliconius* butterflies (Dasmahapatra et al. 2012) and high-altitude resistance genes that humans have acquired from Neanderthals (Hedrick 2013; Sams et al. 2016).

In crops, adaptive introgression has contributed new phenotypes favored by human selection. Maize is adapted to cultivation at low altitudes, but gene flow from different wild relatives from high altitudes has given them the ability to grow and give a profitable yield even at high altitudes (Burgarella et al. 2019). However, introgression in the opposite direction, from crops to wild plants, has also been reported and it usually has devastating effects. An example is the transfer of a herbicide resistance gene from genetically modified rice cultivars into weedy rice populations, which has caused a significant economic crisis for farmers in Brazil (Merotto et al. 2016; Burgarella et al. 2019).

The effect of adaptive introgression in speciation depends on the nature of the adaptive variation. If hybridization and subsequent introgression are introducing selectively favored alleles or alleles that have already been tested by natural selection, then these adaptive combinations of alleles will be likely be collected in one place in the genome (Harrison and Larson 2014). Sometimes, however, the favored phenotype is a result of multiple loci scattered across on the genome, and therefore the receiver population would need continued introgression in order to receive all the alleles needed for the desired phenotypic effect. But even in such situations, introgression will act much faster than mutation. The outcome of introgression may be the appearance of transgressive phenotypes due to transgressive segregation, which can be an evolutionary novelty that will allow the introgressed population to move into other niches. However, we should not forget that hybridization tends to make populations more similar and that population differentiation is difficult in the face of continued gene flow (Abbott et al. 2013; Goulet et al. 2017; Mitchell et al. 2019a).

6. Genetic techniques for studying hybridization

Before the arrival of genetic techniques, the use of morphological characters was the main means of hybrid detection and the study of hybridization. A morphological comparison with the putative parents and the occurrence of an intermediate morphology formed the routine way of hybrid detection (Wilson 1992; Rieseberg et al. 2007; López-Caamal and Tovar-Sánchez 2014). However, this method required a large data set with easily distinguishable morphological characters (Wilson 1992). In addition, scientists had to take into account the correlation among morphological traits and their dependency on the environment (Jost 2017). In some systems, hybridization leads to a mosaic of forms or even to extreme or novel characters, therefore it becomes difficult to distinguish intermediacy from interpopulation variation (Rieseberg 1997; Wei et al. 2015). In the recent decades, genetic approaches have opened new horizons for hybrid detection and introgression studies. Karyotyping and chromosome counting is one of the important tools for hybrid detection in polyploid hybrids (López-Caamal and Tovar-Sánchez 2014). Allozyme markers were also a common tool for quantifying interspecific gene flow in plants and identifying hybrids such as Hawkweed (Mráz et al. 2005) and *Arisaema ehimense* (Maki and Murata 2001). After that, the first DNA fingerprinting techniques such as Restriction Fragment Length Polymorphism (RFLPs) (Jeffreys et al. 1985) were employed. The invention of the polymerase chain reaction in 1986 (PCR; Mullis et al. 1986), led to the use of techniques such as Random Amplified Polymorphic DNA (RAPD; Williams et al. 1990), Amplified Fragment Length Polymorphism (AFLP; Vos et al. 1995) and Simple Sequence Repeats (SSR or microsatellites; Jeffreys et al. 1985) for hybrid detection. These diagnostic genetic tools are still the most widely markers used to detect hybrid and introgression at the population level (López-Caamal and Tovar-Sánchez 2014). The technological advances in sequencing have provided several new markers with a higher level variation and discriminatory capability based on DNA sequences in comparison to allozymes (Sanz et al. 2009). Nevertheless, difficulties to find diagnostic or species-specific markers, which can differentiate between hybridizing species, is an issue that can hamper the use of these methods (Twyford and Ennos 2012). More recently, advances in sequencing technology and the improvement of computing power have given us the ability to obtain and analyze large quantities of nucleotide sequence data from species without a reference genome. Next Generation Sequencing (NGS) now provides an

opportunity to study the dynamic nature of hybridization and introgression with a good coverage of markers distributed across the genome, in order to accurately detect introgressed linkage blocks (Twyford and Ennos 2012). With the advent of long read technologies, such as Pacbio and Nanopore, researchers can now provide genomic resources for non-model groups more quickly, and their availability is increasing sharply. One of the NGS-based methods, in the category genotyping-by-sequencing (GBS), that has been used in quite a number of genetic population analyses, is double digest restriction-site associated DNA-sequencing (ddRADseq; Peterson et al. 2012). In this technique, genomic DNA is fragmented using restriction enzymes and sequenced using next-generation sequencing technologies to obtain Single Nucleotide Polymorphism (SNP) markers that are located in the sequenced fragments. As there are far fewer sequences than in whole-genome sequencing, this strategy increases the coverage of fragments and provides reliable data for more samples, also for species with large genomes. This method has the advantage of having a high repeatability, no need for a reference genome and the ability to analyze many samples at once (Peterson et al. 2012). The main principle of ddRADseq is illustrated in Figure 1.2. First, DNA samples are digested with two different restriction enzymes and afterwards ligation with adaptors containing barcodes is done, which allows for 48 samples to be pooled together. Then DNA samples will be size-selected according to the previously assessed number of fragments in each size range, and size-selected samples are indexed by PCR and primers containing specific indexes. The prepared library can contain sequences with different combinations of barcodes and indexes. After sequencing, quality checking, demultiplexing and cleaning is done for the reads. One of the advantages of this method is that for data analysis a reference genome is not needed and a reference catalog can be constructed out of reliable loci. After that, the reads of each sample can be aligned with the provided catalog and SNP information is collected.

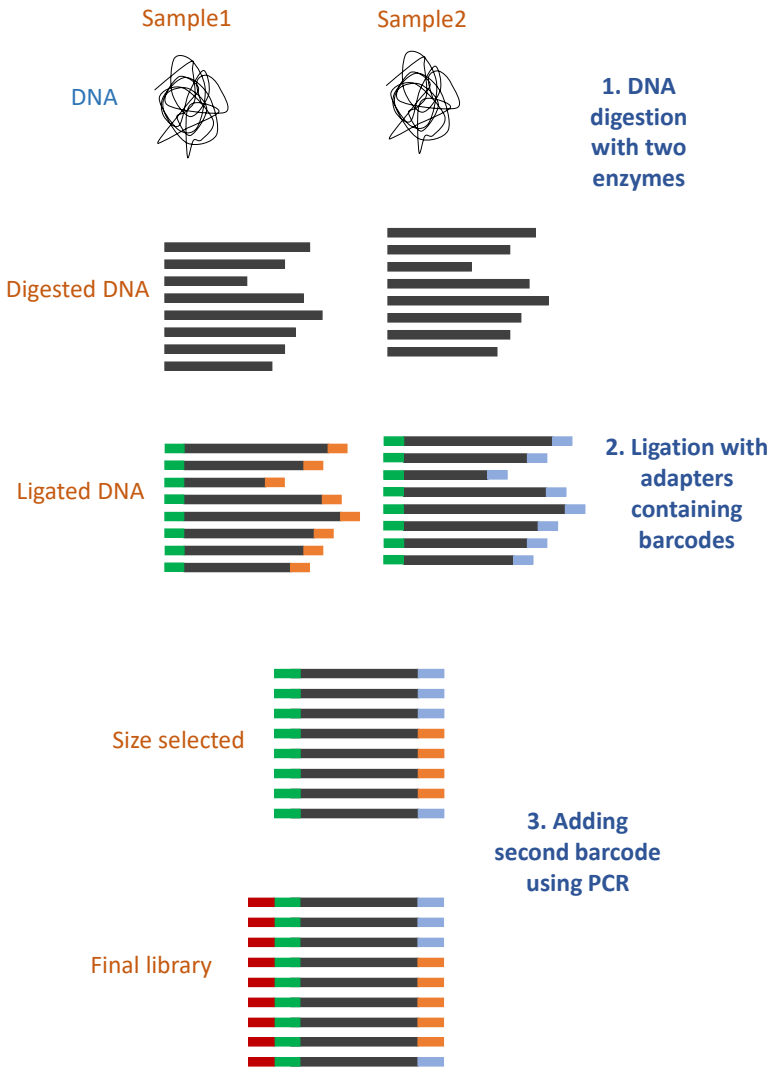


Figure 1.2. Main concept of ddRADseq and library preparation.

6.1. Understanding the mechanisms behind barriers to hybridization

Hybridization usually leads to genomic and epigenetic changes in the offspring and new phenotypes can arise from changes in protein sequences. Merging two divergent genomes during hybridization can result in the remodeling of parental gene expression patterns in hybrids, a phenomenon called transcriptome shock (Adams and Wendel 2005; Buggs et al. 2011). Therefore, there are opportunities for many genes to be expressed and they might finally contribute to transgressive and epistatic phenotypes or maybe

hybrid inviability (Seidl et al. 2019). The interactions between *cis*- and *trans*-regulatory elements usually govern patterns of gene expression and consequently the differences in phenotype between species. *Cis*-acting factors are mechanisms that affect gene expression only on the nearby genes on the same chromosomal allele, such as enhancers, whereas *trans*-factors act equally on both alleles in the diploid hybrid nucleus, like transcription factors and long noncoding RNAs (Reuveni et al. 2018). Both *cis*-acting and *trans*-acting factors have the potential to regulate gene expression patterns in hybrids, depending on the proportion received from each parental genome or their levels of divergence (Tautz 2000; Wittkopp et al. 2004). Evolution in both *cis*- and *trans*-regulatory elements can lead to adaptive morphological changes, and a change in the relative frequency of *cis*- and *trans*-regulatory elements plays an important role in the formation of divergent expression patterns in hybrids (Tautz 2000; Wittkopp et al. 2004; Zhang et al. 2019). Studies on *Drosophila* (Landry et al. 2005) have revealed that *cis*-regulatory elements cause additive inheritance of gene expression, while in *Cirsium arvense* hybrids (Bell et al. 2013), *trans*-regulatory changes contributed more to additive gene expression.

Another mechanism that can cause differential gene expression in hybrids is related to transposable elements, due to their quantitative variation and differential expression in hybrids in comparison to the parents (Labrador et al. 1999; Usai et al. 2020). The dynamics of long-terminal-repeat retrotransposons in the interspecific hybrid of two poplar species that is known for having superior performance compared to its parents, have shown that the genomic abundance and transcription levels of many retrotransposons in the hybrid are higher or lower than in the parental species (Usai et al. 2020). Investigations in *Drosophila* have shown that interspecific hybridization induces the mobilization of transposable elements and that this can be maladaptive (Guerreiro 2014). In interspecific hybrids of two wallaby species, reproductive isolation was facilitated due to the mobilization of retrotransposons, alongside chromosomal expansion and karyotype re-patterning (O'Neill et al. 1998). The formation of hybrid species in sunflower (*Helianthus*) has been accompanied by the amplification of retrotransposons and differences in genome size (Ungerer et al. 2006). Interestingly, intraspecific crosses may also reactivate the mobilization of previously inactive transposons, as has been observed in crosses between several rice cultivars (Kantama et al. 2013).

7. *Rhinanthus minor* and *Rhinanthus major* as model species

7.1. General biology

Our model species are *Rhinanthus major* Ehrh. and *Rhinanthus minor* L. from the Orobanchaceae family (order of Lamiales). The Orobanchaceae family is the largest parasitic plant family with more than 2,060 species in about 90–115 genera and a hybridization rate of 0.59% (Whitney et al. 2010; McNeal et al. 2013). Species in the Orobanchaceae family have different degrees of nutritional dependency, ranging from from non-photosynthetic parasites (holoparasitic) that are completely dependent on their host to photosynthetic parasitic (hemiparasitic) and non-parasitic species (Joel et al. 2013). The genus *Rhinanthus* is one of the hemiparasitic groups of in the Orobanchaceae family, with the ability to parasitize nearly 50 plant species. They attach their roots to the host's roots through haustoria and extract nutrition from the xylem. In general, germination of both species starts in early spring when there are enough developed host roots to attach (Cameron and Seel 2007). However, before germination the seeds require several weeks of cold stratification (Westbury 2004a; Ter Borg 2005), which in nature corresponds with the winter period, but it can also be mimicked in the lab by putting he seeds on moist filter paper at 4°C for about 5–8 weeks (Ter Borg 2005; Natalis and Wesselingh 2012; Ducarme and Wesselingh 2013). Germination of *R. minor* seeds usually starts after 38 days, while *R. major* seeds start germinating after 50 days under lab conditions (Wesselingh et al. 2019). Both *Rhinanthus* species are able to infect several other plant species, but graminoid and legume species are the most suitable hosts, due to their inability to mount cellular-level defenses at the site of haustorial intrusion (Matthies 2017). Both species are annual plants and their persistence and dispersal depend entirely on yearly seed production, due to the lack of a persistent seed bank (Coulson et al. 2001). Among the ± 25 species in the genus *Rhinanthus* , *R. minor* and *R. major* are the most widely distributed species in north and central Europe, where their distribution ranges show a large overlap (Figure 1.5) (Bullock et al. 2003; Westbury 2004b). Both species can grow in a diverse range of open habitats and they used to be very abundant in Europe, but nowadays, due to intensive agriculture and frequent mowing, in western Europe they are mostly restricted to nature reserves. Open areas with low to medium nutrient levels, from sea level to alpine zones, such as pastures, road sides, rocky outcrops, dunes and marshes are suitable habitat (Westbury 2004b; Ducarme et al. 2010). *Rhinanthus minor* is more often found in drier places, whereas *R. major* tends

to grow in more mesic conditions with high host densities. In transient patches, the frequency of *R. major* is low because of the local decrease of hosts and nutrients, and *R. minor* can grow in poorer regions where it is less optimal for *R. major* (Westbury 2004b; Bullock and Pywell 2005; Ducarme et al. 2010).

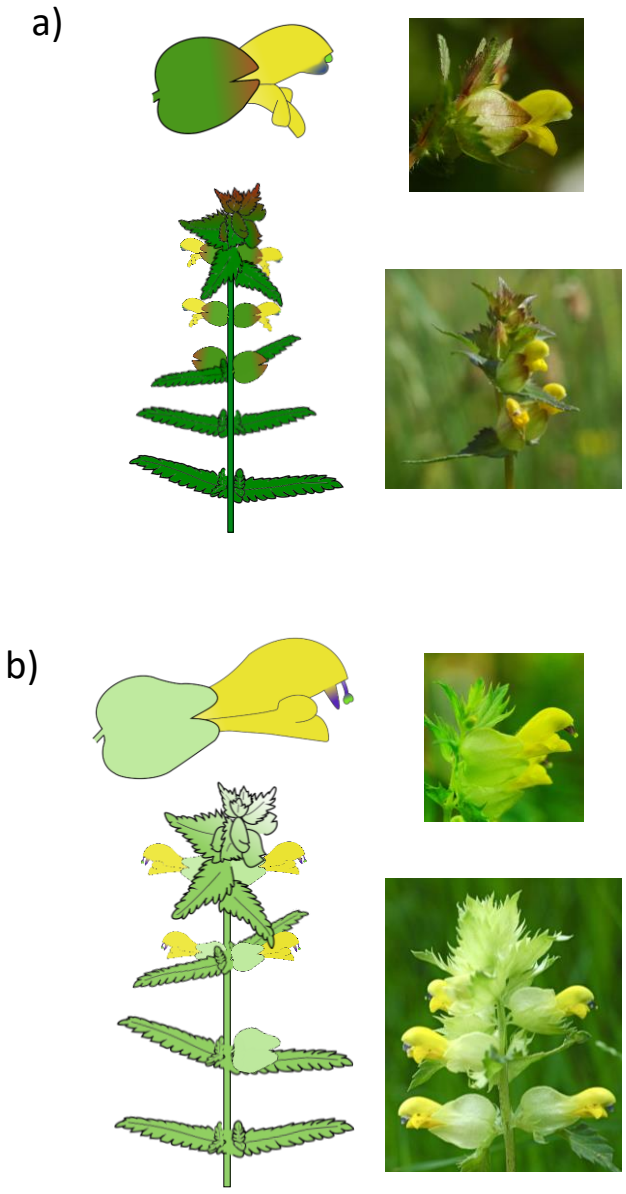


Figure 1.4. Inflorescences and flowers of *Rhinanthus minor* (a) and *R. major* (b). These illustrations were taken from the PhD thesis of Laurent Natalis (2013).

a)



b)



Figure 1.5. Distribution of *R. minor* (a) and *R. major* (b) in Europe. These illustrations were taken from the PhD thesis of Laurent Natalis (2013).

Table 1.1. Differentiating characters in *R. minor* and *R. major*

Character	*R. minor*	*R. major*	Reference
Corolla length	10–14 mm	16–23mm	Ducarme and Wesselingh 2005
Corolla tube shape	straight	curved	Kwak 1978
Tooth length	0.4–1.1 mm	1.5–2.4 mm	Ducarme and Wesselingh 2005
Stigma position	Enclosed in corolla	Protruding	Kwak 1978
Style length	10.50 ± 0.13 mm	16.52 ± 0.17 mm	Natalis and Wesselingh 2012
Calyx colour	green-reddish	whitish-pale green	Kwak 1980
N pollen grains per flower	$1.3*10^5$	$2.7*10^5$	Kwak 1979
Nectar production (µL/4h)	0.094 ± 0.017	0.230 ± 0.035	Natalis & Wesselingh, unpubl. data
Nectar sugar concentration (µg sucrose/µL)	183 ± 18	357 ± 36	Natalis & Wesselingh, unpubl. data
Nectar sugar composition (F:G:S)*	1: 0.4 : 0.3	1 : 0.2 : 0.9	Kwak et al. 1985
Species-specific volatile compounds	none	2-hydroxyacetophenone acetophenone	Mirzaei, Lorgnay & Wesselingh, unpubl. data
Fruit set of untouched flowers (autonomous self-pollination)	95.2%	17.4%	Ducarme and Wesselingh 2013

7.2. Patterns of hybridization

Germination of *R. minor* and *R. major* starts in February after a period of cold stratification (ter Borg 2005), and flowering occurs from May until the end of June with a large overlap in flowering time (Ducarme and Wesselingh 2009). Both species have yellow, tubular, zygomorphic flowers which are mainly pollinated by bumblebees, such as *B. hortorum, B. terrestris, B. pratorum* and *B. hypnorum* (Natalis and Wesselingh 2012). *Rhinanthus major* is taller and has pale bracts in its inflorescences, while *R. minor* is smaller with more reddish-green bracts (Kwak 1980, Figure 1.4 and Table 1.1). The co-occurrence of both species in proximity of each other provides opportunities for hybridization, due to the fact that their reproductive barriers are not complete (Ducarme and Wesselingh 2005). The hybrid offspring is viable, can successfully complete its life cycle and is compatible with the parental species (Kwak 1980). A genetic study of several mixed populations of both species at different ages has shown that in a mixed population 4.5-4.9 % of the plants in the first generation after mixing were F_1 hybrids (Ducarme and Wesselingh 2005; Ducarme et al. 2010). In older populations, 2.2% of assessed plants had intermediate hybrid genotypes, while advanced hybrids (beyond the F_1, generally closer to one of the parents) made up 13-26% of the analyzed individuals (Ducarme et al. 2010). Although only a limited number of species-specific makers could be used in these studies, the number of advanced hybrids close to *R. major* was significantly higher in old mixed populations compared to hybrids close to *R. minor*. These results reveal that unilateral introgression of *R. minor* genetic material into *R. major* populations is the main consequence of hybridization between these species (Ducarme and Wesselingh 2005; Ducarme et al. 2010).

7.3. Reproductive barriers

A plant's mating system can influence the occurrence and frequency of hybridization. Hybridization has a higher probability of occurring in outcrossing species compared to species with a high selfing rate (Ellstrand et al. 1996). Although *Rhinanthus minor* and *R. major* are similar in many characters, they differ in their mating system, which could provide a barrier for hybridization. Both species are self-compatible, but the observed outcrossing rate in *R. major* (0.76) is much higher than the outcrossing rate found for *R. minor* (0.13) (Ducarme and Wesselingh 2013). The probability of autonomous self-pollination in *R. minor* is high, due to the short distance

between stigma and anthers in comparison to *R. major*. However, there are still opportunities for *R. minor* for outcrossing, especially when the flower has just opened and the stigma is still protruding from the corolla (Ducarme and Wesselingh 2013). Several species of bumblebees visit *R. major* and *R. minor*, but most of them have a clear preference for *R. major*, which produces more pollen and nectar. Therefore, when *R. minor* is introduced into a *R. major* population, it is almost completely ignored, while bumblebees switch much more frequently, thereby promoting interspecific crossing, when *R. major* is introduced into a *R. minor* population (Natalis & Wesselingh 2013). F_1 hybrids also receive more visits than *R. minor* (Natalis & Wesselingh 2012), therefore unilateral introgression from *R. minor* into *R. major* is facilitated (Ducarme and Wesselingh 2013, Natalis & Wesselingh 2013).

Hybridization between divergent lineages often results in reduced hybrid viability, which can appear in just one of the two cross directions. This type of isolation, known as isolation asymmetry, can be traced back to Dobzhansky-Muller incompatibilities (DMIs): profound genetic modifications and remodeling of parental gene expression in the F_1 (Turelli and Moyle 2007). Under laboratory conditions, a significant difference in germination rate has repeatedly been observed between the reciprocal F_1 hybrids of our *Rhinanthus* species. The germination rate of hybrids formed on *R. major* (F_{1a} hybrids) is only 5-30%, whereas hybrids formed with *R. minor* as the maternal species (F_{1m} hybrids) is nearly 100%, which is a form of isolation asymmetry (Ducarme and Wesselingh 2013). In an experience under field conditions, this asymmetry was not found (Wesselingh et al. 2019), which suggests that fluctuating temperatures during stratification could play a role in reducing the difference between the reciprocal hybrids.

7.4. Chronology of the work on introgression and reproductive barriers between *R. minor* and *R. major* at UCLouvain

Two PhD students completed a thesis on introgressive hybridization between the two *Rhinanthus* species using genetic techniques in the period 2001–209, and in order to understand the origin of the questions asked in this thesis, it is useful to summarize their work on this theme in chronological order.

Véronique Ducarme (2001–2008). In 2001, Ducarme and Wesselingh (2005) studied two mixed populations of both species with different ages

using species-specific RAPD and ISSR markers. In a mixed but formerly pure population of *R. minor*, where *R. major* had been introduced the year before, both individuals of the two parental species and F_1 hybrids were found using both morphological and genetic analysis, showing that hybridization occurred rapidly. In a Dutch population that was mixed in 1978 (Kwak 1980), but consisted entirely of *R. major* when sampled in 2001, the genetic markers showed traces of past hybridization and introgression. Unidirectional introgression of *R. minor* into *R. major* was observed even after the complete disappearance of *R. minor* from this population. Ducarme et al. (2008) developed 15 polymorphic microsatellite loci which have been useful tools to study mating system, gene flow and hybridization in the genus *Rhinanthus*. Using only eight species-specific RAPD and ISSR markers, Ducarme et al. (2010) studied patterns and dynamics of three mixed populations of both species during four consecutive years (2001–2004). In all studied populations, most of the hybrids were close to *R. major*, a sign that unilateral introgression from *R. minor* into *R. major* was occurring. Ducarme and Wesselingh (2013) using microsatellite marker showed that outcrossing rate of *R. minor* (0.13) is lower than outcrossing rate of *R. major* (0.76). They also summarized the literature on germination rates of F_1 hybrids, showing that the low germination rate for F_{1a} hybrids is found systematically.

Jérôme Vrancken (2003–2009). Vrancken et al. (2009, 2012), using chloroplast DNA and 181 nuclear AFLP markers, provided a phylogeographic study at the European scale for both *Rhinanthus* species. Non-symmetrical introgression of *R. minor* into *R. major* was confirmed.

8. Objectives of this thesis

The general objectives of this thesis are 1) to use modern genetic techniques to study the previously observed asymmetric introgression from *R. minor* into *R. major* in much more detail, and 2) to better understand the observed asymmetric, post-zygotic barrier related to hybrid seed germination. The main objectives for the first part are: 1a) to develop a new, much larger set of genetic markers; 1b) to compare the results obtained on the occurrence of introgression with the new marker set with the results from the old, limited marker system using the same samples as Ducarme *et al.* 2010; 1c) to study introgression in newly sampled populations from different habitats in Belgium.

For the second part, the main objective was to find the cause of the low germination rate in F_1 hybrids with *R. major* as maternal parent. If the low germination rate is the result of a cytonuclear incompatibility during stratification, this should be detectable in the transcriptome, by differential gene expression of the plastid genome.

The first objective of this thesis project was to develop the ddRAD seq technique for our study species and obtain a large number of species-specific SNP markers (Chapter 2). We analyzed several individuals of *R. major* and *R. minor* from pure populations in order to obtain species-specific SNP markers. For developing these markers, we also added *R. minor* samples from several mixed populations of both species in Belgium in order to have sufficient diversity for the highly inbreeding *R. minor*. Since we knew that introgression is occurring from *R. minor* into *R. major,* the *R. minor* samples included were not affected by introgression from *R. major*. We managed to obtain over 1100 SNP markers that had species-specific, diagnostic alleles.

In chapter 3, we re-used samples of hybrids close to *R. major* from previous studies done in 2001–2004 by Ducarme and Wesselingh (2005) and Ducarme et al (2010). These studies used four dominant, species-specific markers (RAPD and ISSR) to detect introgression of *R. minor* into *R. major*, and the plants chosen for analysis carried a single *R. minor* marker. We used the new SNP panel to study levels of introgression and correlations among introgressed loci in more detail.

In chapter 4, the introgression analysis was widened with new samples that were collected in 2018 from mixed populations in different natural habitats, in order to investigate the effect of contrasting natural habitats on the (dis)similarity of introgression patterns. These results could make it possible to find traces of adaptive introgression from *R. minor* into *R. major*.

Finally, the only known post-zygotic barrier in *Rhinanthus* hybridization was further investigated in chapter 5, by comparing gene expression of reciprocal hybrid seeds during dormancy and germination. In this chapter we provide the first RNA reference for *Rhinanthus* and show up- and down-regulation of genes related to germination and maternal effects.

In chapter 6, the obtained results are discussed and placed in perspective and suggestions for further research are made.

Chapter 2

Development of a large set of diagnostic SNP markers[1]

Abstract
Rhinanthus major and *R. minor* are two annual plant species that are known to hybridize in nature and in which unilateral introgression is likely to occur. Here we used double-digest restriction site-associated DNA sequencing technology (ddRAD-seq) to detect 17,144 genome-wide SNPs in *R. major* and *R. minor* individuals from 9 populations. After screening these SNP markers across both species, we selected 1106 putative loci that contain diagnostic, species-specific SNPs, which can be used for assessing and monitoring hybridization and introgression between *R. major* and *R. minor*.

Key words: ddRAD-seq, diagnostic markers, SNP, *Rhinanthus*

[1]Adapted from Mirzaei, K. and Wesselingh, R.A., 2021. Development of a large set of diagnostic SNP markers using ddRAD-seq to study hybridization in *Rhinanthus major* and *R. minor* (Orobanchaceae). Conservation Genetics Resources, 13(1), pp.31–33.

Introduction

Hybridization and introgression are evolutionary phenomena that take place in around 25% of plant species and are likely to be one of the most important sources of evolution and domestication of flowering plants (Baack and Rieseberg 2007). When hybridization and especially introgression occur, accurate identification of species can be significantly impaired (Mallet 2005; Soltis and Soltis 2009).

Rhinanthus major Ehrh. (syn. *R. angustifolius* C.C. Gmelin) and *R. minor* L. (both 2n = 2x = 22: Hambler 1954) are root hemiparasitic annual plants occurring in a diverse range of grassland habitats throughout northern and central Europe. Wherever agricultural practices have been intensified, they have become less common and are often only found in nature reserves. Because they parasitize dominant grass species, they are important tools in the restauration of species-rich grasslands (Westbury et al. 2006, Těšitel et al. 2018). Their ecological niches overlap, so the species can co-occur (Westbury 2004; Ducarme et al. 2010).

There are only a few morphological features that can be used for distinguishing the two species, including bract color and corolla shape and size (Kwak 1978). Although some prezygotic barriers against hybridization exist (Natalis and Wesselingh 2012), hybrid formation occurs readily and F_1 hybrids perform as well as the parental species (Wesselingh et al. 2019). In later generations, an overall pollinator preference for *R. major* (Natalis and Wesselingh 2013; Ducarme and Wesselingh 2013) causes asymmetric introgression from *R. minor* into *R. major* (Ducarme et al. 2010). To study introgression in detail, we needed to greatly expand the limited set of markers we used previously (Ducarme & Wesselingh 2005, Ducarme et al. 2008).

Materials, methods and results

We used a total of 76 samples previously collected in the field or grown from seed in the greenhouse from our two seed source populations DB and HD. The 25 *Rhinanthus major* samples came from nature reserve Doode Bemde (province of Flemish Brabant). For *Rhinanthus minor* we used 51 samples, of which 26 came from two pure populations: 11 from nature reserve Housta-Dardenne (HD; Braine-le-Château, province of Walloon Brabant), and 15 from Jardin Massart (JM; Auderghem, Brussels). We also sampled in six mixed populations and selected samples that were pure *R. minor*: 4 from

Karthuizerduinen (KD; Nieuwpoort, province of West Flanders), 5 from Achter Schoonhoven (AS; Schoonhoven, province of Flemish Brabant), 5 from Ter Yde (TY, Oostduinkerke, province of West Flanders), 4 from Messelbroek (MB; province of Flemish Brabant), 4 from Kalkense Meersen (KM; Uitbergen, province of East Flanders) and 3 from Kijkverdriet (KV; Ravels, province of Antwerpen).

DNA was extracted from young leaves using the CTAB protocol (Doyle and Doyle 1987) and the ddRAD library was prepared based on the protocol described by Peterson et al. (2012). We tested different combinations of restriction enzymes and chose *Eco*RI and *Nla*III for double digestion. The ligated fragments were size-selected (325–425bp) using a Blue Pippin (BDF2010, Cassette type: 2% DF Marker Q2). Illumina indexes were added through PCR and the pooled libraries were paired-end sequenced (PE150) on the Illumina Hi-Seq 2500 platform at the Genomics Core of KULeuven and with Novaseq at Genewiz. Quality control checks on raw sequence data were performed using FASTQC (Andrews 2010). Filtering, demultiplexing and SNP detection were done with STACKS v. 2.1 (Minimum stack depth 5, Distance between stacks 3, Distance between catalog loci 3) (Catchen et al. 2013).

In total, 402,520,746 raw sequencing reads with high quality were generated from the 76 samples with an average of 5,296,325 reads per individual. Depth of coverage for the processed samples varied from 8x to 20x, with an average of 12x per individual. We initially obtained 11,171 loci that were shared by at least 70% of all individuals. After removing loci containing only SNPs with a minor allele frequency below 0.05, 8,545 loci with 16,932 SNPs in total were retained for further analysis. Of these, 57.7% were confirmed to be transitions, and the observed transition-transversion ratio between the two species was 1.36. Principal component analysis (PCA) of all discovered SNP markers divided the nine populations into four distinct groups (Figure 2.1).

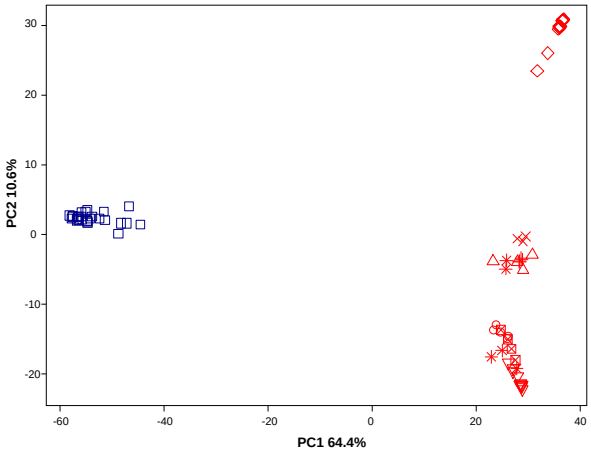


Figure 2.1. Principal component analysis (PCA) based on 16,932 SNP markers. *R. minor* populations are indicated by red shapes: AS (○), KD (Δ), TY (✳), MB (⊠), KM (+), KV (×), HD (◇), JM (▽) and *R. major* is indicated in blue (□).

Genetic diversity analysis performed in STACKS (Catchen et al. 2013) showed that the number of private alleles and the observed heterozygosity were higher in *R. major* than in *R. minor* (Table 2.1). The fixation index (F_{ST}) between the two species was 0.51, and a higher inbreeding coefficient was found for *R. minor* (0.22) compared to *R. major* (0.0468). This is in line with the differences in breeding system between the two species. Ducarme and Wesselingh (2013) found an average outcrossing rate of 0.76 for *R. major*, while this was only 0.13 for *R. minor*. The observed and expected heterozygosity in that study, which used six polymorphic microsatellite markers (Ducarme et al. 2008), were 0.717 and 0.670 for *R. major* and 0.124 and 0.329 for *R. minor*, respectively. This is much higher than for the SNPs, and this can be explained by the much higher variability that is inherent to microsatellites (Fischer et al. 2017).

Table 2.1. Values of genetic diversity within species and populations. A_P = number of private alleles, H_{obs} = observed heterozygosity, H_{exp} = expected heterozygosity, F_{IS} = inbreeding coefficient.

Species	*N*	A_P	H_{obs}	H_{exp}	F_{IS}
R. major	25	6038	0.165±0.001	0.171±0.001	0.046±0.03
R. minor	51	4393	0.062±0.001	0.144±0.001	0.225±0.01

We selected a subset of 1106 loci with species-specific SNPs (1800 SNPs in total) that were fixed completely for both species that were found in at least 70% of individuals as our diagnostic loci (Online Resource 1). These can now be used to study introgression in this species pair in great detail in mixed populations in Belgium and beyond.

Chapter 3

Introgression patterns in mixed *Rhinanthus* populations: increasing detectability with 1100 species-specific SNP markers

Abstract

Hybridization is a common phenomenon among plants and it can provide an important source of genetic novelty for selection to drive populations toward new ecological niches. *Rhinanthus major* and *R. minor* are two annual plant species that are known to hybridize in nature. In naturally mixed populations of both species, hybridization leads predominantly to introgression from *R. minor* into *R. major*, detected with morphological and molecular markers. Here we applied a set of 1106 previously developed species-specific SNP markers to make a detailed study of the patterns of introgression in advanced (F_2, F_3 or backcrosses) hybrids close to *R. major* in four previously studied mixed populations. Among the studied *R. minor* SNP markers, 2 to 5 showed a high frequency in the introgressed individuals in all populations, which suggest that they may be linked to a genomic region that confers a fitness advantage, while 256 *R. minor* markers were never found to introgress. These common patterns could point to intrinsic genetic (in)compatibilities among genomic regions of the two species. Beside the similarities of introgressing and non-introgressing loci among populations, unique non-introgressing and introgressing markers were also observed in each population. A population-specific presence or absence of certain markers could point at linked traits for which fitness in the *R. major* genome is habitat-specific.

Keywords: Introgression, Hybridization, Adaptation, SNP

Introduction

Hybridization and subsequent introgression, which has been described as 'an invasion of the genome' (Mallet 2005), is a fairly widespread phenomenon among plants. In many encounters between related species, reproductive barriers are the most difficult to overcome for the production of an F_1 hybrid generation (Mallet 2005b; Widmer et al. 2009b). Despite the problematic aspects of hybridization for species conservation (Vilà et al. 2000; Fitzpatrick et al. 2015), hybridization and consequently introgression are key evolutionary mechanisms, and the majority of plant species are considered to be outcomes of past interspecific hybridization events (Barton 2001a; Grant and Grant 2019). If hybrids are not reproductively isolated from the parental populations, inevitably their presence will prevent or revert speciation of the parental populations (Abbott et al. 2013). By providing a possibility for diverged species to mate again, hybridization can generate enough genetic novelty for selection to drive populations toward unoccupied niches and habitats (Kagawa and Takimoto 2018). This process, which also known as hybrid adaptive radiation, predicts that hybridization between different evolutionary lineages can cause rapid phenotypic diversification and facilitate adaptive radiation (Grant and Grant 2019). Genetic studies have recently confirmed such adaptive radiation following hybridization in numerous organisms (Litsios and Salamin 2014; Mitchell et al. 2019b). Hybrids often show new or extreme phenotypes compared to their parents, due to heterosis and transgressive segregation, which allows them to live in intermediate habitats or extreme environments where the parents cannot persist (Rieseberg et al. 2000; Barton 2001a). Hybridization can also produce reproductively isolated offspring, which can be the start of a speciation process known as hybrid speciation (Abbott et al. 2013). In most plants, hybrid speciation is a result of allopolyploidy, which rapidly isolates offspring from their parental species, but there are also examples of hybrid speciation without a change in ploidy level, homoploid hybrid speciation (Gross and Rieseberg 2005). Hybrids with a very low fitness, which cannot go beyond the F_1 generation, will not have an impact on

their parents' populations, but hybrids with high fitness, sometimes higher than their parents, will cause interspecific introgression of genetic material (Rieseberg et al. 2000; Barton 2001a; Grant and Grant 2019). If hybrid fitness is lower than that of the parents, reinforcement can lead to the evolution of prezygotic barriers, which protect against the wasteful formation of hybrid zygotes (Barton 2001a; Abbott et al. 2013).

The ubiquity of hybridization in many species and its function, through introgression, in the process of evolution and adaptation have been an interesting subject to study for many scientists, especially the detection of genes that pass across species boundaries and their interaction with the recipient genome. Molecular markers with a reliable distribution across the genome are needed for estimating interspecific gene flow and detection of introgressed linkage blocks (Twyford and Ennos 2012). With the development of PCR-based markers, introgression among many species has been proven, but they can underestimate or overestimate the occurrence of introgression (Baack and Rieseberg 2007). The main restriction of these markers is that they usually target less than a few hundred loci or, like cpDNA and mtDNA markers, depend on maternally inherited DNA (Lammerts van Bueren et al. 2010; Twyford and Ennos 2012). Therefore, because introgression can occur at very low frequencies, more markers are needed to have precise estimates of this event. The advent of technologies such as Next Generation sequencing (NGS), which generate a large quantity of nucleotide sequence data, promises to vastly improve our ability to study hybridization and introgression, by allowing new genomic tools to be generated for organisms with no prior sequence data (Twyford and Ennos 2012). Among methods derived from NGS technology, double digest restriction-site associated (ddRAD) sequencing has been introduced as a promising method to detect large numbers of single nucleotide polymorphism (SNP) markers across the genome without full genome information (Peterson et al. 2012).

Rhinanthus major Ehrh. and *R. minor* L. are annual facultative hemiparasitic plants from *Orobanchaceae* family with yellow, tubular, zygomorphic flowers with large overlap of ecogeographical niches (Hambler 1954a; Westbury 2004b; Ducarme et al. 2010). They parasitize the root system of their host and extract nutrition from the xylem (Cameron and Seel 2007). Although 50 plant species have been reported to be their hosts, graminoid and legume species are preferred hosts due to their inability to mount cellular-level defenses at the site of haustorial intrusion (Matthies 2017). *Rhinanthus minor* and *R. major* are able to produce viable and fertile F_1 hybrids alongside fertile F_2 and backcrosses. Both species are self-compatible, but a higher rate of outcrossing has been observed in *R. major*, which produces more pollen and nectar, while *R. minor* has a high propensity for autonomous self-pollination (Ducarme and Wesselingh 2013b). In presence of equal numbers of the two species, *R. minor* is often ignored by pollinating bumblebees, which makes that F_1 hybrids predominantly cross with *R. major* (Natalis and Wesselingh 2012). Using morphological characters like tooth and corolla length and eight

dominant species-specific genetic markers (ISSR and RAPD) in mixed populations of both species, several studies have shown that hybrids close to *R. major* always occurred in higher frequencies than hybrids close to *R. minor* (Kwak 1980; Ducarme and Wesselingh 2005; Ducarme et al. 2010). During four consecutive years of study in three mixed *Rhinanthus* populations in Belgium (Ducarme et al. 2010a), patterns of marker presence and absence suggested that advanced-generation hybrids are mainly formed by backcrossing with *R. major*. Although introgression from *R. minor* to *R. major* is clearly happening whenever the two species meet, the low number of species-specific markers made it impossible to shed light on the magnitude and nature of this introgression. We therefore, set out to apply our newly acquired panel of diagnostic SNP markers by ddRAD sequencing (Mirzaei and Wesselingh 2021) to track introgression in the same individuals that were studied by Ducarme and Wesselingh (2005) and Ducarme et al. (2010).

Materials and methods

Sampling

Of all the samples collected in four mixed populations in 2001–2004, we selected 152 samples. We used 24 samples from a population named Smalbroek (SB), part of the State Nature Reserve "Stroomdallandschap Drentsche A" near Assen, the Netherlands. This population had already been studied in 1977 and 1978 by Kwak (1980) when *R. minor* and *R. major* were both present. In 2001, when the samples were collected, all plants had a *R. major* morphology, and in only five of the sampled plants one *R. minor* RAPD marker (out of four possible *R. minor* markers) was detected. These five plants were included in this study, as well as 19 randomly chosen samples that did not show signs of introgression in the study by Ducarme and Wesselingh (2005). It is not known when *R. minor* disappeared completely from this population.

From the three populations studied by Ducarme et al. (2010), we re-used all the *R. major* samples in which the presence of a single *R. minor* marker had been detected, as well as 5 more samples of 'pure' *R. major*. This led to a total of 53 samples from the mixed population Kijkverdriet (KV; province of Antwerpen, Belgium) and 54 samples from a mixed population located in the nature reserve Kalkense Meersen near Uitbergen (KM; province of Oost-Vlaanderen, Belgium). Both these populations had been mixed since at least 10 years when the study started in 2001. Finally, we used 21 samples from an

experimental garden (JE) at the Jardin Massart of the Université Libre de Bruxelles in Auderghem (Brussels, Belgium). In this population, *R. minor* was already established and *R. major* had been introduced only one year prior to the start of the study in 2001, so the maximum number of generations of backcrossing was 3 (in 2004). We used 5 'pure' *R. major* samples from 2004 and 16 samples with one *R. minor* marker: 11 from 2004, 4 from 2003 and the single plant that carried a *R. minor* marker in 2001 (probably a trace of introgression in the source population of *R. major*). Further details about the sampling scheme can be found in Ducarme et al. (2010a).

Library preparation

DNA was extracted from frozen young leaves kept at –80°C using the CTAB protocol (Doyle and Doyle 1987) and quality and quantity of the extracted DNA were determined by agarose gel and fluorometer (Invitrogen Qubit). The double digestion was prepared (Peterson et al. 2012) using 600 ng DNA with EcoRI and NlaIII enzymes. After adapter ligations and pooling samples, purification was done using a Serapure magnetic bead suspension (Rohland and Reich 2012). Fragments with sizes between 325 bp and 425 bp were selected with a Blue Pippin (BDF2010, Cassette type: 2% DF Marker Q2) and ten replicates of PCR amplification in volume of 20 µl with concentration of 1X Phusion buffer, 2.5 mM dNTP, 1 pmol RAD primers (for each), 0.2 u of Phusion™ Polymerase and 1 µl of the size-selected DNA were performed at the following conditions: a starting temperature of 98°C (30 sec), followed by 11 cycles of 98°C (10 sec), 60°C (30 sec) and 72° (30 sec), followed by storage at 4°C. Then PCR products were pooled and purified using the Serapure magnetic bead suspension and the prepared library was paired-end sequenced (PE150) on the Hi-Seq 4000 platform at GENEWIZ.

Bioinformatics analysis

After checking the quality of the raw sequence data using FASTQC (2020), Trimmomatic-0.39 (Bolger et al. 2014a) was used to exclude reads with a quality less than Q20. Demultiplexing of high quality reads was done using STACKS v. 1.48 (Catchen et al. 2013). The *denovo_map* module was then used with the following criteria: a maximum distance of 3 allowed between stacks (-M), a minimum depth of 5 for coverage to create a stack (-m) and 2 mismatches allowed between sample loci (-n). Finally, the *populations* module was used for the loci that appeared in ≥ 70% of individuals, (in order to avoid loci with a high amount of missing data, which can cause bias) and

only SNPs with a global minor allele frequency (MAF) of 0.05 or higher were kept for further analysis.

Data analysis

Using the filtered SNPs from previous step, a principal component analysis (PCA) of the 152 *R. major* samples from four (formerly) mixed populations were performed using Tassel software (Bradbury et al. 2007) and the R-package ggplot (Warnes et al. 2009). Population structure analysis was done using STRUCTURE version 2.3.4 (Hubisz et al. 2009) using 2 as the number of groups, in order to obtain the admixture rate for each sampled individual. Therefore, three independent runs with K = 2 were performed with a burn-in length of 10,000 followed by 100,000 iterations. Finally, *R. major* individuals with less than 5% admixture with *R. minor* were separated for further analysis of introgressed loci. Using just *R. major* individuals with less than 5% admixture rate for detecting loci with a high or low rate of introgression is justified by the fact that after 6-7 generations of backcrossing a genetic trait can be fixed in the gene pool (3.125% admixture after four generations of backcrossing, see Fig. 1.4). Individuals with high admixture rates might be the result of recent hybridizations or very early advanced hybrids (first-generation backcross).

In order to detect clusters of markers with the same introgression patterns, we performed a principal coordinate analysis (PCoA) on the species-specific markers that appeared in more than 60% of the studied samples. *Rhinanthus major* markers were coded as 1 and introgressed *R. minor* markers (both homozygous and heterozygous) were coded as 2. The dissimilarity matrix of markers was calculated with the Simple Matching Coefficient using the "smc" method in the R-package "scrime" (Fritsch 2009), and a PCoA biplot was drawn using the R-package ggplot (Warnes et al. 2009) to determine how markers cluster according to their patterns of presence/absence of the *R. minor* allele, including the four original RAPD/ISSR markers for *R. minor*.

Results

In total 1.314 billion paired-end raw reads were generated and nearly 98% of them with an average of ~8.47 million paired-end reads per individual retained for further analysis. The average of sequencing depths for each sample was between 7.21 and 24.76X. In total 2,026,999 stacks loci were detected from all individuals among, of which 8,837 loci containing 13,228 SNPs with MAF > 0.05 were shared between more than 70% of all individuals. The transition-

transversion ratio of the obtained SNPs was 1.32, with 56.79% transitions. General population-genetic information is given in Table 3.1 and Table 3.2.

Table 3.1. Values of genetic diversity within populations. A_P = number of private alleles, H_{obs} = observed heterozygosity, H_{exp} = expected heterozygosity, (F_{IS} = inbreeding coefficient).

Population	A_P	H_{obs}	H_{exp}	F_{IS}
SB	194	0.198±0.003	0.20±0.03	0.03±0.019
JE	48	0.236±0.06	0.219±0.03	0.0023±0.004
KM	69	0.203±0.04	0.221±0.02	0.098±0.04
KV	165	0.238±0.01	0.242±0.01	0.046±0.04

Table 3.2. Pairwise comparison of genetic distances between populations using F_{ST} values

	JE	KM	KV	SB
JE		0.059	0.076	0.124
KM			0.058	0.084
KV				0.096

A PCA of the four studied populations clustered them into four groups (Figure 3.1) with total variation of 18.02% explained by the first three principal axes (PC1: 9.72%, PC2: 8.29% and PC3: 5.18%). The Smalbroek population from the north of the Netherlands was clearly separated from the three Belgian populations.

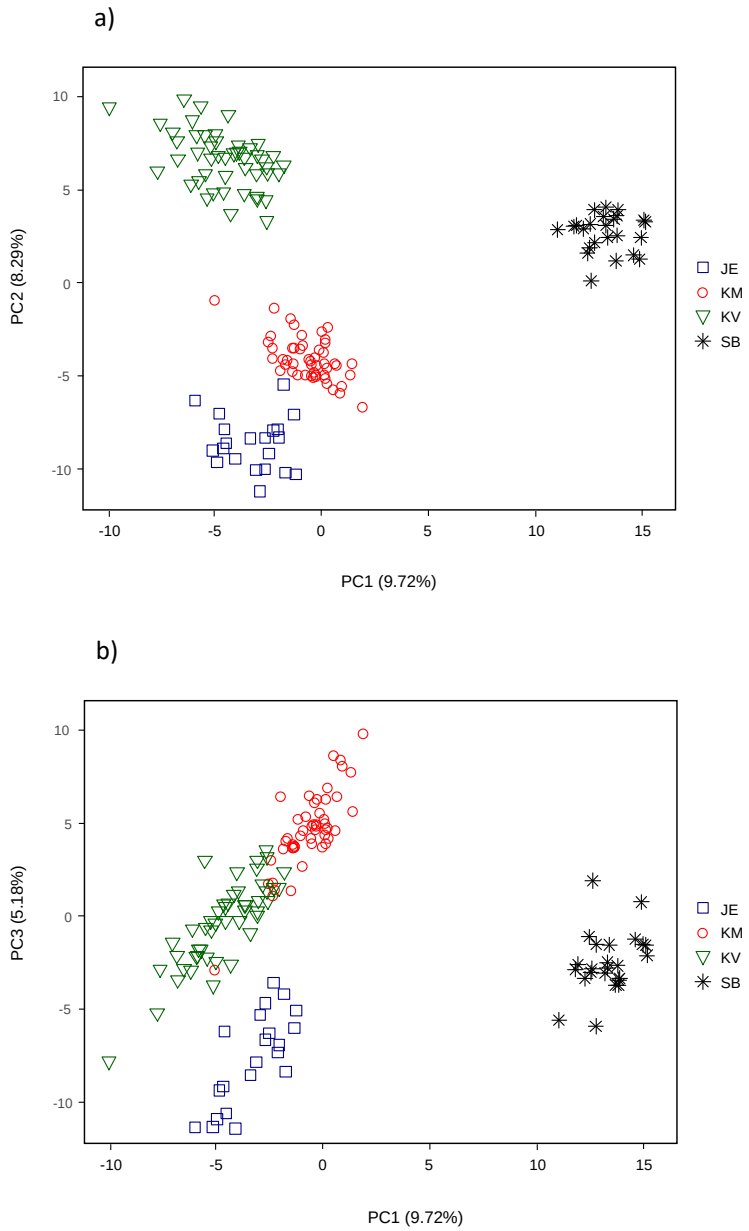


Figure 3.1. Principal component analysis (PCA) of the 152 *R. major* samples from four (formerly) mixed populations, showing the first three axes.

The distribution of admixture rates varied among the populations (Figure 3.5). In the Smalbroek population, only three samples had admixture rates higher than 5%, while this was less than 5% in the other 21 individuals. In the other three populations, 9, 39 and 23 *R. major* individuals, respectively from JE,

KM and KV, had less than 5% admixture (Table 3.3). In the samples from JE, the median percentage *R. major* was close to the expected value for just two generations of backcrossing (87.5%). The number of loci detected at least in one individual out of the 1106 species-specific SNPs in this library varied between 1039 (KV) and 1097 (JE).

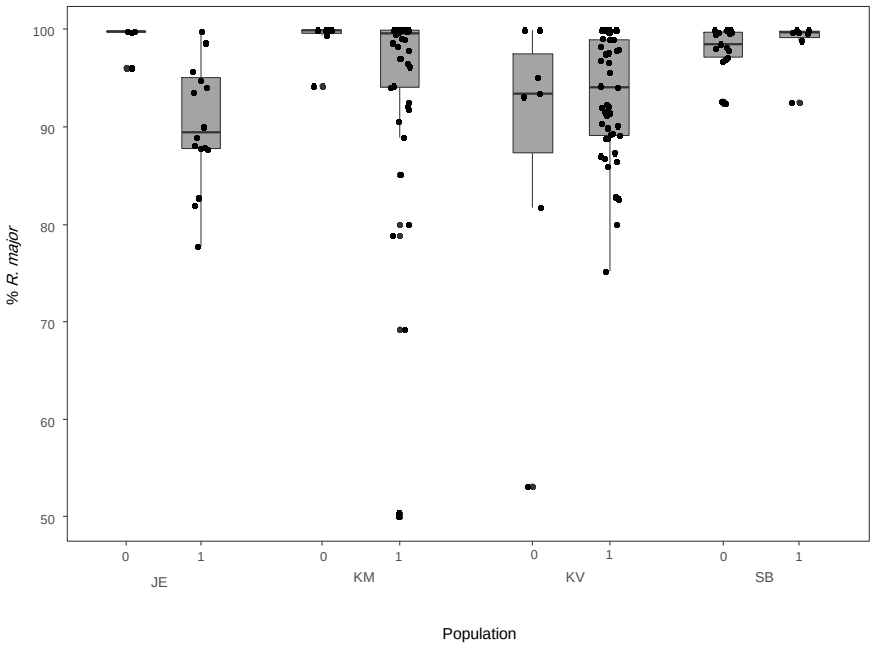


Figure 3.5. The % of *R. major* markers present (= 100 – admixture rate) for each sample in the four populations. Samples are grouped by the number of *R. minor* RAPD/ISSR markers (0 or 1) detected by Ducarme and Wesselingh (2005) and Ducarme et al. (2010).

Table 3.3. Number of sequenced individuals in each population with number of individuals with low admixture rate

	JE	KM	KV	SB
Number of samples	21	53	54	24
Individual with admixture rate < 5% *R. minor*	9	39	23	21
Number of introgressed loci	514	518	597	258
Number of non-introgressing loci	583	548	442	833
Number of introgressed loci with high frequency*	5	2	2	3

Tracking the *R. minor* markers only in individuals with an admixture rate less than 5% showed that the KV population contained the highest number of these *R. minor*-specific markers, 597, while the SB population, with 258 loci, had the lowest number. The similarity analysis of these introgressed loci showed that 111 of them were shared among all populations, while KV had the highest number of unique introgressed loci, namely 109 (Figure 3.2). According to the Venn diagram for the similarity analysis of non-introgressed loci (Figure 3.3), 256 loci were shared among the studied populations and SB, with 220 loci, had the highest number of unique non-introgressed loci. The highest number of similar non-introgressing loci, 488, was shared between SB and KM, which represented respectively 58.6% and 89.1% of their non-introgressed loci. The highest similarity of introgressed loci were the 420 loci shared between KV and KM, which equalled nearly 80% of their introgressed loci. Five (JE), two (KM), two (KV) and three (SB) introgressed loci appeared in more than 70% of studied individuals with a low admixture rate. Among these introgressed loci with high frequency, only two were shared among populations: one was found in JE, KM and KV, while the other was shared between JE and SB. In the PCoA, the markers clustered in two major groups, a small and a large cluster, the latter spread along the second axis (Figure 3.4).

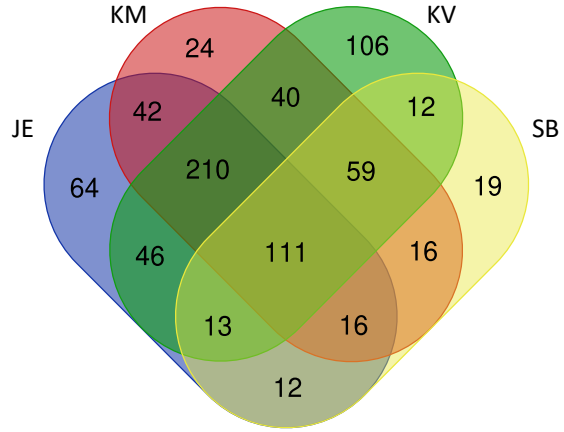


Figure 3.2. Number of similar and unique introgressed loci in each population

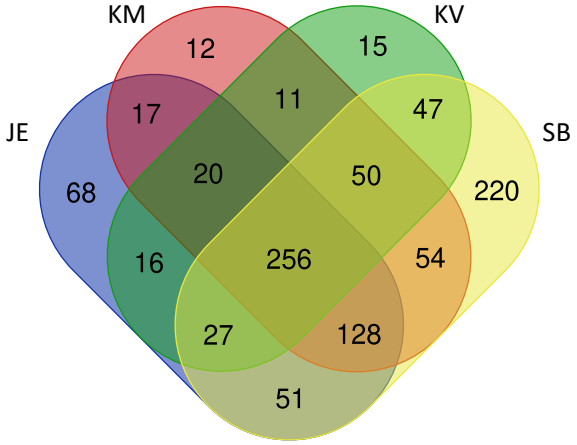


Figure 3.3. Number of similar and unique non-introgressing loci in each population

Discussion

In all four mixed populations, asymmetric hybridization with different levels of introgression was found and this is in accordance with the obtained results of Kwak (1980) using morphological markers, and the studies by Ducarme and Wesselingh (2005) and Ducarme et al (2010) using ISSR and RAPD molecular markers. However, all the *R. major* individuals that were considered to be pure in the two previous studies showed some level of introgression with our SNP data. Since the genome of *Rhinanthus* is quite large (1.4 Gb; Castro et al. 2012), the use of just a few markers is clearly not

enough to correctly estimate the degree of presence of the *R. minor* genome in *R. major* individuals, and the use of a large panel of SNP markers is clearly needed to obtain a much more accurate estimate.

The genetic distance of the Dutch SB population, which is located far from KV, KM and JE populations, logically was higher than the genetic distances among the three Belgian populations. The SB population also showed a higher number of private alleles, which can again be explained by its higher genetic distance. Interestingly, even though there were no *R. minor* individuals present at the moment of sampling in the SB population, there were still *R. minor* markers present due to backcrosses that took place between hybrids and pure individuals of *R. major* in the past. In the other populations, where *R. minor* was also present, some individuals (KV-21, KM-15 and KM-23), despite carrying at most only one *R. minor* marker in the previous study, had a ± 50% admixture rate based on our SNP panel, which demonstrates how inaccurate estimating introgression based on only four *R. minor* markers can be. The levels of admixture rate in the populations that still contained *R. minor* varied more than in the SB population, in which *R. minor* was absent. We do not know if the markers in our SNP panel are scattered evenly across the genome and we have no information on their linkage, but thanks to the high number of SNP markers, these introgression estimates should be much more reliable than in the previous studies with few molecular markers. In the youngest mixed population, JE, the median value for the percentage R. major markers was just under 90%, which is expected to be reached after only 2 generations of backcrossing. This suggests that not only the F_1 hybrids present in 2001, but also F_1 hybrids flowering in 2002 contributed to subsequent backcrossing towards *R. major*.

All loci that were identified by Mirzaei and Wesselingh (2021) as loci containing species-specific markers (Table 3.3) were also detected in this study. This confirms the repeatability of this method and its suitability for studying introgression and hybridization, also in a population located outside Belgium. The PCA for plants with a low admixture rate clearly separated the samples into four groups according to their population of origin. Therefore, introgression of the *R. minor* genome into these *R. major* populations did not influence the genetic distances among populations, even though the levels of introgression were different.

The PCoA showed that the markers were divided into two major groups, based on the similarity of their patterns of introgression in the studied samples (Figure 3.4). These patterns might be indicative of linkage groups or blocks of introgressing genomic regions, but this can only be verified by mapping the

markers onto a reference genome. The identification of markers with very similar introgression patterns was hampered by the amount of missing loci, which was quite high in some samples. The four original RAPD/ISSR *R. minor* markers were spread out in the PCoA, which was to be expected since all the samples used in this study had at most one *R. minor* marker, so the presence of one marker automatically meant the absence of the three others.

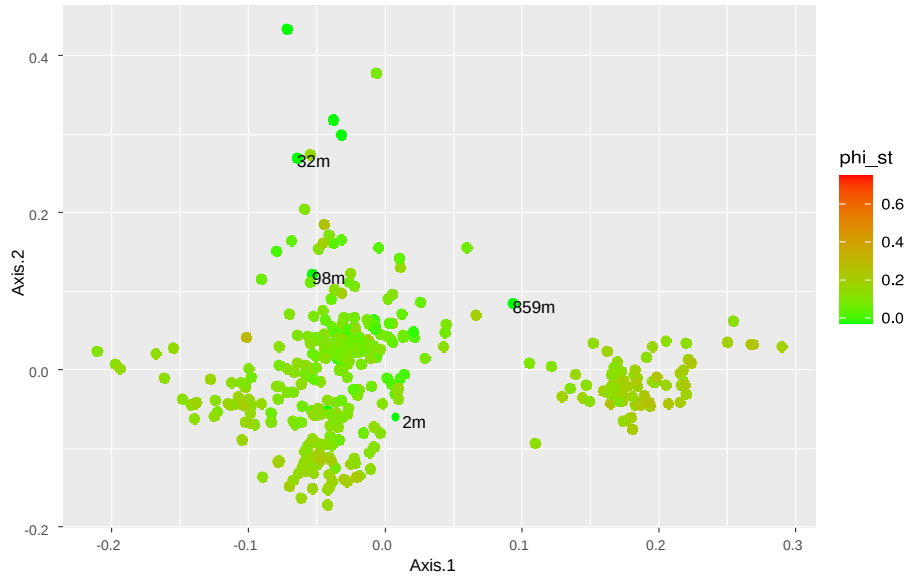


Figure 3.4. PCoA for the markers with less than 40% missing data in all studied individuals of each population. Each dot is representing a single marker, markers with low phi_st (red) have a high introgression rate and markers with high phi_st (green) have low introgression rates. 32m, 98m, 859m and 2m are the RAPD markers have been used in Ducarme and Wesselingh (2005) and Ducarme et al (2010) studies.

The percentage of individuals with a low (< 5%) admixture rate was different in each population, ranging from 42.85% in JE to 87.5% in SB. These individuals with a low admixture rate are the result of several backcrosses of hybrids with pure individuals or with almost pure *R. major* with some fixed *R. minor* markers. Therefore, they provide a good opportunity to know which parts of the *R. minor* genome is more likely to persist in the *R. major* genome. Among the observed introgressed loci, a large number of the loci were common in either all of the populations or in two or three of them. These loci

are likely to be free of conflicts with the *R. major* genome and thus able to introgress in any mixed population. Only a very low number was involved in introgression in only a single population, which might indicate that these loci confer specific advantages for that habitat and population only. The similarity analysis of non-introgressed loci showed that 23.14% of them was never found in individuals with low admixture rates. Some of these loci were common between two or three populations, but in each population, unique loci that were not involved in introgression have been observed. These loci may have been removed from the predominantly *R. major* individuals with only a few generations of backcrosses, due to their incompetency or by chance. The number of loci with a high frequency was different in each population and only a few of them were shared among several populations. The occurrence of these markers with high frequency in each population might be linked to genes that confer general fitness advantages for *R. major*. Unique introgressed markers with a high frequency could be related to loci that give an advantage in that specific population only.

Conclusion

Our SNP marker panel works well in mixed *Rhinanthus* populations from two different countries and gives an accurate estimate of the admixture rate, based on 0.14% of the genome. The similarities in introgression patterns of groups of markers across populations, which represent independent introgression events, suggests that these markers are physically close to each other in the genome and/or closely linked to blocks of coding loci that confer an intrinsic or extrinsic advantage. This can only be proven with a linkage map or a full reference genome, both of which are currently in the making.

Chapter 4

Introgression from *Rhinanthus minor* into *R. major* in various habitats

Abstract

Rhinanthus major and *R. minor* are two annual plant species that are known to hybridize in nature. In naturally mixed populations of these two hemiparasitic species, most of the advanced-generation hybrids found are close to *R. major*. This is likely due to its higher attractiveness to the pollinating bumblebees, so introgression of genetic material takes place from *R. minor* into *R. major*. Using 1106 diagnostic SNP markers, we tracked patterns of introgression using ddRADseq in 157 samples from five mixed populations of both species in different habitats in Belgium. Over 95% of the loci containing species-specific markers were detected again in the studied populations. A one-by-one assessment of the introgressed loci showed that many of them are present in each population. Among the studied markers, only a few showed a high frequency in hybrids close to *R. major,* and most of them had a low frequency in the studied populations. We did not find similarities among the population in the introgressed loci present in high frequency except two that were shared between two pairs of populations.

Keywords: Hybridization, Introgression, Adaptation, *Rhinanthus*

Introduction

Hybridization or interspecies mating and consequently introgression have been reported among many species. When mating between species occurs without facing reproductive barriers, it can provide an important source of new allelic variation within the receiving populations (Seehausen 2004; Mallet 2005b). Like migration, which also introduces new genetic material in a population, introgression is quicker than mutation (Ellstrand et al. 1999). An abundance of genetic novelties in the source population may increase of fitness and adaptation in the introgressed individuals (Barton 2001b). Although there are prezygotic and postzygotic mechanisms preventing hybridization, the exchange of genetic material between related species has happened in a wide range of species and in 25% of plant species (Mallet 2005b; Widmer et al. 2009b). This type of gene flow will also provide an opportunity for accumulation of favourable epistatic genes and promote rapid adaptation (Porter and Johnson 2002; Sackman and Rokyta 2013). Moreover, gene flow by hybridization has the potential to enter large sets of new alleles simultaneously, which allows adaptation for polygenic traits and further acceleration of species evolution (Barton 2001b; Sackman and Rokyta 2013). Introgressed loci are expected to be either neutral or adaptive (Suarez-Gonzalez et al. 2018b, b). Adaptive loci, which tend to introgress more frequently than loci involved in reproductive isolation, will be maintained by selection, so they will have a chance to become fixed in the receiving population, but neutral introgressed loci might be removed by drift over several generations (Holderegger et al. 2006; Kirk and Freeland 2011; Suarez-Gonzalez et al. 2018b; Schrider et al. 2018).

Rhinanthus major Ehrh. and *R. minor* L. are root hemiparasitic annual plants from the largest parasitic angiosperm family, the Orobanchaceae. Both of them are diploid with 2n=22 chromosomes and they are the most widely distributed species in the genus, and can be found in north and central Europe and north America (Hambler 1954a; Westbury 2004b; Ducarme et al. 2010). They can grow in a diverse range of open habitats, with low to moderate nutrient levels, from sea level to alpine zones.

Hybridization takes place between *R. major* and *R. minor* whenever they occur together. In newly mixed populations, 2.2-4.5% of individuals in the first generation are F_1 hybrids, and in later generations, 13–26% are advanced hybrids close to *R. major*, due to asymmetric introgression from *R. minor* into *R. major* (Kwak 1980; Ducarme and Wesselingh 2005; Ducarme et al. 2010). A higher rate of autogamous self-pollination in *R. minor* (Ducarme

and Wesselingh 2013) and pollinator preference for *R. major* and intermediate hybrids are the main causes of this asymmetric introgression (Natalis and Wesselingh 2012).

The use of molecular markers has allowed to describe patterns of hybridization and introgression among many species, but the number of markers that could be used was limited before the arrival of high-throughput sequencing methods (Twyford and Ennos 2012). Using ddRADseq (Peterson et al. 2012), we developed a panel of more than 1100 SNP markers that reliably distinguish *Rhinanthus minor* and *R. major* (Mirzaei & Wesselingh 2021, Chapter 2), and have used them successfully to determine the level of admixture and patterns of introgression in previously detected advanced hybrids close to *R. major* (Chapter 3). In this chapter, we extend the use of the markers to new populations, without prior knowledge of the hybrid status of the samples, and from ecologically contrasting habitats. Presence of SNP markers from *R. minor* with a high frequency in *R. major* in all populations can be related to genomic regions with a general fitness advantage. *R. minor* SNPs that are absent or found in low frequencies in the studied populations are likely to be linked to loci with disadvantageous alleles for hybrid fitness. SNP markers with a high frequency in some habitats and absent or in low frequency in other habitats are indicative of introgressing alleles that only convey a fitness advantage in certain habitats.

Material and methods

Plant material

In June 2018, we visited several sites in Flanders based on information obtained from https://www.observations.be that suggested that both species were present, and randomly sampled young leaves on a total of 355 plants, for which we noted the approximate phenotype (*R. minor*, *R. major*, or hybrid) based on flower morphology and bract colour. For this study, we selected populations that were mixed and large enough, and chose to use samples that had *R. major*-like and hybrid phenotypes. In total, we used 157 samples from five populations (Table 4.1 and Figure 4.1).

Table 4.1. The five sampled populations with the numbers of samples collected and used in this study

Code	Geographic coordinates	Full name	Province	Habitat	Total number of samples	Number of samples used in this study
TY1	51° 01' 07" N 2° 41' 32" E	Ter Yde 1	West Flanders	Mesic dune grassland	87	24
TY2	51° 01' 07" N 2° 41' 32" E	Ter Yde 2	West Flanders	Dry sand dune, wet dune grassland	111	67
KD	50° 08' 27" N 2° 42' 42" E	Karthuizerduinen	West Flanders	Dune valley	22	12
AS	50° 59' 05" N 4° 51' 05" E	Achter Schoonhoven	Flemish Brabant	Brook valley	64	29
MB	50° 59' 44" N 4° 56' 14" E	Messelbroek	Flemish Brabant	Dry grassland	71	25

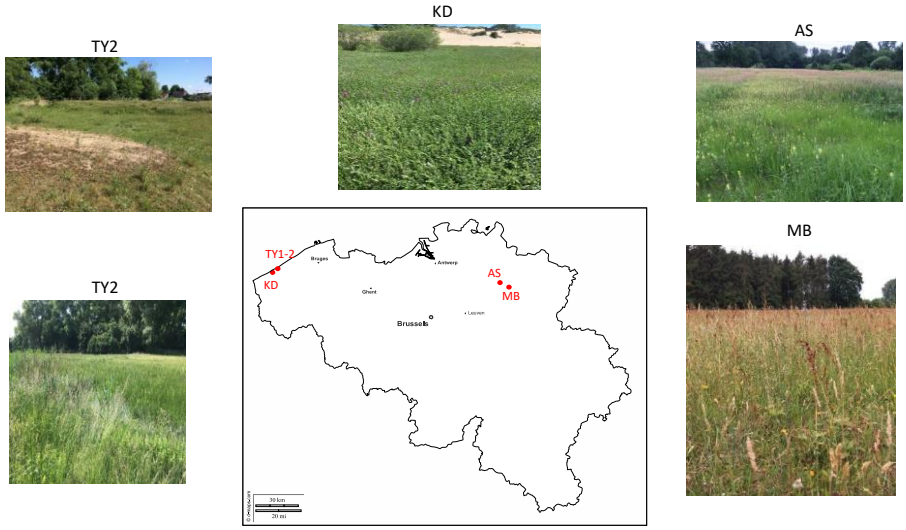


Figure 4.1. Locations of the sampled populations and photos taken in the respective populations.

Library preparation

DNA was extracted from frozen young leaves using the CTAB protocol (Doyle and Doyle 1987) with the following modifications. During DNA precipitation, 450 µl isopropanol and 50 µl sodium acetate (3.0 M, pH 5.2) were mixed with the supernatant and samples were kept for 5 min at room temperature. The samples were then centrifuged for 10 min (11,000 rpm at 4°C). After removing the supernatant, the pellet was washed twice with ethanol 70% and centrifuged for 5 min (7,000 rpm at 4°C). After decanting the ethanol from the washed pellets and drying the samples in a Speed-Vac vacuum concentrator, TE buffer was added to the samples. Quality and quantity of the extracted DNA were determined by agarose gel and fluorometer (Invitrogen Qubit).

The double digest RAD library was prepared based on the protocol described by Peterson et al. (2012). 600 ng DNA of each sample was digested in 15 µl of reaction volume containing 1.5µl Cutsmart buffer (10X), 10 u EcoRI enzyme and 10 u NlaIII enzyme (New England Biolabs) during 7 h at 37°C without heat killing of the enzymes. 200 ng of double digested DNA for each individual was ligated in a total volume of 10 µl containing the two adaptors (2.5 pmol for each), 1 µl buffer (10X) and 50 u T4 DNA ligase (New England

Biolabs) in 16°C for 6 h. Then samples were pooled and purified using Serapure magnetic bead suspension (Rohland and Reich 2012). Then size selection was done by running pooled library on agarose gel and sizes between 325 bp and 425 bp were purified from the gel using agarose gel DNA extraction kit (Sigma-Aldrich).

Ten replicates of PCR amplification in volume of 20 µl with concentration of 1X Phusion buffer, 2.5 mM dNTP, 1 pmol RAD primers (for each), 0.2 u of Phusion™ Polymerase and 1 µl of the size-selected DNA were performed at the following conditions: a starting temperature of 98°C (30 sec), followed by 11 cycles of 98°C (10 sec), 60°C (30 sec) and 72° (30 sec), followed by storage at 4°C. After checking for PCR amplification on agarose gels, the PCR products were pooled and purified using Serapure magnetic bead suspension. The prepared library was paired-end sequenced (PE150) on the NovaSeq platform at GENEWIZ Company.

Bioinformatics analysis

First quality control checks on raw sequence data were performed using FASTQC (Andrews, 2010) and reads with low quality were excluded using Trimmomatic-0.39 (Bolger et al. 2014a). Demultiplexing of reads were done using the *process_radtags* module in STACKS v. 1.48 (Catchen et al. 2013). Using the *ustacks* module, putative stacks (loci) were detected with a maximum distance of 3 allowed between stacks (-M) and a minimum depth of coverage of 5 to create a stack (-m). The stacks catalog was made using *cstacks* with default adjustment except for the number of mismatches allowed between sample loci (-n), for which we used 2. The *sstacks* module was used to match each sample against the obtained catalog, and *populations* module was then used for the loci that appeared in ≥ 70% of individuals. SNP loci with a global minor allele frequency (MAF) of less than 0.05 were discarded to avoid false SNP identification. A PCA was done for all of the samples using TASSEL (Bradbury et al. 2007) and the R-package ggplot (Wickham 2016).

The vcf files obtained from the *populations* module were visualized with TASSEL (Bradbury et al. 2007) to detect introgressed loci in plants that had a *R. major* phenotype. R version 4.0.3 (R Core Team 2020) was used for making plots and VCFtools (Danecek et al. 2011) for calculating the transition vs. transversion (ts/tv) ratio. For population structure analysis, we used a model-based genetic clustering algorithm as implemented in STRUCTURE version 2.3.4 (Hubisz et al. 2009). The number of groups was determined using the ad-hoc statistical method, based on the rate change in the log probability of data between successive K values (ΔK). Five independent runs for K values

ranging from 1 to 3 were performed with a burn-in length of 10,000 followed by 100,000 iterations. We then used STRUCTURE with three replicate runs and K = 2 to estimate the admixture proportions for each populations. *R. major* individuals with ≤ 5% of admixture proportions with *R. minor* were included for further analysis of introgressed loci.

We performed a principal coordinate analysis (PCoA) for species-specific markers that appeared in more than 60% of studied individuals. In this analysis, *R. major* markers were coded as 1 and introgressed *R. minor* markers (homozygous and heterozygous) were coded as 2. The similarity matrix of markers was calculated with the "smc" function in the R-package "scrime" (Fritsch 2009) and the biplot was drawn with the R-package ggplot (Wickham 2016). The PCoA performed for all the populations and each time one population was dropped in order to see effect of population on relatedness of studied markers (Fritsch 2009).

Results

In total, ~1.484 billion paired-end raw reads were generated from the 157 samples with an average of ~9.45 million reads per individual. After removing sequences with low quality, ~1.465 billion (98.71%) paired-end reads were retained for further analysis. Depth of sequencing for the processed samples varied between 6.73X to 11.58X with an average of 8.82X per individual. Initially, 6,511,544 stacks loci were obtained from all individuals and among them 15,688 loci, corresponding to a genomic size of 0.168% of the *R. major* genome (Ellstrand et al. 1999), were shared among at least 70% of sequenced individuals in each population.

After removing loci with MAF < 0.05, 8,596 loci containing 7,986 SNPs were retained for further analysis. Out of the obtained SNPs we found that 54.38% of them were transitions, and the ts/tv ratio was 1.327.

Genetic diversity analysis showed that the number of private alleles (A_P) observed in TY2 (405) was higher than in the other populations, while KD, AS and MB had no private alleles. The observed heterozygosity (H_{obs}) varied between 0.114±0.002 (KD) to 0.276±0.003 (TY2) and the lowest and highest expected heterozygosity (H_{exp}) were observed respectively in KD (0.127±0.005) and TY1 (0.319±0.04) (Table 4.2). A higher inbreeding (F_{IS}) coefficient was observed in TY2 (0.08) in comparison with other populations, and the highest and lowest fixation indices (F_{ST}) were detected in MB and KD (0.119), and AS and MB (0.042), respectively (Table 4.3). The PCA analysis of obtained SNP's for all of the individuals clustered them into three groups (Figures 4.2 and 4.3). In total 32.91% of variation in the data was explained

by the first three principal coordinates (PC1: 18.88%, PC2: 7.23% and PC3: 6.8%).

Table 4.2. Values of genetic diversity within populations. A_P = number of private alleles, H_{obs} = observed heterozygosity, H_{exp} = expected heterozygosity, (F_{IS} = inbreeding coefficient).

Population	A_P	H_{obs}	H_{exp}	F_{IS}
AS	0	0.205±0.005	0.20±0.005	0.03±0.002
KD	0	0.114±0.002	0.127±0.005	0.07±0.009
MB	0	0.254±0.002	0.289±0.005	0.0273±0.007
TY1	2	0.253±0.003	0.319±0.004	0.0703±0.001
TY2	405	0.276±0.003	0.289±0.004	0.0803±0.001

Table 4.3. Pairwise comparison of genetic distances between populations using F_{ST} values.

	KD	TY1	TY2	MB
AS	0.095	0.088	0.067	0.042
KD		0.065	0.064	0.119
TY1			0.058	0.080
TY2				0.046

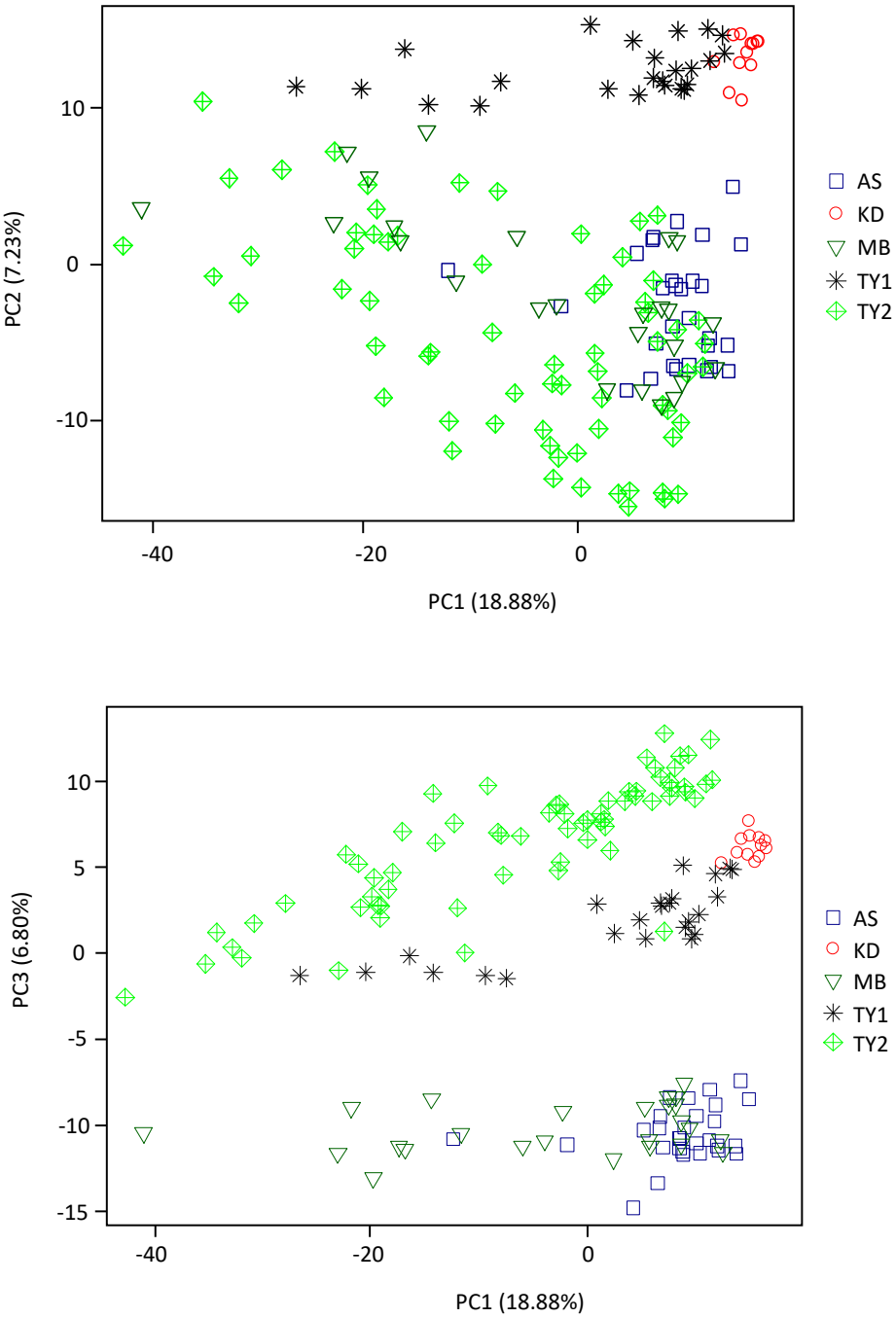


Figure 4.2. Principal component analysis (PCA) of all 157 samples in five mixed populations. **a)** PC1-PC2. **b)** PC1-PC3.

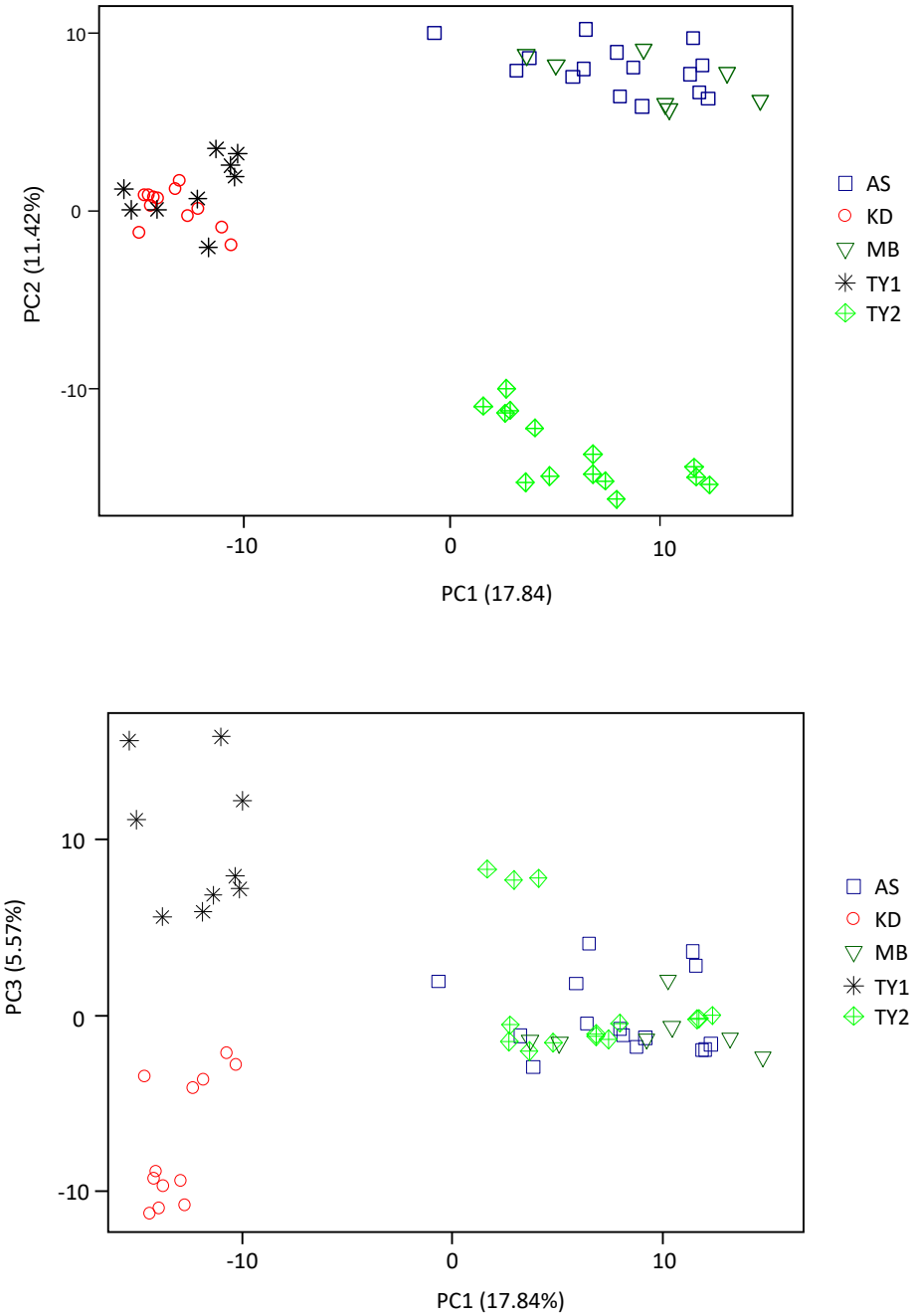


Figure 4.3. Principal component analysis (PCA) on 57 *R. major*-like samples with ≤ 5% *R. minor* admixture in five mixed populations. **a)** PC1-PC2. **b)** PC1-PC3.

The admixture rate (% of *R. minor* markers present) ranged from 16.8 to 99.99% (Figure 4.7). The samples collected in KD had the lowest admixture rate in comparison with other populations. After excluding samples with more than 5% admixture rate with *R. minor* markers, 58 samples (16 AS, 12 KD, 7 MB, 9 TY1 and 14 TY2) remained for further analysis of introgression patterns. Among these samples, the number of loci containing species-specific SNP markers that appeared in at least one individual varied from 1067 (AS) to 1096 (KD) out of 1106 in our SNP panel. The KD population had the highest number of non-introgressed loci (loci without *R. minor* alleles), 994, while TY2 had the lowest (302). The number of introgressed loci with a high frequency among the individuals studied varied from 1 in TY1 and TY2 to 5 in AS (Table 4.4). The highest similarity of non-introgressed loci was detected between TY2 and KD, which shared 93.95% and 73.44%, respectively, of their non-introgressed loci (Table 4.5). Among the introgressed loci, the highest similarity was detected between MB and AS (Table 4.6).

Table 4.4. Number of sequenced individuals in each population with number of individuals with low admixture rate.

	AS	KD	MB	TY1	TY2
Total number of samples used	29	12	25	24	67
Samples with maximum 5% *R. minor* admixture	16	12	7	9	14
Number of introgressed loci	501	102	549	319	302
Number of non-introgressing loci	566	994	528	766	777
Number of introgressed loci with high frequency*	5	2	4	1	1

**R. minor* markers that appeared in ≥ 70% of *R. major*-like individuals in a population

Table 4.5. Number of shared non-introgressing loci among individuals with at least 95% of *R. major* markers present. The numbers in parentheses represent the total number of non-introgressing loci found in each population.

	KD (994)	MB (528)	TY1 (766)	TY2 (777)
AS (591)	554	339	474	470
KD (994)		485	729	730
MB (528)			386	397
TY1 (766)				564

Table 4.6. Number of shared introgressed loci among individuals with at least 95% of *R. major* markers present. The numbers in parentheses represent the total number of introgressed loci found in each population.

	KD (102)	MB (549)	TY1 (319)	TY2 (302)
AS (476)	71	303	209	171
KD (102)		66	64	58
MB (549)			185	175
TY1 (319)				108

Overall, 228 non-introgressed and 27 introgressed loci were shared by all the populations (Figures 4.4 and 4.5). Out of the *R. minor* markers with high frequency in *R. major* individuals, only one was shared, between AS and MB. The PCA analysis of individuals with a low admixture rate with *R. minor* clustered them into four groups (Figures 4.2 and 4.3), and the first three principal coordinates explained 34.87% of their variation. In the PCoA on the markers, four clusters were formed (Figure 4.6).

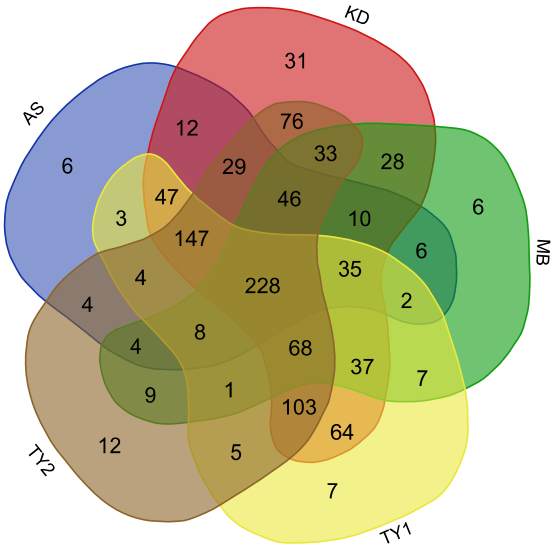


Figure 4.4. Numbers of shared and unique non-introgressing loci among in each population.

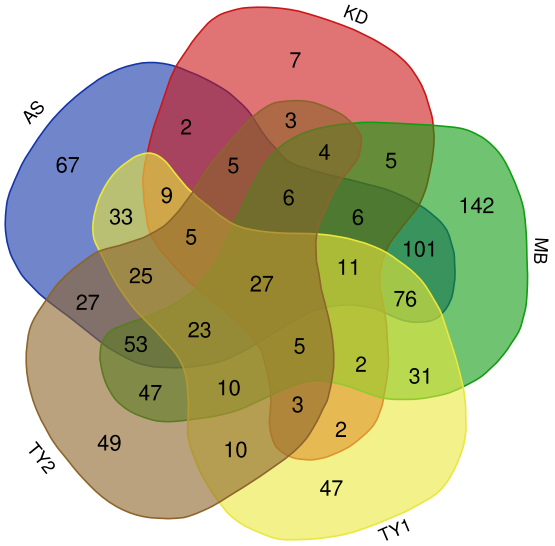


Figure 4.5. Numbers of shared and unique introgressing loci among in each population.

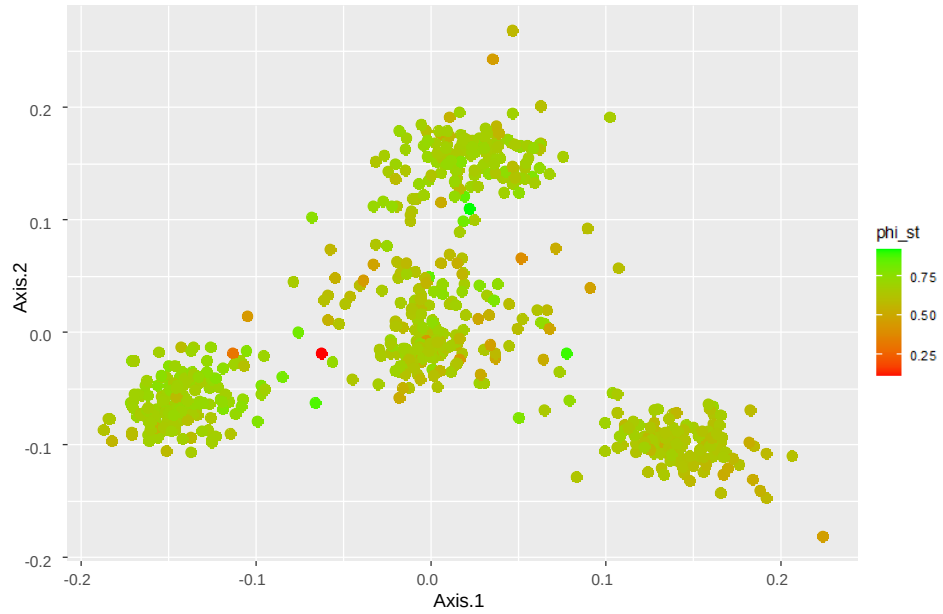


Figure 4.6. PCoA on the species-specific markers with less than 40% missing values in all studied individuals of each population, for all five populations together. Each dot represents one marker, and markers that are close together have similar patterns of presence/absence. Markers with low phi_st have a high rate of introgression.

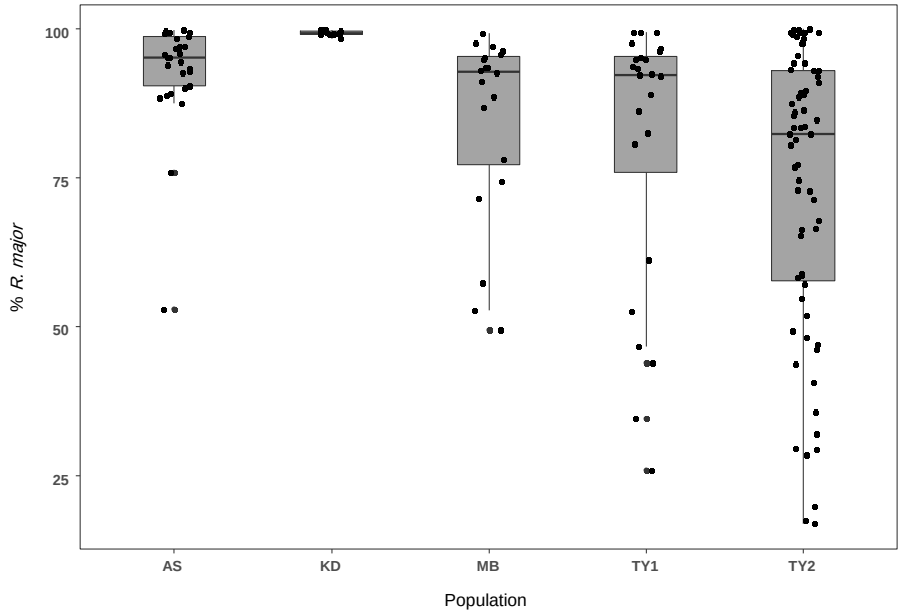


Figure 4.7. The distribution of admixture rates as determined by STRUCTURE in the five populations. Number of individuals in each population with different levels of *R. major* markers or levels of admixture rate. * In this picture individuals with high admixture rate is samples that is more close to *R. major*.

Discussion

Our SNP panel (Mirzaei and Wesselingh 2021, Chapter 2) was again highly reproducible in these newly sampled populations: out of 1106 loci with species-specific SNP markers, we were able to obtain over 96% in the studied populations. The low inbreeding coefficient (F_{IS}) and high observed (H_{obs}) and expected heterozygosity (H_{exp}) observed in the studied populations are in line with previous studies on *R. major* and can be explained by its high outcrossing rate (Ducarme and Wesselingh 2013). The ranges of H_{obs} and H_{exp} in this study are 0.114–0.276 and 0.2–0.319, respectively, while in Chapter 3 these ranges were 0.196–0.238 and 0.208–0.234. The observed F_{IS} is between 0.02 and 0.08, very similar to what we found in Chapter 3: 0.02–0.09. Similar ranges for H_{obs}, H_{exp} and F_{IS} were also detected in Chapter 2 (Mirzaei and Wesselingh 2021). The low levels of heterozygosity in the KD population might be related to low levels of admixture with *R. minor*. The low numbers of private alleles and low levels of F_{ST} among studied populations suggest that some of these

populations might have had opportunities to exchange genetic material through pollinators or seed dispersal from one population to another. In the PCA analysis, the presence of two clusters is based on geographical distance: AS and MB are in the east of Flanders, while KD, TY1 and TY2 are in West Flanders, on the North Sea coast. Within these provinces, grasslands in nature reserves are increasingly mown by machinery that is shared over a large territory. This has already be shown to lead to *Rhinanthus* seed dispersal on mowing machines (Strykstra et al. 1996) among previously isolated populations in the Netherlands. This might explain the similarity among populations within the same province. The PCA analysis on only *R. major*-like samples, with a low level of admixture from *R. minor*, split the samples into four clusters according to their location, but these had low F_{ST} and A_p indexes. The transition /transversion mutation (ts/tv) rate has been used as an indication of genetic distance between species like peanut (Gupta et al. 2015) and *Arabidopsis* (Ossowski et al. 2010). The high level of the ts/tv ratio (1.327) was in line with the low level of genetic diversity among populations and similar with the ts/tv ratio observed in Chapter 3.

The levels of genetic admixture between *R. minor* and R. *major* varied widely among the populations. According to our results, almost all of the loci with species-specific markers were involved in introgression in *R. major* individuals with low levels of admixture in at least one population. Only 27 loci were involved in introgression in all of the studied populations, which were different from the shared introgressed loci in Chapter 3. Loci with introgressing *R. minor* markers that had a high frequency in one population were also involved in introgression in the other populations, but they never showed high frequencies there. When comparing the introgressing markers with high frequency to those appearing in hybrids close to *R. major* in the populations studied in Chapter 3, we did not find any similarity between them, while 156 loci that were not involved in introgression were shared with non-introgressing loci in Chapter 3.

Not all alleles or loci of one species have the same chance to participate in introgression, because across the genome introgression is heterogeneously expanded. In the genome there are regions, like functional regions or regions related to species boundaries, that are less likely to participate in introgression (Payseur 2010; Harrison and Larson 2016; Schrider et al. 2018). Beside a general low tendency of some regions to introgress, the rate of introgression can be affected by the intensity of natural selection in each population (Schrider et al. 2018). Therefore, alleles with advantages for the receiving population tend to have a higher frequency than alleles linked to

disadvantageous or neutral characters (Harrison and Larson 2014). Another mechanism that might have an effect on the frequency of the studied alleles is genetic drift. Genetic drift is a process in which allele frequencies within a population change by chance alone as a result of sampling error from generation to generation. Population size has a strong influence on the proportion of allele frequencies that deviate from neutral expectation (Star and Spencer 2013).

The similarity of introgression patterns in the species-specific markers was analysed by PCoA, which yielded four distinct clusters. These patterns might be caused by genetic linkage among these loci, or by linkage with *R. minor* alleles for functional traits that convey an advantage in the recipient *R. major* genome.

Conclusion

In this study, we again identified a series of *R. minor* markers that have introgressed into *R. major,* suggesting that there might be both adaptive introgression and neutral introgression going on in these mixed populations. However, the amount of missing data (non-amplified loci) was quite high in some individuals. In order to decrease the amount of missing data, the use of methods with the capacity to capture all targeted loci is needed in future studies. Because of the large diversity in habitats among the sampled populations, varying patterns of introgression among these populations might be related to the adaptive value of linked functional alleles obtained from *R. minor*. This can only be proven if we can place these loci onto an annotated genome of *R. minor*, which is not yet available.

Chapter 5

Transcriptome profiling and comparison of *Rhinanthus major* and *R. minor* reciprocal F_1 hybrids during seed stratification and germination

Abstract

Germination is a vital stage in a plant's life cycle, and a different germination behavior of offspring in comparison to their parents can have fitness consequences. In studies on hybridization between *R. minor* and *R. major*, low germination rates of F_1 hybrids with *R. major* as the maternal parent have often been reported. In contrast, the F_{1m} hybrid, with *R. minor* as the maternal parent, germinates readily and rapidly. In order to find the cause of this difference, we used RNA-Seq to obtain transcriptome profiles of F_{1a} and F_{1m} seeds during stratification at 4°C and just after germination, after 40 days of stratification for the F_{1m} seeds and 60 days for the F_{1a} seeds. A comparison of the transcriptome of F_{1a} seeds that had just germinated (60 days) with non-germinated F_{1a} seeds after 40 and 60 days revealed 2918 and 1349 differentially expressed (DE) genes, respectively. For F_{1m} seeds, 958 genes showed differential expression in germinated and non-germinated seeds after 40 days. The DE genes of F_{1a} and F_{1m} hybrids clustered into two separate groups, even though they had the same parents, and no differentially expression was found for plastid genes. Non-germinated F_{1a} seeds had an abundance of enzymes and proteins associated with peroxidase activity, peroxiredoxin activity and nutrient reservoir activity. Expression of genes related to seed germination and seed development increased in non-germinated F_{1a} hybrid seeds between 40 and 60 days of cold stratification. F_{1a} seeds that had germinated showed an upregulation of genes related to the gibberellic acid-mediated signaling pathway and response to gibberellin, along with a low expression of DELLA superfamily. Although the results demonstrated strong differences in gene expression during stratification

between the reciprocal hybrids, we could not identify its cause, since no plastid genes were differentially expressed. It is possible that differences in embryo development after seed formation and before stratification play a role, including epigenetic imprinting.

Keywords: Hybridization, Reproductive barrier, Maternal effect, Transcriptome, RNA-seq

Introduction

Natural hybridization is a source of novel genetic material for evolution and it can occur among species due to a lack of reproductive barriers. It is more common among plants than animals (Rieseberg 1997), and although it is rare in some families, in others, such as *Dennstaedtiaceae*, it is very common phenomena (Whitney et al. 2010). When hybridization occurs, the fate of the F_1 hybrid can vary from showing hybrid vigor—which has greatly benefited agriculture—to hybrid breakdown, which can be a key factor in speciation (Rieseberg 1997; Arnold et al. 1999; Johansen-Morris and Latta 2006). Hybrid survival can play an important role in determining the effect of hybridization on the future composition of population, given that a hybrid can be a bridge for introgression or become a new species (Stukenbrock 2016; Grant and Grant 2019). By providing opportunities for gene flow between sister species, hybridization can cause a decrease or an increase (reinforcement) of reproductive barriers (Pickup et al. 2019).

Most flowering plant species are hermaphrodite, and therefore the formation of reciprocal hybrids is often possible, but these reciprocal hybrids do not always have the same performance (Christopher et al. 2019). Asymmetry in the fitness of reciprocal F_1 hybrids has been reported in many plant species (Tiffin et al. 2001; Turelli and Moyle 2007). This type of isolation, known as isolation asymmetry, is different from Dobzhansky–Muller incompatibilities (DMIs) because the reciprocal F_1 hybrids carry the same autosomal genotype (Turelli and Moyle 2007). The combination of two divergent genomes through hybridization will introduce immediate and profound genetic modifications and remodeling of parental gene expression in the F_1, which can cause genome shock and have wide effects on species establishment and diversification (Paun et al. 2007; Kerbs et al. 2017).

Due to extrinsic or intrinsic factors, reciprocal hybrids may show differences in fitness and can be different from their parents in the same environment (Paun et al. 2007). According to Darwin's corollary to Haldane's rule, isolation asymmetry can be an outcome of different relative rates of evolution in the parents of the F_1 hybrid (Turelli and Moyle 2007). *Rhinanthus major* Ehrh. and *R. minor* L. (both $2n = 2x = 22$: Hambler (1954b)) are annual plants in the Orobanchaceae family, occurring in a diverse range of open habitats (Westbury 2004a; Ducarme et al. 2010). Both species are self-compatible, but a higher outcrossing rate has been documented for *R. major* (76%) in

comparison with *R. minor* (13%; (Ducarme et al. 2010)). Both species are hemiparasites that take organic carbon and mineral nutrients from the root systems of their hosts, even though they are able to photosynthesise (Rümer et al. 2007). Their roots have found to be attached to nearly 50 plants species from 18 families, but the most suitable hosts are members of the Poaceae and Fabaceae (Gibson and Watkinson 1991).

Seed germination as a first stage of plant transition to development phase is composed of a series of steps, from water absorbance to embryo development and digestion of the starch and protein reserves of the seed (Bentsink and Koornneef 2008). If seed germination is mainly controlled by dormancy (primary or secondary dormancy), this can provide the best timing for plant growth and development to reduce extrinsic mortality (Hoyle et al. 2015). Differences in seed dormancy within the same species can place the earliest and the latest plant in different environments with different challenges and this can provide an opportunity for adaptive divergence (Donohue et al. 2005; Hoyle et al. 2015). Naturally, seeds of *Rhinanthus* species have physiological dormancy that can be broken by several weeks of cold stratification (Westbury 2004a; Ter Borg 2005; Marin et al. 2019). Under laboratory conditions, a significant difference in germination rate has repeatedly been observed between the reciprocal F_1 hybrids. The germination rate of hybrids formed on *R. major* (F_{1a} hybrids) is only 5-30%, whereas for hybrids formed with *R. minor* as the maternal species (F_{1m} hybrids) it is nearly 100%, often even better than the parental species (Kwak 1979; Campion-Bourget 1980; Natalis and Wesselingh 2012; Ducarme and Wesselingh 2013). However, this difference was not observed in a field experiment (Wesselingh et al. 2019), which seems to indicate that the experimental conditions used for germination in the laboratory are not representative for the germination behaviour of the hybrids under more natural conditions. Marin et al. (2019) have shown for *R. minor* that seed features related to quality such as germination capacity and seed vigor can play an important role in plant emergence, establishment and performance. They also found that elongation and growth of *R. minor* embryo continued after the seeds are released from the fruit. Anatomical studies on *R. major* seed showed that the seed contained a well-developed embryo before the start of stratification (Tiagi 1966).

Gene expression undergoes important changes during seed development, and comparative studies of transcriptomes of seeds during dormancy and germination have helped scientists to understand the processes that are taking place. Among expression profiling technologies, RNA sequencing

(RNA-Seq), a next–generation sequencing (NGS) technology, is a comprehensively informative technique to monitor wide transcriptional changes during various stages of plant life. Therefore, we used RNA-Seq to study the gene expression during stratification and germination of both F_{1a} and F_{1m} hybrids.

Material and methods

Plant materials and RNA extraction

Hybrid seeds were produced by reciprocal hand pollination on plants grown from seed in the greenhouse in 2019, applying *R. major* pollen onto *R. minor* stigmas after emasculation to produce F_{1m} hybrid seed, and *R. minor* pollen onto *R. major* stigmas for F_{1a} hybrid seeds. Seeds were collected from dry capsules in June-July 2019 and stored at room temperature. Thirty seeds in three replicates from each cross type were placed in petri dishes for germination in a refrigerator (± 4°C) in October 2019. We checked the dishes regularly for germination, which was scored as soon as the radicle was visible. We sampled germinated seeds at an early stage, with less than 1 mm of the radicle protruding from the seed, and sampled non-germinated seeds at the same time, after 40 days and 60 days of stratification. RNA was extracted from seeds using the QIAGEN RNeasy Mini Kit (3 seeds per sample) and the sample characteristics are provided in table 5.1.

Table 5.1. Characteristics of the samples used for RNA-seq

Samples	ID
Non-germinated F_{1m} after 40 days	F1m_40_NG
Germinated F_{1m} after 40 days	F1m_40_G
Non-germinated F_{1a} after 40 days	F1a_40_NG
Non-germinated F_{1a} after 60 days	F1a_60_NG
Germinated F_{1a} after 60 days	F1a_60_G

The columns used for RNA extraction were washed again by TE buffer and centrifuged for 5 minutes at 10000 rpm to remove the DNA. The retrieved DNA was used for PCR (20 µl final volume, 2 µl of 10 X buffer, 0.4 mM

dNTPs, 1.5 mM MgCl2, 2.0 U Taq polymerase, 5 pM forward primer and 5 pM reverse primer) with species-specific primers (5'CACCCTGATTTCTCTTTCTTCAA, 5'TTAAGACCCCATAAAAAGGAGGA) and obtained PCR products were digested by RsaI enzyme to confirm if they were indeed hybrid (Wesselingh et al. 2019; Mirzaei and Wesselingh 2021). Finally, the RNA of 15 hybrid samples was sent to GENEWIZ for sequencing on Illumina NovaSeq (2*150 bp) platform.

De novo assembly and sequence annotation

Preliminary quality checks of all sequenced samples were performed using FastQC and quality trimming and adaptor removal were done using Trimmomatic-0. 39 (Bolger et al. 2014b). *In silico* normalization of reads was done using the normalization option in Trinity package and *de novo* transcriptome assembly was performed using three assemblers: Trinity (Grabherr et al. 2011), Trans-ABySS (Robertson et al. 2010) and SPAdes-rna (Bushmanova et al. 2019). A first appraisal of the quality of the assembled transcriptomes was performed by estimating representation read counts in each assembly using bowtie2 (Langmead and Salzberg 2012). Transcriptome completeness was explored using BUSCO v.4.1.2 (Benchmarking Universal Single Copy Orthologs) (Simão et al. 2015) to obtain the percentage of single-copy orthologues represented. Plotting of the BUSCO results was performed using the ggplot package in R v4.03 (R Core Team 2020). Functional annotation of transcriptome was conducted using the Trinotate (Bryant et al. 2017). Prediction of coding regions in transcripts and open reading frames (ORFs) were performed by TransDecoder v5.5.0 (Grabherr et al. 2011). UniProtKB/Swiss-Prot and plant NCBI NR (NCBI non-redundant protein database) databases were used for homology searches (Ye et al. 2006) and HMMER v.3 (Finn et al. 2011) and Pfam (Punta et al. 2012) were used for protein domain identification. Signal peptide predictions was obtained using signalP v.445 (Peterson et al. 2012) and transmembrane regions were predicted using the tmHMM v.246 server (Krogh et al. 2001). Ribosomal RNA genes were detected with RNAMMER v.1.247 (Lagesen et al. 2007) and finally annotation outputs were loaded into a Trinotate SQLite Database.

Transcript abundance and differential expression analysis

To estimate transcript abundance, we used the alignment-based RSEM as abundance estimation method. Therefore, we used bowtie2 for alignment, and the sorted alignment file in BAM format was generated by SAMtools-1.3.1

(Li et al. 2009) and used for RSEM program (Li and Dewey 2011). Finally, a matrix of normalized expression values were estimated with the abundance_estimates_to_matrix.pl script in the Trinity package. The resulting matrix of normalized expression values was fed into the count_matrix_features_given_MIN_TPM_threshold.pl script for ExN50 analysis. The ExN50 statistics was calculated by Trinity accessory scripts contig_ExN50_statistic.pl and plotted with ggplot in R. Differential expression was preformed using adapted pipeline in Trinity which included Bioconductor v3.4, edgeR v4.0 (Robinson et al. 2010), Limma (Ritchie et al. 2015), ctc (Lucas and Jasson 2006), Biobase (Huber et al. 2015), gplots (Warnes et al. 2009). The transcript differential expression analysis was performed on the matrix of raw read counts using the edgeR R package. The false discovery rate raw *p*-values were adjusted for multiple comparisons by the Benjamini-Hochberg method (Haynes 2013). A false discovery rate [FDR] < 0.05 and $|\log2FC| \geq 2$ (positive or negative used for over- or under-expression, respectively) were used as criteria for identifying significant differences in expression. Gene set enrichment analyses of GO terms were conducted on each set of differentially expressed transcripts to determine over-represented functional pathways. GO enrichment analysis of differentially expressed (DE) genes was performed using Fisher exact test, P-value ≤ 0.05 and the hypergeometric Fisher exact test ($P < 0.05$) and Benjamini (FDR < 0.05) were used to detect statistically significant enrichment of the KEGG pathway (Kanehisa et al. 2016).

Results

Germination and RNA-sequencing

In the F_{1m} hybrid seeds, germination started after 34 days and after 70 days 90% of seeds had germinated. On the contrary, F_{1a} hybrid seeds started to germinate after 39 days and even after 120 days in the refrigerator the germination rate was still under 17%. The mRNA libraries of the five treatments generated over ~296 million raw paired-end reads, out of which ~263 million high quality paired-end reads ($Q \geq 20$) remained with ~17.53 million reads per sample. The final lengths of remaining reads were between 36 and 151bp, with a GC content of approximately 48% in all samples (Table S5.1).

Assembly

The Trinity assembly, with more complete reads (91.8%) and a higher representation of sequenced reads (97.53%) had a better quality in comparison

with the two other assemblies (Table 5.2). The total length of assembled reads, number of transcripts and the N50 index of Trinity assembly were also higher than in the Trans-ABySS and SPAdes-rna assemblies. Therefore we used the Trinity assembly for further analysis.

The Trinity assembly contained a total of 356379 transcripts (206,255 trinity genes) with an average length of 861bp and N50 value of 1,558bp (Table 5.2). The lengths of the assembled transcripts ranged from 200 to 22,700 bp and about 74.34% of the transcripts were in the range of 201–1000 bp (~1 kb), 23.98% transcripts were 1001–4000 bp (1.1 to 4.0 kb) and 1.68% was longer than 4001 bp (>3 kb) (Figure S5.1). Plotting transcript expression (Ex) against ExN50 value-as another overall quality checking of assembly- revealed that saturation point of the assembly was at 92% of the total expression with length of 1884 bp (Figure S5.2). According to ExN50 value, deeper sequencing is very unlikely to provide longer reads for our transcriptomes (Figure S5.2).

Table 5.2. Summary of assembly statistics generated by various pipelines. BUSCO Eukarya database OrthoDB v.10 busco genes.

Workflow	**Trinity**	**Trans-ABySS**	**SPAdes-rna**
Representation of RNA-seq reads	97.53%	82.8%	74.59%
BUSCO Results			
Complete	91.8%	87.3%	58.1%
Fragmented	3.1%	8.6%	22.3%
Missed	5.1%	4.1%	19.6
General assembly metrics			
Length, Mbp	306	298	163
Number of transcripts	356379	317485	169473
N50, bp	1558	1461	1272

Annotation

In total 173,638 ORFs were predicted, including: 92241 (53.12%) as complete ORFs contained a starting codon for methionine and ending stop codon, 32193 (18.54%) as 5´ partial ORF which lacked the start codon, 18825 (10.84%) as 3´-partial ORF and 30329 (17.46%) as internal ORF which were partial at both 5´ and 3´. Nearly 78% of the ORFs and 66% of the transcripts were matched with the Swiss-Prot database and 58% with the plant NR database. Homologous genes with high probability scores were found for 58.03%

(100775) of predicted ORFs. Taxonomic homology search of ORFs revealed that the highest similarity (5%) of the ORFs matched with *Capsicum annuum* (Solanaceae) and *Gossypium raimondii* (Malvaceae) and only 2% (2157) with *Striga asiatica* from Orobanchaceae family (Table S5.2).

Differential gene expression analysis

A preliminary comparison of replicates across all samples was performed using a PCA on normalized read counts (Figure 5.1). The first principal component, which explained 27.80% of the observed variation, clearly separated the two reciprocal hybrids. The second component (13.5%) allowed to distinguish between germinated and non-germinated seeds of the F_{1a} hybrid. A clear separation between 40 and 60 days of stratification for the non-germinated F_{1a} hybrid seeds is visible on the third principal axis (11.93%). Correlation analysis of differentially expressed (DE) genes clearly revealed two groups, again corresponding to the two reciprocal hybrids (Figure 5.2). The highest number of DE genes (4165) was detected between non-germinated F_{1a} seeds at 40 days and germinated F_{1a} seed at 60 days with respectively 1634 and 2531 upregulated genes (Table 5.3; Figure 5.3; Tables S5.5). Forty-three upregulated genes such as phytochrome B (phyB), NAC domain-containing protein, and oil body-associated proteins were shared among non-germinated seeds, while 399 shared upregulated genes were observed in germinated seeds (Figure 5.3).

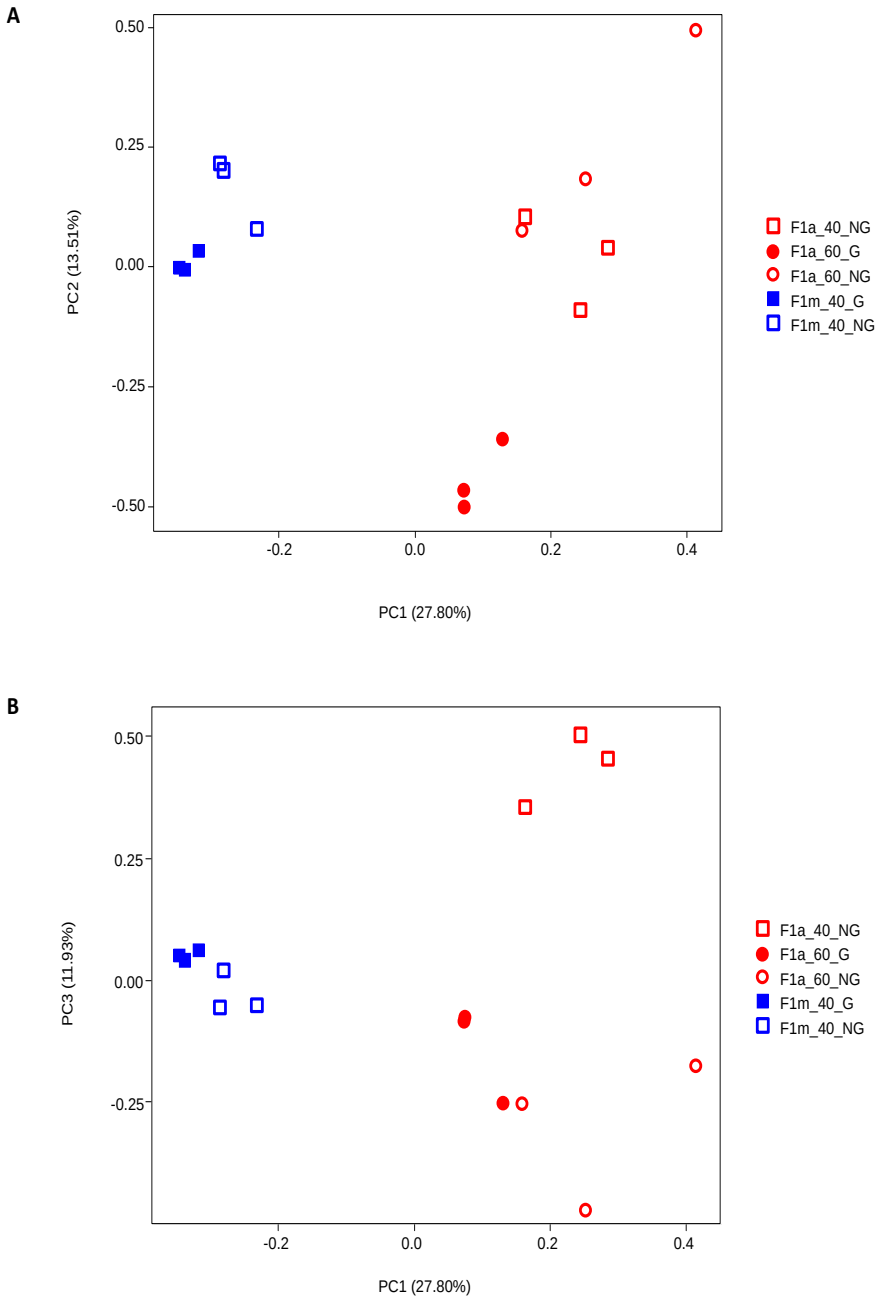


Figure 5.1. Principal component analysis of RNA-seq samples of *Rhinanthus* hybrid seeds using variance-stabilized estimated raw counts. a) PC1-PC2, b) PC1-PC3

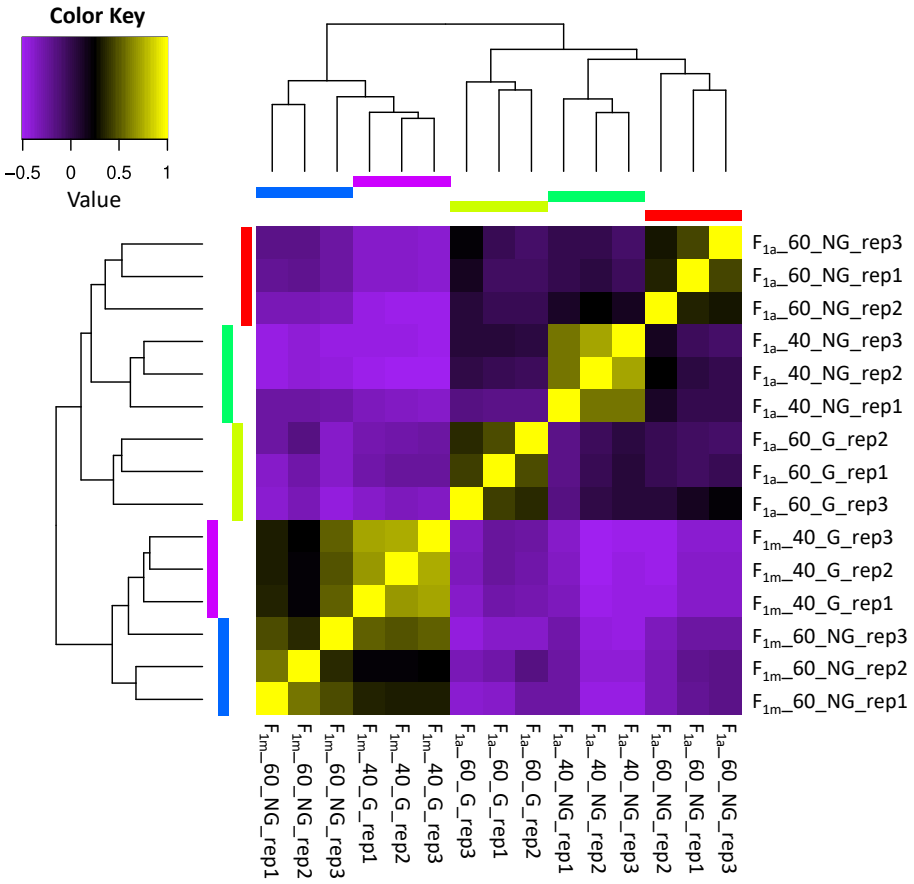
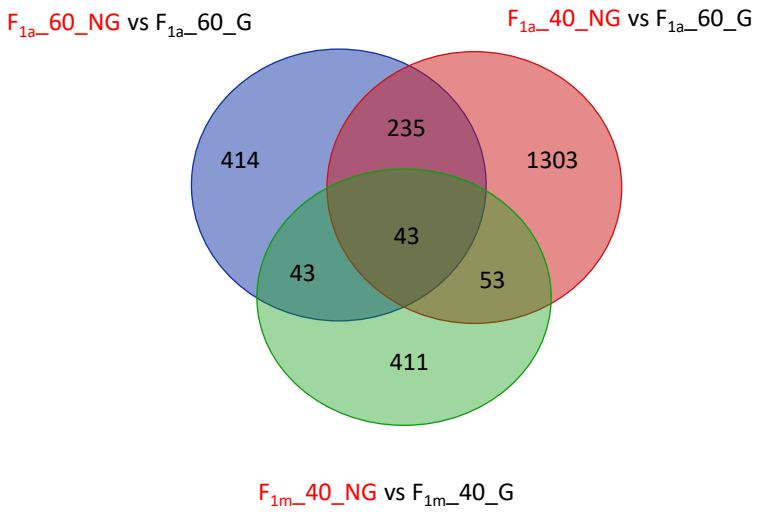


Figure 5.2. Comparison among replicates in each group of germinated and non-germinated seeds of reciprocal F_1 hybrids between *Rhinanthus major* (a) and *R. minor* (m).

A



B

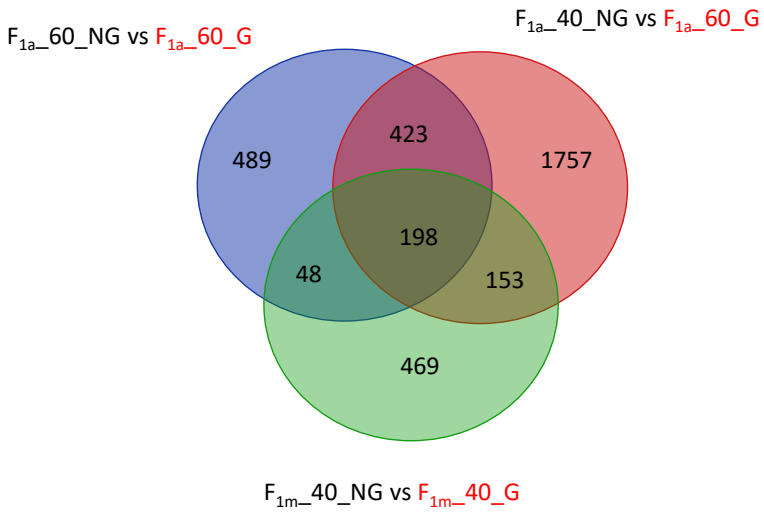


Figure 5.3. Venn diagram showing the numbers of unique and shared upregulated genes in each comparative analysis between non-germinated (NG) seeds and germinated seeds (G) for A) the non-germinated seeds (F_{1a}_60_NG, F_{1a}_40_NG and F_{1m}_40_NG), and B) the germinated seeds (F_{1a}_60_G and F_{1m}_40_G).

Table 5.3. Number of differentially expressed transcripts in three treatment comparisons ([FDR] < 0.05 and |log2FC| ≥ 2) in F_1 hybrids between *Rhinanthus major* (a) and *R. minor* (m), the letter indicating the maternal parent. NG = non-germinated seeds, G = germinated seeds, after 40 or 60 days on wet filter paper at 4°C.

Treatment	**F_{1a}_60_NG vs F_{1a}_60_G**		**F_{1a}_40_NG vs F_{1a}_60_G**	
Number of upregulated genes	735	1158	1634	2531
Treatment	**F_{1m}_40_NG vs F_{1m}_40_G**		**F_{1a}_60_NG vs F_{1a}_40_NG**	
Number of upregulated genes	550	868	1053	1363

In total 176380 (49.50%) and 167885 (47.10%) transcripts were assigned to GO terms and KEGG pathways, respectively. GO enrichment analysis of DE genes in the non-germinated seeds assigned them into 71, 81 and 70 significant GO functional groups, respectively, within three main categories: molecular function, biological process, and cellular components (Tables S5.3). The GO molecular function category related to nutrient reservoir activity (GO:0045735) and oxidoreductase activity (GO:0016491) ,with the highest number of upregulated genes that were significantly enriched in F_{1a}_40_NG vs F_{1a}_60_G and in F_{1a}_60_G vs F_{1a}_40_NG (Figures 5.4A, 5.4B). Alcohol dehydrogenase (NAD+) activity (GO:0004022) and oxidoreductase activity (GO:0016491) were found to be significantly enriched in F_{1m}_40_NG vs F_{1m}_40_G upregulated genes (Figure 5.4C). In terms of biological process, response to salt stress (GO:0009651) and oxidation-reduction process (GO:0055114) were enriched in F_{1m}_40_NG vs F_{1m}_40_G, with the highest number of upregulated genes, whereas among the upregulated genes in F_{1a}_40_NG (Figures 5.4A, 5.4B, 5.4C) we did not have any candidates in terms of biological process and cellular components. In terms of cellular components, extracellular region (GO:0005576) was shared between F_{1a}_60_G and F_{1m}_40_G (Tables S5.3).

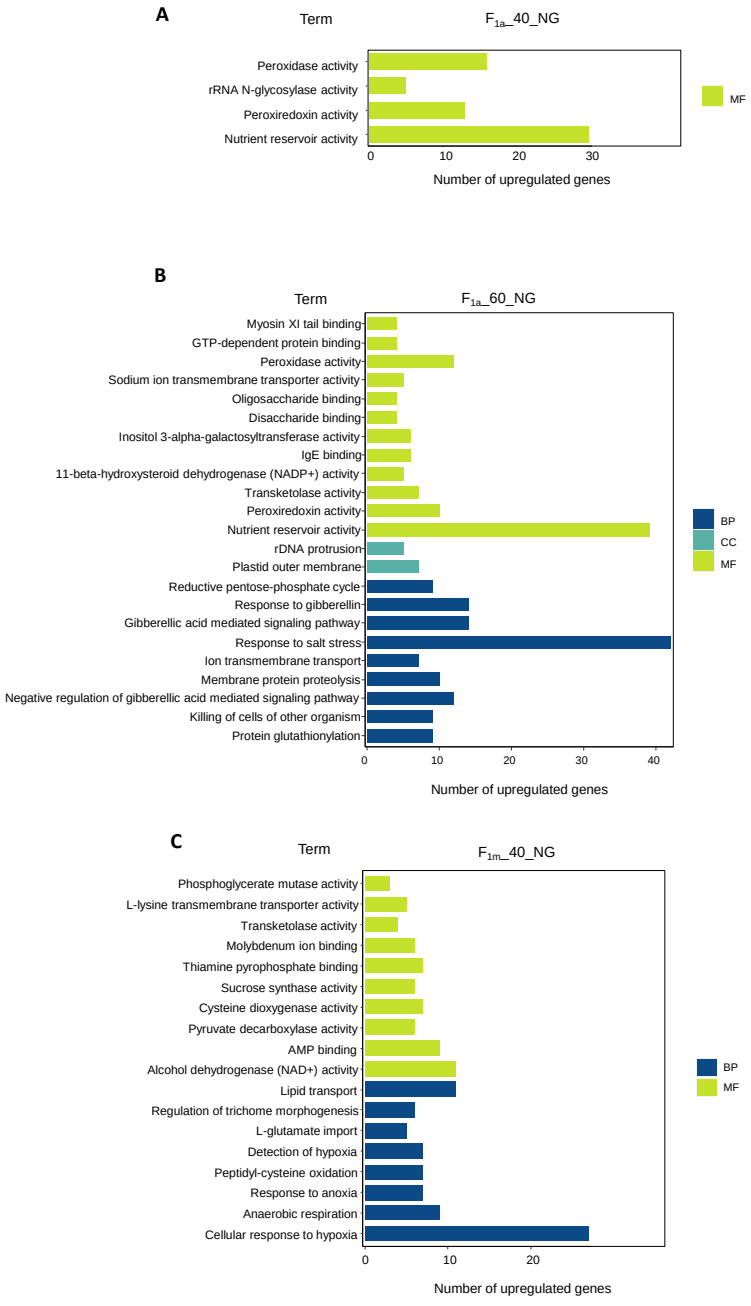


Figure 5.4. GO pathway enrichment analysis, based on upregulated genes in a) F_{1a}_40_NG vs F_{1a}_60_G, b) F_{1a}_60_NG vs F_{1a}_60_G and c) F_{1m}_40_NG vs F_{1m}_40_G. BP = Biological process, CC= Cell cycle, MF= Molecular function

In total 20 unique KEGG pathways displayed significant changes (P-value ≤0.05) in all compared treatments. "Metabolic pathways", "plant hormone signal transduction" and "biosynthesis of secondary metabolites" were represented among all the treatments. "Ubiquitin mediated proteolysis" and "amino sugar and nucleotide sugar metabolism" pathways were also common between F_{1a}_60_NG vs F_{1a}_60_G and F_{1a}_40_NG vs F_{1a}_60_G treatments. "Linoleic acid metabolism" and "starch and sucrose metabolism" pathways were found to be unique in F_{1m}_40_NG vs F_{1m}_40_G (Figure 5.5; Table S5.4).

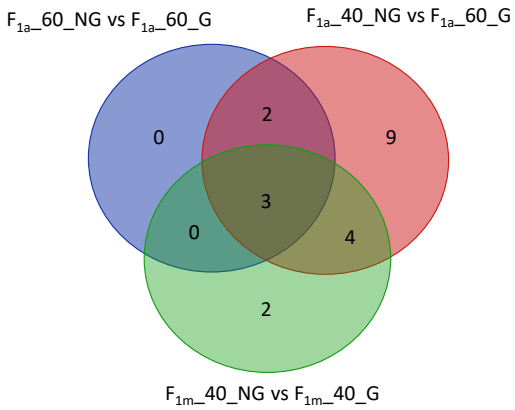


Figure 5.5. Venn diagram showing the number of unique and shared KEGG pathways related to DE genes in germinated and non-germinated reciprocal F_1 hybrid seeds.

Discussion

Overall, the gene expression profiles differed markedly between the two reciprocal hybrids, both before and after germination. Within the F_{1a} hybrids, the differences in gene expression between germinated and non-germinated seeds were stronger than for the F_{1m} hybrid seeds, and they also changed over the course of the stratification period.

DE genes related to seed development and osmotic stresses

Among the upregulated genes in the non-germinated F_{1a} seeds in comparison with the germinated seeds, peroxidase activity, peroxiredoxin activity and nutrient reservoir activity were common. Plant peroxidases, mostly known for reactive oxygen species (ROS) metabolism, are broadly active in various stages of plant life, from plant development to seed germination (Syros et al. 2005). ROS can cause oxidation damage to cell components, but recently it has been proven that ROS can play a key signaling role during germination or dormancy release (Oracz et al. 2007; Sarath et al. 2007; El-Maarouf-Bouteau and Bailly 2008). One of the sources of ROS in plant cells is peroxisome (glyoxysome) activity (Sandalio and Romero-Puertas 2015). The most significantly enriched upregulated genes in the non-germinated F_{1a} seeds after 60 days were related to glyoxysome, malate synthase activity and the glyoxylate cycle. Peroxiredoxin (Prx) proteins are types of plant antioxidants that protect lipids, enzymes, and DNA against ROS (Rouhier et al. 2001). During late seed development and dormancy in mature seeds they are highly expressed, and by sensing harsh conditions they are involved in the maintenance of dormancy while protecting the embryo from damage caused by ROS (Haslekås et al. 2003). It has been shown that Prx genes are up–regulated by ABA and osmotic stresses during dormancy and suppressed by gibberellic acid upon germination (Aalen 1999). A recent study in *Arabidopsis thaliana* has shown that *AtPER1*, a seed-specific peroxiredoxin, is involving in enhancing primary seed dormancy by eliminating ROS to suppress ABA catabolism and GA biosynthesis (Chen et al. 2020). The numbers of upregulated genes related to nutrient reservoir activity, which provides the proteins required for the development or growth of seeds, were high in non-germinated F_{1a} seeds, both after 40 and 60 days, whereas only one gene from this category (*VCL21*) was highly expressed in the non-germinated seeds of the reciprocal F_{1m} hybrid at 40 days. Additionally, GO enrichment resulted in significant representation of rRNA N-glycosylase activity genes in the non-germinated F_{1a} seeds after 40 days. rRNA N-glycosylase activity is a type of toxic activity which depurinates rRNAs and arrests protein synthesis during translation. They have been widely detected in plants and mainly act as antifungal, antibacterial and antiviral agents (Sharma et al. 2004; Zhu et al. 2018). Glutathionylation activity protects cells against oxidative stress, especially heavy metal stress, and is also involved in many other processes, including cell cycle and cell differentiation, symbiosis and flowering (Rouhier et al. 2008; Gao et al. 2009; Yadav 2010). Nine genes related to this activity were upregulated after 60 days in non-germinated F_{1a} seeds. It has been shown

that this protein is involved in post-transitional modification under oxidative stress conditions and can act as a redox signaling mechanism for helping the cells to sense and signal harmful stress conditions and trigger appropriate responses against stress (Gao et al. 2009). Alongside the common GO term enrichment for upregulated genes in the non-germinated F_{1a} seeds in comparison with the germinated F_{1a} seeds, we found the gibberellic acid-mediated signaling pathway, response to gibberellin and negative regulation of gibberellic acid-mediated signaling pathway GO enrichment only after 60 days in the non-germinated F_{1a} seeds.

DE genes related to phytohormone signal transduction

Maintenance and release of dormancy depend on the intrinsic balance between abscisic acid (ABA) and gibberellic acid (GA). While the maintenance of dormancy depends on high ABA/GA ratios, release of dormancy implies an increased biosynthesis of GA and degradation of ABA, resulting in low ABA/GA ratios (Kermode 2005). The enrichment pattern in non-germinated F_{1a} seeds after 60 days showed an increase in the expression of enzymes and proteins associated with the gibberellic acid-mediated signaling pathway and response to gibberellin in comparison to 40 days. In contrast, genes related to the negative regulation of the gibberellic acid-mediated signaling pathway and response to salt stress were highly abundant in non-germinated F_{1a} seeds after 60 days. These genes, like membrane-bound NAC transcription factor (*NTL8*), gibberellin 2-oxidases (*GA2oxs*) and ABSCISIC ACID-INSENSITIVE 5 (*ABI5*) are likely to be linked to the maintenance of dormancy in these seeds (Kim et al. 2008b; Lo et al. 2008; Kim and Park 2008). NAC transcription factors (*NTLs*) are related to the stress response and they have negative regulatory effects on seed germination (Kim and Park 2008). *GA2oxs* through 2β-hydroxylation can inactivate GA and increase the ABA/GA ratio (Sakamoto et al. 2004). *ABI5* from bZIP transcription factor family has a key role in ABA signaling and inhibiting seed germination (Finkelstein and Lynch 2000).

All of the enriched KEGG pathways for DE genes in non-germinated F_{1a} seeds after 60 days were shared with the same class after 40 days. After 40 days, however, there were 13 more enriched pathways compared to germinated F_{1a} seeds after 60 days, and they were related to lipid metabolism, arginine biosynthesis and glutathione metabolism. Genes clustered in this pathway were mainly upregulated in both germinated and non-germinated seeds after 60 days. Genes related to the "circadian rhythm" pathway were only enriched in upregulated genes in germinated seeds (F_{1m} after 40 days and F_{1a} after 60

days). Although "metabolic pathways" and "biosynthesis of secondary metabolites" were also quite abundant in the other treatments, here we mainly focused on "plant hormone signal transduction" pathway. Phytohormones like GAs promote germination but ABA and Auxin (Aux) are hormones known to induce and maintain seed dormancy (Liu et al. 2013; Tuan et al. 2018). In germinated F_{1a} seeds, upregulation of Auxin transporter protein 1 (*AUX1*) and Auxin/Indole-3-Acetic Acid (*Aux/IAA*) was detected, while in non-germinated F_{1a} seeds an upregulation of Transport inhibitor response 1 (*TIR1*) was observed. Studies have shown that Aux/IAA proteins act as positive regulators during seed germination and *TIR1* act as negative regulator for germination (Liu et al. 2013; Hussain et al. 2020). High Aux/IAA protein expression during germination will promote germination through the inhibition of *ABI3* transcription (Hussain et al. 2020). Upregulation of *ABI5* was observed in all of the non-germinated seeds and in non-germinated F_{1a} seeds at 40 days there was an high expression of pyrabactin resistance 1(PYR1)/PYR1-like2 (*PYL2*), *PYL4* and *GAI* with *RGA* (members of the DELLA superfamily). *PYL* receptors are in the ABA signaling pathway and DELLA proteins are key negative regulators of the GA signaling pathway. It has been shown that a high expression of these genes can induce dormancy and prevent germination (Tyler et al. 2004; Tuan et al. 2018). The high abundance of *ABI5* and members of the DELLA superfamily in non-germinated seeds could be the main reason of low germination rate of F_{1a} seeds, since we also detected a high expression of *ABI5* in non-germinated F_{1m} seeds after 40 days.

Maternal effects on F_1 germination

Large differences in gene expression were observed when we compared F_1 germinated and non- germinated seeds with different maternal parents in comparison with groups sharing the same maternal parent species. The reciprocal F_1 seeds differed in the number of days to germination and in the final germination percentage, and this was also conspicuous in the gene expression analysis, which showed clear maternal effects, leading to a prolonged dormancy in the hybrids with *R. major* as the maternal parent.

We only found differential gene expression for nuclear genes, not for mitochondrial or chloroplast genes. Dormancy, germination and seedling establishment can be maternally controlled through maternal tissues surrounding the embryo, such as the endosperm and seed coat (Debeaujon et al. 2007; Iwasaki et al. 2019). Maternal effects can be caused by the maternal environment and by the maternal genome (Chiang et al. 2011; Fernández Farnocchia et al. 2019). The parents of our F_1 hybrids were grown

simultaneously in the same conditions and with the same host plant species, so the observed maternal effects are most likely maternal genetic effects. Even in germinated seeds, we found strong differences in gene expression between the reciprocal hybrids, which indicates the potential long-term nature of these maternal effects.

Although differences in gene expression patterns were clear between the reciprocal hybrids, we cannot yet pinpoint which pathway or mechanism is the trigger of the prolonged dormancy observed. A possible explanation could lie in embryo development before stratification. Embryo morphology has strong effects on dormancy and germination, and seeds with morphological dormancy require a period of ripening or embryo maturation prior to germination (Forbis et al. 2002). *Rhinanthus* seeds collected directly fruit ripening have a low germination percentage (Ter Borg 2005), suggesting that these also need a period of after-ripening. Marin *et al*. (2019) found that *R. minor* seeds from seed lots with smaller embryos (relative to endosperm size) germinated more slowly. If F_{1a} hybrid seeds have smaller embryos that are relatively underdeveloped, this could be a potential mechanism that delays germination. This would require a study of embryo size and gene expression directly after seed formation and during after-ripening.

Studies have revealed that epigenetic process like DNA methylation, histone modifications, and small RNAs play an important role in the different stages of seed development, including embryogenesis, seed maturation, and germination (Greaves et al. 2015; Wang and Köhler 2017; Eriksson et al. 2020). This contribution of epigenetic regulation has also been reported in the fitness of F_1 hybrids and heterosis or hybrid vigour (Groszmann et al. 2013; Greaves et al. 2015). The dormancy of seed can be effected by epigenetic process: it has been shown in *Arabidopsis* that dormancy levels are inherited from the mother through inactivation of the paternal allele of the promoter region of allantoinase (*ALN*) gene by DNA methylation (Iwasaki et al. 2019). It seems likely that internal factors, including plant hormones and embryo development, in combination with epigenetic changes may explain the difference in germination behavior of the *Rhinanthus* reciprocal F_1 hybrids.

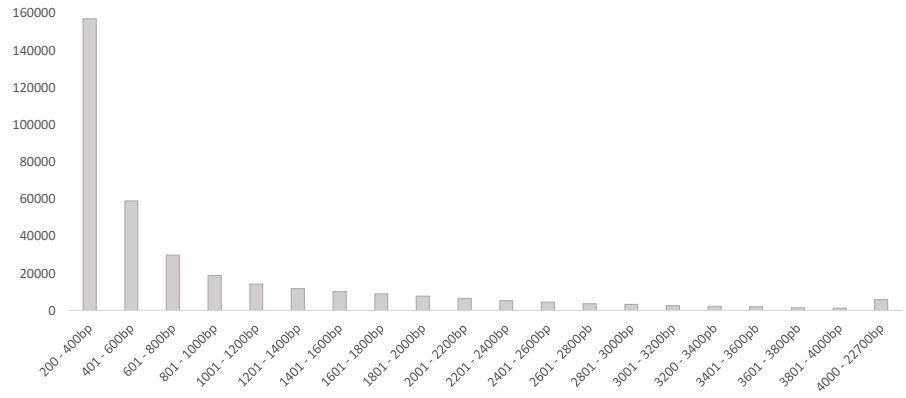


Figure S5.1. The length distribution of assembled transcriptomes of F_1 seeds.

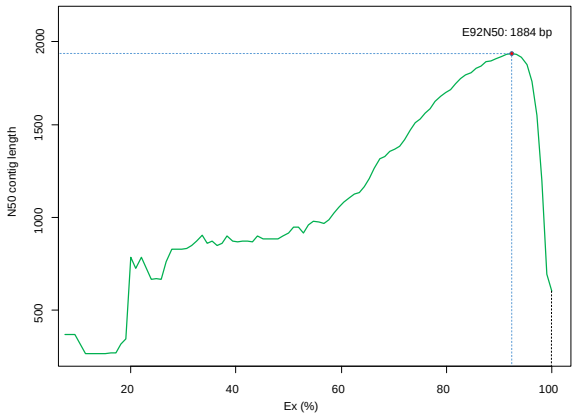


Figure S5.2. Expression percentage by N50 (ExN50) calculated as a fraction of the total expressed data (Ex). Red dot is ExN50 at the point of assembly saturation (92%) showing that 16,358 transcripts are covered by 92% of reads and N50 of this assembly subset is equal to 1,884 bp.

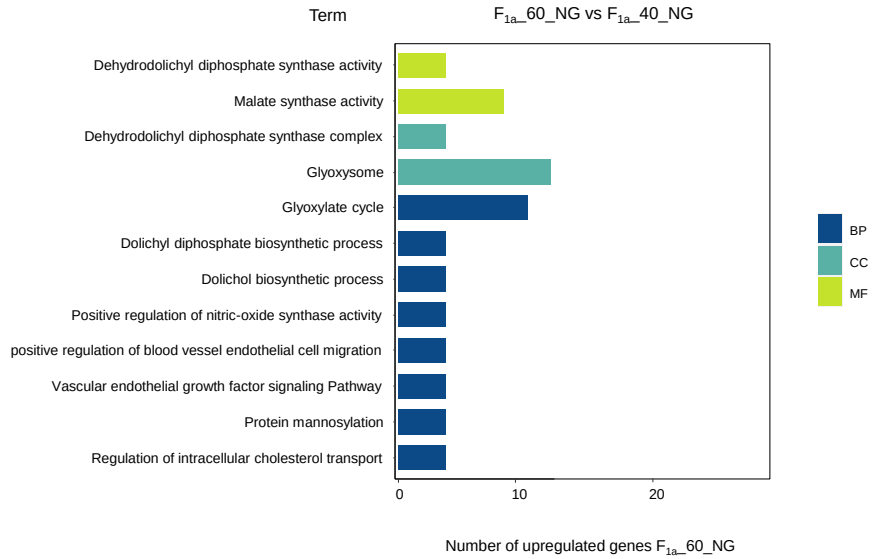


Figure S5.3. GO pathway enrichment analysis, based on upregulated genes in F_{1a}_40_NG vs F_{1a}_60_NG, GO term classification of upregulated genes in F_{1a}_60_NG. BP = Biological process, CC= Cell cycle, MF= Molecular function

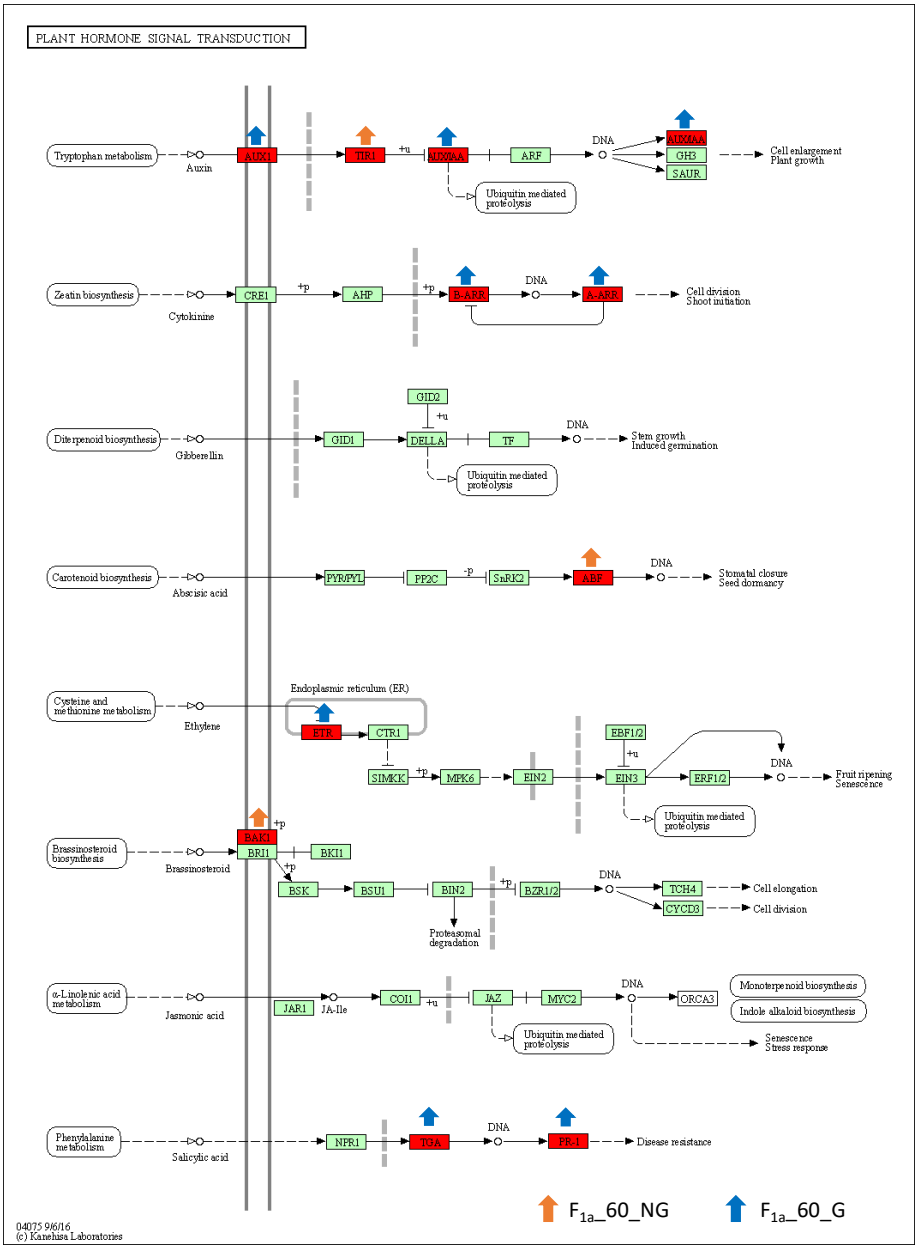


Figure S5.4A. KEGG pathway enrichment results of DE genes in F1a_60_NG vs F1a_60_G.

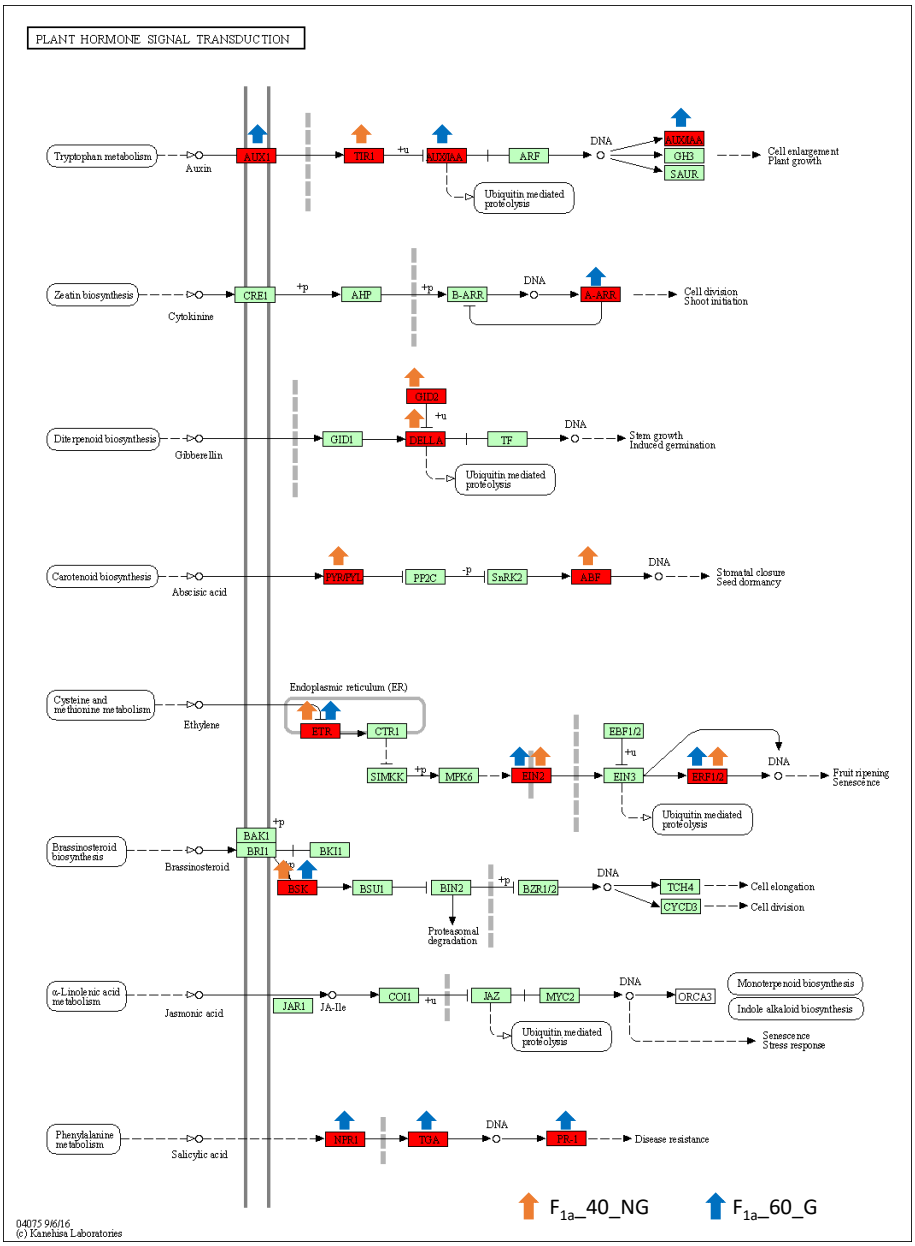


Figure S5.4B. KEGG pathway enrichment results of DE genes in F1a_40_NG vs F1a_60_G.

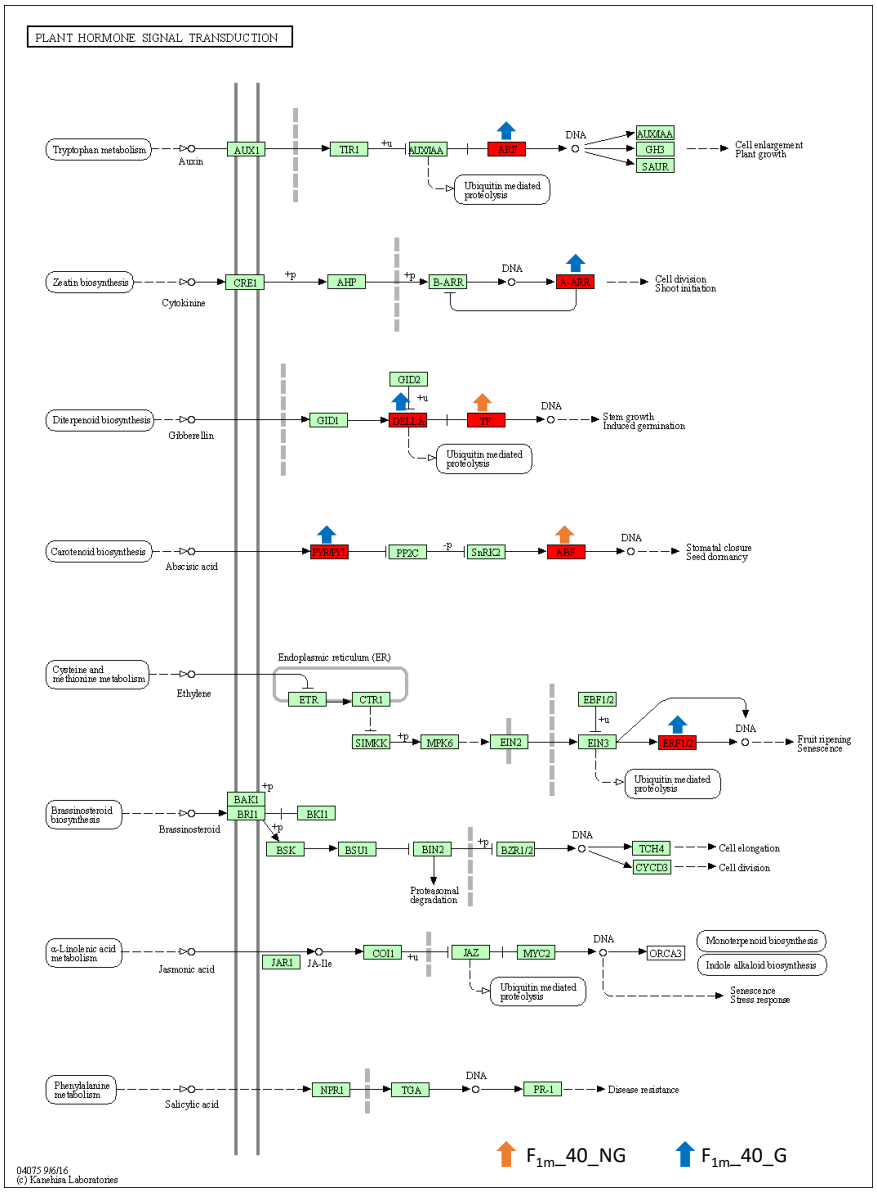


Figure S5.4C. KEGG pathway enrichment results of DE genes in F1m_40_NG vs F1m_40_G.

Table S5.1. Number of sequenced reads per samples details and number of high quality reads (Q>20) have been used for assembly and DE analysis.

Sample	Raw PE reads	High quality reads	percentage
F_{1a}_60_NG_rep1	19439586	17020996	87.55%
F_{1a}_60_NG_rep2	21555530	18653980	86.53%
F_{1a}_60_NG_rep3	19523841	17641208	90.35%
F_{1a}_60_G_rep1	18970053	16856742	88.85%
F_{1a}_60_G_rep2	18985001	17244304	90.83%
F_{1a}_60_G_rep3	19472891	17531514	90.03%
F_{1a}_40_NG_rep1	20137400	18533775	92.03%
F_{1a}_40_NG_rep2	19132134	17054228	89.13%
F_{1a}_40_NG_rep3	19434401	17137427	88.18%
F_{1m}_40_G_rep1	22805432	20163127	88.41%
F_{1m}_40_G_rep2	18919276	16697648	88.25%
F_{1m}_40_G_rep3	19025330	16603091	87.26%
F_{1m}_40_NG_rep1	20383643	17931419	87.96%
F_{1m}_40_NG_rep2	19130607	16871810	88.19%
F_{1m}_40_NG_rep3	19297860	17064342	88.42%

Table S5.2. Significant BLASTp matches of *Rhinanthus* hybrid seed ORFs with other plant species in the Plant NR database. Significant with a BitScore>100.

Species	Family	Match number	Parentage
Capsicum annuum	Solanaceae	5059	5.02%
Gossypium raimondii	Malvaceae	4552	4.52%
Cynara cardunculus	Asteraceae	3737	3.71%
Vitis vinifera	Vitaceae	3517	3.49%
Juglans regia	Juglandaceae	3511	3.48%
Lupinus angustifolius	Fabaceae	3476	3.45%
Arabidopsis thaliana	Brassicaceae	3308	3.28%
Actinidia chinensis	Actinidiaceae	2718	2.7%
Vitis riparia	Vitaceae	2664	2.64%
Noccaea caerulescens	Brassicaceae	2284	2.27%
Citrus clementina	Rutaceae	2181	2.16%
Striga asiatica	Orobanchaceae	2157	2.14%
Others	-	61611	61.14%

Table S5.3A. GO terms classification of DE transcripts in F_{1a}_60_NG vs F_{1a}_60_G.

Category	GO ID	Term	Number of genes
MF	GO:0045735	Nutrient reservoir activity	43
CC	GO:0005576	Extracellular region	140
MF	GO:0004497	Monooxygenase activity	34
MF	GO:0004601	Peroxidase activity	27
BP	GO:0002253	Activation of immune response	14
CC	GO:0009505	Plant-type cell wall	40
BP	GO:0055114	Oxidation-reduction process	149
BP	GO:0005975	Carbohydrate metabolic process	82
CC	GO:0048046	Apoplast	57
MF	GO:0020037	Heme binding	54
MF	GO:0016705	Oxidoreductase activity	36
CC	GO:0005618	Cell wall	80
BP	GO:0019253	Reductive pentose-phosphate cycle	18
BP	GO:0045490	Pectin catabolic process	17
BP	GO:2000039	Regulation of trichome morphogenesis	10
MF	GO:0004802	Transketolase activity	7
BP	GO:0031640	Killing of cells of other organism	10
BP	GO:0010731	Protein glutathionylation	9
MF	GO:0051920	Peroxiredoxin activity	10
MF	GO:0016491	Oxidoreductase activity	91
MF	GO:0005506	Iron ion binding	45
MF	GO:0004553	Hydrolase activity	42
MF	GO:0045330	Aspartyl esterase activity	13
MF	GO:0030599	Pectinesterase activity	14
BP	GO:0042545	Cell wall modification	14
MF	GO:0051119	Sugar transmembrane transporter activity	10
BP	GO:0009415	Response to water	8
MF	GO:0052716	Hydroquinone:oxygen oxidoreductase	13
BP	GO:0009938	Negative regulation of gibberellic acid	7
BP	GO:0009739	Response to gibberellin	23
BP	GO:0046274	lignin catabolic process	7
BP	GO:0010597	Green leaf volatile biosynthetic process	7
MF	GO:0004857	Enzyme inhibitor activity	14
MF	GO:0070524	11-beta-hydroxysteroid dehydrogenase	5
BP	GO:0006873	Cellular ion homeostasis	4

MF	GO:0047216	Inositol 3-alpha-galactosyltransferase	7
BP	GO:0009664	Plant-type cell wall organization	18
MF	GO:0016984	Ribulose-bisphosphate carboxylase activity	7
CC	GO:0016021	Integral component of membrane	429
MF	GO:0033879	Acetylajmaline esterase activity	5
MF	GO:0046910	Pectinesterase inhibitor activity	10
BP	GO:0009651	Response to salt stress	70
MF	GO:0016711	Flavonoid 3'-monooxygenase activity	4
BP	GO:0034220	Ion transmembrane transport	7
BP	GO:0009820	Alkaloid metabolic process	9
MF	GO:0016985	Mannan endo-1,4-beta-mannosidase activity	7
MF	GO:0004322	Ferroxidase activity	9
BP	GO:0033619	Membrane protein proteolysis	10
MF	GO:0015250	Water channel activity	11
BP	GO:0055085	Transmembrane transport	78
MF	GO:0019863	IgE binding	6
MF	GO:0016165	Linoleate 13S-lipoxygenase activity	7
CC	GO:0009527	Plastid outer membrane	7
BP	GO:0051603	Proteolysis	22
MF	GO:0070492	Oligosaccharide binding	4
MF	GO:0048030	Disaccharide binding	4
MF	GO:0050247	Raucaffricine beta-glucosidase activity	5
MF	GO:0050506	Vomilenine glucosyltransferase activity	5
BP	GO:0009821	Alkaloid biosynthetic process	6
CC	GO:0030875	rDNA protrusion	5
BP	GO:0006952	Defense response	77
MF	GO:0015081	Sodium ion transmembrane transporter	17
BP	GO:0009740	Gibberellic acid mediated signaling pathway	5
MF	GO:0045544	Gibberellin 20-oxidase activity	4
MF	GO:0102911	(-)-secoisolariciresinol dehydrogenase	4
BP	GO:0042744	Hydrogen peroxide catabolic process	17
MF	GO:0035673	Oligopeptide transmembrane transporter	7
BP	GO:0010262	Somatic embryogenesis	12
MF	GO:0030742	GTP-dependent protein binding	4
MF	GO:0080115	Myosin XI tail binding	4
BP	GO:0048316	Seed development	21

Table S5.3B. GO terms classification of DE transcripts in F_{1a}_40_NG vs F_{1a}_60_G.

Category	GO ID	Term	Number of genes
BP	GO:0055114	Oxidation-reduction process	136
MF	GO:0008422	Beta-glucosidase activity	37
BP	GO:0071456	Cellular response to hypoxia	26
MF	GO:0004022	Alcohol dehydrogenase (NAD+) activity	11
BP	GO:0009061	Anaerobic respiration	10
MF	GO:0004553	Hydrolase activity	39
MF	GO:0015250	Water channel activity	15
BP	GO:0034059	Response to anoxia	8
CC	GO:0005576	Extracellular region	90
MF	GO:0016208	AMP binding	10
MF	GO:0045330	Aspartyl esterase activity	12
MF	GO:0030599	Pectinesterase activity	13
BP	GO:0006952	Defense response	60
BP	GO:0005975	Carbohydrate metabolic process	15
BP	GO:0045490	Pectin catabolic process	71
BP	GO:0042545	Cell wall modification	41
MF	GO:0020037	Heme binding	13
MF	GO:0004601	Peroxidase activity	18
MF	GO:0004737	Pyruvate decarboxylase activity	6
BP	GO:0002253	Activation of immune response	11
MF	GO:0102483	Scopolin beta-glucosidase activity	17
MF	GO:0016702	Oxidoreductase activity	12
MF	GO:0004857	Enzyme inhibitor activity	13
BP	GO:0018171	Peptidyl-cysteine oxidation	7
MF	GO:0102311	8-hydroxygeraniol dehydrogenase activity	8
MF	GO:0015267	Channel activity	12
BP	GO:0070483	Detection of hypoxia	7
MF	GO:0004568	Chitinase activity	9
BP	GO:0006032	Chitin catabolic process	9
MF	GO:0016705	Oxidoreductase activity	25
BP	GO:0009415	Response to water	7
MF	GO:0016831	Carboxy-lyase activity	7
CC	GO:0005618	Cell wall	9
MF	GO:0046910	Pectinesterase inhibitor activity	8
BP	GO:0016998	Cell wall macromolecule catabolic process	59

BP	GO:0051938	L-glutamate import	5
MF	GO:0004190	Aspartic-type endopeptidase activity	30
MF	GO:0017172	Cysteine dioxygenase activity	7
MF	GO:0008061	Chitin binding	8
CC	GO:0099503	Secretory vesicle	22
MF	GO:0010333	Terpene synthase activity	6
MF	GO:0004497	Monooxygenase activity	18
BP	GO:0000272	Polysaccharide catabolic process	9
BP	GO:2000039	Regulation of trichome morphogenesis	6
MF	GO:0030976	Thiamine pyrophosphate binding	8
MF	GO:0038023	Signaling receptor activity	10
MF	GO:0033879	Acetylajmaline esterase activity	4
BP	GO:0009808	Lignin metabolic process	4
BP	GO:0055085	Transmembrane transport	56
MF	GO:0004807	Triose-phosphate isomerase activity	6
CC	GO:0048046	Apoplast	37
MF	GO:0004802	Transketolase activity	4
MF	GO:0102406	Omega-hydroxypalmitate	3
MF	GO:0050734	Hydroxycinnamoyltransferase activity	3
MF	GO:0016157	Sucrose synthase activity	6
BP	GO:0015074	DNA integration	20
CC	GO:0042597	Periplasmic space	7
MF	GO:0051213	Dioxygenase activity	17
MF	GO:0004519	Endonuclease activity	21
BP	GO:0016114	Terpenoid biosynthetic process	7
MF	GO:0004619	Phosphoglycerate mutase activity	3
MF	GO:0030151	Molybdenum ion binding	6
MF	GO:0004351	Glutamate decarboxylase activity	3
BP	GO:0010262	Somatic embryogenesis	60
MF	GO:0016491	Oxidoreductase activity	5
BP	GO:0010345	Suberin biosynthetic process	6
BP	GO:0009664	Plant-type cell wall organization	5
MF	GO:0015189	L-lysine transmembrane transporter activity	13
BP	GO:0006869	Lipid transport	18
BP	GO:0042744	Hydrogen peroxide catabolic process	14

Table S5.3C. GO terms classification of DE transcripts in F_{1m}_40_NG vs F_{1m}_40_G.

Category	GO ID	Term	Number of genes
BP	GO:0055114	Oxidation-reduction process	303
CC	GO:0048046	Apoplast	121
BP	GO:0002253	Activation of immune response	239
CC	GO:0005576	Extracellular region	24
MF	GO:0004553	Hydrolase activity	89
MF	GO:0004497	Monooxygenase activity	51
MF	GO:0016491	Oxidoreductase activity	186
MF	GO:0045735	Nutrient reservoir activity	38
BP	GO:0009415	Response to water	15
CC	GO:0005618	Cell wall	38
MF	GO:0004601	Peroxidase activity	160
BP	GO:0045490	Pectin catabolic process	29
BP	GO:0006952	Defense response	174
MF	GO:0051119	Sugar transmembrane transporter	22
MF	GO:0045330	Aspartyl esterase activity	17
BP	GO:0031408	Oxylipin biosynthetic process	30
MF	GO:0051920	Peroxiredoxin activity	16
BP	GO:0005975	Carbohydrate metabolic process	141
MF	GO:0020037	Heme binding	86
MF	GO:0030599	Pectinesterase activity	22
MF	GO:0004474	Malate synthase activity	14
BP	GO:0042545	Cell wall modification	22
BP	GO:0019752	Carboxylic acid metabolic process	13
MF	GO:0050614	Delta24-sterol reductase activity	36
BP	GO:0010200	Response to chitin	6
CC	GO:0009514	Glyoxysome	41
MF	GO:0016747	Transferase activity	33
MF	GO:0016984	Ribulose-bisphosphate carboxylase	11
MF	GO:0004857	Enzyme inhibitor activity	23
CC	GO:0009505	Plant-type cell wall	55
MF	GO:0016630	Protochlorophyllide reductase activity	8
MF	GO:0004740	Pyruvate dehydrogenase	6
BP	GO:1904183	Negative regulation of pyruvate	6
BP	GO:0055085	Transmembrane transport	147
MF	GO:0016762	Xyloglucan:xyloglucosyl transferase	19
MF	GO:0016705	Oxidoreductase activity	48

MF	GO:0052716	Hydroquinone:oxygen oxidoreductase	8
CC	GO:0005615	Extracellular space	36
MF	GO:0050661	NADP binding	4
BP	GO:0000717	Nucleotide-excision repair	4
BP	GO:0000019	Regulation of mitotic recombination	35
BP	GO:0009607	Response to biotic stimulus	17
BP	GO:0009664	Plant-type cell wall organization	29
BP	GO:0046274	Lignin catabolic process	8
MF	GO:0046910	Pectinesterase inhibitor activity	15
BP	GO:0006097	Glyoxylate cycle	33
BP	GO:0009695	Jasmonic acid biosynthetic process	20
CC	GO:0016020	membrane	312
MF	GO:0016985	Mannan endo-1,4-beta-mannosidase	6
MF	GO:0016207	4-coumarate-CoA ligase activity	10
MF	GO:0016165	Linoleate 13S-lipoxygenase activity	13
BP	GO:0009820	Alkaloid metabolic process	11
BP	GO:0044262	Cellular carbohydrate metabolic process	6
MF	GO:0004351	Glutamate decarboxylase activity	5
BP	GO:0048316	Seed development	39
MF	GO:0015250	Water channel activity	18
BP	GO:0006073	Cellular glucan metabolic process	13
BP	GO:0009861	Jasmonic acid and ethylene-dependent	12
MF	GO:0033879	Acetylajmaline esterase activity	6
MF	GO:0050275	Scopoletin glucosyltransferase activity	5
BP	GO:0006979	Response to oxidative stress	66
BP	GO:2000899	Xyloglucan catabolic process	11
BP	GO:0085030	Symbiotic process benefiting host	11
MF	GO:0050506	Vomilenine glucosyltransferase activity	7
MF	GO:0050247	Raucaffricine beta-glucosidase activity	7
CC	GO:0099503	Secretory vesicle	45
BP	GO:0042593	Glucose homeostasis	6
BP	GO:0009809	Lignin biosynthetic process	23
BP	GO:0045951	mitotic recombination	4
BP	GO:0071446	Response to salicylic acid stimulus	6
BP	GO:0019253	Reductive pentose-phosphate cycle	20
MF	GO:0022857	Transmembrane transporter activity	91
MF	GO:0005199	Structural constituent of cell wall	9
MF	GO:0030598	rRNA N-glycosylase activity	6

MF	GO:0008171	O-methyltransferase activity	11
CC	GO:0042807	Central vacuole	6
MF	GO:0016758	Transferring hexosyl	33
MF	GO:0070006	Metalloaminopeptidase activity	10
BP	GO:0015824	Proline transport	4
BP	GO:0006829	Zinc ion transport	7
BP	GO:0009741	Response to brassinosteroid	19

Table S5.4. KEGG pathway enrichment results of DE genes in F_{1a}_60_NG vs F_{1a}_60_G, F_{1a}_40_NG vs F_{1a}_60_G and F_{1m}_40_NG vs F_{1m}_40_G.

Term	ID	Number of genes	P-Value
F_{1a}_60_NG vs F_{1a}_60_G			
Metabolic pathways	ath01100	47	0.000298
Ubiquitin mediated proteolysis	ath04120	8	0.004205
Plant hormone signal transduction	ath04075	11	0.004205
Biosynthesis of secondary metabolites	ath01110	25	0.005014
Amino sugar and nucleotide sugar metabolism	ath00520	6	0.04045
F_{1a}_40_NG vs F_{1a}_60_G			
Metabolic pathways	ath01100	105	1.33E-12
Plant hormone signal transduction	ath04075	23	2.26E-06
Glycerolipid metabolism	ath00561	10	4.5E-05
Biosynthesis of secondary metabolites	ath01110	46	0.00027
Glycerophospholipid metabolism	ath00564	10	0.001224
Amino sugar and nucleotide sugar metabolism	ath00520	11	0.002452
Plant-pathogen interaction	ath04626	12	0.004828
alpha-Linolenic acid metabolism	ath00592	6	0.005911
MAPK signaling pathway - plant	ath04016	10	0.007646
Ubiquitin mediated proteolysis	ath04120	10	0.010951
Glycosphingolipid biosynthesis - globo and isoglobo series	ath00603	3	0.013633
Glutathione metabolism	ath00480	8	0.014248
Fatty acid biosynthesis	ath00061	5	0.020304
Fatty acid degradation	ath00071	5	0.026145
Fatty acid metabolism	ath01212	6	0.026145
Fatty acid elongation	ath00062	4	0.044855

Circadian rhythm - plant	ath04712	4	0.044855
Arginine biosynthesis	ath00220	4	0.044855
F_{1m}_40_NG vs F_{1m}_40_G			
Metabolic pathways	ath01100	43	6.93E-08
Biosynthesis of secondary metabolites	ath01110	19	0.00694
alpha-Linolenic acid metabolism	ath00592	4	0.00694
Plant hormone signal transduction	ath04075	8	0.011712
Circadian rhythm - plant	ath04712	3	0.024828
Linoleic acid metabolism	ath00591	2	0.024828
MAPK signaling pathway - plant	ath04016	5	0.024828
Glycerophospholipid metabolism	ath00564	4	0.041279
Starch and sucrose metabolism	ath00500	5	0.045252

Table S5.5. Number of differentially expressed transcripts in all treatment comparisons ([FDR] < 0.05 and |log2FC| ≥ 2) in F_1 hybrids between *Rhinanthus major* (a) and *R. minor* (m), the letter indicating the maternal parent. NG = non-germinated seeds, G = germinated seeds, after 40 or 60 days on wet filter paper at 4°C.

	F_{1a}_60_NG	F_{1a}_60_G	F_{1a}_40_NG	F_{1m}_40_NG
F_{1a}_60_NG	-			
F_{1a}_60_G	1893	-		
F_{1a}_40_NG	2416	4165	-	
F_{1m}_40_NG	4421	5189	6052	-
F_{1m}_40_G	8096	5762	8969	1420

Chapter 6

General discussion

Nowadays, molecular genetic tools provide important and unique resources for studying evolutionary events such as hybridization, introgression and adaptations. Before the onset of NGS, population-genetic studies were labor-intensive and expensive, and it was not efficient for non-model species with insufficient genomic resources. The majority of studies on population genetics and evolutionary events were based on a limited number of markers such as RAPD, SSR and AFLP. Additionally, despite the fact that some markers were not informative, in some cases, like RAPDs, the results were not repeatable from one lab to another. Due to the low number of markers in each of those systems, understanding of the full evolutionary history and population structure of species could not be complete.

Although whole-genome sequencing of tens or hundreds of individuals remains expensive and beyond the means of most researchers, NGS-based methods such as ddRAD seq (Peterson et al. 2012) provide a high uniformity and replicability across samples, and higher sequencing depths at each locus to ensure that the polymorphisms discovered represent true sequence variants rather than sequencing errors.

Using the ddRAD-seq technique, for the first time a genome-wide set of SNP markers with reliable coverage in *Rhinanthus minor* and *R. major* was generated. The loci obtained in chapters 2–4 covered 0.103%, 0.14% and 0.168% of the genome, respectively (Ellstrand et al. 1996). Although this is not a high coverage, the loci are likely to be well distributed across the genome, due to the fact that the same method and size selection have provided a high-density genetic linkage map with precisely 11 linkage groups in *R. major*, which corresponds to their haploid chromosome number (Mirzaei and Wesselingh, unpublished data). However, this can be investigated in more

detail by mapping these loci to the whole genome sequence. The genetic diversity indices, such as transition/transversion ratio and F_{ST} values, showed low genetic distance between *R. minor* and *R. major* in all studied populations. The 1106 putative loci with species-specific SNP markers provide a valuable resource for assessing and monitoring hybridization and introgression between *R. major* and *R. minor*. Our SNP panel has already been used to develop a PCR + restriction test based on a single SNP to distinguish F_1 hybrids from the parental species (Wesselingh et al. 2019). This is a great improvement in comparison to the eight species-specific RAPD/ISSR-markers (Ducarme and Wesselingh 2005) and 15 microsatellite markers without diagnostic alleles (Ducarme et al. 2008) that have been used before (Ducarme et al. 2010). It should however be noted that these diagnostic SNP markers were obtained from a limited number of populations and only in Belgium, although they also proved useful in a population from the north of the Netherlands (Smalbroek, chapter 3). Our extensive panel of diagnostic SNP markers could be useful for obtaining a higher resolution in phylogeographical patterns in *R. major* and *R. minor* discovered in previous studies by Vrancken et al. (2012) using AFLP, which yielded 181 polymorphic markers. In red mangroves, a phylogeographic discontinuity at the southern tip of Florida was found that was not detected in previous studies using SSRs (Hodel et al. 2017). The use of ddRADseq in yellowfish (*Labeobarbus natalensis*) has provided a higher resolution in their phylogeography (Stobie et al. 2018). However, any diagnostic genotyping assays with our provided dataset in other regions should include sufficient samples from pure populations to test the validity of the diagnostic markers.

In general, the number of loci that contained SNP markers provided in this study was much higher than other studies on hybridization in plants, such as willow hybrids (*Salix purpurea* L. x *S. helvetica* Vill) (Gramlich et al. 2018) and oak (*Quercus ilex* x *Q. suber*) hybrids (López de Heredia et al. 2020). It is worth to mention that the genome sizes of willow (~421 Mb) and oak (~730 Mb) in these studies are much smaller than the *R. major* genome (~1400 Mb). Beside genetic differences between the studied species, one of the critical steps of ddRAD is related to the range of size selection, the output capacity of the sequencing machine and number of samples assembled in one library for sequencing, which have an effect on the number of obtained shared loci between studied individuals. The use of different bioinformatics strategies and restriction for obtained loci is an effective factor in this sort of comparison. Our strategy for defining a locus as species-specific was to only pick loci for which all the SNPs that were found, between 1 and 7 per locus, had a different

allele in the two species. Other studies have used a single SNP as representative of the entire sequence, regardless of the alleles in other SNPs present.

In several studies, the occurrence of organelle sequences in ddRAD libraries has been reported, for example 23% of the loci in a study on the orchid *Cypripedium calceolus* (Gargiulo et al. 2021) were organelle sequences. In our ddRAD data, the percentage of chloroplast DNA in a sample varied between 0 and 25%, but none of the differentially expressed loci shared among more than 70% or even 50% of the studied individuals were found to be located in the chloroplast DNA. Aligning the 1106 loci containing species-specific markers to the chloroplast genomes of 31 plants from the Orobanchaceae also showed no match. In the last decades, chloroplast DNA has been important for the study of phylogeny and population genetics in plants, but if the nuclear DNA and chloroplast DNA have different evolutionary histories, this may result in different tree topologies (Wang et al. 2014). Therefore, reconstructing the phylogeny and evolutionary history of the genus *Rhinanthus* should be done by a combined analysis of nucleus and chloroplast DNA data sets with caution to avoid incongruence between data sets. If significant incongruence exists between both data sets, they should not be combined without further justification, otherwise the estimated phylogenetic tree will represent an oversimplification of the evolutionary history (Mason-Gamer and Kellogg 1996; Wang et al. 2014).

6.1 Similarity and dissimilarity of introgression in studied populations

In all of the studied populations in chapter 3 and 4, we observed again the asymmetric introgression with variable admixture rates that has been reported by Kwak (1980), Ducarme and Wesselingh (2005), and Ducarme et al (2010) using morphological and molecular markers. In the studies by Ducarme and Wesselingh (2005), and Ducarme et al (2010), the admixture rate of most individuals was underestimated or overestimated due to the limited number of markers. With four dominant *R. minor* RAPD markers, introgression is only detected when at least one of these markers is present, which represents an admixture rate of 12.5%. In all of the 'pure' *R. major* individuals in these studies, we found some *R. minor* SNPs, showing that they too are introgressed. Using just a few markers is sufficient to give a global idea of the representation of the *R. minor* genome in *R. major* individuals, but by using a large panel of SNP markers, we could estimate the levels of asymmetric introgression much better. How many markers do we actually need in order to have the correct

estimate of introgression between this two species? Since introgressing alleles are often concentrated in restricted regions of the genome (Harrison and Larson 2014; Suarez-Gonzalez et al. 2018a), we would need to have the full sequences of both genomes in order to detect these hot spots in the genome that are involved in introgression.

6.2 Was it a good idea to use ddRAD seq for studying introgression?

Although we were able to recover a very high number of targeted loci for more than 70% of individuals in each population, we still had some missing loci. In chapter 3, we used Blue Pippin for size selection during library preparation, but in chapter 4 we cut bands from an agarose gel and used DNA purification for size selection, which helped us to obtain a recovery rate of targeted loci of nearly 65%, while by using Blue Pippin we had a 51% recovery rate. This can be improved by enrichment techniques for particular genome regions or loci of interest. Enrichment methods, such as DNA hybridization capture methods, which require designing capture baits, can be used to select these loci. This will increase cost efficiency and the representation of sequences of interest. In reduced genome representation method such as ddRADseq, obtaining candidate loci for classifying introgressed individuals or hybrids requires enough reads in order to provide reliable coverage of the genome. Providing several types of defining loci and filtering in order to minimize the potential sources of error in marker identification could be another key factor. Since we did not have whole genome sequences of the investigated species, prior *in silico* analysis of enzymatic digestion was not possible, therefore different enzyme combinations, including EcoRI, NlaIII, MluCI, HindIII, MspI and SphI, were used to count the number of digested fragments and optimize the sequencing results according to budget, number of samples and level of coverage. It has been stressed that reducing the number of missing loci by increasing the depth of sequencing using less samples in each library can remove biases (Twyford and Ennos 2012; López de Heredia et al. 2020). Therefore, a precise estimate of the admixture rate in these populations can benefit from preliminary information from whole genome data, and the number of missing loci could be reduced by increasing the depth of sequencing, using less of samples and using equal amounts of input DNA at each round of sequencing.

We performed PCoA in order to identify clusters of *R. minor* SNP markers with similar introgression patterns (Chapters 3 and 4), and we found different patterns in each study. Their similarity of introgression due to

presence/absence in the studied samples clustered them in two and four groups, respectively. The diverse patterns of clustering could be related to missing data or to concurrent selection on loci linked to these markers. These patterns might thus be indicative of linkage groups or blocks of introgressing genomic regions, but this can only be verified by mapping the markers onto a reference genome. However, their level of introgression into *R. major*, which is indicated by *Phi_st*, shows that different levels of introgression are found within clusters. PCoA treats missing data more satisfactorily than PCA (Rohlf 1972) and it is recommended over PCA when there are lots of missing data and when there are fewer individuals than characters (Mohammadi and Prasanna 2003), but in our data the patterns changed when we excluded individual samples with high levels of missing data. Therefore providing enough data in order to estimate the connection between loci that are involved in introgression is needed.

In general, the population dynamics of hybrids and introgressed individuals will be affected by responses to selection pressure. As selection can act evenly or unevenly on specific phenotypes in different conditions, different rates of introgression are possible across different habitats. It is known that genes or alleles related to low viability and fertility are under strong selection, therefore they will have limited introgression. But in the opposite case, genes or genomic regions which convey a superior or equal fitness as their parents can spread quickly through a population depending on levels of outbreeding. Understanding the overall patterns of introgression in *Rhinanthus* will be easier by identifying the genome regions which are under selection. By quantifying the strength of natural selection on individuals of *Rhinanthus* including *R. minor*, *R. major* and their hybrids in each habitat we can define how much of the ongoing introgression is adaptive, leading towards strengthening the reproduction barrier, or other evolutionary events. But the limitations for studying adaptive introgression in *R. minor* and *R. major* are related to the large size of both species' genomes, the unavailability of a reference genome, a wide range of niches and habitats, the long seed stratification period, difficulties with cultivation in a greenhouse and their dependency on host species. Although the effect of these parasites on their hosts has been studied (Press and Phoenix 2005), the effect of the host species on *Rhinanthus* fitness is still not very clear.

Cytoplasmic genetic material, which is found in plastids and mitochondria, can be directly transferred from the maternal plant to the offspring and this transmission is independent of nuclear genes (Roach and Wulff 1987). These

maternal elements have effects on offspring traits and they may decrease over time or persist during the whole life cycle and sometimes continue through several generations. Maternal elements in plants directly or indirectly influence morphology or physiology of their offspring, such as seed mass/size or late germination (Videvall et al. 2016). The main interaction of chloroplast or mitochondrial genes with the nuclear genome is through signaling pathways (Koussevitzky et al. 2007). Our experiment in chapter 5 was conducted only during stratification and germination of hybrid seeds, while maternal effects might last much longer. Therefore, further experiments are required to determine whether the observed maternal effect persists over time and to what degree it leads to differences in the hybrid phenotype. It has been demonstrated that the experience of environmental conditions in which mother plants have been grown can be transferred to their offspring through organelle DNA or genomic imprinting, which can be beneficial for the next generation (Uller 2008). It has been shown in *Arabidopsis* that seasonal seed dormancy and timing of germination can be affected by the environmental conditions experienced by mother plants (Munir et al. 2001).

Cyto-nuclear incompatibilities are genetic incompatibilities due to improper interactions between mitochondrial or chloroplast and nuclear genomes that affect fitness traits, such as viability and fertility of the hybrids (Barnard-Kubow et al. 2016). Cyto-nuclear incompatibilities can happen between and within species and often cause an asymmetrical decrease in fitness in offspring of reciprocal crosses (Fishman and Willis 2006).

Parasitic plants are categorized in photosynthetic parasites (hemiparasites) and nonphotosynthetic parasites (holoparasites). Hemiparasitic plants possess varying rates of photosynthetic ability, while holoparasitic plants obtain all nutrients from their host and show complete loss of photosynthetic ability (Irving and Cameron 2009). A reduction in chloroplast genome size or even the complete loss of chloroplasts has been observed in holoparasitic plants (Barrett et al. 2014). In some cases, like *Cassytha filiformis* and *Striga hermonthica*, the quadripartite genome structure, which is a conserved part of the chloroplast, has been lost or perturbed (Wu et al. 2017). *Rhinanthus* is hemiparasitic, but we do not have their chloroplast genome information available in order to check the level of size decrease or lost genes. Although we found a low germination rate in F_{1a} and no significant upregulation or down-regulation of cytoplasmic gene expression, due to their influence on nuclear gene expression they should not be ignored and more studies are needed to explore the importance of cytoplasmic genes in *Rhinanthus* hybrids. Even though deterioration is commonly happening in the chloroplast of

parasitic plants (Irving and Cameron 2009), the mitochondrial genome could be more involved than the chloroplast in cyto-nuclear interactions, which requires more investigation.

6.3 Conclusions and perspectives

This work sets the next step towards understanding hybridization and introgression between *Rhinanthus major* and *R. minor*. We succeeded to establish a large panel of species-specific markers using ddRADseq to discriminate *R. minor* from *R. major*. Furthermore, we determined the degree of introgression of these loci across different mixed populations in order to detect similarities and differences in introgression in those populations. We were not yet able to find out why some of the loci had more introgression and how they are linked. However, this can be investigated by using the whole genome sequence of *R. minor*, which is currently being annotated. After this, the effects of different environments on introgression between both species can be studied in more detail. Even though we did not have linkage information of these species-specific loci, using PCoA gave us some indications of linkage between them. This is just a first estimation and it needs to be studied more in detail by looking at the genome or a genetic map based on hybrids. During my work, I have tried to make an F_2 population of selfed F_1 hybrids in order to find QTLs related to flower shape, pollen and nectar production, leaf structure and volatile compounds, but because of problems during cultivation there were not enough individuals to make a reliable genetic linkage map. This will be continued, and the output of this project could also help us to find other loci related to asymmetric introgression between *R. minor* and *R. major* and their relation with our species-specific loci. Since ddRADseq has been set up in our lab, more than 70 samples of *Rhinanthus* species from different countries in Europe have been sequenced in order to provide more resolution in phylogeographical patterns in *R. major* and *R. minor* which have been studied by Vrancken et al (2009, 2012), and to establish a phylogeny of the whole genus. Fitness analysis of F_1 hybrids showed that they have similar fitness as their parents (Wesselingh et al. 2019), but the fitness of F_2 and backcross hybrids is still under study. The SNP panel will also be used to see if hybrids with high fitness have specific gene combinations. Our result of gene expression analysis in reciprocal hybrids showed that the maternal genome has an effect on seed germination and dormancy, but more studies are needed in order to find the mechanism behind this differential gene expression. All in all, this thesis is just a start, but it has

opened the way towards a full understanding of hybridization and introgression between *Rhinanthus major* and *R. minor*.

References

Aalen RB (1999) Peroxiredoxin antioxidants in seed physiology. Seed Science Research 9:285–295. https://doi.org/10.1017/S096025859900029X

Abbott R, Albach D, Ansell S, et al (2013) Hybridization and speciation. Journal of Evolutionary Biology 26:229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. TAXON 59:1375–1386. https://doi.org/10.1002/tax.595005

Adams KL, Wendel JF (2005) Novel patterns of gene expression in polyploid plants. Trends in Genetics 21:539–543. https://doi.org/10.1016/j.tig.2005.07.009

Aldhebiani AY (2018) Species concept and speciation. Saudi J Biol Sci 25:437–440. https://doi.org/10.1016/j.sjbs.2017.04.013

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (Pat) (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Annals of Botany 120:183–194. https://doi.org/10.1093/aob/mcx079

Anderson E, Hubricht L (1938) Hybridization in tradescantia. Iii. the evidence for introgressive hybridization. American Journal of Botany 25:396–402. https://doi.org/10.1002/j.1537-2197.1938.tb09237.x

Anderson E, Stebbins GL (1954) Hybridization as an evolutionary stimulus. Evolution 8:378–388. https://doi.org/10.2307/2405784

Arnold ML (1997) Natural hybridization and evolution. Oxford University Press, Oxford, New York

Arnold ML, Bulger MR, Burke JM, et al (1999) Natural hybridization: how low can you go and still be important? Ecology 80:371–381. https://doi.org/10.1890/0012-9658(1999)080[0371:NHHLCY]2.0.CO;2

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D (2015) The origins of reproductive isolation in plants. New Phytol 207:968–984. https://doi.org/10.1111/nph.13424

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17:513–518. https://doi.org/10.1016/j.gde.2007.09.001

Barnard-Kubow KB, So N, Galloway LF (2016) Cytonuclear incompatibility contributes to the early stages of speciation. Evolution 70:2752–2766. https://doi.org/10.1111/evo.13075

Barrett CF, Freudenstein JV, Li J, et al (2014) Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Mol Biol Evol 31:3095–3112. https://doi.org/10.1093/molbev/msu252

Barton NH (2001) The role of hybridization in evolution. Molecular Ecology 10:551–568. https://doi.org/10.1046/j.1365-294x.2001.01216.x

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annual Review of Ecology and Systematics 16:113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Bell GDM, Kane NC, Rieseberg LH, Adams KL (2013) RNA-Seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. Genome Biology and Evolution 5:1309–1323. https://doi.org/10.1093/gbe/evt072

Bentsink L, Koornneef M (2008) Seed dormancy and germination. arbo.j 2008:. https://doi.org/10.1199/tab.0119

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bridle JR (2002) Genes, categories and species: the evolutionary and cognitive causes of the species problem. Heredity 88:488–488. https://doi.org/10.1038/sj.hdy.6800082

Bryant DM, Johnson K, DiTommaso T, et al (2017) A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Reports 18:762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Buggs RJA, Soltis PS, Soltis DE (2009) Does hybridization between divergent progenitors drive whole-genome duplication? Molecular Ecology 18:3334–3339. https://doi.org/10.1111/j.1365-294X.2009.04285.x

Buggs RJA, Zhang L, Miles N, et al (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Current Biology 21:551–556. https://doi.org/10.1016/j.cub.2011.02.016

Bullock JM, Moy IL, Coulson SJ, Clarke RT (2003) Habitat-specific dispersal: environmental effects on the mechanisms and patterns of seed movement in a grassland herb Rhinanthus minor. Ecography 26:692–704. https://doi.org/10.1034/j.1600-0587.2003.03525.x

Bullock JM, Pywell RF (2005) Rhinanthus: a tool for restoring diverse grassland? Folia Geobot 40:273–288. https://doi.org/10.1007/BF02803240

Burgarella C, Barnaud A, Kane NA, et al (2019) Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. Front Plant Sci 10:. https://doi.org/10.3389/fpls.2019.00004

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8:. https://doi.org/10.1093/gigascience/giz100

Cameron DD, Seel WE (2007) Functional anatomy of haustoria formed by *Rhinanthus minor*: linking evidence from histology and isotope tracing. New Phytologist 174:412–419. https://doi.org/10.1111/j.1469-8137.2007.02013.x

Campion-Bourget F (1980) Autogamy and allogamy in the five main species of French *Rhinantus* L. Revue de cytologie et de biologie vegetales - le botaniste 3:

Castro M, Castro S, Loureiro J (2012) Genome size variation and incidence of polyploidy in Scrophulariaceae sensu lato from the *Iberian Peninsula*. AoB Plants 2012:pls037. https://doi.org/10.1093/aobpla/pls037

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124–3140. https://doi.org/10.1111/mec.12354

Chen H, Ruan J, Chu P, et al (2020) AtPER1 enhances primary seed dormancy and reduces seed germination by suppressing the ABA catabolism and GA biosynthesis in Arabidopsis seeds. The Plant Journal 101:310–323. https://doi.org/10.1111/tpj.14542

Chiang GCK, Bartsch M, Barua D, et al (2011) DOG1 expression is predicted by the seed-maturation environment and contributes to geographical variation in germination in Arabidopsis thaliana. Mol Ecol 20:3336–3349. https://doi.org/10.1111/j.1365-294X.2011.05181.x

Christie K, Strauss SY (2019) Reproductive isolation and the maintenance of species boundaries in two serpentine endemic Jewelflowers. Evolution 73:1375–1391. https://doi.org/10.1111/evo.13767

Christopher DA, Mitchell RJ, Trapnell DW, et al (2019) Hermaphroditism promotes mate diversity in flowering plants. Am J Bot 106:1131–1136. https://doi.org/10.1002/ajb2.1336

Chunco AJ (2014) Hybridization in a warmer world. Ecol Evol 4:2019–2031. https://doi.org/10.1002/ece3.1052

Coulson SJ, Bullock JM, Stevenson MJ, Pywell RF (2001) Colonization of grassland by sown species: dispersal versus microsite limitation in responses to management. Journal of Applied Ecology 38:204–216. https://doi.org/10.1046/j.1365-2664.2001.00585.x

Coyne JA, Orr HA (2004) Speciation. Sinauer

D. Whitney K, A. Randell R, H. Rieseberg L (2015) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist. https://doi.org/10.1086/504606

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dasmahapatra KK, Walters JR, Briscoe AD, et al (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98. https://doi.org/10.1038/nature11041

De Queiroz K (2007) Species concepts and species delimitation. Systematic Biology 56:879–886

Debeaujon I, Lepiniec L, Pourcel L, Routaboul J-M (2007) Seed coat development and dormancy. In: Annual plant reviews volume 27: seed development, dormancy and germination. John Wiley & Sons, Ltd, pp 25–49

Devos N, Tyteca D, Raspé O, et al (2003) Patterns of chloroplast diversity among western European Dactylorhiza species (Orchidaceae). Plant Syst Evol 243:85–97. https://doi.org/10.1007/s00606-003-0068-0

Donohue K, Dorn L, Griffith C, et al (2005) Niche construction through germination cueing: life-history responses to timing of germination in *Arabidopsis thaliana*. Evolution 59:771–785

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin

Ducarme V, Risterucci AM, Wesselingh RA (2008) PERMANENT GENETIC RESOURCES: Development of microsatellite markers in *Rhinanthus angustifolius* and cross-species amplification. Molecular Ecology Resources 8:384–386. https://doi.org/10.1111/j.1471-8286.2007.01966.x

Ducarme V, Vrancken J, Wesselingh RA (2010) Hybridization in annual plants: patterns and dynamics during a four-year study in mixed *Rhinanthus* populations. Folia Geobotanica 45:387

Ducarme V, Wesselingh RA (2013) Outcrossing rates in two self-compatible, hybridising *Rhinanthus* species: implications for hybrid formation. Plant Biology 15:541–547. https://doi.org/10.1111/j.1438-8677.2012.00668.x

Ducarme V, Wesselingh RA (2009) Performance of two Rhinanthus species under different hydric conditions. Plant Ecol 206:263. https://doi.org/10.1007/s11258-009-9640-1

Ducarme V, Wesselingh RA (2005) Detecting hybridization in mixed populations of *Rhinanthus minor* and *Rhinanthus angustifolius*. Folia Geobot 40:151–161. https://doi.org/10.1007/BF02803231

East EM (1936) Heterosis. Genetics 21:375–397

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539–563. https://doi.org/10.1146/annurev.ecolsys.30.1.539

Ellstrand NC, Whitkus R, Rieseberg LH (1996) Distribution of spontaneous plant hybrids. Proc Natl Acad Sci U S A 93:5090–5093

El-Maarouf-Bouteau H, Bailly C (2008) Oxidative signaling in seed germination and dormancy. Plant Signaling & Behavior 3:175–182. https://doi.org/10.4161/psb.3.3.5539

Eriksson MC, Szukala A, Tian B, Paun O (2020) Current research frontiers in plant epigenetics: an introduction to a Virtual Issue. New Phytologist 226:285–288. https://doi.org/10.1111/nph.16493

Fernández Farnocchia RB, Benech-Arnold RL, Batlla D (2019) Regulation of seed dormancy by the maternal environment is instrumental for maximizing plant fitness in Polygonum aviculare. Journal of Experimental Botany 70:4793–4806. https://doi.org/10.1093/jxb/erz269

Field DL, Ayre DJ, Whelan RJ, Young AG (2011) Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregata* and common *E. rubida*. Heredity 106:841–853. https://doi.org/10.1038/hdy.2010.127

Finkelstein RR, Lynch TJ (2000) The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. The Plant Cell 12:599–609. https://doi.org/10.1105/tpc.12.4.599

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39:W29–W37. https://doi.org/10.1093/nar/gkr367

Fishman L, Willis JH (2006) A cytonuclear incompatibility causes anther sterility in Mimulus hybrids. Evolution 60:1372–1381. https://doi.org/10.1554/05-708.1

Fitzpatrick BM, Ryan ME, Johnson JR, et al (2015) Hybridization and the species problem in conservation. Current Zoology 61:206–216. https://doi.org/10.1093/czoolo/61.1.206

Forbis TA, Floyd SK, de Queiroz A (2002) The evolution of embryo size in angiosperms and other seed plants: implications for the evolution of seed dormancy. Evolution 56:2112–2125. https://doi.org/10.1111/j.0014-3820.2002.tb00137.x

Fritsch A (2009) Package "scrime" analysis of high-dimensional categorical data such as snp data. Accessed 24 Apr 2021

Gao X-H, Bedhomme M, Veyel D, et al (2009) Methods for analysis of protein glutathionylation and their application to photosynthetic organisms. Molecular Plant 2:218–235. https://doi.org/10.1093/mp/ssn072

Gavrilets S (2000) Waiting time to parapatric speciation. Proc Biol Sci 267:2483–2492

Genovart M (2008) Natural hybridization and conservation. Biodivers Conserv 18:1435. https://doi.org/10.1007/s10531-008-9550-x

Gibson CC, Watkinson AR (1991) Host selectivity and the mediation of competition by the root hemiparasite Rhinanthus minor. Oecologia 86:81–87. https://doi.org/10.1007/BF00317393

Goulet BE, Roda F, Hopkins R (2017) Hybridization in plants: old ideas, new techniques. Plant Physiology 173:65–78. https://doi.org/10.1104/pp.16.01340

Grabherr MG, Haas BJ, Yassour M, et al (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Gramlich S, Wagner ND, Hörandl E (2018) RAD-seq reveals genetic structure of the F2-generation of natural willow hybrids (*Salix* L.) and a great potential for interspecific introgression. BMC Plant Biology 18:317. https://doi.org/10.1186/s12870-018-1552-6

Grant PR, Grant BR (1992) Hybridization of bird species. Science 256:193–197. https://doi.org/10.1126/science.256.5054.193

Grant PR, Grant BR (2019) Hybridization increases population variation during adaptive radiation. PNAS 116:23216–23224. https://doi.org/10.1073/pnas.1913534116

Greaves IK, Gonzalez-Bayon R, Wang L, et al (2015) Epigenetic changes in hybrids. Plant Physiology 168:1197–1205. https://doi.org/10.1104/pp.15.00231

Gross BL (2012) Genetic and phenotypic divergence of homoploid hybrid species from parental species. Heredity (Edinb) 108:157–158. https://doi.org/10.1038/hdy.2011.80

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. Journal of Heredity 96:241–252. https://doi.org/10.1093/jhered/esi026

Groszmann M, Greaves IK, Fujimoto R, et al (2013) The role of epigenetics in hybrid vigour. Trends in Genetics 29:684–690. https://doi.org/10.1016/j.tig.2013.07.004

Guerreiro MPG (2014) Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in Drosophila. Mob Genet Elements 4:. https://doi.org/10.4161/mge.34394

Guo Q (2014) Plant hybridization: the role of human disturbance and biological invasion. Diversity and Distributions 20:1345–1354. https://doi.org/10.1111/ddi.12245

Gupta SK, Baek J, Carrasquilla-Garcia N, Penmetsa RV (2015) Genome-wide polymorphism detection in peanut using next-generation restriction-site-associated DNA (RAD) sequencing. Mol Breeding 35:145. https://doi.org/10.1007/s11032-015-0343-0

Hämälä T, Mattila TM, Leinonen PH, et al (2017) Role of seed germination in adaptation and reproductive isolation in Arabidopsis lyrata. Molecular Ecology 26:3484–3496. https://doi.org/10.1111/mec.14135

Hambler DJ (1954) Cytology of the Scrophulariaceae and Orobanchaceae. Nature 174:838–838. https://doi.org/10.1038/174838a0

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105:795–809. https://doi.org/10.1093/jhered/esu033

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol Ecol 25:2454–2466. https://doi.org/10.1111/mec.13582

Haslekås C, Viken MK, Grini PE, et al (2003) Seed 1-cysteine peroxiredoxin antioxidants are not involved in dormancy, but contribute to inhibition of

germination during stress. Plant Physiol 133:1148–1157. https://doi.org/10.1104/pp.103.025916

Haynes W (2013) Benjamini–Hochberg Method. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H (eds) Encyclopedia of Systems Biology. Springer, New York, NY, pp 78–78

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22:4606–4618. https://doi.org/10.1111/mec.12415

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. Genetica 139:617–638. https://doi.org/10.1007/s10709-011-9547-3

Hodel RGJ, Chen S, Payton AC, et al (2017) Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Scientific Reports 7:17598. https://doi.org/10.1038/s41598-017-16810-7

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecol 21:797–807. https://doi.org/10.1007/s10980-005-5245-9

Hoyle GL, Steadman KJ, Good RB, et al (2015) Seed germination strategies: an evolutionary trajectory independent of vegetative functional traits. Front Plant Sci 6:. https://doi.org/10.3389/fpls.2015.00731

Huang X, Yang S, Gong J, et al (2016) Genomic architecture of heterosis for yield traits in rice. Nature 537:629–633. https://doi.org/10.1038/nature19760

Huber W, Carey VJ, Gentleman R, et al (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods 12:115–121. https://doi.org/10.1038/nmeth.3252

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Hussain S, Kim SH, Bahk S, et al (2020) The auxin signaling repressor IAA8 promotes seed germination through down-regulation of ABI3 transcription in Arabidopsis. Front Plant Sci 11:. https://doi.org/10.3389/fpls.2020.00111

Irving LJ, Cameron DD (2009) You are what you eat: Interactions between root parasitic plants and their hosts. In: Advances in Botanical Research. Academic Press, pp 87–138

Ishida S, Takahashi A, Matsushima N, et al (2011) The long-term consequences of hybridization between the two Daphnia species, D. galeata and D. dentifera, in

mature habitats. BMC Evolutionary Biology 11:209. https://doi.org/10.1186/1471-2148-11-209

Iwasaki M, Hyvärinen L, Piskurewicz U, Lopez-Molina L (2019) Non-canonical RNA-directed DNA methylation participates in maternal and environmental control of seed dormancy. eLife 8:e37434. https://doi.org/10.7554/eLife.37434

Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79. https://doi.org/10.1038/316076a0

Joel DM, Gressel J, Musselman LJ (eds) (2013) Parasitic Orobanchaceae: parasitic mechanisms and control strategies. Springer-Verlag, Berlin Heidelberg

Johansen-Morris AD, Latta RG (2006) Fitness consequences of hybridization between ecotypes of Avena barbata: hybrid breakdown, hybrid vigor, and transgressive segregation. Evolution 60:1585–1595

Jost J (2017) Relations and dependencies between morphological characters. Theory Biosci 136:69–83. https://doi.org/10.1007/s12064-017-0248-z

Kagawa K, Takimoto G (2018) Hybridization can promote adaptive radiation by means of transgressive segregation. Ecology Letters 21:264–274. https://doi.org/10.1111/ele.12891

Kantama L, Junbuathong S, Sakulkoo J, et al (2013) Epigenetic changes and transposon reactivation in Thai rice hybrids. Mol Breeding 31:815–827. https://doi.org/10.1007/s11032-013-9836-x

Kerbs B, Ressler J, Kelly JK, et al (2017) The potential role of hybridization in diversification and speciation in an insular plant lineage: insights from synthetic interspecific hybrids. AoB PLANTS 9:. https://doi.org/10.1093/aobpla/plx043

Kermode AR (2005) Role of abscisic acid in seed dormancy. J Plant Growth Regul 24:319–344. https://doi.org/10.1007/s00344-005-0110-2

Kim M, Cui M-L, Cubas P, et al (2008a) Regulatory genes control a key morphological and ecological trait transferred between species. Science 322:1116–1119. https://doi.org/10.1126/science.1164371

Kim S-G, Lee A-K, Yoon H-K, Park C-M (2008b) A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination. The Plant Journal 55:77–88. https://doi.org/10.1111/j.1365-313X.2008.03493.x

Kim S-G, Park C-M (2008) Gibberellic acid-mediated salt signaling in seed germination. Plant Signaling & Behavior 3:877–879. https://doi.org/10.4161/psb.3.10.6247

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutral markers in molecular ecology. Int J Mol Sci 12:3966–3988. https://doi.org/10.3390/ijms12063966

Koussevitzky S, Nott A, Mockler TC, et al (2007) Signals from chloroplasts converge to regulate nuclear gene expression. Science 316:715–719. https://doi.org/10.1126/science. 1140516

Krieger U, Lippman ZB, Zamir D (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nature Genetics 42:459–463. https://doi.org/10.1038/ng.550

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. Journal of Molecular Biology 305:567–580. https://doi.org/10.1006/jmbi.2000.4315

Kwak MM (1980) Artificial and natural hybridization and introgression in *Rhinanthus* (*Scrophulariaceae*) in relation to bumblebee pollination. Taxon 29:613–628. https://doi.org/10.2307/1220333

Kwak MM (1979) Effects of bumblebee visits on the seed set of pedicularis, Rhinanthus and Melampyrum (scrophulariaceae) in the netherlands. Acta Botanica Neerlandica 28:177–195. https://doi.org/10.1111/j.1438-8677.1979.tb00333.x

Labrador M, Farré M, Utzet F, Fontdevila A (1999) Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol 16:931–937. https://doi.org/10.1093/oxfordjournals.molbev.a026182

Lafon-Placette C, Johannessen IM, Hornslien KS, et al (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between Arabidopsis lyrata and Arabidopsis arenosa in Central Europe. PNAS 114:E1027–E1035. https://doi.org/10.1073/pnas.1615123114

Lafon-Placette C, Köhler C (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. Molecular Ecology 25:2620–2629. https://doi.org/10.1111/mec.13552

Lagesen K, Hallin P, Rødland EA, et al (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35:3100–3108. https://doi.org/10.1093/nar/gkm160

Lammerts van Bueren ET, Backes G, de Vriend H, Østergård H (2010) The role of molecular markers and marker assisted selection in breeding for organic agriculture. Euphytica 175:51–64. https://doi.org/10.1007/s10681-010-0169-0

Landry CR, Wittkopp PJ, Taubes CH, et al (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. Genetics 171:1813–1822. https://doi.org/10.1534/genetics.105.047449

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. https://doi.org/10.1186/1471-2105-12-323

Li H, Handsaker B, Wysoker A, et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Litsios G, Salamin N (2014) Hybridisation and diversification in the adaptive radiation of clownfishes. BMC Evolutionary Biology 14:245. https://doi.org/10.1186/s12862-014-0245-5

Liu X, Zhang H, Zhao Y, et al (2013) Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. PNAS 110:15485–15490. https://doi.org/10.1073/pnas.1304651110

Lo S-F, Yang S-Y, Chen K-T, et al (2008) A novel class of Gibberellin 2-Oxidases control semidwarfism, tillering, and root development in Rice. Plant Cell 20:2603–2618. https://doi.org/10.1105/tpc.108.060913

López de Heredia U, Mora-Márquez F, Goicoechea PG, et al (2020) ddRAD sequencing-based identification of genomic boundaries and permeability in Quercus ilex and Q. suber hybrids. Front Plant Sci 11:. https://doi.org/10.3389/fpls.2020.564414

López-Caamal A, Tovar-Sánchez E (2014) Genetic, morphological, and chemical patterns of plant hybridization. Revista Chilena de Historia Natural 87:16. https://doi.org/10.1186/s40693-014-0016-0

Lowry DB, Modliszewski JL, Wright KM, et al (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. Philosophical Transactions of the Royal Society B: Biological Sciences 363:3009–3021. https://doi.org/10.1098/rstb.2008.0064

Lucas A, Jasson S (2006) Using amap and ctc packages for huge clustering. R News 58–60

Lukhtanov VA, Shapoval NA, Anokhin BA, et al (2015) Homoploid hybrid speciation and genome evolution via chromosome sorting. Proceedings of the Royal Society B: Biological Sciences 282:20150157. https://doi.org/10.1098/rspb.2015.0157

Lyu J (2016) Crop evolution: after allopolyploidization. Nature Plants 2:1–1. https://doi.org/10.1038/nplants.2016.156

Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biological Journal of the Linnean Society 82:453–466. https://doi.org/10.1111/j.1095-8312.2004.00332.x

Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. Journal of Zoology 284:151–182. https://doi.org/10.1111/j.1469-7998.2011.00829.x

Mackay IJ, Cockram J, Howell P, Powell W (2021) Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. Plant Biotechnology Journal 19:26–34. https://doi.org/10.1111/pbi.13481

Maki M, Murata J (2001) Allozyme analysis of the hybrid origin of *Arisaema ehimense* (Araceae). Heredity 86:87–93. https://doi.org/10.1046/j.1365-2540.2001.00813.x

Mallet J (1995) A species definition for the modern synthesis. Trends in Ecology & Evolution 10:294–299. https://doi.org/10.1016/0169-5347(95)90031-4

Mallet J (2010) Why was Darwin's view of species rejected by twentieth century biologists? Biol Philos 25:497–527. https://doi.org/10.1007/s10539-010-9213-7

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 363:2971–2986. https://doi.org/10.1098/rstb.2008.0081

Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20:229–237. https://doi.org/10.1016/j.tree.2005.02.010

Marin M, Laverack G, Matthews S, Powell AA (2019) Germination characteristics of Rhinanthus minor influence field emergence, competitiveness and potential use in restoration projects. Plant Biology 21:470–479. https://doi.org/10.1111/plb.12707

Martin NH, Bouck AC, Arnold ML (2006) Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics 172:2481–2489. https://doi.org/10.1534/genetics.105.053538

Mason-Gamer RJ, Kellogg EA (1996) Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Systematic Biology 45:524–545. https://doi.org/10.1093/sysbio/45.4.524

Matthies D (2017) Interactions between a root hemiparasite and 27 different hosts: growth, biomass allocation and plant architecture. Perspectives in Plant Ecology, Evolution and Systematics 24:118–137. https://doi.org/10.1016/j.ppees.2016.12.006

McNeal JR, Bennett JR, Wolfe AD, Mathews S (2013) Phylogeny and origins of holoparasitism in Orobanchaceae. American Journal of Botany 100:971–983. https://doi.org/10.3732/ajb.1200448

Mendelson TC (2003) Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: Etheostoma). Evolution 57:317–327. https://doi.org/10.1111/j.0014-3820.2003.tb00266.x

Merotto A, Goulart ICGR, Nunes AL, et al (2016) Evolutionary and social consequences of introgression of nontransgenic herbicide resistance from rice to weedy rice in Brazil. Evolutionary Applications 9:837–846. https://doi.org/10.1111/eva.12387

Mirzaei K, Wesselingh RA (2021) Development of a large set of diagnostic SNP markers using ddRAD-seq to study hybridization in Rhinanthus major and R. minor (Orobanchaceae). Conservation Genetics Resources. https://doi.org/10.1007/s12686-020-01181-5

Mitchell N, Campbell LG, Ahern JR, et al (2019a) Correlates of hybridization in plants. Evolution Letters 3:570–585. https://doi.org/10.1002/evl3.146

Mitchell N, Owens GL, Hovick SM, et al (2019b) Hybridization speeds adaptive evolution in an eight-year field experiment. Scientific Reports 9:6746. https://doi.org/10.1038/s41598-019-43119-4

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Science 43:1235–1248. https://doi.org/10.2135/cropsci2003.1235

Mráz P, Chrtek J, Fehrer J, Placková I (2005) Rare recent natural hybridization in Hieracium s. str. – evidence from morphology, allozymes and chloroplast DNA. Plant Syst Evol 255:177–192. https://doi.org/10.1007/s00606-005-0329-1

Munir J, Dorn LA, Donohue K, Schmitt J (2001) The effect of maternal photoperiod on seasonal dormancy in Arabidopsis thaliana (Brassicaceae). Am J Bot 88:1240–1249

Natalis L (2013) Reproductive barriers and their role in asymmetric Rhinanthus hybridization. UCL - Université Catholique de Louvain

Natalis LC, Wesselingh RA (2012) Post-pollination barriers and their role in asymmetric hybridization in Rhinanthus (Orobanchaceae). American Journal of Botany 99:1847–1856

Natalis LC, Wesselingh RA (2012) Shared pollinators and pollen transfer dynamics in two hybridizing species, *Rhinanthus minor* and *R. angustifolius*. Oecologia 170:709–721. https://doi.org/10.1007/s00442-012-2346-4

Nieto Feliner G, Álvarez I, Fuertes-Aguilar J, et al (2017) Is homoploid hybrid speciation that rare? An empiricist's view. Heredity 118:513–516. https://doi.org/10.1038/hdy.2017.7

O'Neill RJW, O'Neill MJ, Graves JAM (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–72. https://doi.org/10.1038/29985

Oracz K, Bouteau HE-M, Farrant JM, et al (2007) ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. The Plant Journal 50:452–465. https://doi.org/10.1111/j.1365-313X.2007.03063.x

Ortego J, Gugger PF, Sork VL (2017) Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. New Phytologist 213:942–955. https://doi.org/10.1111/nph.14182

Ossowski S, Schneeberger K, Lucas-Lledó JI, et al (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92–94. https://doi.org/10.1126/science.1180677

Ostevik KL, Andrew RL, Otto SP, Rieseberg LH (2016) Multiple reproductive barriers separate recently diverged sunflower ecotypes. Evolution 70:2322–2335. https://doi.org/10.1111/evo.13027

Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annual Review of Genetics 34:401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Paun O, Fay MF, Soltis DE, Chase MW (2007) Genetic and epigenetic alterations after hybridization and genome doubling. Taxon 56:649–56

Paun O, Forest F, Fay MF, Chase MW (2009) Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol 182:507–518. https://doi.org/10.1111/j.1469-8137.2009.02767.x

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources 10:806–820. https://doi.org/10.1111/j.1755-0998.2010.02883.x

Peterson BK, Weber JN, Kay EH, et al (2012) Double Digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLOS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pickup M, Brandvain Y, Fraïsse C, et al (2019) Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. New Phytol 224:1035–1047. https://doi.org/10.1111/nph.16180

Porter AH, Johnson NA (2002) Speciation despite gene flow when developmental pathways evolve. Evolution 56:2103–2111

Press MC, Phoenix GK (2005) Impacts of parasitic plants on natural communities. New Phytologist 166:737–751. https://doi.org/10.1111/j.1469-8137.2005.01358.x

Punta M, Coggill PC, Eberhardt RY, et al (2012) The Pfam protein families database. Nucleic Acids Research 40:D290–D301. https://doi.org/10.1093/nar/gkr1065

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing

Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the Monkeyflowers Mimulus Lewisii and M. Cardinalis (phrymaceae). Evolution 57:1520–1534. https://doi.org/10.1111/j.0014-3820.2003.tb00360.x

Reuveni E, Getselter D, Oron O, Elliott E (2018) Differential contribution of cis and trans gene transcription regulatory mechanisms in amygdala and prefrontal cortex and modulation by social stress. Scientific Reports 8:6339. https://doi.org/10.1038/s41598-018-24544-3

Rieseberg LH (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28:359–389. https://doi.org/10.1146/annurev.ecolsys.28.1.359

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity 83:363–372. https://doi.org/10.1038/sj.hdy.6886170

Rieseberg LH, Baird SJ, Gardner KA (2000) Hybridization, introgression, and linkage evolution. Plant Mol Biol 42:205–224

Rieseberg LH, Carney SE (1998) Plant hybridization. New Phytologist 140:599–624. https://doi.org/10.1046/j.1469-8137.1998.00315.x

Rieseberg LH, Kim S-C, Randell RA, et al (2007) Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129:149–165. https://doi.org/10.1007/s10709-006-9011-y

Ritchie ME, Phipson B, Wu D, et al (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43:e47–e47. https://doi.org/10.1093/nar/gkv007

Roach DA, Wulff RD (1987) Maternal effects in plants. Annual Review of Ecology and Systematics 18:209–235

Robertson G, Schein J, Chiu R, et al (2010) De novo assembly and analysis of RNA-seq data. Nature Methods 7:909–912. https://doi.org/10.1038/nmeth.1517

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 22:939–946. https://doi.org/10.1101/gr.128124.111

Rohlf FJ (1972) An empirical comparison of three ordination techniques in numerical taxonomy. Systematic Zoology 21:271–280. https://doi.org/10.2307/2412165

Rothfels CJ, Johnson AK, Hovenkamp PH, et al (2015) Natural hybridization between genera that diverged from each other approximately 60 million years ago. The American Naturalist 185:433–442. https://doi.org/10.1086/679662

Rouhier N, Gelhaye E, Sautiere P-E, et al (2001) Isolation and characterization of a new peroxiredoxin from poplar sieve tubes that uses either glutaredoxin or thioredoxin as a proton donor. Plant Physiology 127:1299–1309. https://doi.org/10.1104/pp.010586

Rouhier N, Lemaire SD, Jacquot J-P (2008) The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. Annu Rev Plant Biol 59:143–166. https://doi.org/10.1146/annurev.arplant.59.032607.092811

Rümer S, Cameron DD, Wacker R, et al (2007) An anatomical study of the haustoria of Rhinanthus minor attached to roots of different hosts. Flora - Morphology, Distribution, Functional Ecology of Plants 202:194–200. https://doi.org/10.1016/j.flora.2006.07.002

Ruse M (1969) Definitions of species in biology. The British Journal for the Philosophy of Science 20:97–119

Sackman AM, Rokyta DR (2013) The adaptive potential of hybridization demonstrated with bacteriophages. J Mol Evol 77:221–230. https://doi.org/10.1007/s00239-013-9586-8

Sakamoto T, Miura K, Itoh H, et al (2004) An overview of Gibberellin metabolism enzyme genes and their related mutants in Rice. Plant Physiology 134:1642–1653. https://doi.org/10.1104/pp.103.033696

Sams AJ, Dumaine A, Nédélec Y, et al (2016) Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biology 17:246. https://doi.org/10.1186/s13059-016-1098-6

Sandalio LM, Romero-Puertas MC (2015) Peroxisomes sense and respond to environmental cues by regulating ROS and RNS signalling networks. Annals of Botany 116:475–485. https://doi.org/10.1093/aob/mcv074

Sanz N, Araguas RM, Fernández R, et al (2009) Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. Conserv Genet 10:225–236. https://doi.org/10.1007/s10592-008-9550-0

Sarath G, Hou G, Baird LM, Mitchell RB (2007) Reactive oxygen species, ABA and nitric oxide interactions on the germination of warm-season C4-grasses. Planta 226:697–708. https://doi.org/10.1007/s00425-007-0517-z

Schrider DR, Ayroles J, Matute DR, Kern AD (2018) Supervised machine learning reveals introgressed loci in the genomes of Drosophila simulans and D. sechellia. PLOS Genetics 14:e1007341. https://doi.org/10.1371/journal.pgen.1007341

Schumer M, Rosenthal GG, Andolfatto P (2014) How common is homoploid hybrid speciation? Evolution 68:1553–1560. https://doi.org/10.1111/evo.12399

Scopece G, Widmer A, Cozzolino S (2008) Evolution of postzygotic reproductive isolation in a guild of deceptive orchids. The American Naturalist 171:315–326. https://doi.org/10.1086/527501

Seehausen O (2004) Hybridization and adaptive radiation. Trends in Ecology & Evolution 19:198–207. https://doi.org/10.1016/j.tree.2004.01.003

Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. Molecular Ecology 17:30–44. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Seidl F, Levis NA, Jones CD, et al (2019) Variation in hybrid gene expression: Implications for the evolution of genetic incompatibilities in interbreeding species. Molecular Ecology 28:4667–4679. https://doi.org/10.1111/mec.15246

Serrato-Capuchina A, Matute DR (2018) The role of transposable elements in speciation. Genes (Basel) 9:254. https://doi.org/10.3390/genes9050254

Sharma N, Park S-W, Vepachedu R, et al (2004) Isolation and characterization of an RIP (ribosome-inactivating protein)-like protein from tobacco with dual enzymatic activity. Plant Physiol 134:171–181. https://doi.org/10.1104/pp.103.030205

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annual Review of Plant Biology 60:561–588. https://doi.org/10.1146/annurev.arplant.043008.092039

Song Y, Endepols S, Klemann N, et al (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. Curr Biol 21:1296–1301. https://doi.org/10.1016/j.cub.2011.06.043

Star B, Spencer HG (2013) Effects of genetic drift and gene flow on the selective maintenance of genetic variation. Genetics 194:235–244. https://doi.org/10.1534/genetics.113.149781

Stobie CS, Oosthuizen CJ, Cunningham MJ, Bloomer P (2018) Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. Ecol Evol 8:2326–2342. https://doi.org/10.1002/ece3.3821

Strykstra RJ, Bekker RM, Verweij GL (1996) Establishment of *Rhinanthus angustifolius* in a successional hayfield after seed dispersal by mowing machinery. Acta Botanica Neerlandica 45:557–562

Stukenbrock EH (2016) The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology® 106:104–112. https://doi.org/10.1094/PHYTO-08-15-0184-RVW

Suarez-Gonzalez A, Lexer C, Cronk QCB (2018) Adaptive introgression: a plant perspective. Biol Lett 14:. https://doi.org/10.1098/rsbl.2017.0688

Sun Y, Lu Z, Zhu X, Ma H (2020) Genomic basis of homoploid hybrid speciation within chestnut trees. Nature Communications 11:3375. https://doi.org/10.1038/s41467-020-17111-w

Suni SS, Hopkins R (2018) The relationship between postmating reproductive isolation and reinforcement in Phlox. Evolution 72:1387–1398. https://doi.org/10.1111/evo.13507

Syros TD, Yupsanis TA, Economou AS (2005) Expression of peroxidases during seedling growth in Ebenus cretica L. as affected by light and temperature treatments. Plant Growth Regul 46:143–151. https://doi.org/10.1007/s10725-005-8087-1

Tautz D (2000) Evolution of transcriptional regulation. Current Opinion in Genetics & Development 10:575–579. https://doi.org/10.1016/S0959-437X(00)00130-1

Ter Borg SJ (2005) Dormancy and germination of six Rhinanthus species in relation to climate. Folia Geobot 40:243–260. https://doi.org/10.1007/BF02803238

Tiagi B (1966) Development of the seed and fruit in Rhinanthus Major and R. serotinus. American Journal of Botany 53:645–651. https://doi.org/10.2307/2439740

Tiffin P, Olson S, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. Proceedings of the Royal Society of London Series B: Biological Sciences 268:861–867. https://doi.org/10.1098/rspb.2000.1578

Todesco M, Pascual MA, Owens GL, et al (2016) Hybridization and extinction. Evol Appl 9:892–908. https://doi.org/10.1111/eva.12367

Tuan PA, Kumar R, Rehal PK, et al (2018) Molecular mechanisms underlying abscisic acid/gibberellin balance in the control of seed dormancy and germination in cereals. Front Plant Sci 9:. https://doi.org/10.3389/fpls.2018.00668

Turelli M, Moyle LC (2007) Asymmetric postmating isolation: darwin's corollary to haldane's rule. Genetics 176:1059–1088. https://doi.org/10.1534/genetics.106.065979

Turissini DA, McGirr JA, Patel SS, et al (2018) The rate of evolution of postmating-prezygotic reproductive isolation in Drosophila. Molecular Biology and Evolution 35:312–334. https://doi.org/10.1093/molbev/msx271

Twyford AD, Ennos RA (2012) Next-generation hybridization and introgression. Heredity 108:179–189. https://doi.org/10.1038/hdy.2011.68

Tyler L, Thomas SG, Hu J, et al (2004) DELLA proteins and gibberellin-regulated seed germination and floral development in Arabidopsis. Plant Physiology 135:1008–1019. https://doi.org/10.1104/pp.104.039578

Uller T (2008) Developmental plasticity and the evolution of parental effects. Trends Ecol Evol 23:432–438. https://doi.org/10.1016/j.tree.2008.04.005

Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16:R872-873. https://doi.org/10.1016/j.cub.2006.09.020

Usai G, Mascagni F, Vangelisti A, et al (2020) Interspecific hybridisation and LTR-retrotransposon mobilisation-related structural variation in plants: A case study. Genomics 112:1611–1621. https://doi.org/10.1016/j.ygeno.2019.09.010

Vallejo-Marin M (2012) Mimulus peregrinus (Phrymaceae): A new british allopolyploid species. PhytoKeys 14:1–14. https://doi.org/10.3897/phytokeys.14.3305

van Hengstum T, Lachmuth S, Oostermeijer JGB, et al (2012) Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. Plant Syst Evol 298:1119–1131. https://doi.org/10.1007/s00606-012-0624-6

Videvall E, Sletvold N, Hagenblad J, et al (2016) Strong maternal effects on gene expression in Arabidopsis lyrata hybrids. Molecular Biology and Evolution 33:984–994. https://doi.org/10.1093/molbev/msv342

Vilà M, Weber E, Antonio CMD (2000) Conservation implications of invasion by plant hybridization. Biological Invasions 2:207–217. https://doi.org/10.1023/A:1010003603310

Vos P, Hogers R, Bleeker M, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Vrancken J, Brochmann C, Wesselingh RA (2012) A European phylogeography of Rhinanthus minor compared to Rhinanthus angustifolius: unexpected splits and signs of hybridization. Ecology and Evolution 2:1531–1548. https://doi.org/10.1002/ece3.276

Wang G, Köhler C (2017) Epigenetic processes in flowering plant reproduction. Journal of Experimental Botany 68:797–807. https://doi.org/10.1093/jxb/erw486

Wang W, Li H, Chen Z (2014) Analysis of plastid and nuclear DNA data in plant phylogenetics—evaluation and improvement. Sci China Life Sci 57:280–286. https://doi.org/10.1007/s11427-014-4620-7

Warnes GR, Bolker B, Bonebakker L, et al (2009) gplots: various R programming tools for plotting data. R package version 2:1

Wei L, Li Y-F, Zhang H, Liao W-J (2015) Variation in morphological traits in a recent hybrid zone between closely related Quercus liaotungensis and Q. mongolica (Fagaceae). Journal of Plant Ecology 8:224–229. https://doi.org/10.1093/jpe/rtv023

Wesselingh RA, Hořčicová Š, Mirzaei K (2019) Fitness of reciprocal F1 hybrids between Rhinanthus minor and Rhinanthus major under controlled conditions and in the field. Journal of Evolutionary Biology 32:931–942. https://doi.org/10.1111/jeb.13492

Westbury DB (2004) *Rhinanthus minor* L. Journal of Ecology 92:906–927. https://doi.org/10.1111/j.0022-0477.2004.00929.x

Whitney KD, Ahern JR, Campbell LG, et al (2010) Patterns of hybridization in plants. Perspectives in Plant Ecology, Evolution and Systematics 12:175–182. https://doi.org/10.1016/j.ppees.2010.02.002

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag New York

Widmer A, Lexer C, Cozzolino S (2009) Evolution of reproductive isolation in plants. Heredity 102:31–38. https://doi.org/10.1038/hdy.2008.69

Williams JG, Kubelik AR, Livak KJ, et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535. https://doi.org/10.1093/nar/18.22.6531

Wilson P (1992) On inferring hybridity from morphological intermediacy. TAXON 41:11–23. https://doi.org/10.2307/1222481

Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. Nature 430:85–88. https://doi.org/10.1038/nature02698

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. Conservation Biology 15:1039–1053. https://doi.org/10.1046/j.1523-1739.2001.0150041039.x

Wood TE, Takebayashi N, Barker MS, et al (2009) The frequency of polyploid speciation in vascular plants. PNAS 106:13875–13879. https://doi.org/10.1073/pnas.0811575106

Wu C-S, Wang T-J, Wu C-W, et al (2017) Plastome evolution in the sole hemiparasitic genus Laurel Dodder (Cassytha) and insights into the plastid phylogenomics of Lauraceae. Genome Biol Evol 9:2604–2614. https://doi.org/10.1093/gbe/evx177

Yadav SK (2010) Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. South African Journal of Botany 76:167–179. https://doi.org/10.1016/j.sajb.2009.10.007

Ye J, McGinnis S, Madden TL (2006) BLAST: improvements for better sequence analysis. Nucleic Acids Research 34:W6–W9. https://doi.org/10.1093/nar/gkl164

Zhang M, Tang Y-W, Qi J, et al (2019) Effects of parental genetic divergence on gene expression patterns in interspecific hybrids of Camellia. BMC Genomics 20:828. https://doi.org/10.1186/s12864-019-6222-z

Zhang Z, Chen J, Li L, et al (2014) Research advances in animal distant hybridization. Sci China Life Sci 57:889–902. https://doi.org/10.1007/s11427-014-4707-1

Zhu F, Zhou Y-K, Ji Z-L, Chen X-R (2018) The plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. Front Plant Sci 9:. https://doi.org/10.3389/fpls.2018.00146

(2020) Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 29 Nov 2020