Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Marco Pellino

Born in: Napoli (Italy)

12 /01/1983

Oral-examination

Heidelberg 2015

Gene expression analysis and transcriptome evolution in apomicts: a case study in *Boechera* and *Ranunculus*

Referees: Prof. Dr. Timothy Sharbel

              Prof. Dr. Marcus Koch

Dedicated to all

DECLARATION

This dissertation is the result of my own account of research I carried out at the Apomixis Research Group of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany and contains as its main content work which has not been previously submitted for a degree in any educational institution.

Signed:_____

Date:_____

Marco Pellino

PUBLICATIONS AND PRESENTATIONS (DURING PHD STUDY)

Publications

Pellino, M., Sharbel, T.F., Mau, M., Amiteye, S., and Corral, J.M. (2011). Selection of reference genes for quantitative real-time PCR expression studies of microdissected reproductive tissues in apomictic and sexual *Boechera*. BMC Research Notes 4, 303

Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H., and Sharbel, T.F. (2013). Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. Molecular Ecology 22, 5908–5921.

Hojsgaard, D., Greilhuber, J., **Pellino, M.**, Paun, O., Sharbel, T.F., and Hörandl, E. (2014). Emergence of apospory and bypass of meiosis via apomixis after sexual hybridisation and polyploidisation. New Phytol. 204 1000–1012.

Pellino,M. Hojsgaard, D. Hörandl, E and Sharbel, T.F. (2015). Chasing the apomictic factor with gene expression analysis of live microdissected ovules in *Ranunculus.* (in review)

Oral and poster presentations

TALK: A selfish origin of recombination. – IPK

TALK: Analysis of gene expression pathways in apomictic and sexual individuals. – IPK

TALK: Gene expression comparisons of live-microdissected ovules from apomictic and sexual *Ranunculus.* – IPK

TALK: Gene expression analysis and evolution in *Ranunculus.* – Plant breeding and reproduction workshop, Ghent University, Belgium

TALK: Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. – RED BIO Conference 2013, Mar de Plata, Argentina

TALK: Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. – Göttingen University

TALK: Evolution and transcriptome analysis in the apomictic *Ranunculus auricomus* complex. – University of Bahia Blanca , Argentina

POSTER: Comparative transcriptomal profiling of live-microdissected ovules from sexual and apomictic *Ranunculus.* – IPK Institute day

POSTER: Asexual genome evolution in the apomictic *Ranunculus auricomus* complex : examine the effect of hybridization and mutation accumulation. – SMBE 2013 conference, Dublin, Ireland

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AI	Aposporous initial
AFLP	Amplified fragment length polimorphism
BLAST	Basic Local Alignment Tool
BLASTN	Basic Local Alignment Tool using nucleotide database
BLASTX	Basic Local Alignment Tool using protein database
cDNA	Complementary DNA
dN	Non-synonimous substitution
dS	Synonimous substitution
EBN	Endosperm Balance Number
FDR	False discovery rate
GO	Gene onthology
HKG	Housekeeping gene
InDel	Insertion/Deletion
Kb	Kilobases
MMC	Megaspore Mother Cell
mRNA	Micro RNA
Nt	Nucleotide
ORF	Open Reading Frame
PCA	Principle component analysis
PE	Paired end
PMC	Pollen mother cell
qRT-PCR	Quantitative real time PCR
RNAseq	RNA sequencing
rRNA	Ribosomial RNA
SE	Single end
SNP	Single nucleotide polymorphism
TAIR	The *Arabidopsis* Information Resources
TBALSTX	Basic Local Alignment Tool comparinges six frame translations of nucleotide sequence against to six frame translations of nucleotide database
TE	Transposable element

ABSTRACT

Apomixis refers specifically to asexual reproduction through seed in plants. Like other modes of asexual reproduction it has received much attention from evolutionary biologists and has been subject of many studies throughout the last decades. At the same time, it attracts interest from an economic point of view, as the long-term goal to technologically induce apomixis in major crop plants offers the prospect of a potential agricultural revolution.

Hence, interests have been growing in the scientific community in order to elucidate the evolution and underlying molecular genetic mechanisms of apomixis. Here I present a multifaceted approach to the problem by (1) the development of biotechnological tools in order to (2) apply molecular evolutionary methods to narrow down the possible causes and consequences of asexual reproduction in plants. In this work, representatives of two genera (*Ranunculus* L. and *Boechera* Á. Löve & D. Löve) were studied in order to advance current apomixis knowledge from different perspectives.

In the framework of a microarray based transcriptomic analysis of ovules extracted from sexual and apomictic *Boechera*, a list of housekeeing genes (HKGs) was selected based on a stability of expression analysis subsequently conducted in different vegetative and reproductive tissues of apomictic and sexual species. Using a GeNorm algorithm, different combinations of HKGs were identified, including Ribosomal Subunit 18 (*BoechRPS18*), Elongation Factor Alpha 1 (*BoechEf α 1*), Actine 2 (*BoechACT2*) and Ubiquitine (*BoechUBQ*), based on their pairwise stability in ovules, anthers, and vegetative tissue in apomictic and sexual species. These genes, specifically chosen to be reproduction mode- and tissue-specific, have subsequently been used for normalization in the experimental validation of two candidates genes related to apomixis in *Boechera*

Next, molecular evolutionary causes and consequences of apomixis were investigated by analyzing the transcriptomic effects of asexual reproduction and its correlated traits (*i.e.* hybridization, polyploidy and mutation accumulation). Flower-specific RNA from sexual and apomictic species of the wild apomictic *Ranunculus auricomus* complex was used for high throughput transcriptome sequencing (RNAseq). The first *de novo* assembled transcriptome for these species was used as a reference sequence for mining Single Nucleotide Polymorphism (SNP) and insertion-deletion

(indel) polymorphisms. The data were further used to design and manufacture a custom 3 x 1.4 million spot expression microarray. Comparative SNP analysis between apomictic and sexual individuals (specifically, two apomicts from two populations and three sexuals from three populations) corroborated the hybrid origin of apomictic *Ranunculus*, as proposed by Paun *et al.* (2006b), and could furthermore elucidate the Pleistocene origin and subsequent divergence of the apomictic individuals. In addition, sites of divergent selection were detected with the analysis of non-synonymous (dN) to synonymous (dS) substitution rates, strengthening the idea of rapid divergence in the hybrids.

Finally, the custom microarray was used for the hybridization of RNA from live-microdissected ovules (four developmental stages) from the three apomictic and four sexual individuals used in the SNP analysis. The comparative stage specific transcriptome analysis was used to detect stage specific differentially expressed genes in ovules, in order to identify signatures of apomixis and to produce a list of potential candidates underlying the reproductive switch. 555 stage specific genes were found to be differentially expressed throughout ovule development, and eight genes showed a significant shift in expression pattern throughout ovule development in apomicts compared to sexuals.

A further classification was conducted following the predictions made from Nogler's extensive work in *Ranunculus* in which different genetic factors were proposed for the induction and penetrance of apomixis. In that light, differentially expressed homoeologous genes were classified into three categories based on their relative expression in apomicts compared to their phylogenetic sexual parent, with the final aim of classifying the number of genes potentially responsible for apomixis. In doing so, we have provided a solid base for future studies in wild (i.e with little or no genetic information available) *Ranunculus* species.

By developing biotechnical tools for their study, identifying genes potentially involved in the establishment of apomixis, and analyzing their evolutionary history, this work presents an important step towards a more comprehensive understanding of the processes and patterns connected to apomixis in model and non-model plants, and has far-reaching potential for agricultural use.

ZUSAMMENFASSUNG

Der Begriff Apomixis bezeichnet die pflanzliche vegetative Vermehrung durch Samen und erweckt wie andere Arten der ungeschlechtlichen Vermehrung seit mehreren Jahrzenten großes Interesse bei Evolutionsbiologen. Gleichzeitig erntet Apomixis verstärkt Aufmerksamkeit von wirtschaftlicher Seite, da das langfristige Ziel der technologischen Einbringung von Apomixis in weit verbreitetete Nutzpflanzen das Potenzial einer modernen landwirtschaftlichen Revolution birgt.

Hiermit lege ich einen vielseitigen Ansatz vor, der (1) die Entwicklung biotechnologischer Werkzeuge zur (2) Anwendung molekulargenetischer Methoden zur Eingrenzung der möglichen Ursachen und Folgen von asexueller Vermehrung in Pflanzen beinhaltet. Um den derzeitigen Kenntnisstand von Apomixis aus unterschiedlichen Perspektiven zu erweitern, wurden in dieser Arbeit zwei Gattungen (*Ranunculus* L. und *Boechera* Á. Löve & D. Löve) untersucht.

Vor dem Hintergrund einer Microarray-basierten Transkriptomanalyse von Samenanlagen sexueller und apomiktischer Vertreter der Gattung *Boechera* und einer anschließenden Analyse der Expressionsstabilität in verschiedenen vegetativen und reproduktiven Geweben wurde eine Liste von Haushaltsgenen (HKGs), die hohe Expressionsstabilität in verschiedenen generativen und reproduktiven Geweben aufweisen, erstellt. Unter Verwendung des GeNorm Algorithmus wurden basierend auf einem paarweise angelegten Vergleich ihrer Stabilität in Samenanlagen, Antheren und vegetativen Geweben verschiedene Kombinationen von HKGs identifiziert, zu denen z.B. Elongation Factor Alpha 1 (*BoechEfα1*), Actine 2 (*BoechACT2*) und Ubiquitine (*BoechUBQ)* gehörten. Diese Gene, die eigens ihrer Reproduktionsmodus- und Gewebespezifität ausgewählt wurden, sind seither von anderen Autoren für die Normalisierung der experimentellen Validierung zweier Kandidatengene bezüglich der Ausprägung von Apomixis in der Gattung *Boechera* eingesetzt worden.

Im Weiteren wurden mittels einer Analyse der Transkriptomeffekte von asexueller Fortpflanzung sowie der damit korrelierten Eigenschaften (Hybridisierung, Polyploidie und Mutations-Akkumulation) evolutive Ursachen und Folgen von Apomixis untersucht. Hochdurchsatz-Transkriptomsequenzierung (RNAseq) wurde für blütenspezifische RNA von generativen und apomiktischen Blüten des natürlich vorkommenden *Ranunculus auricomus* Komplexes durchgeführt.

Das davon ausgehend erstellte, erste *de novo* assemblierte Transkriptom dieser Arten wurde als Referenz für die Identifizierung von Einzelnukleoid-Polymorphismen (SNPs) und Insertion-Deletion-Polymorphismen (Indels) genutzt. Im Folgenden dienten die Transkriptomdaten zur Erstellung eines maßgefertigten 3 x 1.4 Millionen Testfeld Expressions-Microarrays. Eine vergleichende SNP Analyse von drei generativen und zwei apomiktischen Individuen aus jeweils unterschiedlichen Populationen bekräftigte die Annahme, dass die apomiktischen *Ranunculus* Stammlinien aus einer ursprünglichen Hybridisierung hervorgegangen sind und deutete auf einen Pleistozänen Ursprung mit nachfolgender Divergenz der apomiktischen Individuen hin. Zusätzlich wurde durch die Analyse des Verhältnisses der nicht-synonymen (dN) zu synonymen (dS) Substitutionsraten divergierende Selektion festgestellt, wodurch die Idee einer rapiden Divergenz der Hybridarten bestärkt wird.

Schließlich wurde der erstellte Microarray für die Hybridisierung von RNA lebend-mikrodissektierter Samenanlagen vier verschiedener Entwicklungsstadien der generativen und apomiktischen Individuen aus der SNP Analyse genutzt. Die nachfolgende, vergleichende stadiumsspezifische Transkriptomanalyse diente der Identifikation stadiumspezifisch differentiell exprimierter Gene in den Samenanlagen und damit dem Nachweis von Spuren von Apomixis sowie der Erstellung einer Liste potenzieller Genkandidaten, die in den Wechsel des Reproduktionsmodus involviert sind. Insgesamt zeigten 555 entwicklungsstadium-spezifische Gene differenzielle Expression während der Entwicklung der Samenanlagen und bei acht dieser Gene konnte während der Entwicklung eine signifikante Verlagerung der Expressionsmuster von apomiktischen verglichen mit sexuellen Individuen festgestellt werden.

Zusätzlich wurde eine Klassifizierung in Bezug auf Noglers umfangreicher Arbeit über *Ranunculus* (Nogler, 1984b), die verschiedene genetische Faktoren bezüglich der Induktion und Manifestation von Apomixis enthält, durchgeführt. Hierzu wurden differentiell exprimierte Gene in drei Kategorien gemäß ihrer Expression im apomiktischen Individuum verglichen zu den vegetativen Eltern eingeteilt, um jene Gene, die potenziell mit Apomixis in Verbindung stehen, näher zu klassifizieren. Dadurch wurde eine solide Basis für zukünftige Studien von wilden *Ranunculus* Arten, für die spärliche oder gar keine genetische Informationen verfügbar sind, geschaffen.

Durch die Entwicklung biotechnologischer Werkzeuge, die Identifikation von Genen, die potenziell an der Ausprägung von Apomixis beteiligt sind, sowie der Analyse ihrer Evolutionsgeschichte stellt diese Arbeit einen wichtigen Schritt auf dem Weg zu einem umfassenderen Verständnis der Prozesse und Muster in Zusammenhang von Apomixis in Model- und Nichtmodel-Arten dar und hat weitreichendes Potenzial für landwirtschaftliche Nutzung.

Table of Contents

CHAPTER 1: LITERATURE REVIEW

1.1. Apomixis and sexual reproduction

1.1.1. Sexual reproduction: Double fertilization and embryogenesis

If one had a closer look at the flower of a sexually reproducing plant, zooming in deeper into its female reproductive organs (gynoecia), one would see at its base the functional ovary. It is inside the ovary that the female haploid gametophyte (embryo sac), enclosed within the ovule, develops and is eventually fertilized. In most (70%) angiosperms, including *Arabidopsis* and maize (Mansfield & Briarty, 1991; Russell, 1979), the fully developed embryo sac consists of seven cells whose structure, organization and development is known as the *Polygonum* type. (Reiser & Fischer, 1993).

The first step of the normal female sexual developmental process (megasporogenesis) occurs in the ovule where a premeiotic cell (the megaspore mother cell or megasporocyte) begins to develop within the megasporangium. The diploid megasporocyte enlarges, and undergoes two meiotic divisions, developing into four haploid megaspores aligned along the micropyle-chalazal axis. The megaspore closest to the chalaza typically receives most of the organelles, survives and starts to enlarge while the three more distant ones degenerate and eventually disappear. At this point megasporogenesis is concluded, and with the beginning of the second phase (megagametogenesis) the surviving functional megaspore undergoes three rounds of mitosis to produce the eight-nucleate embryo sac. During the second mitotic division the four nuclei separate into couples and migrate to the opposite points of micropylar-chalazal axis. Of the pair at the micropylar end, one divides to produce two synergid cells, while the other produces the egg cell as well as one of the two cells of the binucleate central cell. At the opposite end, the remaining two migrated nuclei form, after division, the three antipodals and the second binucleate central cell. At this point the embryo sac is fully developed and ready to be fertilized.

On the male side, sperm cells are found in pollen that are produced in the anthers, the male reproductive organs. Pollen formation can be divided in two main parts. In the first step the gametophyte initial cell or pollen male cell (PMC) undergoes one round of DNA replication followed by one meiotic and one mitotic division. In contrast to female development, the four (meiotically reduced) products of this step

form male gametophytes, consisting of two large vegetative and two smaller generative cells. During the second step the four haploid cells separate from each other to follow different developmental destinies: the small generative cells undergo mitosis to form two reproductive sperm cells (Lord & Russell, 2002), while the two larger vegetative cells maintain their unreduced state and eventually form the pollen tube that will guide the two sperm cells into the embryo sac.

Pollination of the flower is followed by the formation of a pollen tube which is guided through the pistil into the embryo sac through the micropilar entrance by chemical attractants produced by the synergid cells (Higashiyama *et al.*, 2001). Prior to the arrival of the pollen tube, one of the two synergid cells degenerates while the two haploid sperm cells fertilize (1) the haploid egg cell and (2) the binucleate central cell (Lord & Russell, 2002; Russell, 1979), a process referred to as double fertilization. The products of this event are the diploid zygote and triploid functional endosperm tissue.

Any alteration from this normal sexual development usually results in abortion and degeneration of the embryo sac (Koltunow & Grossniklaus, 2003a).

1.1.2. Apomictic reproduction: asexual reproduction through seeds

The term "apomixis" was used for the first time in 1908 by Winkler in order to define a wide range of asexual reproduction, from viviparism to vegetative propagation (Winkler, 1908a). Today, apomixis is used *sensu stricto* to refer to asexual reproduction through seeds.

It is well accepted that sexuality is the original state of plant reproduction (Karron *et al.*, 2012). Nevertheless, ca. 400 different species distributed over 40 families have been reported to be able to reproduce apomictically (Carman, 1997; Hojsgaard *et al.*, 2014a; Ozias-Akins, 2006). In most of these examples asexuality and sexuality co-exist within the same family, genus, population and even individual. The close relationship between sexual and apomictic individuals, and the diversity of plants that show the ability of asexual seed production suggest that apomixis has developed repeatedly and independently from sexual ancestors in different lineages (Carman, 1997a; van Dijk & Vijverberg, 2005a; Hojsgaard *et al.*, 2014a):

Besides rare exceptions, there are three main components that deviate from the normal sexual pathway which characterize apomictic development (Nogler, 1984a): (1) apomeiosis - circumvention of the two mitotic divisions by the megasporocyte to

produce a fully developed and unreduced megaspore, (2) parthenogenesis - the development of the unreduced megaspore into a functional embryo sac without fertilization by the male gamete, and (3) autonomous or fertilization dependent (*i.e.* pseudogamous) formation of the endosperm (see Figure 1.1).

It is important to note that all three factors must occur together in order to produce functional maternal offspring. For example, fertilization of an unreduced egg cell (*i.e.* apomeiosis and no parthenogenesis) results in an aberrant 2n + n hybrid characterized by a ploidy increase. Alternatively, parthenogenesis of the reduced egg cell (*i.e.* haploid parthenogenesis) results in a n + 0 (haploid) apomictic embryo which does not perpetuate (non recurrence). Finally, in most cases fertilization of the central cell is essential for maintaining the maternal-paternal genome ratio of the endosperm (Endosperm Balance Number – EBN (Haig & Westoby, 1989a)). According to the hypothesis of parent specific gene expression in the endosperm, this ratio plays a major role in the correct development of the seed. A conflict between mother, father and progeny for resource allocation in the endosperm has been driven by selection for an optimal ratio of two maternal (m) and one paternal (p) genome contribution. Supporting evidence in maize (Haig & Westoby, 1989a; Lin, 1984a) has shown how deviation from the optimal 2m:1p balance results in abnormal (*i.e.* smaller kernel) or absent development of the endosperm. Nevertheless, some apomicts (mostly Asteraceae) do not require fertilization of the central cell (autonomous endosperm formation; (Hojsgaard *et al.*, 2014a))

Even though the result (*i.e.* genetically identical offspring to the mother plant) of each of three above mentioned components are conserved through different plant groups, different types of apomictic reproduction exist in the wild, and are classified based on the origin of the embryo and the role of the male gamete.

In "adventitious embryony" (Bicknell & Koltunow, 2004; Koltunow, 1993) the embryo (i.e the sporophyte) arises and develops from a somatic cell of the ovule, *de facto* bypassing the gamethopyte generation. Adventitious embryony is found in *Citrus* (Frost, 1968), *Opuntia* (Asker & Jerling, 1992), *Nigritella* (Teppner, 1996), *Tulipa* and *Lilium* (Marasek *et al.*, 2004).

The second mode, "gametophytic apomixis" (review in (Nogler, 1984a)), occurs when the embryo sac arises from an unreduced embryo sac initial. Gametophytic

apomixis can be further subdivided based on the origin of the initial (the cell that will give rise to the apomictic embryo):

a) Diplospory -where the initial is a generative cell that undergoes mitosis (*Antennaria*-type), as in *Poa alpina* and *Eragrostis*, or undergoes a modified meiosis (*Taraxacum*-type), as in *Boechera* (Nogler, 1984a);

b) Apospory -where the initial is a somatic cell, usually from the nucellus of the ovule (*Hieracium*-Type), as in *Ranunculus*, *Poa pratensis* and *Hypericum* (Nogler, 1984b)

The final classification is related to the fertilization of the binucleate central cell by the male pollen. When full and functional development of the endosperm requires the polar nuclei to be fertilized it is referred to as "pseudogamy". Alternatively, in "autonomous apomixis" both egg cell and endosperm develop parthenogenetically (*i.e.* without fertilization).

4
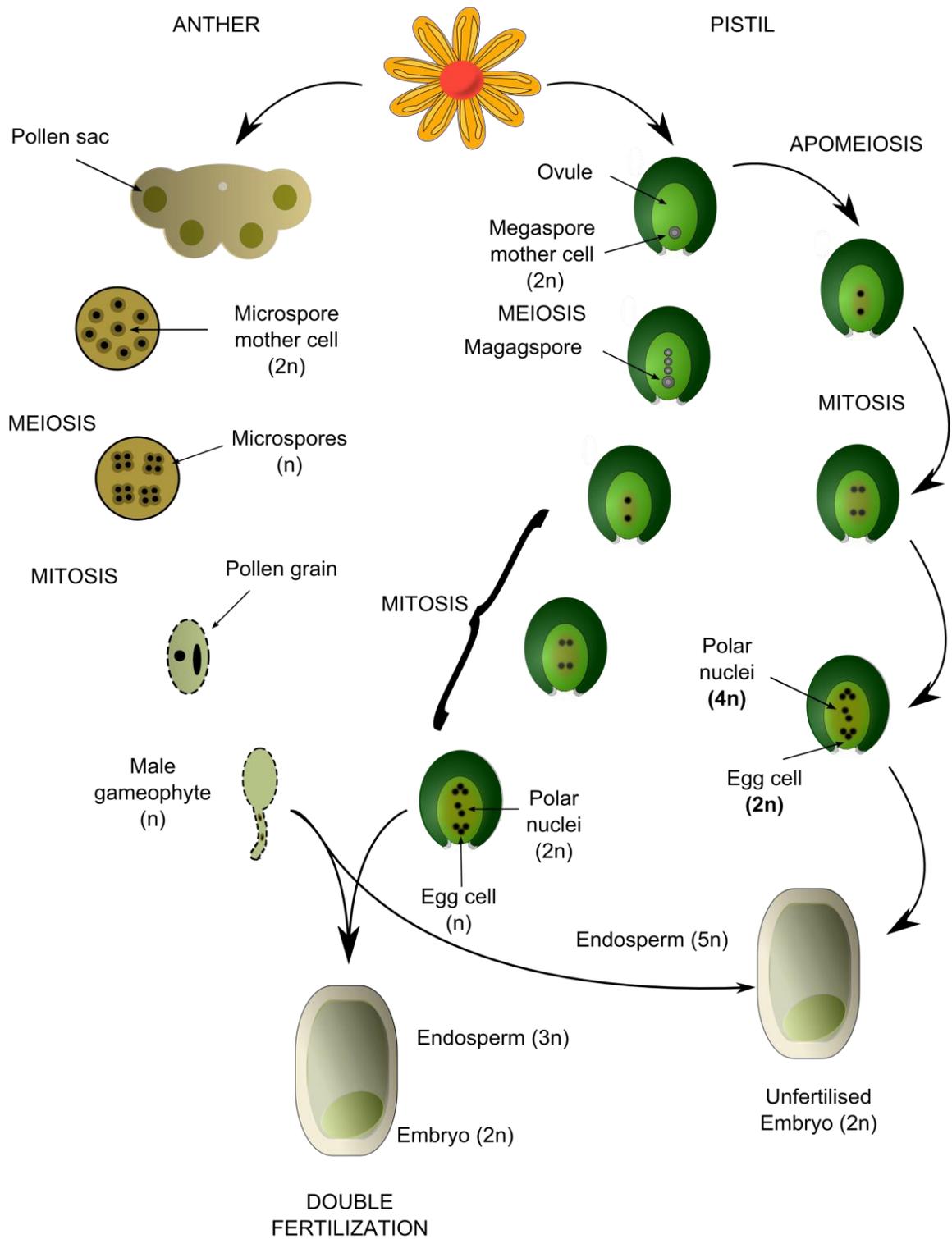
Figure 1.1: Schematic representation of sexual and apomeiotic ovule development

1.2. The genetic mechanisms of apomixis

1.2.1. Hypotheses

Throughout the last 30 years apomixis has been repeatedly demonstrated to be under genetic control (Barcaccia *et al.*, 1998; Corral *et al.*, 2013; Grimanelli *et al.*, 1998; Morell, 1999; Nogler, 1984a; Noyes & Rieseberg, 2000; Ozias-Akins *et al.*, 1998; Pessino *et al.*, 1998; Savidan, 1982), but characterization of the genetic mechanisms underlying its regulation is still at a very young stage. Since all three main components (see above) are required for stable apomixis (see 2.1.2), the possibility that they evolved independently as a coordinated polygene regulation event is very unlikely (Mogie, 1992). Additionally, the observation in *Panicum*, *Ranunculus* and *Hieracium,* of single Mendelian traits responsible for the inheritance of apomixis (Koltunow, 2000; Nogler 1984; Savidan, 1982) lead to the hypothesis that apomixis is controlled by a single master regulatory gene, or a tandem group of linked genes (Grossniklaus *et al.*, 2001). Conversely, for *Taraxacum*, *Erigeron*, *Hieracium*, and *Boechera* a model involving 2-3 co-segregating genes controlling each of the three main components of apomixis has been proposed as result of segregation analyses (van Dijk *et al.*, 1999; Noyes & Rieseberg, 2000; Schranz *et al.*, 2005; Kotani *et al.*, 2014). While this model challenges the single factor theory, co-adapted genes could be tightly linked in a non-recombining region, while the single gene discovered in *Panicum*, *Ranunculus* and *Hieracium* could have a master regulatory function controlling a cascade of other elements regulating asexual reproduction (Grossniklaus *et al.*, 2001). In addition to the view of genetic factor(s) regulating apomixis, another theory (the hybridization-derived floral asyncrony - HFA) proposes that apomixis arises from the sexual developmental pathway through deregulation of duplicated gene sets necessary for normal female gamete development, with asynchronous changes in gene expression to be a consequence of polyploidization and/or hybridization (Carman, 1997). It is supported by work in *Tripsacum* spp. (Grimanelli *et al.*, 2003) and *Boechera* spp. (Sharbel *et al.*, 2010) where, heterochronic megaspore formation and heterochronic gene expression patterns in developing ovules have been observed.

1.2.2. The genetic factors

Different experimental techniques have been used to investigate the genetic factor(s) regulating apomixis. The application of mapping strategies to naturally

occurring apomictic species has led to the discovery of a range of molecular markers linked to apospory, with resultant support for both the hypotheses of a single master regulatory gene and multi locus control of the trait. The finding of a recessive lethal A⁻ allele in *Ranunculus auricomus* (Nogler 1984) has been followed by similar discoveries in *Paspalum notatum* (Martínez *et al.*, 2001), *Panicum maximum* (Ebina *et al.*, 2005), *Taraxacum officinalis* (Van Dijk & Bakx-Schotman, 2004) and *Hypericum perforatum* (Schallau *et al.*, 2010), while findings in *Erigon annuus* support control by two loci (Noyes & Rieseberg, 2000). The scenario in *Poa pratensis* is even more complex as a combination of five different loci was hypothesized to induce stable asexual seed formation (Matzk *et al.*, 2005). Although several factors been discovered, the final task of pinpointing genes responsible for apomixis using mapping strategies is made arduous by the fact that in most of the analyzed species, these factors are located in regions of the genome where recombination is suppressed.

Recently, developing experimental techniques for gene expression analysis have permitted the comparison of sexual and asexual species at the transcriptomic level. Using different molecular approaches, several candidate genes have been discovered to be differentially expressed between apomicitc and sexual accessions (Albertini *et al.*, 2004; Pessino *et al.*, 1998; Polegri *et al.*, 2010; Rodrigues *et al.*, 2003; Singh *et al.*, 2007; Vielle-Calzada *et al.*, 1996). In parallel, implementation of micromanipulation tools like laser-assisted microdissection or live tissue microdissection have been used to access the female and male gametophyte of the plant, allowing tissue specific transcriptomic profiling (Corral *et al.*, 2013; Day *et al.*, 2005; Mau *et al.*, 2013; Schmid *et al.*, 2012; Schmidt *et al.*, 2014)).

1.3. Evolution of apomixis in natural populations

1.3.1. Hybridization and polyploidization may induce apomixis

Elaborating his hypothesis in 1997, Carman suggested that the asynchronous expression of genes involved in female gamete development could explain apomixis induction, and proposed hybridization and/or polyploidization as triggers since both phenomena have genome-wide effects. This idea is backed by the fact that, beside few exceptions (Kantama *et al.*, 2007), almost all apomictic plants are polyploid (Koltunow & Grossniklaus, 2003) and show elevated heterozygosity compared to their sexual relatives (Gornall, 1999; but see Lovell *et al.*, 2013).

In fact, polyploidization leads to massive and wide spread genomic changes, both in chromosome structure and on the level of gene regulatory networks. The fate of duplicate genes can vary: genes may be silenced or undergo neofunctionalization or subfunctionalization, retaining only a partial function compared to the ancestral gene or being expressed only in particular tissues (Adams *et al.*, 2004; Lynch *et al.*, 2001). Cases of duplicated gene loss following polyploidization have been reported in maize and cereals (Ilic *et al.*, 2003), and studies in *Arabidopsis thaliana* have revealed that the loss of duplicated genes does not occur randomly (Blanc & Wolfe, 2004). Another effect can be regulatory modifications in gene expression. Dosage effects influenced by copy number variation (CNV) and transcription factors have been reported in maize (Guo *et al.*, 1996). Also, changes in regulatory networks affecting normal meiotic and mitotic processes have been proposed (Comai, 2005)

In addition to polyploidization, hybridization is also known to induce a plethora of changes in the hybrid compared to its parental taxa, including chromosome structural rearrangements (Metcalfe *et al.*, 2007; Rieseberg *et al.*, 2003), gene expression changes (Albertin *et al.*, 2006; Hegarty & Hiscock, 2005), silencing (Adams & Wendel, 2005), transposon activation (Liu & Wendel, 2000) and unequal expression of duplicated genes (Adams, 2007).

Modifications associated with polyploidization and/or hybridization could therefore theoretically produce a shift or deregulation in the expression profile of genes involved in the necessary steps of sexual development. Alternatively, the hybrid/polyploid plant could experience asynchronous expression of meiotic and gametogenesis specific genes, resulting in premature gametogenesis without meiosis (Grimanelli *et al.*, 2001)

1.3.2. Hybridization and polyploidization as indirect consequences of apomixis

Even though hybridization and polyploidization are common features of almost all apomictic lineages, their evolutionary role is still disputed. On the one hand polyploidization may be the trigger that induces apomixis, but in some instances it may also represent a secondarily derived factor required to stabilize apomixis. In *Ranunculus auricomus*, for example, the recessive lethal factor A$^-$, which is associated with apomixis induction, cannot be inherited by monoploid gametes (Nogler, 1984b) due to its lethality in the hemizygous state. Hence polyploid gametes which are

8

heterozygous at this locus (*i.e.* containing both the recessive and wild-type allele) are hypothetically selected for in order to enable transmission of the factor.

Along these lines, the observed higher heterozygosity in apomicts (Hojsgaard *et al.*, 2014a) could be the result of different processes: 1) independent allelic divergence and mutation accumulation (Meselson effect; (Mark Welch & Meselson, 2000)) and 2) mutation accumulation *per se* (Muller ratchet; (Muller, 1964)), both phenomena that may occur when meiosis and recombination are suppressed (Birky, 1996), or 3) interspecific hybridizazion (Paun *et al.*, 2006b).

Regardless of the question concerning the role of hybridization and polyploidization with respect to the origin of apomixis, both traits have strong evolutionary consequences. Hybridization between distantly-related sexual parents results in phenotypically-variable offspring with different fitness and adaptation abilities (Barton, 2001). In some cases the newly formed hybrid can show extreme characteristics compared to its parental taxa, an effect known as transgressive segregation, heterosis or hybrid vigor (Lippman & Zamir, 2007). The increased variation and/or higher vigor can allow the newly formed hybrid to better adapt to particular environments or to colonize new habitats (Lewontin, 1966).

The gene redundancy and increased number of potential mutation sites which characterize polyploidy genomes can have positive consequences on the organism, for example by buffering the effects of deleterious mutations, because the extra copy of a deleteriously mutated gene will still be functional (Comai, 2005). Polyploidy furthermore provides the potential for duplicate genes to acquire a new function (neofuntionalization) (Comai, 2005). Other advantages of polyploidy are connected with hybridization and the occurrence of positive mutations. Heterosis can be enforced by polyploidy because the latter can fix parental genomes in the allopolyploid state, with heterozygosity levels being prevented by preventing intergenomic recombination (Gerstein & Otto, 2009), whie positive mutations are expected to have a stronger effect on polyploid populations when the population is small (Otto & Whitton, 2000)

In spite of the potential advantages of polyploidization and hybridization in creating variability and improved adaptability to the environment, there are a number of problems that must be overcome for a hybrid and/or polyploid to establish. Reproductive isolation of the newly formed hybrid must be acquired to prevent the new gene combinations from being lost by gene flow from the parent populations

(Mayr, 1963). Moreover, newly formed hybrids require finding a mating partner to insure hybrid speciation (Smith, 1993). Establishment of a newly-formed hybrid (hybrid speciation) can be attained via chromosome duplication (polyploidy) in the same (autopolyploidy) or between species (allopolyploidy). Polyploidization not only allows isolation from parental taxa of lower ploidy, but in connection with acquired or already-present selfing it makes hybrid/polyploid reproduction possible (Hörandl, 2010). For example, polyploidization allowed the spontaneous generation of hybrids in cultivated *Primula* (Ramsey & Schemske, 2002). A second route of polyploidization is via the triploid bridge, where a triploid occasionally produces an unreduced gamete that can contribute to the formation of stable tetraploid progeny (Yamauchi *et al.*, 2004).

Nevertheless, after both polyploids and hybrids arise, the challenge of epigenetic instabilities must be faced, even though it is not yet clear whether these may play a role in the switch to asexuality (Comai, 2005). Studies on the changes in methylation patterns in plant cytosine have suggested that epigenetic changes are a consequence of environmental or genomic challenges, and enable an individual to regulate physiological responses (Liu & Wendel, 2003). In this sense hybridization and polyploidization can be seen as challenges (genomic shock) that cause epigenetic instabilities. This hypothesis has been confirmed by several experiments where neoallopolyploids indeed showed heritable repatterned DNA methylation (Adams & Wendel, 2005). Increased phenotypic variability connected with poor fertility has also been documented in neoallopolypoids of *Arabidopsis* (Comai *et al.*, 2000; Madlung *et al.*, 2005). In addition to epigenetic changes, meiotic disturbances also represent a problem to overcome. In autopolyploids there is an elevated chance of multivalent formation (*i.e.* three or more chromosomes synapsed during prophase), while in allopolyploidy incorrect pairing between non-homologous chromosomes is possible (Comai, 2005). To solve this problem, autopolyploids must efficiently resolve multivalent chromosome formation, or enforce genetic mechanisms for bivalent pairing. In allopolyploidy, correct pairing must be ensured by genetic mechanisms. Evidence suggests that adaptation for these mechanisms is a requirement for polyploidy stabilization (Comai, 2005).

Considering the challenges facing a newly-formed hybrid and/or polyploid, apomixis could be seen as solution which enables the maintenance of the evolutionary

advantages of each trait (Kearney, 2005). For example, apomixis by definition causes immediate reproductive isolation, and in some instances polyploidization is associated with the loss of self-incompatibility (Comai, 2005). The apomeiotic megaspore development typical of apomictic reproduction can also serve as a path to bypass the meiotic disturbances that may occur during hybrid and polyploid sexual reproduction (Hojsgaard *et al.*, 2014b). Finally, regardless of contrasting hypotheses regarding epigenetic changes, apomixis could hypothetically overcome the disruptive effects that such changes produce on regulatory patterns (Comai, 2005).

1.3.3. Origin of apomixis

Deciphering the origin and age of apomixis in plants is difficult, and so far has been subject to a great deal of speculation. The most intriguing hypothesis connects glaciations with the rise of apomixis/parthenogenesis, as most apomictic plants exist at higher latitudes and are generally found in places strongly affected by the glacial periods of the Late Pleistocene. Recurrent periods of glaciation occurred for the first time 2.5 – 0.9 million of years ago and have since then recurred with a frequency of approximately 100 000 years (Kearney, 2005). During glacial periods species distributions would contract toward lower latitude and glacial refugia, leaving behind extensive harsh and arid areas (Kearney, 2005). Post glacial recolonization towards milder environments would lead to events of recurrent hybridization that may have given rise to apomictic lineages (Hörandl, 2006), while the distribution of sexual ancestors would reflect that of glacial refugia (Kearney, 2005).

These speculations have recently been supported by molecular dating techniques in insects (Law & Crespi, 2002), mollusks (Johnson & Bragg, 1999) and reptiles (Strasburg & Kearney, 2005).

Apomictic plants are thought to be evolutionarily recent and are found scattered along the tips of phylogenies, however empirical age calculation is problematic (Hörandl & Hojsgaard, 2012). In apomictic plants the calculation of estimated origin is biased by the potential reversal of reproduction and lack of fossil records in which the reproductive modus can be inferred (Schwander & Crespi, 2009).

The advantage for a clonally reproducing species in colonizing niches and relatively empty environments is summarized in the "general purpose genotype" theory (Lynch, 1984), whereby selection favors a single clone that has "captured" the adapted genotype for that particular environment. This is especially true when an

asexual lineage starts with a higher fitness compared to its sexual relatives (*e.g.* due to hybridization higher heterozygosity – see above).

An additional competitive advantage compared to sexual species during colonization of new habitats is the intrinsic short-term cost of sexual reproduction, namely the two-fold cost of sex (Kondrashov, 1993). Briefly, this cost arises because sexuals must invest their energy in producing males and females, while asexuals can allocate 100% of their resources in the production of females, since asexuality is for the most part by definition a female trait. Several conditions can reduce this advantage, including self-fertilizing hermaphroditic sexuals, parental care, and isogamy (Martens, 1998)

Even though apomictic populations have the theoretical ability to occupy diverse niches (Tangled-Bank theory (Bell, 1982; Ghiselin, 1974)), evolutionary theory also predicts that on the long term the disadvantage of asexuality should increase to such a degree that asexuals will eventually go extinct (but see the rare exceptions of "asexual scandals" I.e Bdelloid rotifers (Judson & Normark, 1996; Mark Welch & Meselson, 2000) – but see (Birky, 2004)). Among these a significant disadvantage is the inability for selection to efficiently eliminate deleterious mutations in asexual organisms, an idea first suggested by Muller (*i.e.* Muller's ratchet; Muller, 1964). Lack of recombination in an asexual population with a mutation distribution having a minimum number of deleterious mutations (n), in conjunction with a low probability of back-mutation, means that an individual with n-1 deleterious mutations is unlikely to arise. Genetic drift leads to the loss of the n mutational class, leading to a shift in mutation distribution towards n + 1 (*i.e.* a "click" in the ratchet; Muller, 1964). Assuming that most mutations are deleterious or slightly deleterious (Smith, 1978), this ratchet-like mechanism will lead to a reduction in fitness and the eventual extinction of the clonal lineage (Lynch *et al.*, 1993). In contrast for sexuals, mutation occurring independently in different genotypes can be brought together, which when lethal, lead to the death of that individual and the loss of those mutations (Kondrashov, 1982).

A second disadvantage exists for asexual populations considering rapidly changing environments or coevolution, for example with pathogens. Because of the lack of recombination, asexuals are expected to have a slower rate of evolution compared to sexuals (Martens, 1998). For example, if A and B are two distinct favorable alleles with

12

an additive effect, the only way an asexual AB individual could arise is if one mutation occurred when the previous one is already present. Alternatively in a sexual population, the two mutations could occur independently in different individuals and then, by recombination, segregate into the same genotype. Consequently, response to changes in the direction of selection will be more efficient in sexual populations, allowing them to adapt more efficiently to changing environments (Fisher-Müller Accelerated Evolution Theory (Fisher, 1930; Muller, 1932)) and to better respond to arms races with pathogens (Red Queen theory) (Van Valen, 1973)

The common geographical distribution of parthenogenetic taxa, mostly located in temperate climatic areas (Hojsgaard *et al.*, 2014a), and at the same time, areas with a lower incidence of parastisms are supportive evidence of these mechanisms (Martens, 1998).

1.4. Apomixis in crop plants

Breeding programs in crop plants mostly target the fixation of traits of interests, and heterosis through laborious and expensive F1 breeding schemes. However, breeding companies and farmers alike face the problem that the obtained F1 generation (exhibiting the desired phenotype), if crossed, would eventually lose that trait because of allelic segregation (Spillane *et al.*, 2004). Hence farmers must purchase new seeds every farming season, rather than crossing their own plants. The induction of apomixis into crops could potentially fix heterozygosity – and therefore hybrid vigor – and associated traits of interest for an endless number of generations to allow farmers to perpetuate their own cultivars (Spillane *et al.*, 2004). For breeders, apomixis could be used to stabilize heterozygous crop plants to reduce the costs and time of breeding programs (estimation of the benefit for rice alone has been calculated to be $2.5 billion per year; McMeniman & Lubulwa, 1997). Single generation hybrids could hypothetically be produced, opening the door to niche breeding for hybrids adapted to a particular environment (Spillane *et al.*, 2004). In addition, vegetatively-propagated crop plants could be stored at their seed stage, decreasing the negative effects of pathogen transmission (*e.g.* being able to clonally propagate potato and cassava would result in an estimated saving of $3.3 billion per year; Savidan *et al.*, 2001).

With the exception of *Tripsacum* (Bicknell & Koltunow, 2004) and *Boechera* (Beilstein *et al.*, 2006) there are no natural apomicts related to major crop species.

Attempts to implement apomixis technology into sexual crops have been conducted by back crossing apomixis components from wild relatives into sexual crop plants using maize, pearl millet (Savidan *et al.*, 2001) and wheat (Liu *et al.*, 1994). These approaches have encountered difficulties connected to male sterility and progenies with low agronomic potential. On the other hand, approaches based on mutagenesis have been able to produce a proportion of clonal seeds (< 50%) via crosses between mutants at loci controlling meiosis and chromosome segregation in maize (Marimuthu *et al.*, 2011).

CHAPTER 2: SELECTION OF REFERENCE GENES FOR QUANTITATIVE REAL-TIME PCR EXPRESSION STUDIES OF MICRODISSECTED REPRODUCTIVE TISSUES IN APOMICTIC AND SEXUAL *BOECHERA*

2.1. Summary

Apomixis, a natural form of asexual seed production in plants, is considered to have great biotechnological potential for agriculture. It has been hypothesised that de-regulation of the sexual developmental pathway could trigger apomictic reproduction. The genus *Boechera* represents an interesting model system for understanding apomixis, having both sexual and apomictic genotypes at the diploid level. Quantitative qRT-PCR is the most extensively used method for validating genome-wide gene expression analyses, but in order to obtain reliable results, suitable reference genes are necessary. In this work we have evaluated six potential reference genes isolated from a 454 (FLX) derived cDNA library of *Boechera*. RNA from live-microdissected ovules and anthers at different developmental stages, as well as vegetative tissues of apomictic and sexual *Boechera*, were used to validate the candidates. Based on homologies with *Arabidopsis*, six genes were selected from a 454 cDNA library of *Boechera*: RPS18 (Ribosomal sub protein 18), Efalpha1 (Elongation factor 1 alpha), ACT 2 (Actin2), UBQ (polyubiquitin), PEX4 (Peroxisomal ubiquitin conjugating enzyme) and At1g09770.1 (*Arabidopsis thaliana* cell division cycle 5). Total RNA was extracted from 17 different tissues, qRT-PCRs were performed, and raw Ct values were analyzed for primer efficiencies and gene ratios. The geNorm and normFinder applications were used for selecting the most stable genes among all tissues and specific tissue groups (ovule, anthers and vegetative tissues) in both apomictic and sexual plants separately. Our results show that BoechRPS18, BoechEf α 1, BoechACT2 and BoechUBQ were the most stable genes. Based on geNorm, the combinations of BoechRPS18 and BoechEf α 1 or BoechUBQ and BoechEf α 1 were the most stable in the apomictic plant, while BoechRPS18 and BoechACT2 or BoechUBQ and BoechACT2 performed best in the sexual plant. When subgroups of tissue samples were analyzed, different optimal combinations were identified in sexual ovules (BoechUBQ and BoechEf α 1), in anthers from both reproductive systems (BoechACT2 and BoechEf α 1), in apomictic vegetative

tissues (BoechEfα1 and BoechACT2) and sexual vegetative tissues (BoechRPS18 and BoechEfα1). NormFinder ranked BoechACT2 as the most stable in the apomictic plant, while BoechRPS18 was the best in the sexual plant. The subgroups analysis identified the best gene for both apomictic and sexual ovules (BoechRPS18), for anthers from both reproductive system (BoechEfα1) and for apomictic and vegetative tissues (BoechACT2 and BoechRPS18 respectively). From a total of six tested genes, BoechRPS18, BoechEfα1, BoechACT2 and BoechUBQ showed the best stability values. We furthermore provide detailed information for the accurate normalization of specific tissue gene expression analyses of apomictic and sexual *Boechera*.

2.2. Introduction

Sexual reproduction in plants is a highly regulated process in which meiosis and syngamy initiate embryo and seed development. Aberrations in any step typically lead to abortion of seed development (Yang *et al.*, 2005). In contrast, apomixis (asexual reproduction through seeds (Asker & Jerling, 1992)) is an alternative reproductive strategy in which aberrations to normal sexual processes are viable (Aliyu *et al.*, 2010), and is found naturally in more than 400 species. Compared to sexual reproduction, apomixis is characterized by three developmental steps: the production of a meiotically unreduced egg cell (apomeiosis), parthenogenetic development of this egg cell without fertilization, and production of a functional endosperm with (pseudogamy) or without (autonomous) fertilization of the binucleate central cell of the ovule (Grimanelli *et al.*, 2001). Importantly, apomictic seeds have embryos which are genetically identical to the mother plant. Hence, the successful introgression of apomixis into crop plants would greatly facilitate the fixation and propagation of genetic heterozygosity and associated hybrid vigour over successive generations, and could significantly reduce costs associated with hybrid seed production (Spillane *et al.*, 2004). The biotechnological potential of apomixis has thus raised tremendous research interest.

Apomixis has repeatedly evolved from sex, and while the evolutionary origin and molecular nature of apomixis remain enigmatic, various hypotheses regarding specific genetic mechanisms have been proposed. One possible mechanism is de-regulation in the timing of sexual reproductive genes or pathways (Grimanelli *et al.*, 2001). The switch from sexual to apomictic reproduction has also been associated with gene dosage effects during endosperm development (Morgan *et al.*, 1998).

Furthermore, the global regulatory effects of polyploidy and hybridity, both of which characterize virtually all asexual plants (and parthenogenetic animals), have been proposed as possible triggers for the switch from sex to apomixis (Carman, 1997; Grimanelli *et al.*, 2001). More specifically, hybridity has been hypothesized to induce asynchronous expression of sexual reproduction genes to lead to apomixis (Carman, 1997).

Understanding patterns of differentially expressed genes is crucial for disentangling the complex regulatory networks which characterize sexual and apomictic seed production. Advances in cell isolation methods, in conjunction with next generation sequencing technology, have enabled global comparisons of gene expression patterns between sexual and apomictic reproductive tissues, and have provided support for deregulation of reproductive pathways in the switch from sex to apomixis (Sharbel *et al.*, 2009, 2010). The analysis of gene expression, however, requires sensitive, precise, and reproducible measurements for particular mRNA sequences in specific tissues. In this regard, quantitative real-time PCR (qRT-PCR) is at present the most extensively used method for validating genome-wide (*e.g..* microarray) expression data (Sharbel *et al.*, 2010), due to its high sensitivity, specificity, and broad quantification range (Hong *et al.*, 2008). Although it is an extremely powerful technique, qRT-PCR requires strict normalization steps to compensate for several experimental variables that cannot be completely controlled (*e.g.* amount of starting material, enzymatic efficiency, differences in the transcription activity between cell or tissues) and which can influence reproducibility between experiments (Vandesompele *et al.*, 2002). Accurate normalization of qRT-PCR results is thus essential for precise comparisons between samples. The standard approach for normalization of qRT-PCR data is the use of internal control or reference genes, often referred to as housekeeping genes (HKGs (Bustin *et al.*, 2005)). This class of genes encodes proteins that typically function in basic cell metabolism or maintenance, with constant expression levels and low levels of fluctuation between most tissues. Currently, the most common and well-described housekeeping genes used for the normalization of gene expression data include actin (Czechowski *et al.*, 2005), glyceraldehyde-3-phosphate dehydrogenase (G3PDH (Czechowski *et al.*, 2005)), ribosomal genes, cyclophilin, elongation factor 1-a (Efα1 (Czechowski *et al.*, 2005; Dean *et al.*, 2002; Sturzenbaum & Kille, 2001; Thomas *et al.*, 2003)), adenine phosphoribosyl transferase (aprt (Orsel *et al.*, 2002)) and tubulin

(Ozturk *et al.*, 2002). Recently, Silveira *et al.* (Silveira *et al.*, 2009) established BbrizUBCE, BbrizE1F4A and BbrizEF1 as the best reference genes for analyses of sexual and apomictic ovary tissues of the monocotyledon *Brachiaria*. Many studies have shown that standard housekeeping genes used as internal standards for the quantification of mRNA expression can indeed vary with the experimental conditions (Sturzenbaum & Kille, 2001; Thellin *et al.*, 1999; Warrington *et al.*, 2000). A well-tested housekeeping gene showing significant expression stability in a plant species or tissue type might not show the same stability if used in different experimental situations, species or tissues (Bustin *et al.*, 2005; Sturzenbaum & Kille, 2001). Reference genes therefore need to be properly validated for specific species, tissue types or reproductive modes when designing quantitative gene expression studies (Schmittgen & Zakrajsek, 2000). Furthermore, the use of a single housekeeping gene for qRT-PCR normalization is not recommended due to potential error, and it has been proposed that at least two or three housekeeping genes should be used in parallel as internal standards (Andersen *et al.*, 2004; Thellin *et al.*, 1999; Vandesompele *et al.*, 2002). Thus it is essential that prior validation of all reference genes is performed to confirm their expression stability in particular experimental conditions or tissues/cells, in order to prevent inaccurate data interpretation and subsequent false conclusions.

The genus *Boechera* (Brassicacea) is becoming a model system for studying apomixis, being composed of both sexual and apomictic genotypes, the latter of which display quantitative variation for levels of apomictic seed production (Aliyu *et al.*, 2010). Importantly, the occurrence of diploid apomictic forms (Boecher, 1951) in *Boechera* makes it possible to compare differences in gene expression between apomictic and sexual individuals without the concomitant effects of polyploidy. Moreover, as wild relatives of *Arabidopsis thaliana*, molecular genetic studies in *Boechera* are facilitated by the extensive genetic resources which have been developed for this model plant (Naumova *et al.*, 2001). In addition, *Boechera* species have been used for comparative genomic analysis, including partial genome sequencing (Windsor *et al.*, 2006), genetic map construction (Schranz *et al.*, 2007) and transcriptome sequencing (Sharbel *et al.*, 2009, 2010), and the entire genomes of *B. stricta* and *B. divaricarpa* are being sequenced (DOE Joint Genome Institute; http://www.jgi.doe.gov webcite).

18

Considering the growing importance of this genus for evolutionary functional genomics, the aims of this work are to (1) validate the stability of some commonly used housekeeping genes, and (2) evaluate a new housekeeping gene for qRT-PCR analyses of microdissected reproductive tissues in apomictic and sexual members of the genus *Boechera*. We have identified one putative new HKG from our transcriptomic analyses of sexual and apomictic ovules at different developmental stages (Sharbel *et al.*, 2009, 2010), and have additionally tested five known HKGs from *Arabidopsis* and other plant genera [Actin 2, s18 rRNA, elongation factor 1-α (Efα1), Pex 4 and Polyubiquitin 10; Table 2.1]. All HKGs were tested for stable gene expression patterns in both sexual and apomictic *Boechera* in various microdissected reproductive tissues including: four ovule stages (Figure 2.1; (Sharbel *et al.*, 2009)), three anther stages (Figure 2.2; (Armstrong & Jones, 2003)) and four different tissues (flowers, leaves, roots, stems; Table2.2)

Table 2.1 *Boechera*-specific qRT-PCT primers for tested HKGs

Gene identification/Gene description	Primer sequence 5'-3' Forward and reverse	Amplicon size (bp)	Amplification efficiency ± SD *	EMBL Accession number
BoechACT2/ Actin 2	GTTCCACCACTGAGCACAATGTTACC AGTCTTGTTCCAGCCCTCTTTTGTG	132	0.94±0.003	FR846456
BoechEF1/ Elongation factor-1 alpha	CCAAGGGTGAAAGCAAGGAGAGC CACTGGTGGTTTTGAGGCTGGTATCT	75	0.96±0.002	FR846458
BoechRPS18/Ribosomal protein S18	GCTGGGGAGTTATCTGCTGCTGAG CTTGCCGTCTTTGTAATCCTTCTGC	117	0.94±0.003	FR846460
BoechPEX4/Peroxin 4	TTTGCAGTTGACAGTTGGATCTTGTTC TCGCTCGTGATGCCTATTCATCATAC	143	0.83±0.009	FR846459
BoechAt1g09770.1/*Arabidopsis thaliana* cell division cycle 5	GCCATGATCTAAAAAGTTGGGACAAA TATTCGTCACAACACATGCAAGGTTTA	145	0.88±0.007	FR846457
BoechUBQ/Polyubiquitine	GGCTAAGATCCAGGACAAGGAAGGTAT CTGGATGTTATAGTCAGCCAAAGTGCG	71	0.94±0.004	FR851958

*Amplification efficiency was calculated using the miner algorithm (Zhao & Fernald, 2005)for the calculation of the means and corresponding SD :

$$\bar{x} = \sigma^2 \cdot \sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i$$

(Hardy et al., 1988)

2.3. Methods

Two *Boechera* accessions, a sexual diploid *B. stricta* and a facultative apomictic diploid *B. divaricarpa* were selected for the analyses (Table 2.2). Seedlings of these accessions were grown and maintained in a phytotron at the IPK under controlled environmental conditions (day: 16 h, 21°C; night: 8 h, 16°C; humidity 70%).

Six candidate housekeeping genes were selected, five previously-described HKGs in other plant genera and one new HKG which appeared to be stably expressed in a SuperSAGE dataset (Table 2.1; (Sharbel *et al.*, 2009, 2010)). The gene to which the selected SuperSAGE tag sequence corresponded was found via a BLAST search (Altschul *et al.*, 1990) to two flower-specific (sexual and apomictic) *Boechera* cDNA libraries which were sequenced using 454 (FLX) technology (Sharbel *et al.*, 2009). The corresponding *Boechera* cDNA (Polyubiquitine, Table 2.1) was annotated using a homology search to the *Arabidopsis* genome (http://www.*Arabidopsis*.org webcite). *Arabidopsis* homologues to the five known HKG's were similarly identified, and these were BLASTed (Altschul *et al.*, 1990) to the *Boechera* cDNA libraries (E-value < 3e-024 and 2e-019 for the apomictic and sexual 454 cDNAs respectively) to obtain corresponding *Boechera*-specific gene sequences. PCR primers were then designed for DNA sequencing of the identified genes using DNASTAR Lasergene® Primer Select (http://www.dnastar.com/products/lasergene.php webcite).

DNA was extracted from 100 mg of leaf tissue from each plant using a Qiagen Dneasy® Plant Mini Kit (QIAGEN, Hilden, DE) according to the manufacturer's instructions. For all HKGs, PCR reactions (10 μl) were mixed as follows: 25 ng of DNA, 1 μl of PCR Buffer II, 10 pmol for each primer, 0.025 U of AccuPrime™ Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA) with 3.5 mM of MgCl2 and 4.95 μl of H2O. PCR reactions were performed in a Mastercycler ep384 (Eppendorf, Hamburg, DE) using the following touchdown thermal cycling profile: 94° for 10 min; 9 cycles of 94° for 15 sec, 65° for 15 sec (1 degree decrease in temperature every cycle with a final temperature of 54°), 72° for 30 sec; 35 cycles of 94° for 30 sec, 57° for 15 sec, 68° for 2 min 30 sec; and a final 68° for 15 min. Each PCR product was cloned into a TOPO TA Cloning® (Invitrogen) vector according to the recommendation of the supplier. Eight clones per product were confirmed by DNA sequencing using Sanger Sequencing

methods on an ABI 3730 xL platform (Applied Biosystem. Carlsbad, California) and analyzed using the DNASTAR Lasergene® SeqBuilder and MegAlign programs.

qRT-PCR primers were designed using DNASTAR Lasergene® PrimerSelect, with all amplification products targeted between 70 and 160 bp, and melting temperatures between 58° to 63° C. The newly-designed primers were checked using the following PCR (20 μl) protocol: 25 μg of genomic DNA, 2 μl of 10 x reaction buffer, 20 pmol for each primer, 0.5 u/μl of BioTAQ DNA Polymerase (Bioline GmbH, Luckenwalde, DE), 2.5 nM of dNTPs, 2 mM of MgCl2, 11.1 μl of water. We used the following thermal cycling profile: 94° C for 3 min, 35 cycles of 94°C for 30 sec, 59°C for 15 sec, 68°C for 1 min and finally 70°C for 7 min. The size of all PCR products was verified on a 1.5% agarose gel.

Total RNA was isolated from four different tissues (leaf, root, stem and flower) harvested from two biological replicates of both *Boechera* accessions (Table 2.2) using the Qiagen Rneasy® Plant Mini Kit following the manufacturer's instructions. The isolated RNA was treated with Qiagen Rnase-Free DNase according to the producer's protocol in order to eliminate any contaminating traces of DNA. A second purification step was performed using a Qiagen® Rneasy Mini Kit to eliminate contaminating polysaccharides, proteins and the DNase enzyme. The final concentration and quality was checked using an Agilent Technologies 2100 Bioanalyzer NanoChip (Agilent Technologies, - Santa Clara, CA, United States).

The gynoecia of sexual and apomictic *Boechera* were dissected from flowers at the megasporogenesis stage in a 0.55 M sterile mannitol solution between 7:30 am and 9:00 am each day. Microdissection was performed in a sterile laminar air flow cabinet under a stereoscopic microscope (Stemi 1000; Carl Zeiss). Ovules at 4 different developmental stages (Table 2.2 and Figure 2.1) and placental tissues were then collected under an inverted microscope (Axiovert 200 M; Carl Zeiss), in sterile conditions using sterile glass needles (self made using Narishige PC-10 puller). For each developmental stage approximately 20 ovules and 1 mm² of ovary tissue were collected in separate sterile Eppendorf tubes containing 200 μl of RNA stabilizing buffer, using a glass capillary (internal diameter 150 μm) interfaced to an Eppendorf Cell Tram Vario. Anthers at corresponding flower developmental stages 8-10 (Armstrong and Jones, 2003) (Figure 2.2) were selected for extraction of total RNA. Approximately 30 anther

heads per sample were dissected from fresh whole flower buds and stored in RNA stabilizing buffer (RNA later; Sigma-Aldrich) under a stereoscopic microscope (Zeiss Stereo Discovery V12) using sterile glass needles. RNA was extracted using a Qiagen PicoPure Isolation Kit and purified of contaminating DNA using Qiagen RNase-Free DNase.

First strand cDNA was synthesised from 10 ng starting RNA with a RevertAid™H Minus First Strand cDNA Synthesis Kit (Fermentas) using an oligo(dT)18 primer following the manufacturer's instructions. The resultant concentration was checked using a PicoGreen® dsDNA Quantitation Kit (Invitrogen) with a NanoDrop® ND-3300 Spectrofluorometer (NanoDrop). qRT-PCR reactions were performed on an ABI-PRISM 7700 HT FAST Real-Time PCR System (Applied Biosystems) with the following cycling profile: 50°C for 2 min, 95°C for 10 min; 40 cycles of 95° C for 15 sec, 60° C for 1 minute. 10 μl reactions were performed using the following master mix: 5 μl of SYBR I Master Mix buffer, a total of 16.6 pmol for both, sense and anti-sense primers, 2.5 μl of water and 1.5 μl of cDNA. A melting curve gradient was obtained from the product at the end of the amplification for checking amplicon quality. cDNA samples derived from somatic tissues (leaf, root, stem and flower) were run in a serial dilution range of 5, 2.5, 1.25, 0.625 and 0.312 ng. All samples were run in triplicate with the control gene included in each plate. Due to low amounts of starting cDNA material from the microdissected ovules, a dilution range of 1, 0.5, 0.25, 0.125, 0.062 ng was used. Candidate and control genes were run simultaneously in two replicates with 4 ovule stages and 3 anther head stages for both sexual and apomictic accessions. *Boechera* Polyubiquitin 10 was selected as control gene due to its extensive use and proven reliability as a reference control in *Boechera* (Sharbel *et al.*, 2009), *Arabidopsis* (Czechowski *et al.*, 2005) and other plants (Salmona *et al.*, 2008).

Considering instrument background fluorescence, Crossing Point (Cp) is defined as the point at which sample fluorescence rises significantly above the background fluorescence characteristic of a particular detection system, and it is used as a measure for the starting copy numbers of the target gene. For every cDNA, the mean expression level and standard deviation for each set replicate was calculated. In cases where Cp values between replicates of the same gene diverged by more than one unit, as measured from cDNAs extracted from microdissected tissues, two additional replicates

of that particular gene were performed under the same experimental conditions. The corresponding qPCR efficiencies were determined by the Miner algorithm (Zhao and Fernald, 2005). To quantify gene expression in comparison to a reference gene, the relative expression ratio (R) was determined using the $\Delta\Delta$Ct method as described by Pfaffl (Pfaffl *et al.*, 2004). The obtained R values for all the genes were transferred into the geNorm program (http://medgen.ugent.be/~jvdesomp/genorm/ webcite) for calculation of the expression stability as described by Vandesompele *et al.* (Vandesompele *et al.*, 2002)

For validation of the best HKGs, four SuperSAGE Tags that had shown reproductive mode-specific expression in cDNA between ovules at the second and fourth developmental stage were selected (see additional file 2). qRT-PCR reactions of cDNA from apomictic and sexual ovules at stages 2 and 4 were performed on an ABI-PRISM 7700 HT FAST Real-Time PCR System (Applied Biosystems) with the following cycling profile: 50° C for 2 min, 95° C for 10 min; 40 cycles of 95° C for 15 sec, 60° C for 1 minute. 10 μl reactions were performed using the following master mix: 5 μl of SYBR I Master Mix buffer, a total of 16.6 pmol for both sense and anti-sense primers, 2.5 μl of water and 1.5 μl of cDNA diluted to 0.5 ng. Genes of interest and HKGs were run simultaneously in triplicate. For every cDNA, the mean expression level and standard deviation for each set replicate was calculated. The corresponding qPCR efficiencies were determined using the Miner algorithm (Zhao & Fernald, 2005). The expression data were normalized according to the REST algorithm using the REST2009 software (Pfaffl *et al.*, 2002)

2.4. Results

2.4.1. *Boechera* homologues of candidate HKGs

In order to identify optimal reference (HKG) genes in the genus *Boechera*, one candidate gene from a SuperSAGE dataset which was found to be uniformly expressed between apomictic and sexual *Boechera* accessions (Sharbel *et al.*, 2009; Sharbel *et al.*, 2010), and five previously-described HKGs from other plant species (Actin 2, RPS18, Elongation factor 1-α, Pex 4 and UBQ) were selected. Using the *Arabidopsis* genome database, in addition to two flower-specific *Boechera* cDNA libraries, we were able to design PCR primers in order to amplify, clone and sequence sections of the following six

genes   BoechACT2,   BoechPEX4,   BoechPEX4,   BoechRPS18,   BoechEf α 1,
BoechAt1g09770.1 and *Boechera* Polyubiquitin 10 (Table 2.1).

Figure 2.1: Live microdissections of *Boechera* ovules at multiple developmental stages. (a) 1I to 1II; (b) 2II to 2IV; (c) 3II to 3III; (d) 3V to 4I (Schneitz *et al.*, 1995) Bars = 10 μm.

2.4.2. HKG stability analysis using qRT-PCR

The six candidate reference genes were evaluated for gene expression stability by qRT-PCR, using qRT-PCR primers designed in exonic regions from the cloned and sequenced *Boechera* homologues (Table 2.1). Based on cDNA analysis and the dissociation curve for each of the primer sets tested, a single PCR product with the expected size was amplified HKG stability across different tissues, both vegetative and microdissected reproductive tissues, shows relatively tight Ct distributions for all six genes (Figure. 2.3). Based upon the distribution of Ct values across different tissues, the HKGs could be split into low (BoechRPS18, BoechEf α 1, BoechACT2 and BoechUBQ) and high (BoechAt1g09770.1 and BoechPEX4) values (Figure. 2.1). Using the Miner algorithm (Zhao & Fernald, 2005), amplification efficiencies (E) were calculated to range between 0.74 ± 0.03% and 1.01± 0.1%. Expression ratios (R) were calculated,

and amplification efficiencies and Ct values exported to the geNorm program as described by Vandesompele *et al.* (2002). In order to evaluate gene stability, geNorm relies on the principle that two ideal control genes have the same expression ratio in all the samples despite cell type or experimental conditions. The program calculates two variables: the pairwise variation (V), which indicates the minimum number of reference genes required for a precise normalization, and the average pairwise variation of a particular gene compared to that of all other genes (M). Genes with the lowest M values have the most stable expression (Table 2.3).

2.4.3. Apomictic and sexual specific HKGs

R values of all the samples were calculated according to reproductive mode in order to test for differences in reference genes between apomictic and sexual *Boechera*. The results show that three genes, BoechRPS18, BoechACT2 and BoechEfα1, are the most stable in all tissues of both the apomictic and sexual accessions (Figure. 2.4). The pairwise variation (V) values showed that for accurate normalisation, the two most suitable stable genes to employ are BoechRPS18 and BoechEfα1 for the apomictic accession, and BoechRPS18 and BoechAct2 for the sexual accession. With the addition of one more gene, pairwise variation (V2/3) values of 0.042 for the apomictic and 0.078 for the sexual accessions were obtained (Figure. 2.5), values far below the cut-off of 0.15 suggested by Vandesompele *et al.* (2002)
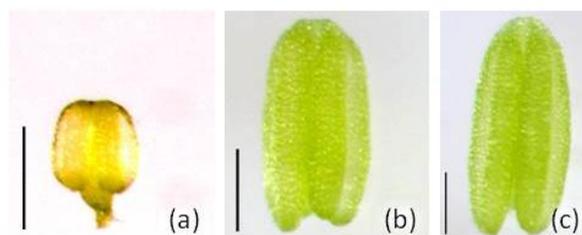
Figure 2.2 Live microdissected of *Boechera* anthers at multiple developmental stages (a) 7-8; (b) 9-10; (c) 11;(Armstrong and Jones, 2003) Bar = 200 μm

Table 2.2 *Boechera* **accessions, ovule and anther stage-specific developmental characteristics and ploidy of cells**

Species[a]	ID	Collection locality	Reproduction	Ovule stage[c]	Ovule development[d]	Anther stage[g]	Anther development[h]
B. stricta	MT49	Sagebrush Meadow, MT	Sex	A	Nucellus	7-8	Premeiotic PMC
				B	MMC[e] formation	9-10	Meiotic PMC[f]
				C	Tetrad degeneration	to 11	Microspore formation
				D	fertilised ovules		
B.divaricarpa	MT15	Vipond Park, MT	Apomixis	A	Nucellus	7-8	Premeiotic PMC
				B	MMC[e] formation	9-10	Meiotic PMC
				C	Tetrad degeneration	to 11	Microspore formation
				D	fertilized ovules		

(a) Species identification was based upon silique orientation, trichome morphology, and cpDNA sequences (Kiefer *et al.*, 2009) (b) Reproductive mode was confirmed using the flow cytometric seed screen(Matzk *et al.*, 2000; Sharbel *et al.*, 2009) (c) See Figure 1 for images of each stage. (d) According to Schneitz *et al* (1995) (e)MMC: Megaspore mother cellf PMC: Pollen mother cell g,h According to Armstrong (Armstrong & Jones, 2003)

2.4.4. Vegetative and reproductive tissue-specific HKGs

In order to identify the best reference gene suitable for specific tissues, separate analyses were performed independently for vegetative tissues, and microdissected live ovules (all stages) and anthers (all stages). As expected, BoechRPS18, BoechACT2 and Efα1 again exhibited high stability, but interestingly the stability values varied between tissue groups (Table 2.3). To validate this result, the analyses were repeated but this time the two least stable genes (BoechPEX4 and BoechAt1g09770.1) were removed from the set. In doing so, similar stability values (*i.e.* no effect on M) for BoechRPS18, BoechACT2 and Efα1 to those observed in the earlier analyses were obtained. These data thus show that BoechRPS18, BoechACT2 and BoechEfα1 provide the best combination of HKGs for any tissue specific normalisation in *Boechera*.
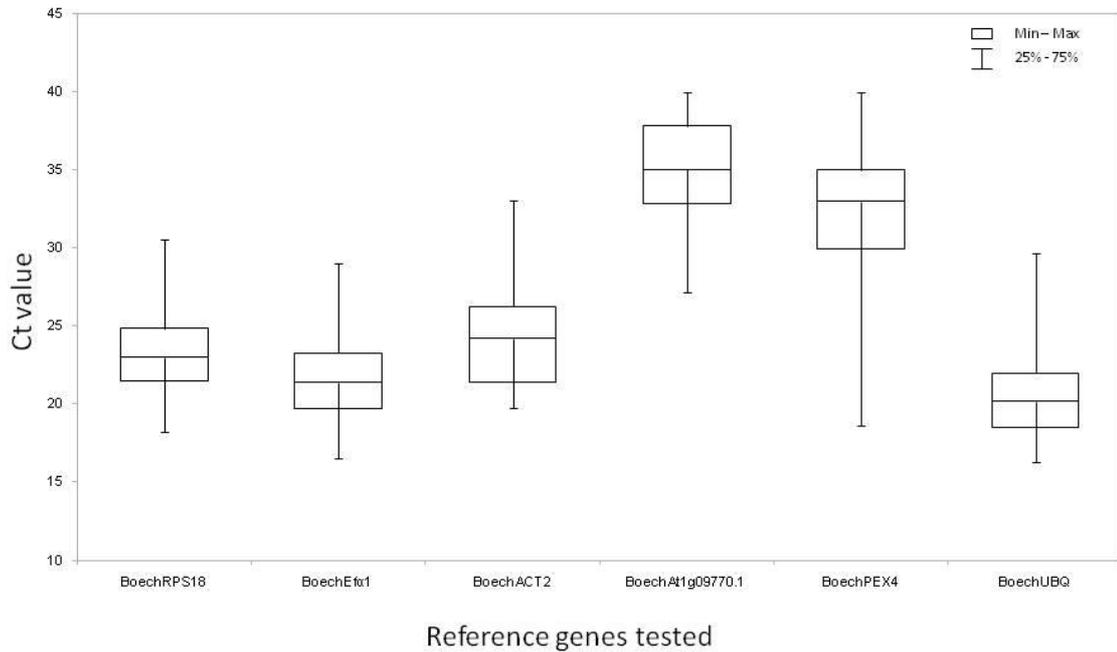
Figure 2.3 Box-whisker plot showing the Ct variation of each candidate reference gene among the different tissue samples

2.4.5. Reproductive mode independent housekeeping genes

Further analyses were done to identify the two best HKGs common to both apomictic and sexual reproductive modes. In removing $Ef\alpha1$ from the gene set of the apomictic accession, BoechRPS18 and BoechACT2 were the most stable genes with M = 0.24 and V2/3 = 0.073. BoechRPS18 and BoechACT2 can therefore be used independently of reproductive mode, and should be chosen in cases where the reproductive mode of the plant under study is uncertain

28

Table 2.3 Summary of the two best HKG combinations for different tissues and reproductive system according to GeNorm.

Tissue	Recommended HKGs	M values	V2/3 values	Recommended HKGs, (*Boech*RPS18 replaced by UBQ)	M values	V2/3 values
Apo All tissues	BoechRPS18 , BoechEfα1	0.130	0.042	BoechUBQ , BoechEfα1	0.120	0.120
Apo vegetative tissues	BoechEfα1, BoechACT2	0.170	0.007	BoechUBQ , BoechEfα1	0.220	0.120
Apo ovules	BoechRPS18, BoechEfα1	0.060	0.040	BoechUBQ, BoechEfα1	0.055	0.131
Apo Anthers	BoechACT2, BoechEfα1	0.020	0.001	BoechACT2, BoechEfα1	0.020	0.090
Sex All tissues	BoechRPS18, BoechACT2	0.017	0.078	BoechUBQ, BoechACT2	0.200	0.100
Sex vegetative tissues	BoechRPS18, BoechEfα1	0.070	0.089	BoechUBQ, BoechACT2	0.220	0.100
Sex ovules	BoechRPS18, BoechACT2	0.010	0.088	BoechUBQ, BoechEfα1	0.010	0.057
Sex Anthers	BoechEfα1, BoechACT2	0.001	0.025	BoechEfα1, BoechACT2	0.001	0.016

2.4.6. Ubiquitin stability analysis

To ascertain the stability and suitability of Ubiquitin (BoechUBQ) in all tissues of *Boechera*, the R values of all test genes were recalculated with BoechRPS18 as control, and analysed using geNorm. As expected the result showed BoechUBQ to be the most stable gene in both apomictic (M=0.11) and sexual (M=0.20) *Boechera* accessions. Based upon calculations of V, the data here show that BoechUBQ and BoechEf α 1 (V2/3 = 0.12) provide accurate normalisation for apomictic genotypes, whereas BoechUBQ and BoechAct2 (V2/3 = 0.10) are more appropriate for sexual *Boechera*.
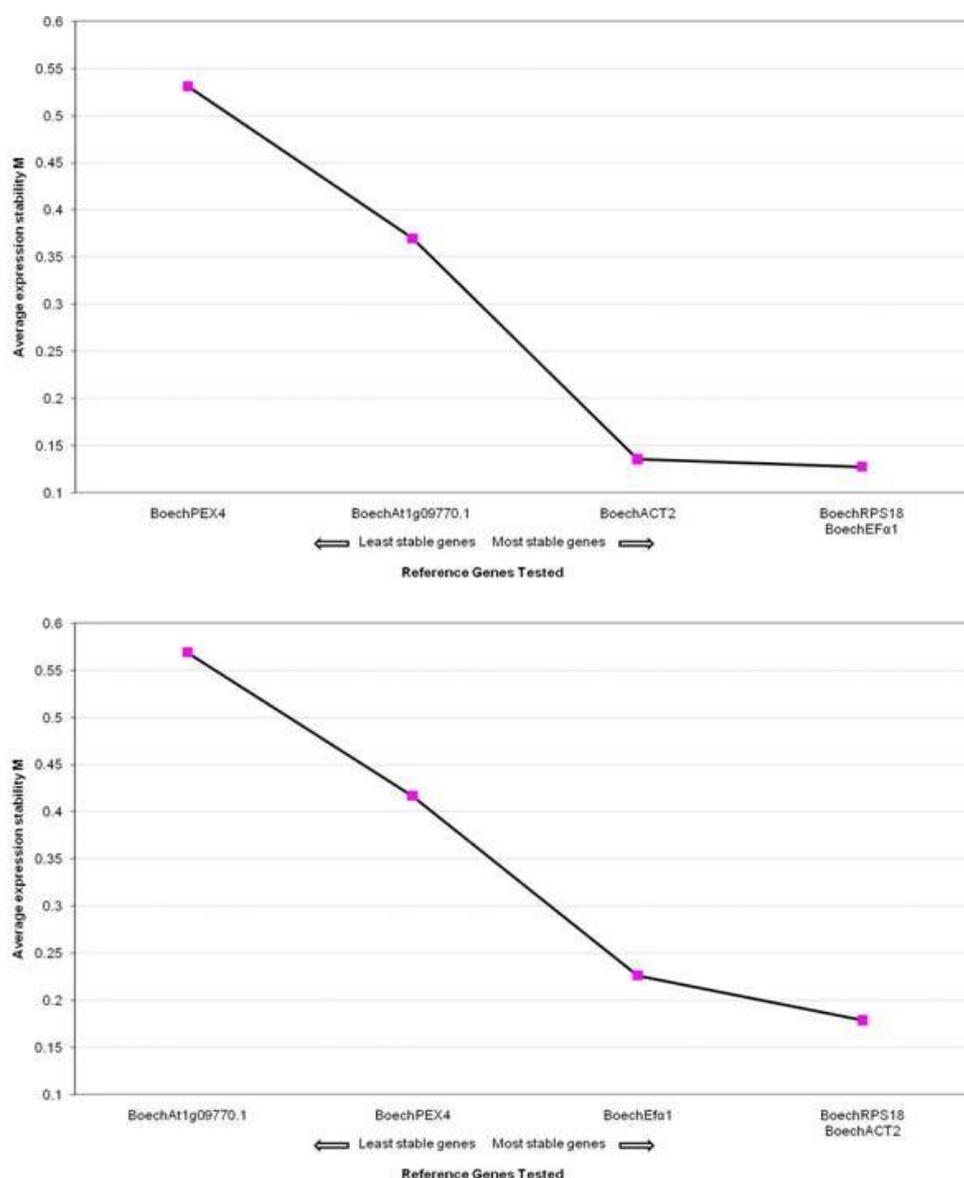
Figure 2.4 Average expression stability values (M) of the control reference genes from GeNorm Plotted from the most stable (right) to the less stable (left) using UBQ as reference, from (a) apomictic vegetative and reproductive *Boechera* tissues, and (b) sexual vegetative and reproductive *Boechera* tissues
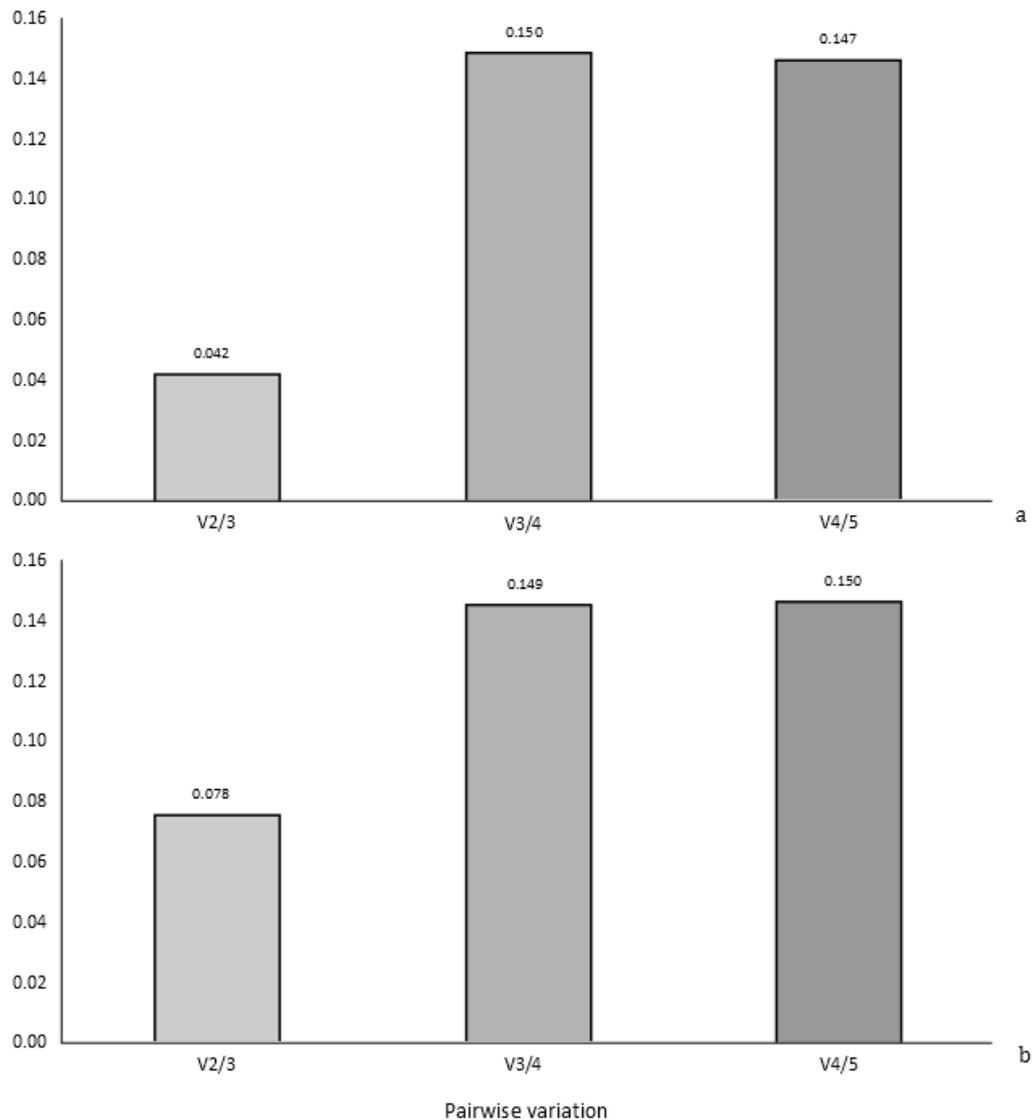
Figure 2.5 Pairwise variation (V) of the selected reference genes in (a) apomictic and (b) sexual *Boechera* calculated on GeNorm, from the most to the less stable according to M values. (a); V2/3 pairwise variation between the two most stable genes (RPS18 and Ef α 1) + the third most stable gene (Act2); V3/4: addition of the fourth most stable gene (at1g09770.1); V4/5: addition of the fifth most stable gene (Pex4). (b); V2/3 RPS18 , Act2 + Ef α 1; V3/4: + Pex4; V4/5: + at1g09770.1

2.5. Discussion

Based upon the transcriptional profiles of the six housekeeping genes tested in this study, and geNorm and normFinder analyses of different vegetative and reproductive tissues of sexual and apomictic *Boechera*, we conclude that ribosomal subunit protein 18 (BoechRPS18), elongation factor-1 (BoechEf α 1), Actin 2 (BoechACT2) and Polyubiquitine (BoechUBQ) are the most stable. Although all four genes show significant stability, their values (M) varied depending on either specific tissue and/or reproductive system. Considering these criteria and pairwise variation

(V), we propose optimal combinations of reference genes for normalization of gene expression data in transcriptome analyses of different tissues (Table 2.3). Using normFinder we were able to verify the stability of our candidates by using a different algorithm. Interestingly, in 11 analyses out of 16, the first HKG to be ranked as most stable by NormFinder was also found in the best combination provided by geNorm. In five cases the second best HKG to be ranked by normFinder was included in the best couple provided by geNorm, while the remaining one was ranked as the third best. These discrepancies could be explained by the fact that geNorm and normFinder use two different algorithms for the evaluation of the best HKGs. geNorm provides the two genes that have the most similar expression profiles through a stepwise elimination of the least stable genes in the sample, while normFinder instead uses a model-based approach to calculate a stability value which represents the expression variation of the gene throughout the sample. Differences in ranking when using these two programs have previously been found (see (Van Hiel *et al.*, 2009; Silver *et al.*, 2008)). Considering the errors that could result from single HKG normalization strategy (Vandesompele *et al.*, 2002) and that at least one gene from the best couple identified by geNorm was always ranked in the first 2 positions by normFinder, we suggest using the specific pair of genes recommended by geNorm (Table 2.3). As validation for our choice of genes, the best couples of HKGs (according to geNorm) were chosen for normalizing the expression of four SuperSAGE tags that had previously showed reproduction specific expression between two ovule developmental stages in *Boechera*. qRT-PCR was performed and the relative expression of the four tags was normalized against BoechRPS18, BoechACT2 and BoechEfα1 in the best combination according to geNorm, using the REST2009 software (Pfaffl *et al.*, 2002). The result of the normalization was consistent with the SuperSAGE expression data (See additional file 2). We hypothesize that interactions between the four best housekeeping genes identified here are minimized since each is involved in independent cellular processes.

CHAPTER 3: ASEXUAL GENOME EVOLUTION IN THE APOMICTIC *RANUNCULUS AURICOMUS* COMPLEX: EXAMINING THE EFFECTS OF HYBRIDIZATION AND MUTATION ACCUMULATION

3.1. Summary

Asexual lineages are thought to be prone to extinction because of deleterious mutation accumulation (Muller's ratchet). Here we analyze genomic effects of hybridity, polyploidy and allelic divergence in apomictic plants, and identify loci under divergent selection among sexual/apomictic lineages. RNAseq was used to sequence the flower-specific transcriptomes of five genotypes of the *Ranunculus auricomus* complex, representing three sexual and two apomictic reproductive biotypes. The five sequence libraries were pooled and *de novo* assembly performed, and the resultant assembly was used as a backbone for a subsequent alignment of each separate library. High quality single nucleotide (SNP) and insertion-deletion (indel) polymorphisms were mined from each library. Annotated genes for which open reading frames (ORF) could be determined were analyzed for signatures of divergent versus stabilizing selection. A comparison between all genotypes supports the hypothesis of Pleistocene hybrid origin of both apomictic genotypes from *R. carpaticola* and *R. cassubicifolius*, with subsequent allelic divergence of apomictic lineages (Meselson effect). Pairwise comparisons of non-synonymous (dN) to synonymous (dS) substitution rate ratios between apomictic and sexual genotypes for 1231 genes demonstrated similar distributions for all comparisons, although 324 genes demonstrated outlier (*i.e.* elevated) dN/dS ratios. Gene ontology analyses of these outliers revealed significant enrichment of genes associated with reproduction including meiosis and gametogenesis, following predictions of divergent selection between sexual and apomictic reproduction, although no significant signal of genome-wide mutation accumulation could be identified. The results suggest that gene function should be considered in order to understand effects of mutation accumulation in asexual lineages.

3.2. Introduction

Asexual reproduction has evolved repeatedly and independently from sexual ancestors in many species of animals and plants (Barton & Charlesworth, 1998; Mittwoch, 1978; Suomalainen, 1950). A number of factors are hypothesized to make

asexuality advantageous, including avoidance of the twofold cost of sex (Maynard Smith, 1978), and adaptation to environmental stability (Bell, 1982). On the other hand, sex can be advantageous in changing environments (Van Valen, 1973), for generating genetic variance (Fisher, 1930), and for purging deleterious mutations (Kondrashov, 1982; Muller, 1964). The lack of meiotic recombination and syngamy defines the asexual genome with a number of effects, including decreased probability of fixing advantageous mutations (Crow & Kimura 1965), the accumulation of deleterious mutations (*i.e.* Muller's ratchet; Felsenstein, 1974; Muller, 1964), and the loss of meiotic recombinational DNA repair mechanisms (Bernstein *et al.*, 1988; Hörandl, 2009), all of which play a central role in the evolutionary theory of sex (Kondrashov, 1993; Maynard Smith, 1978).

Many plants reproduce through apomixis, a reproductive phenomenon that allows the mother to produce clonal progeny via seeds (Nogler, 1984). Apomixis is found naturally in more than 400 plant species, and comprises various forms (http://www.apomixis.uni-goettingen.de; Hojsgaard *et al*, submitted). Gametophytic development is characterized by three steps: the production of a meiotically unreduced egg cell (apomeiosis), parthenogenetic development of this egg cell without fertilization, and production of a functional endosperm with (pseudogamy) or without (autonomous) fertilization of the binucleate central cell of the ovule (Koltunow & Grossniklaus, 2003). While the evolutionary origin and molecular mechanisms behind apomixis remain enigmatic, it is generally accepted that deregulation in the timing of developmental steps characteristic of the sexual reproductive pathway (Grossniklaus *et al.*, 2001; Koltunow, 1993), the result of global gene regulatory effects associated with polyploidy and/or hybridity (Carman, 1997; Grossniklaus, 2001), leads to the induction of apomixis. In fact, almost all apomictic plants are polyploid (Koltunow and Grossniklaus, 2003) with some exceptions of diploid apomictic hybrids (Kantama *et al.*, 2007).

Polyploidy increases the load of deleterious mutations as more potential mutation sites are available; all else being equal, the mean fitness of a population is approximately reduced by deleterious mutations by cU (where c is the ploidy level and U the mutation rate of the haploid genome; Gerstein and Otto, 2009). In the short term, polyploidy can buffer negative effects of the mutational load by masking recessive deleterious mutations, with a temporary rise in the fitness of a newly formed polyploid

(Otto and Whitton, 2000). In the long run, masked mutations persist longer and with higher frequencies in the population before being eliminated by selection (Otto and Whitton, 2000). In sexual lineages, the meiosis-mixis cycle further exposes haploid gametes or gametophytes with deleterious mutations and low fitness to purifying selection (Hörandl, 2009). In polyploid apomicts, this possibility of purging deleterious recessive mutations is often decreased considering that female gametes are by definition meiotically-unreduced, while both meiotically-reduced and –unreduced male gametes characterize some populations. Consequently, asexual polyploid lineages are expected to have short-term benefits, but might in the long run be more prone to decreasing fitness and eventual extinction.

When mutations are beneficial and partially to fully dominant, their effects are expected to have more significant relative positive effects in polyploid populations (Otto and Whitton, 2000). Asexuality is beneficial in newly formed polyploids and hybrids as it helps to bypass the effects of meiotic irregularities during gamete formation (Comai, 2005). In the majority of apomictic plants, the Mendelian factors regulating apomixis are indeed dominant (Ozias-Akins and Van Dijk, 2007). On the other hand, highly heterozygous, vigorous hybrid genotypes could have a selective advantage because of heterosis effects (Comai, 2005); in this case, the heterozygous hybrid genotype is expected to be favored by stabilizing selection (Otto, 2009). Heterozygosity could even be fixed in sexual allopolyploids by preferential homolog chromosome pairing at meiosis (Comai, 2005) rather than being correlated with the shift to asexuality. The expected genomic signature of such stabilizing selection in allopolyploids would be a genome-wide increase of heterozygosity, but not necessarily a rapid evolution in genes associated with apomixis.

All apomictic plants appear scattered on the tips of phylogenies and are thought to be evolutionarily young (Van Dijk and Vijverberg, 2005). However, estimating ages of lineages from phylogenies is biased by potential reversals from apomixis to sexuality (Schwander and Crespi, 2009; Hörandl and Hojsgaard, 2012). For asexual plants, hardly any empirical age estimate is available as reproductive mode cannot be assessed from the fossil record. Ancient asexuality can be inferred from intra-individual genomic divergence of alleles. In the absence of recombination, allele pairs are expected to independently accumulate mutations and diverge from one another; this so-called "Meselson effect" was described from ancient asexual animals (Welch & Meselson,

2000) and has been tested in plants (Corral *et al.*, 2009). Alternatively, similar high levels of heterozygosity may rather reflect the hybrid nature of apomicts (Beck *et al.*, 2012).

The *Ranunculus auricomus* complex is becoming a model system for a comparative genome evolution between sexual and asexual taxa. It comprises hundreds of apomictic species that are divided into two main subcomplexes which constitute morphologically distinct groups: R. *auricomus* and *R. cassubicus* (Hörandl *et al.*, 2009). The "*cassubicus*" subcomplex is a morphologically well-defined subgroup (Hörandl, 1998; Hörandl *et al.*, 2009) of a few sexual diploid and apomictic polyploid cytotypes, from which two sexual species and one apomictic taxon have been particularly studied. *Ranunculus carpaticola* is a diploid sexual species widespread in the Carpathians and central Slovakia, and the closely related sexual, diploid and autotetraploid R. cassubicifolius is distributed from eastern Austria to Switzerland (Hörandl *et al.*, 2009). Hexaploid apomicts, which likely originated from ancient interspecific hybridization between diploid *R. carpaticola* and the autotetraploid cytotype of *R. cassubicifolius* from Lower Austria, are present in central Slovakia (Paun *et al.*, 2006a). Population genetic marker analyses (AFLPs, SSRs), DNA sequence analysis and morphometric data support the hybrid origin hypothesis (Hörandl *et al.*, 2009; Paun *et al.*, 2006a). As with many taxa (Richards, 2003), apomictic *Ranunculus* are thought to have arisen during the last glacial period (115 000 – 15 000 years before present; Paun *et al.*, 2006a). A more distantly and geographically-separated taxon which is also part of the "*auricomus*" subcomplex, *Ranunculus notabilis*, is a sexual diploid species found in southeastern Austria (Hörandl *et al.*, 2009). Genetic distances between populations suggest that *R. notabilis* separated from the *R. cassubicifolius-R. carpaticola* ancestor c. 900 000 years ago, while the latter speciated c. 300 000 years ago (Hörandl, 2004).

Rapid advances in DNA sequencing technology are, by nature of increasing genome coverage per individual, leading to more confident use of single nucleotide polymorphisms (SNPs) in studies of ecology, conservation and evolution. SNPs have been successfully used to uncover individual (Cork & Purugganan, 2005), population (Song *et al.*, 2009), and species specific (Bajgain *et al.*, 2011; Trick *et al.*, 2009) differences, and have furthermore been employed to identify disease-linked markers (Hampe *et al.*, 2007). Next generation sequencing (NGS) technology has been successfully used for detecting SNPs in maize (Barbazuk *et al.*, 2007), (Novaes *et al.*,

2008), *Brassica napus* (Trick *et al.*, 2009) and rye (Haseneyer *et al.*, 2011), to name a few.

SNP analysis of annotated genes provides insights into functional sequence diversity and putative divergent selection (Bajgain *et al.*, 2011). For example, a high ratio of non-synonymous vs. synonymous substitutions could suggest that specific genes are under divergent selection. While sexual species in allopatry are expected to diversify slowly, independently-evolving asexual lineages are expected to rapidly diverge, the result of little to no recombination and syngamy, followed by the accumulation of divergent alleles (Lynch & Connery, 2000; Welch & Meselson, 2000; van Dijk, 2003; Corral *et al.*, 2009). Polyploidy and hybridity would further leave genomic signatures in asexual organisms as a higher number of possible mutational sites are available, while the shift to apomictic reproduction could have pleiotropic effects on functional traits and genes connected to apomixis.

Here we describe high-quality SNP discovery based upon RNAseq data collected from five sexual and apomictic individuals of four taxa of the *Ranunculus auricomus* complex. The goals of this project were to: (1) identify high-quality SNPs using strict quality and coverage parameters, to (2) compare mutation accumulation (*i.e.* Muller's ratchet), divergence and the genomic signature of hybridization between sexual and apomictic forms, and (3) to identify gene classes under divergent selection between sexual species, apomictic lineages and apomictic versus sexual taxa.

3.3. Materials and methods

3.3.1. RNA collection and extraction

Flowers were collected from five different genotypes of the *Ranunculus auricomus* complex (Table 3.1). For each plant total RNA was isolated from five different harvested flower sizes using the Qiagen RNeasy Plant Mini Kit (QIAGEN, www.qiagen.com) following the manufacturer's instructions. The isolated RNA was treated with Qiagen RNase-Free DNase according to the producer's protocol in order to eliminate any contaminating trace of DNA. A second purification step was performed using a Qiagen RNeasy Mini Kit to eliminate contaminating polysaccharides, proteins and the DNase enzyme. The final concentration and quality was checked using an Agilent Technologies 2100 Bioanalyzer NanoChip (Agilent Technologies, - Santa Clara, CA, United States)

Table 3.1 *Ranunculus* samples used in SNP analysis

Taxon	Sample	Ploidy[a]	Reprod.	Locality	Collectors and voucher information
R. carpaticola	REV1	2x	Sex	Slovakia, Slovenské rudohorie,Revúca, hill Skalka (forest)	Hörandl, 8483, 01.05.1998 (WU)
R. carpaticola x *cassubicifolius*	TRE	6x	Apo	Slovakia, Strázovské vrchy (near Trençín), between Kubra and Kubrica, close to the bus-stop Kyselka (margin of Carpinus forest and meadow)	Hörandl, Paun, Mládenková, s.n., 30.04.2004 (SAV)
R. carpaticola x *cassubicifolius*	VRU 2	6x	Apo	Slovakia, Turçianska kotlina,Vrútky-Piatrová, behind cottage (meadow)	Hörandl, Paun, Mládenková, s.n., 01.05 2004 (SAV)
R. cassubicifolius	YBB 1	4x	Sex	Austria, Lower Austria, Wülfachgraben, SE Ybbsitz (forest)	Hörandl, 8472, 12.04.1998 (WU)
R. notabilis	NOT1	2x	Sex	Austria, Burgenland, Strem valley on 1.5 km ENE Strem, 220 m, wet meadow and forest margin	Hörandl 5612+7220. 20.04.1996(WU);

WU: herbarium of the University of ViennaUniversität Wien, SAV: herbarium of the Institute of Botany, Slovak Academy of science, Bratislava. For ploidy identification see Hörandl & Greilhuber, 2002 and Paun *et al.*, 2006

3.3.2. RNA normalization

In order to avoid over representation of the most commonly transcribed genes, full-length enriched, normalized cDNA libraries were generated using a combination of a mint-universal cDNA synthesis kit (Evrogen, Moscow, Russia) and the Trimmer Direct cDNA normalization kit (Evrogen, http://www.evrogen.com) which utilizes the Kamchatka crab duplex specific nuclease method (Zhulidov *et al.*, 2004; Shcheglov *et al.*, 2007). The procedure generally followed the manufacturer's protocol but included several important modifications, as described by Vogel *et al.* (2010).

Each step of the normalization procedure was carefully monitored to avoid the generation of artifacts and overcycling. The optimal condition for ds-cDNA synthesis was empirically determined by subjecting the cDNA to a range of thermocycle numbers and their products checked by electrophoresis. The optimal cycle number was defined as the maximum number of PCR cycles without any signs of overcycling. Optimization of the complete cDNA normalization procedure was essentially performed as described in Vogel & Wheat (2011).

3.3.3. Sequencing, assembly and annotation

RNAseq, using 15 μl of normalized cDNA (200ng/µl), was outsourced to Fasteris (www.fasteris.com). As the genome of *Ranunculus* is uncharacterized, and furthermore since no genomic information was available from any other closely related species, a dual-sequencing approach was taken in order to balance costs and the ability to *de novo* assemble the sequencing data. Thus a single *R. carpaticola* was sequenced using a 108-mer paired-end (PE) approach, while the remaining four samples (*R. cassubicifolius*, *R. notabilis* and the two hybrid apomicts) were analysed by 54-mer single-end (SE) sequencing, both using the HiSeqTM 2000 Sequencing System from Illumina® (http://www.illumina.com/).

CLC genomic workbench (CLC bio version. 4.9, www.clcbio.com) was used for sequence assembly. At first the libraries were trimmed for vector contamination, length and quality score, and two methods were attempted to create the best reference assembly from which SNPs would be inferred. In a first assembly (pooled *de novo*) approach, all libraries were pooled and assembled using the following CLC parameters: nucleotide mismatch cost = 2; insertion-deletion costs = 3; length fraction = 0.5; similarity = 0.9; and any conflict among the individual bases were resolved by voting for the base with highest frequency, while contigs shorter than 300 bp were removed from the final analysis. A second (iterated *de novo*) approach performed a *de novo* assembly on each species (*i.e.* library) separately using the same parameters, followed by a *de novo* assembly of all contigs from each individual libraries together, including the reads that were initially discarded in each library. Once again contigs smaller than 300 bp were removed from the data set. The different assemblies were evaluated using the number of matched reads and N50 values. The final contig set was annotated by searching for similarities against the complete protein database of UniProt (UniProt Release 2012_08), using the 2.2.18 version of Blastx (ftp://ftp.ncbi.nih.gov/blast/executables/). The resultant homology data were filtered based upon E-value, and only hits having E ≤ 1-10 were kept.

Table 3.2 Raw sequencing results from RNAseq runs on five *Ranunculus* genotypes.

	Genotype				
	REV1	TRE(Apo)	VRU 2(Apo)	YBB1	NOT
Sequencing technology	Illumina PE	Illumina SE	Illumina SE	Illumina SE	Illumina SE
Number of reads	137 304 156	22 203 280	21 330 140	20 934 596	19 972 288
Number of reads after quality filtering	96 855 485	8 799 138	8 958 973	8 218 588	8 056 033
Average read length	108	54	54	54	54
Number of matched reads	120 154 734 (91.79 %)				

3.3.4. SNP calling

The transcriptome reads were processed in a long-read alignment to the chosen reference assembly (see above) using BWA (version: 0.5.9-r16, Li & Durbin, 2010). For *R. carpaticola*, reads were mapped in paired end mode (PE), while the remaining genotypes were mapped in single end mode (SE). BWA mapping was performed using the default settings for Illumina sequence reads, except setting the minimal base quality for read trimming to 20 (see table 3.3 & 3.4). Read trimming was performed down to a minimal required read length of 35 bp if necessary. These conservative default settings consider only two nucleotide differences in the seeding process of the initial read placement. Gaps were penalized more strictly in comparison to mismatches. Within the mapping process insertions or deletions (INDELs) with less than 5 bp distance to the contig end were discarded (see BWA documentation).

Putative single nucleotide polymorphisms (SNPs) and INDELs were called using Samtools (version 0.1.18 r982; Li *et al.*, 2009). The raw set of potential variations was generated using VCFtools (version 0.1.17; Danecek *et al.*, 2011). Artificial variations caused by unambiguous bases in the reference (inserted Ns) were removed and remaining variations were filtered using the vcfutils module of VCFtools.

SNP mining was performed for each of the genotypes individually. To access relevant information on SNP calling we developed a Perl script to combine filtered SNPs from different genotypes to identify shared versus genotype-specific SNPs. Variation calling was run under default settings, and considered the following criteria: high-quality bases at SNP position, minimal read mapping quality (10), the probabilities for end distance bias (<0.001) and biases in base quality (1e-100; Li *et al.*, 2012). To exclude SNP calls located in collapsed regions we applied adjustable read

depth settings and discarded positions with extreme read depth. The tablet *al*ignment viewer (Milne *et al.*, 2010) was used for graphical visualization and inspection of putative SNP calls. Developed Perl scripts are freely available by the authors, but should be considered as in developmental stage.

Table 3.3 Default parameters used for vcfutils module of VCFtools

Abbreviation	Type	Description	default
-Q	INT	minimum RMS mapping quality for SNPs	10
-d	INT	minimum read depth	2
-D	INT	maximum read depth	10000000
-a	INT	minimum number of alternate bases	2
-w	INT	SNP within INT bp around a gap to be filtered	3
-W	INT	window size for filtering adjacent gaps	10
-1	FLOAT	min P-value for strand bias (given PV4)	0.0001
-2	FLOAT	min P-value for baseQ bias	1.00E-100
-3	FLOAT	min P-value for mapQ bias	0
-4	FLOAT	min P-value for end distance bias	0.0001
	FLOAT	min P-value for HWE (plus F<0)	0.0001
-p	INT	print filtered variants	1

Table 3.4 Default parameters used for vcfutils module of VCFtools

	Abbreviation	Type	Description	default	applied
GENERAL	-C	INT	parameter for adjusting mapQ; 0 to disable	0	0
	-d	INT	max per-BAM depth to avoid excessive memory usage	250	1000000
	-M	INT	cap mapping quality at INT	60	60
	-q	INT	skip alignments with mapQ smaller than INT	0	0
	-Q	INT	skip bases with baseQ/BAQ smaller than INT	13	0
SNP/INDEL					
	-e	INT	Phred-scaled gap extension seq error probability	20	20
	-F	FLOAT	minimum fraction of gapped reads for candidates	0.002	0.002
	-h	INT	coefficient for homopolymer errors	100	100
	-L	INT	max per-sample depth for INDEL calling	250	250
	-m	INT	minimum gapped reads for indel candidates	1	1
	-o	INT	Phred-scaled gap open sequencing error probability	40	40
	-P	STR	comma separated list of platforms for indels	all	all

42

3.3.5. Open reading frame and synonymous/non-synonymous mutation analysis

In order to identify synonymous and non-synonymous mutations, SNP variation was assigned to codon position based upon open reading frame (ORF) analysis using contigs in which at least one position had a high-quality polymorphism across all the lineages. The sequences were analyzed to find the starting codon and the ORF using the module GetOrf from the EMBOSS package (Version 1.5 Rice *et al.*, 2000). When two ORFs were found overlapping a portion of the same contig, or in the case of an ORF located inside a longer one, then the longest ORF was selected. In case of two similarly long (within 90% of each other) ORFs located in opposite strands of the contigs, the polarity of the ORFs was then compared with the strand orientation from the Blastx analysis and the appropriate one selected.

The proportion of dN/dS for all genotypes was pairwise calculated for all resultant annotated contigs where ORFs could be assigned using the Bio::Align::DNAStatistic BioPerl module (http://www.bioperl.org). In order to correct for cases with ratios having 0 in the denominator or in the numerator, a value of 1 was added to both synonymous and non-synonymous data (dN and dS) before calculating the ratios, as already described by Bajgain *et al.*, 2011 and Novaes *et al.*, 2008. Boxplots of the pairwise distributions of dN/dS between sexual, sexual and apomictic, and apomictic genotypes were generated using SPSS (Version 21.0: http://www-01.ibm.com/software/analytics/spss/products/statistics/), the program marks as outliers values which are equal or higher than one and a half box length, defined as the distance between the first and the third quartile (IQR). Lower outliers are those values which are equal or lower than one and a half IQR .The genes characterized by outlier values were selected for gene ontology (GO) analysis. Gene ontology analyses were performed using Blast2GO (Conesa *et al.*, 2005), using the default annotation parameters of the program. Enrichment analyses were performed using a Fisher´s exact test with FDR (false discovery rate) correction, and the complete set of annotated genes from *Ranunculus* (see above) was used as a reference dataset. A second GO investigation was performed with the web-based program AgriGO (Du *et al.*, 2010) using *Arabidopsis* homologous genes and the default annotation parameters of the program. ORFs were blasted against the *Arabidopsis* database (TAIR: http://www.Arabidopsis.org) and filtered with a BlastX (E-value cutoff of E ≤ 1-5). A

Fisher's exact test with FDR correction was used for defining enriched GO terms, and the AgriGo results were filtered for p value (E ≤ 1-5) and FDR (E ≤ 1-5).

3.3.6. Hybrid Origin

Previous hypothesis that the apomictic accessions would be hybrid derivatives from 2x *R. carpaticola* and 4x *R. cassubicifolius* were based upon extensive AFLP analyses of 450 individuals from the geographical distribution area (Paun *et al.*, 2006a). However, since AFLPs are dominant markers with unknown genetic background, this dataset could not provide information regarding allelic composition and its divergence in apomictic lineages, and hence neither on the relative evolutionary ages of these apomicts. Here we wanted to (1) date the origin of hybrids, and (2) test for single vs. multiple origins and post-hybrid divergence of apomictic lineages. Based on this phylogenetic framework we were able to test whether the SNP dataset can discriminate between heterozygosity derived from hybrid origin and that resulting from putative mutations accumulated after the hybrid origin of each apomictic lineage (see under g below). A phylogenomic analysis was performed on the SNP dataset whereby the presence/absence of SNPs in parents and the 6x hybrids bears information on evolutionary relationships and genomic evolution (Bajgain *et al.*, 2011). The original SNP dataset was converted into a presence/absence matrix of SNPs, considering shared presence only (ie. SNPs appearing in at least two taxa, and differing in at least two taxa). Since our dataset did not provide detailed information on allelic ratios in the different cytotypes, a direct analysis of heterozygosity in hybrids was not feasible. To compare normal SNPs (*e.g.*, A) called from our assembled reference transcriptome with those heterozygous ones (*e.g.*, A/T) at the same site - which likely reflect different alleles in polyploids; Buggs *et al.* 2010; Trick *et al.*, 2012), the latter ones were coded with the standard IUPAC code for ambiguous sites (Bajgain *et al.*, 2011). Indel polymorphisms greater than 2 bp were coded as a single phylogenetic character (ie. equivalent to one site), based on the assumption that these polymorphisms would represent single evolutionary events (Simmons & Ochoterena, 2000). All other SNPs with low coverage were coded as gaps in the data matrix and treated as missing data in the phylogenetic analysis.

The expected conflict of phylogenetic signals in hybrids was resolved by using split network methods (Huson *et al.*, 2010) as implemented in Splitstree 4.0 (Huson & Bryant, 2006). By using uncorrected P-distances, we applied a distance-based method

of analysis to test for the expected reticulate data structure. A Neighbor Net analysis was done, shown as a hybridization network, and rooted with the more distantly related *R. notabilis* as outgroup taxon (see phylogeny by Hörandl *et al.*, 2009). Neighbor Net calculates the support for "splits" (relationships) from genetic distances and displays these splits in a graph (*i.e.* a "splits graph" or "split network"). The algorithm determines a circular ordering of taxa (*i.e.* based on the extent of differences between the SNP data, the taxa are ordered around a circle). The layout on the circle determines what splits occur in the data, and are displayed in a planar graph. The support for each of these splits is then measured using a least squares method that adjusts the lengths of the splits in the splits graph so as to minimize the difference with the pairwise distances in the original data matrix (Huson & Bryant, 2006).

3.3.7. Allelic sequence divergence in apomicts

To test the hypothesis that alleles at a single locus in parthenogenetic lineages are expected to accumulate mutations and diverge from each other in the absence of meiosis (Welch & Meselson, 2000), intra-individual allelic sequence divergence was measured in apomictic lineages after their hybrid origin. We selected a sample including 10 000 annotated SNPs for *R. carpaticola*, *R. cassubicifolius*, and the hybrid *R. carpaticola × R. cassubicifolius*. In order to check for the presence of intra-individual allelic sequence divergence, we looked first for heterozygous SNPs (*e.g.*, A/T at one site) per sequence/contig which were exclusive to either one of the two hexaploid apomictic lineages (see also Trick *et al.*, 2012). Such SNPs could reflect mutation accumulation between homologous chromosomes which were subsequently maintained through a lack of homogenizing mechanisms (*i.e.* homologous pairing and recombinational DNA repair during meiosis). In contrast, heterozygous sites derived from hybrid origin would be located on homologous chromosomes, which usually do not pair and recombine during meiosis (Comai, 2005), and would be shared both between allopolyploid apomicts and in either of their parents (see also Trick *et al.*, 2012). These latter heterozygous sites were not considered for calculating the age of apomictic genotypes.

Finally, as the number of sequence changes should be proportional to the time back to the hybridization event, we estimated the age of the apomictic genotypes by the number of generations needed to obtain the observed values. A theoretical number of expected mutations/neutral substitutions per generation was calculated using: (1) a

standard substitution rate for plant nuclear genes (u=5e-9; Wolfe *et al.*, 1989), (2) the total number of possible mutational sites in our sample (n=10 000), and (3) a multiplying factor of 3 (hexaploids had 3 times the number of possible sites). The generation time in *Ranunculus auricomus* was considered to be three years (from seedling to flowering and reproduction).

3.4. Results

3.4.1. Sequencing, assembly and high-quality SNP calling

Five normalized cDNA libraries were synthesized from RNA extracted from pooled flower buds at five different stages. The one paired-end (PE) and four single-end (SE) RNAseq runs generated a total of 221 744 460 reads, corresponding to a total of 19.5 Gbp of sequence. In detail, the 54 bp SE sequencing run of two apomictic genotypes (TRE and VRU2) and two sexual genotypes (YBB and NOT) generated from 19 to 22 million reads per library, while the 108-mer PE sequencing run of one sexual genotype (REV) generated ca. 138 million reads (Table 3.2).

The five libraries were screened for adaptor contamination, sequence quality and minimum length, resulting in 30 % of the PE library and an average of 60 % for the four SE libraries being trimmed. The pooled *de novo* assembly yielded 62 102 contigs greater than 300 bp in length, and incorporated 91% of all initial reads. The final library had N50 = 1064 bp, with the largest contig = 11 942 bp (Figure. 3.1), while the iterated *de novo* assembly produced 72 915 contigs with N50 = 1039. While the PE R, *carpaticola de novo* assembling incorporated 89% of the PE reads the 4 SE libraries results in 50 % of unassembled reads. Hence the pooled *de novo* assembly was kept for subsequent analyses.
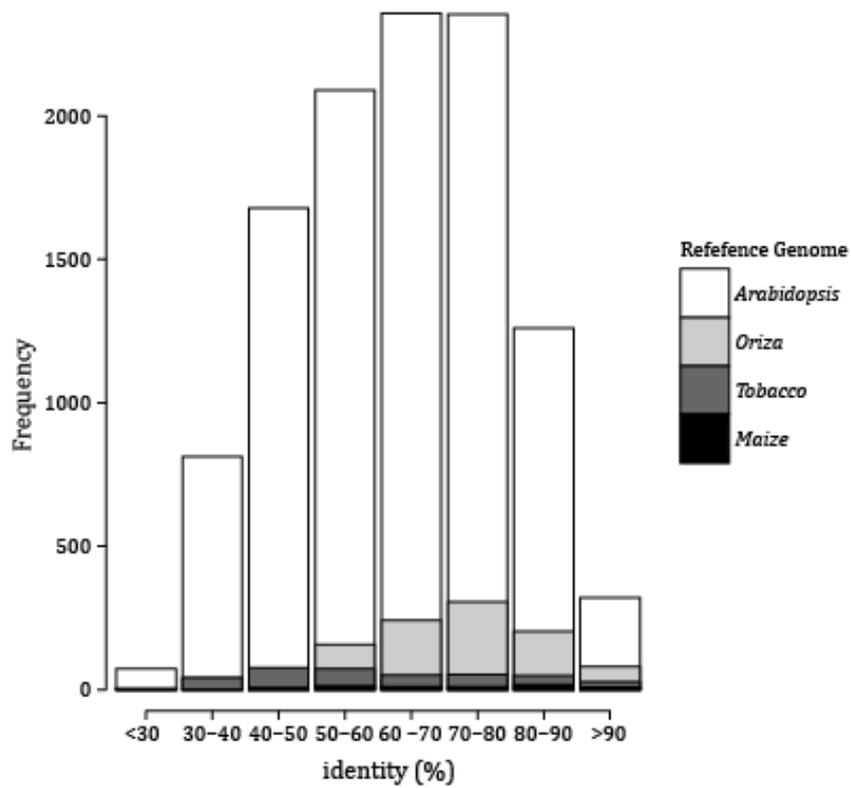
Figure 3.1 Frequency distribution of blasted sequences and their similarity to known genera in %

Table 3.5 Numbers of high-quality SNPs and InDels called from *Ranunculus* RNAseq data

Genotype	SNPs	InDel
REV1	323 622	12 937
TRE(Apo)	34 361	1 564
VRU2(Apo)	30 785	1 360
YBB1	25 976	1 267
NOT	32 684	1 604
Total	447 428	18 732

In total 19 977 contigs could be annotated, with the maximum number of homologies found with *Arabidopsis thaliana* (56 %,), *Oryza* (7.6 %), and *Nicotiana* (2%). The percentage identity ranged from 20% to 100% (figure 3.1). A total of 447 428 high quality SNPs and 18 732 indels were mined from the libraries (Table 3.3). Of these, 73% were detected in the PE *R. carpaticola* (REV1) library, due to the significantly higher number of reads for this sample. The greater number of polymorphisms in the *R. carpaticola* library did not represent a bias as only shared polymorphisms between libraries were considered in downstream analyses.
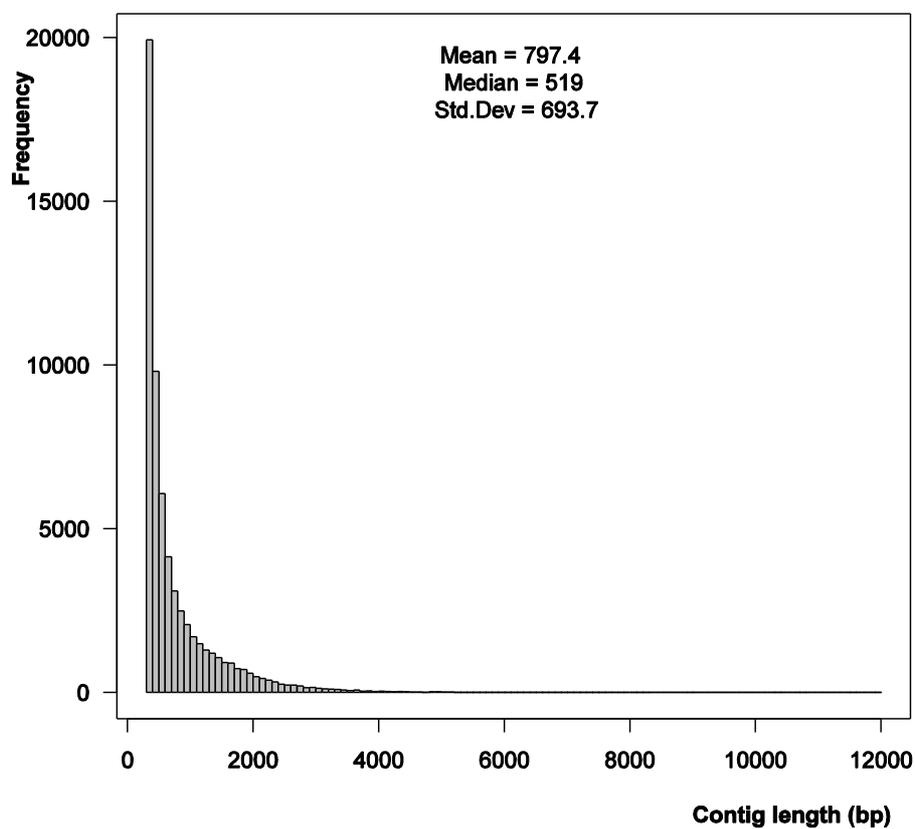
Figure 3.2 Contig length distribution from *de novo* assembly of RNAseq reads from 5 *Ranunculus* cDNA libraries

3.4.2. SNP variation supports Pleistocene hybrid origin of apomictic *Ranunculus*

A total of 1677 characters with SNPs shared by at least two taxa (1408 single-base or ambiguous sites, the rest short-sequence polymorphisms) were identified and used to generate a well-supported hybridization network having a Fit value of 100% and 100% bootstrap support for all nodes (Figure 3.2). Using *R. notabilis* as the outgroup, the analysis shows that the two apomictic genotypes are closely related to one another, and their intermediate relationship between *R. cassubicifolius* and *R. carpaticola* supports the hypothesis of hybrid origin from these two taxa (Figure. 3.2). The hybridization network suggests a single origin of the hybrid with subsequent divergence of the two apomictic lineages. The apomict VRU1 shares more SNPs with *R. cassubicifolius* while the apomictic lineage TRE was closer to *R. carpaticola* (Figure. 3.2). Branch lengths of the hybridization network of the sexual parental taxa slightly exceed that of apomictic derivatives, which is consistent with an earlier evolutionary origin and divergence in allopatry. Divergence of the two apomictic accessions is higher than that between origin and split of the two accessions.

A relatively small number of mutations (and higher than those of the sexual reproducing parents) were accumulated in the asexual hexaploids after hybridization (Table 3.4). The estimated number of generations needed to reach the observed accumulation of multiple SNPs per locus suggests an approximate evolutionary origin of the hybrids to be maximally around 80 000 years before present.
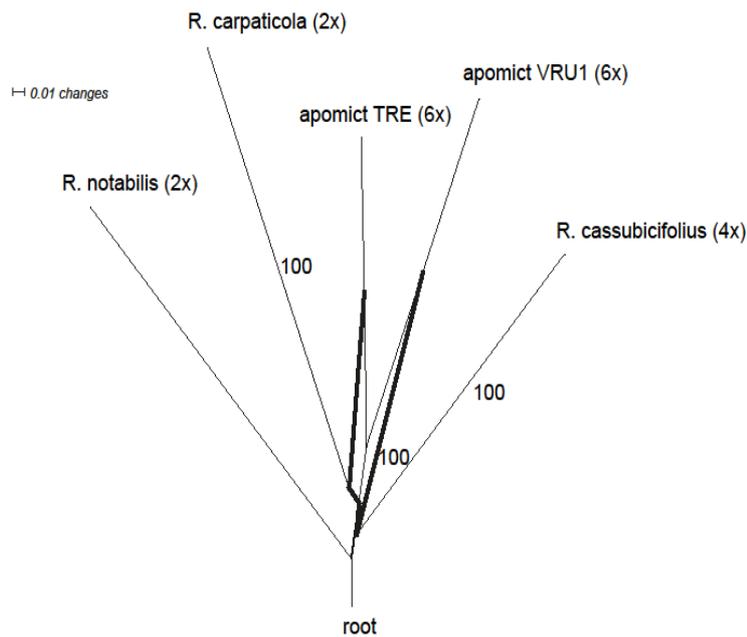
Figure 3.3 Hybridization network (Splitsgraph) of five *Ranunculus* based on 1477 phylogenetically informed SNPs, with *R. notabilis* defined as outgroup (Fit = 100%). Bootstrap percentages are indicated beside branches. Lines in bold indicate the hybrid relationship of the apomictic lineages reconstructed from SNPs shared by apomicts and parents. The numbers in parenthesis indicate the ploidy levels.

3.4.3. Non-Synonymous / Synonymous mutation rates in sexual and apomictic genotypes

The analysis of non-synonymous to synonymous mutation rates were performed on 1231 genes which fulfilled the following criteria: the genes could be annotated, the ORFs could be assigned, and high-quality SNPs could be measured and compared for each genotype across the length of the complete gene. The overall distributions of pairwise comparisons of the dN/dS ratios for apomictic, apomictic and sexual, and sexual genotypes were strikingly similar, with the exception of outlier values (Figure. 3.3). The number of genes showing outlier dN/dS values increased from apomict (n = 163), to sexual (n = 373), to apomictic-sexual (n = 386) comparisons.

A gene ontology (GO) analysis of gene enrichment for outlier values (Figure. 3.3) in all 3 comparisons, using all *Ranunculus* genes which could be annotated (n = 1846) as custom reference set, showed no statistically significant results. Interestingly, genes with elevated outlier dN/dS ratios exclusively found in the apomictic-sexual comparison were associated with processes involved in meiosis and gametogenesis (Figure. 3.4). Moreover the proportion of GO terms associated with reproduction was

50

significantly higher in the apomictic-sexual comparison (Table 3.5). The outliers in the apomict-apomict comparison were associated with "metabolic" and "cellular" GO terms. Finally the comparison between sexuals yielded "metabolic process", "cellular processes" and "response to stimulus" as the three most representative terms (figure 3.6 & 3.7).

Table 3.6 Evolutionary divergence times as calculated from differential mutation accumulation observed between clonal hexaploids *R. carpaticola* × *cassubicifolius* genotypes and their putative parental sexual species

Genotype	mutation rate (m)*	Loci with heterozygous SNPs**	number of generations***	Time (years)
TRE	9,1485e-5	3	20000	60000
VRU2	9,1485e-5	4	26667	80000

*= expected frequency of mutation/substitution per generation

= accumulated after the hybridization event*= number of generations needed to reach the observed number of loci with heterozygous SNPs when lack of meiosis is assumed

Table 3.7 Number of GO terms associated with reproduction in comparisons of genes showing outlier values of dN/dS.

	Comparison	Reproductive GO	Other GO	Total
Apo-Apo	1	6	157	163
Apo-Sex	2	41	345	386
Sex-Sex	3	15	358	363

Fisher exact test of comparisons: 1-2 (P = 0.007), 2-3 (P= 0.0004), 1-3 (P= 1);
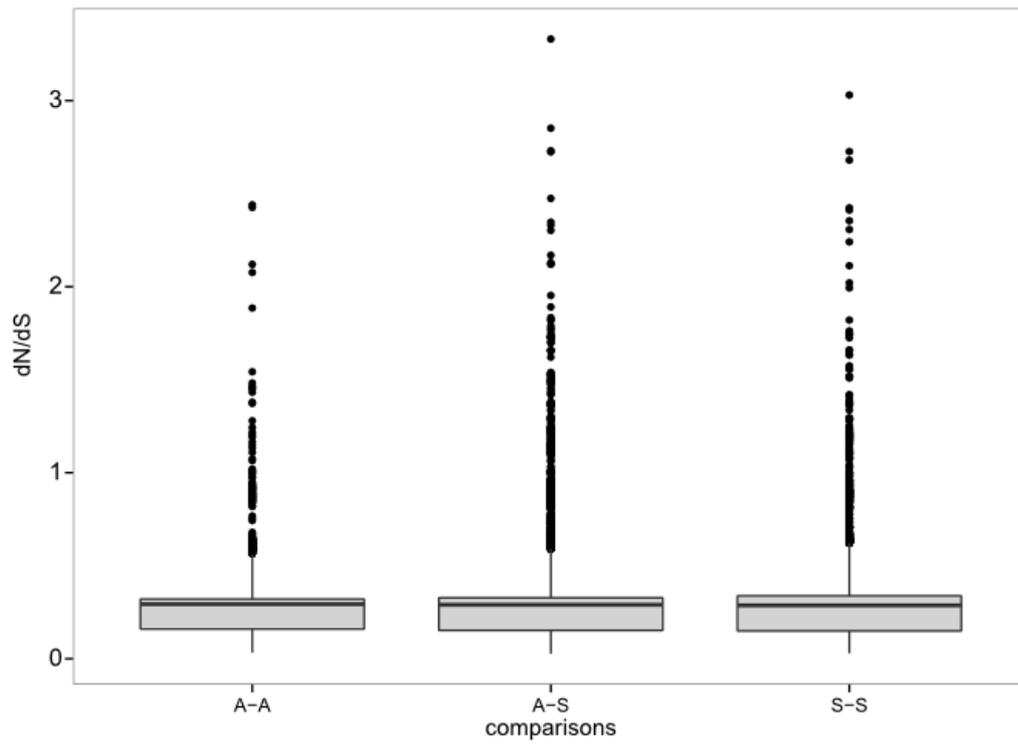
Figure 3.4 Boxplot distributions of pairwise non-synonymous (dN) to synonymous (dS) nucleotide substitution ratios for 574 ORFs which shared at least a single SNP at one common position across all *Ranunculus*. The three distributions represent apomict-apomict (A-A), apomict-sexual (A-S) and sexual-sexual (S-S) pairwise comparisons (circles are outliers values).

3.5.Discussion

Whole transcriptome EST sequencing and gene annotation in non-model organisms, once the extremely expensive products of coordinated efforts between multiple laboratories, are becoming standardized analytical pipelines heralded by rapid advances in NGS and bioinformatics technology (Bajgain *et al.*, 2011; Novaes *et al.*, 2008). Here, we have taken advantage of the high coverage offered by one NGS platform (Illumina HiSeqTM 2000 Sequencing System) to generate and assemble the first transcriptome in the *Ranunculus auricomus* complex, a naturally-occurring biological system for studying apomixis, polyploidy, and hybrid genome evolution. Using a number of methods for qualification and quantification of DNA sequence polymorphisms (see methods), we have furthermore generated a robust dataset of high-quality SNP and indel polymorphisms in order to compare five *Ranunculus* genotypes characterized by sexual and apomictic reproduction, ploidy variation, and hybridization (Table 3.1).

3.5.1. Sequencing strategy, a balance between costs, goals and natural history

The decision to generate both SE and PE libraries from the different *Ranunculus* taxa being analyzed here was based upon consideration of costs (*i.e.* SE libraries are cheaper than PE) and the fact that this genome is uncharacterized (*i.e.* PE libraries can be more successfully *de novo* assembled). We analyzed RNA rather than DNA as we would expect the highest amount of information on effects of mutational load in the coding regions. Furthermore, we used an RNA normalization step, as our ultimate goal was to maximize genic (and allelic) sequence representation for their use in designing gene expression and CGH (comparative genome hybridization; see Aliyu *et al.*, in review) microarrays for future experiments. Working within a finite project budget, this approach has had both advantages and disadvantages.

As *R. carpaticola* is the only diploid progenitor in the hybridization event being studied here (Paun *et al.*, 2006a; Table 3.1), it was chosen for the generation of the PE library since we assumed that (1) it would be characterized by ancestral polymorphisms of the whole "*cassubicus*" group, (2) that the complexity of sequence information resulting from ploidy variation would be minimized, and (3) having a high

coverage from one of the parents would increase the confidence level when searching new single nucleotide variations after hybrid origin. Indeed, the use of a *de novo* assembly strategy whereby trimmed reads from all libraries were pooled to generate a backbone sequence to which each individual library was reference assembled, enabled the assembly of 91% of the original quality-trimmed sequencing reads (Table 3.2). The high proportion of removed reads included reads filteredby size (short reads), quality threshold, and vector contamination (*e.g.* illumina adaptors and those from the cDNA normalization procedure). This has naturally led to a reduction of the gene coverage and SNP calling, but the point of this strict set of criteria was to identify high quality polymorphisms for subsequent analyses. While most (73%) of the high-quality polymorphisms were identified in the *R. carpaticola* (PE) library, only shared polymorphisms between libraries were considered in our subsequent analyses, and hence this did not represent a bias in our interpretations.

Homology to known sequences was found for about one third of the total contig number (19 977), which is likely a reflection of genomic divergence of *Ranunculus* over millions of years to the quite distantly related reference species (*Arabidopsis*) for which significant sequence information exists. *Ranunculus* belongs to basal eudicots which are relatively underrepresented in sequence databases, and thus our data may fill an important gap for understanding of transcriptome evolution in angiosperms. Alternatively, the presence of orphan genes in the genus, as have been found in other plant species with relatively little transcriptome information, for example *Artemisia tridentata* (Bajgain *et al.*, 2011) and *Epimedium sagittatum* (Zeng *et al.*, 2010), could additionally contribute to the low genomic similarity of *Ranunculus* to other databases.

3.5.2. The resolving power of SNPs

A high number of high-quality SNPs were detected among all libraries, based upon their high coverage and quality control filtering applied by the vcfutils module of bcftools. Because of the low number of individuals selected for the analysis, we cannot exclude that a fraction of the SNPs represent individual rather than population specific polymorphisms. Nevertheless, considering the high number of SNPs, a sample of just five individuals representing four taxa comprised sufficient information for the reconstruction of their evolutionary history. In comparison, previous population genetic approaches using AFLPs required 450 individuals out of the complete range of the species to reconstruct the ancient hybridization event between *R. carpaticola* and *R.*

cassubicifolius (Paun *et al.*, 2006a) but could neither resolve timing of events nor divergence between apomictic accessions. Our study shows that large amounts of SNP data allow not only for a reliable reconstruction of phylogenetic relationships with a few individuals, but also provide more information on evolutionary history (*e.g.* mutation accumulation). Moreover, AFLP markers could not provide information on genetic background, and DNA sequence analyses of standard plastid and nuclear markers from a few individuals could not resolve taxonomic relationships (Hörandl *et al.*, 2009). SNP analysis is thus powerful for studying evolutionary processes in groups with young evolutionary histories and shallow phylogenies, even on a minimal number of samples.

3.5.3. Divergence and evolutionary origin

After a single hybrid origin, the apomictic lineages diversified rapidly from each other and in comparison to their sexual progenitors. The branch lengths of apomicts and sexuals are almost the same despite a much younger evolutionary age (max. 80 000 years, while diploid sexuals diverged c. 900 000 years ago; Hörandl, 2004); hence, divergence of apomicts must have happened at a faster rate. This effect is even underestimated in the hybridization network, as only shared SNPs (*i.e.* without autapomorphic SNPs) were considered for this analysis (Figure. 3.3). Analysis of SNPs exclusive to apomicts provided information regarding divergence times, and places the hybridization event reliably into the last glacial period. The method is thus powerful for dating of shallow phylogenies which is otherwise difficult without a fossil as calibration point (*e.g.* Forest, 2009). Overall our age estimate is concordant with the general divergence time estimates of the *Ranunculus* phylogeny (Emadzade & Hörandl, 2011), where the R. *auricomus* complex is placed on terminal nodes (Emadzade *et al.*, 2011). The age estimate for origin of hybrids is younger than the divergence time of sexual parents of the hybrids (c. 0.3 Mill. years; see Hörandl, 2004). Our dating confirms the general assumption that apomictic plant lineages are evolutionarily young and associated with the Pleistocene

In sexually reproducing, outcrossing diploids, allelic variation is re-shuffled within a population to maintain a balance of homo- and heterozygous individuals. Divergence between alleles in the same individual cannot accumulate, being inhibited by the homogenizing effect of homolog pairing and recombination at meiosis. In

contrast, diploid apomictic lineages characterized by decreased frequencies of meiotic crossing-over and chromatid interchange are expected to gradually accumulate inter- and intra-individual allelic sequence variation as a result of independent mutation accumulation. The resultant inter-allelic divergence is known as the Meselson effect, and has been observed in Bdelloid rotifers, with some evidence demonstrated in apomictic *Boechera* (Corral *et al.*, 2009; Welch & Meselson, 2000). Interestingly, in parthenogenetic Bdelloid rotifers divergent alleles at the same locus can acquire distinct but related functions which produce proteins which act synergistically (Pouchkina-Stantcheva *et al.*, 2007). Whether functional divergence between alleles similarly occurs in young asexual *Ranunculus* lineages remains an open question.

Our approach to identify intra-individual allelic divergence at single loci, and its correlation with apomixis in *R. auricomus*, provides evidence that apomictic plants can accumulate Meselson-effect-like changes. In the absence of gene conversion, highly efficient DNA repair or other homogenizing mechanisms, in addition to disturbed DNA repair mechanisms associated with homologous recombination during meiosis, mutations should accumulate faster in apomicts despite low levels of sexual reproduction (Ceplitis, 2003). We suggest that the observed number of divergent gene copies exclusively found in allohexaploid apomicts can be attributed to post-hybridization mutational events associated with the young evolutionary history of this taxon.

The rapid accumulation of changes within a short time period is most likely enhanced by polyploidy, as on the one hand more possible mutational sites are available, while on the other hand the availability of multiple gene copies would mask the effect of deleterious recessive mutations, and consequently, masked mutations could reach higher frequencies before elimination by selection. As neopolyploids, apomictic lineages would benefit from masking effects of deleterious mutations while they do not yet suffer from a highly accumulated mutational load (Gerstein & Otto, 2009). However, detailed models of such effects on hexaploids are still not available. Polyploidy might explain that, in spite of their evolutionarily young age, genomic divergence was observed in the hexaploids analysed here, in contrast to ancient diploid asexual animals (*e.g.* mites) for which no signs of the Meselson effect could be found (Schaefer *et al.*, 2006). However, considering that Oribatid mites reproduce via

automixis and inverted meiosis (Heethoff *et al.*, 2009), genomic evolution in these ancient asexuals may not be directly comparable to plants.

Whether the accumulation of mutational changes measured here reflects positive or negative fitness effects, however, cannot be directly inferred from our dataset. Nonetheless, divergent selection interestingly occurs in genes associated with the cell cycle (see below), an observation which suggests a selective process associated with the shift from sexual to apomictic reproduction.

3.5.4. dN/dS ratios as signals for divergent selection

Considering the dearth of genomic information related to the genus, we approached the estimation of the differences in number of synonymous and non-synonymous mutations between sexual and asexual *Ranunculus* similarly to what has been previously done in other non-model species (*i.e. Eucalyptus grandis*; Novaes *et al.*, 2008). We focused on those ORFs that shared at least one SNP across all the lineages. While this approach has the drawback of loosing information from rare alleles and hotspots of genetic diversity, it minimized false positives due to alignment errors.

The analysis of dN/dS frequencies in pair wise comparisons between sexual and asexual genotypes revealed similar distributions, in addition to 345 outliers within and between reproductive groups (Figure. 3.3). While no significant enrichment was discovered in the GO analysis of outlier genes within each comparison, outliers between different comparisons of apomicts and sexual showed a significant enrichment of genes associated with reproduction (Table 3.5). Considering misregulation of sexual genes as the hypothesized induction mechanism for apomictic seed formation, a hybridization event which brings together relatively divergent alleles (and their associated regulatory factors) of genes associated with reproduction is consistent with the enrichment of such genes with outlier dN/dS ratios in the sex-apomict comparison.

biological_process Level 2

GO-id	GO-term
GO:0051327	M phase of meiotic cell cycle
GO:0000279	M phase
GO:0007128	meiotic prophase I
GO:0007127	meiosis I
GO:0007129	synapsis
GO:0051324	prophase
GO:0009554	megasporogenesis
GO:0070192	chromosome organization involved in meiosis
GO:0009556	microsporogenesis
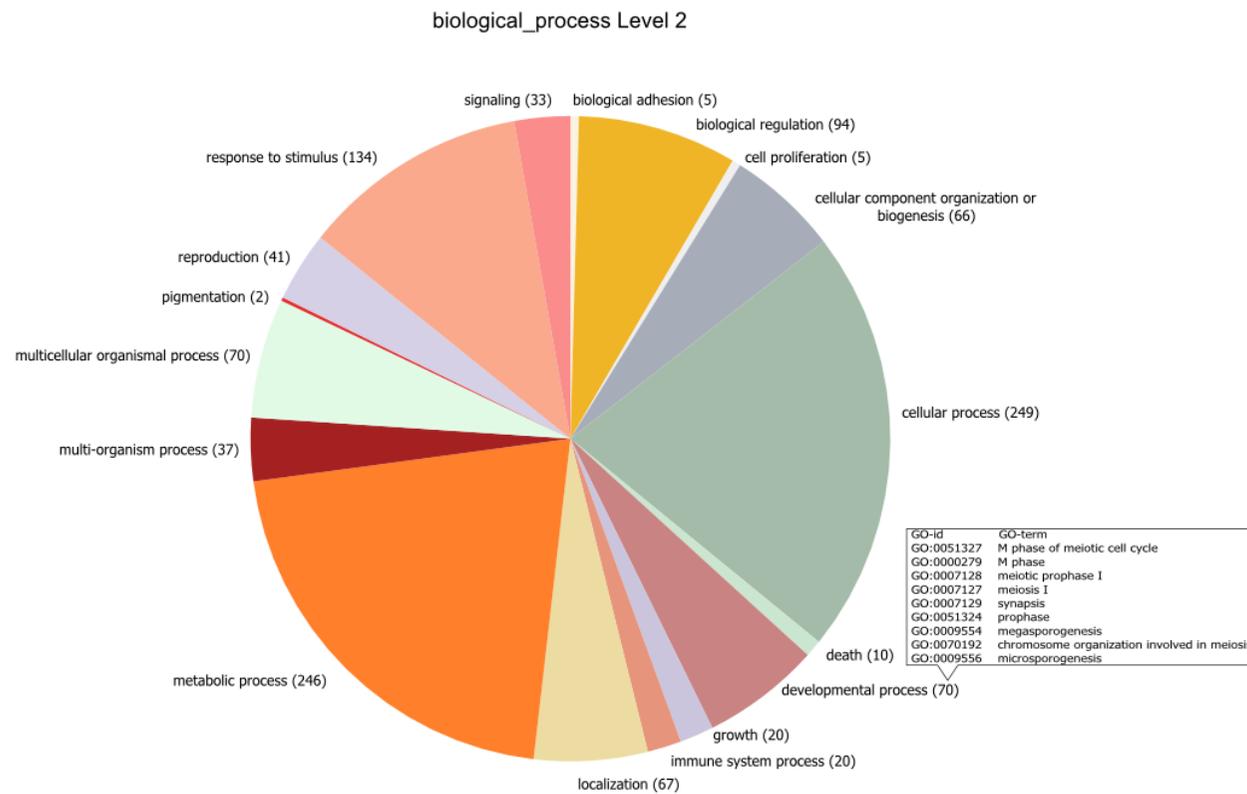
Figure 3.5 Pie chart of second level GO terms (Biological process) of genes showing outlier dN/dS ratios in the apomict-sexual comparison. Numbers in brackets refer to the number of GO terms in each category. Terms listed in the developmental process bubble cloud represent level 5 to 7 GO terms.

58
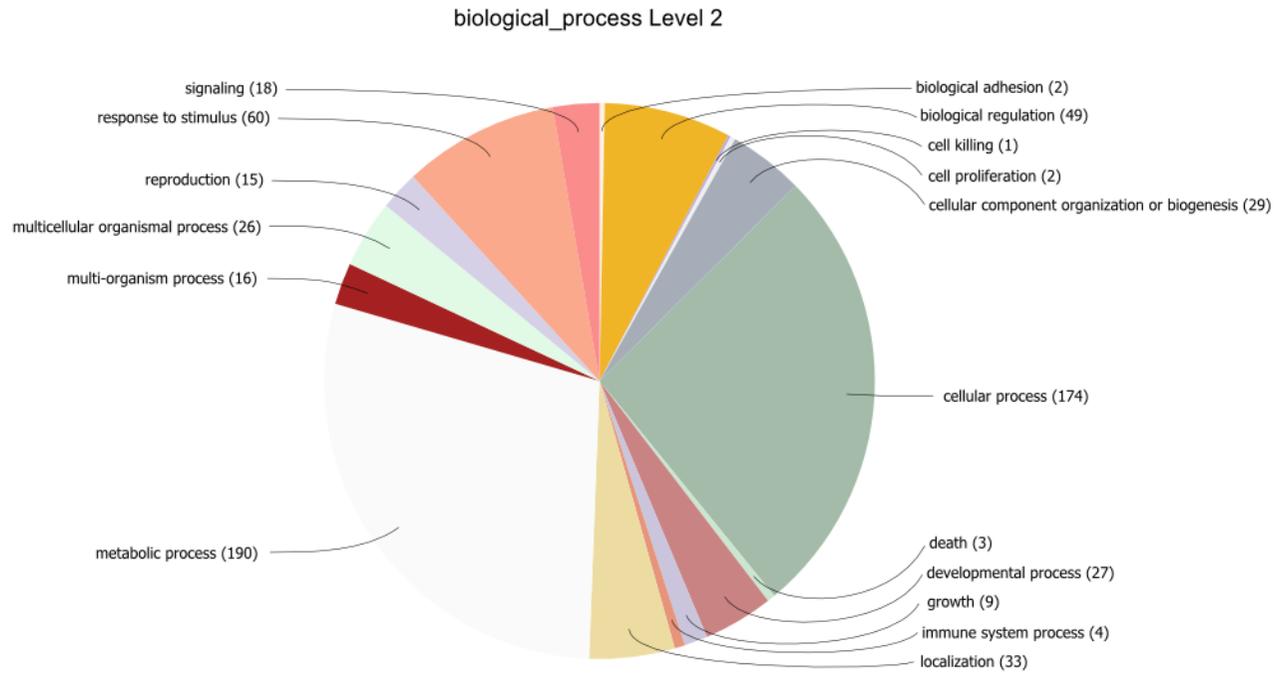
Figure 3.6 Pie chart of second level GO terms (Biological Process) of gene showing outliers dN/dS ratio in the sexual-sexual comparison. Number in the bracket refers to the number of terms in each category
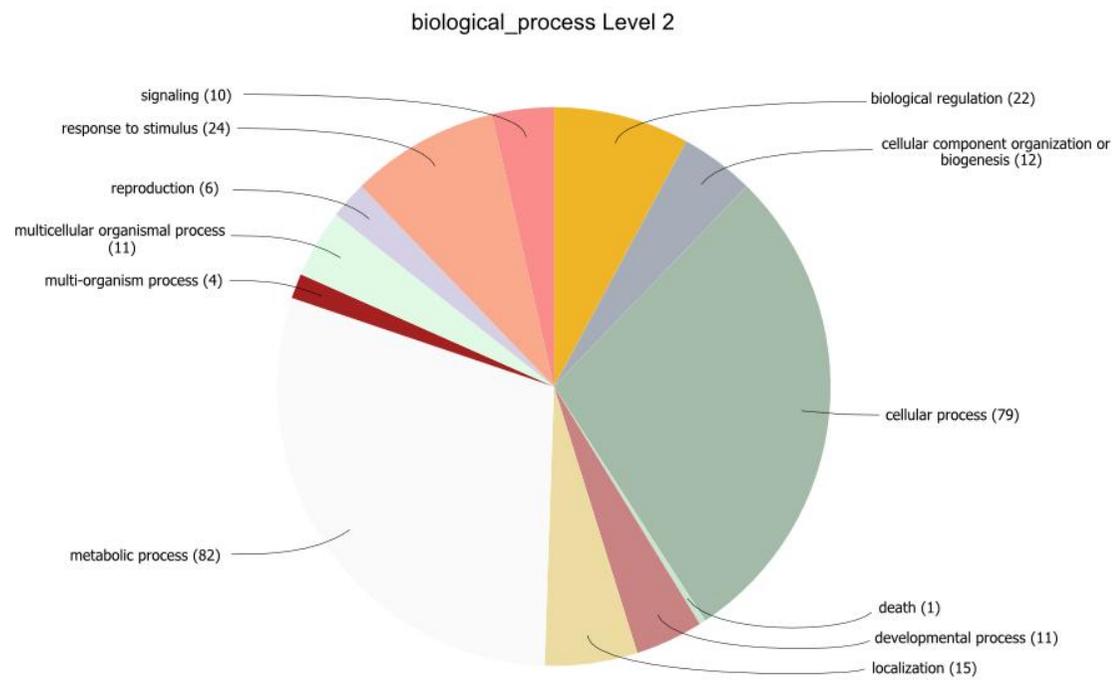
Figure 3.7 Pie chart of second level GO terms (Biological Process) of gene showing outliers dN/dS ratio in the apomictic-apomictic comparison. Number in the bracket refers to the number of terms in each category

Interestingly, outlier genes associated with six meiosis and gametogenesis terms were identified (Figure. 3.4), one of which was homologous to the ASY1 protein from *Arabidopsis* (E-value: 8e-24 with 45% identity). ASY1 is homologous to the *Saccharomyces cerevisiae* HOP1gene (Caryl *et al.*, 2000), which is associated with meiosis, and promotes synapses formation preceding meiotic recombination. HOP1 mutants in yeast show reduced levels of meiotic recombination (Hollingsworth & Byers, 1989) while the mutant for ASY1 in *A. thaliana* showed an almost complete absence of homologous chromosome synapsis during male meiosis (Ross *et al.*, 1996). Indeed, lack of chromosome pairing characterizes sterility and low pollen viability in aposporous apomictic Cassava (Nassar, 2001) as well as other apomictic taxa, included *Ranunculus* and *Paspalum* (Hörandl, 1997; Hörandl, 2008; Izmailow, 1996; Podio *et al.*, 2012). These meiotic disturbances are thought to be a consequence of hybridization and/or polyploidy (Izmailow, 1967; Paun *et al.*, 2006a), while the functional connection to apomixis remains to be studied.

The finding of divergent selection (ie. genes with high dN/dS ratios), albeit in a small number of samples, supports the hypothesis that in the short term, apomixis might be induced or enhanced by selection on particular loci. For example, the hexaploid apomictic lineages are geographically more widespread, occupy more different niches and are more abundant than the sexual progenitors (Paun *et al.*, 2006a; Hörandl *et al.*, 2009), and hence divergent selection could reflect a heterosis like effect in the hybrid-apomicts which is correlated with their evolutionary success. On the other hand, the sexual-apomictic comparison additionally revealed a number of genes with low dN/dS ratios, a result which points to stabilizing selection acting on otherwise highly heterozygous genotypes. Taken together, the mutations observed in these young polyploids may have not necessarily negative effects on fitness, as would be predicted by Muller's ratchet. Divergence among apomictic lineages, as observed in our SNP dataset, might reflect the ecological potential of lineages (*i.e.* niche specialization) as observed in natural populations (Paun *et al.*, 2006b). Therefore, mutation accumulation does not necessarily confirm the doomed view of Stebbins (1950) that asexual plant lineages would rapidly go to extinction and would represent dead ends of evolution. An in-depth analysis of gene families, their evolution and their functions is required to understand the actual effects of mutation accumulation on evolutionary histories.

CHAPTER 4: CHASING THE APOMICTIC FACTOR IN *RANUNCULUS*. REDUCING THE COMPLEXITY OF REPRODUCTION, PLOIDY AND HYBRIDIZATION EFFECT

4.1.Introduction

The evolution of asexuality from sexual progenitors has occurred repeatedly and independently in a wide range of plants and animals (Barton and Charlesworth, 1998; Mittwoch, 1978; Suomalainen, 1950), a shift whose evolutionary consequences have been the subject of many studies exploring factors related to the evolutionary success of asexuality (*e.g.* its reiterated occurrence across time and species). It has been suggested that asexual reproduction is advantageous because of the avoidance of the twofold cost of sex (Maynard Smith, 1978) and because of a stronger adaptability to stable environments (Bell, 1982). On the other hand limitations arising from the lack of meiotic recombination are expected, for example with respect to the fixation of advantageous mutations (Crow and Kimura, 1965) and in meiotic recombinational DNA repair mechanisms (Hörandl, 2009). In addition, asexuals are expected to accumulate deleterious mutations in a non-reversible ratchet-like fashion (Muller, 1964), a process which accelerates in small populations (*i.e.* mutational meltdown; Lynch *et al.*, 1993). In contrast, sexuality could be advantageous in changing environments, or for the production of genetic variance (Fisher, 1930). Sexual recombination furthermore facilitates the purging of deleterious mutations (Kondrashov, 1982; Muller, 1964) and, in a similar fashion, increases the probability of recombining advantageous mutations from two parents into a single offspring.

Asexuality in plants, or apomixis, is a naturally-occurring form of clonal reproduction through seed (Nogler, 1984a). Apomixis has been described in more than 293 genera scattered throughout angiosperms (Carman, 1997 Hojsgaard *et al.*, 2014a). Several types of apomixis have been characterized, as defined by the origin of the meiotically-unreduced megagametophyte. In all classifications three developmental steps must sequentially occur to produce a functional clonal seed: (1) the development of a meiotically unreduced egg cell (apomeiosis), (2) the development of the egg cell without fertilization (parthenogenesis), and (3) the production of a functional endosperm with (pseudogamy) or without (autonomous) fertilization (Koltunow and Grossniklaus, 2003).

Polyploidy and hybridization are common traits of almost all apomictic plants (Koltunow and Grossniklaus, 2003), and as such both have been hypothesized to be triggers for the switch in reproductive mode. For example, it has been proposed that de-regulation of the normal sexual pathway, leading to apomixis induction (Carman, 1997), could arise via the chromosomal (Kantama *et al*., 2007; Sánchez-Morán *et al*., 2001), genomic (Pikaard, 2001; Soltis and Soltis, 1999), and transcriptomic (Adams, 2007; Adams *et al*., 2004) global gene regulatory changes that accompany polyploidization and hybridization. In this light, heterochronic gene expression (*i.e.* deregulation) between sexual and apomictic reproductive tissues has been shown in *Tripsacum* spp. (Grimanelli *et al*., 2003) and *Boechera* spp. (Sharbel *et al*., 2010), observations which support the hypothesis of apomixis induction by (or associated with) a shift in gene expression through time (Grossniklaus, 2001; Koltunow, 1993).

The *Ranunculus auricomus* complex is becoming a model for studying genomic differences between sexual and apomictic taxa. Hundreds of species are grouped into two morphologically distinct subcomplexes: *R. auricomus* and *R. cassubicus* (Hörandl *et al*., 2009). The "*cassubicus*" group is composed of sexual diploid and autotetraploid species, and a few apomictic polyploids. The closely related diploid sexual species *R. carpaticola*, and diploid and autotetraploid *R. cassubicifolius* extend from Switzerland to west Slovakia (Hörandl *et al*., 2009), while hexaploid apomicts are located in central Slovakia. Analyses based upon AFLPs and SSRs, combined with DNA sequence (Hörandl *et al*., 2009; Paun *et al*., 2006) and SNPs (Pellino *et al*., 2013), support a hybrid origin for hexaploid apomicts sometime during the last glacial period (60 000-80 000 years ago). Apomictic *Ranunculus* are aposporous ("*Hieracium* type"), pseudogamous and express apomixis facultatively (Hörandl, 2008; Izmailow, 1965; Nogler, 1984b, Hojsgaard *et al*. 2014b). The aposporous initial (AI) cell arises from a somatic cell of the nucellus during the megaspore tetrad stage, crushing the megaspore mother cell (MMC) as it enlarges. Endosperm development requires fertilization from a meiotically reduced male gamete (pseudogamy) (Izmailow, 1965).

In the framework of an SNP analysis comparing sexual and asexual *Ranunculus* genotypes using Illumina RNAseq data (Pellino *et al*., 2013), we have developed a custom expression microarray that was used to transcriptionally profile ovules microdissected at four developmental stages (ranging from pre- to post-MMC) from four sexual and three apomictic *Ranunculus* genotypes. Our experiment was designed

to overcome expected difficulties of comparing gene expression patterns between the differing ploidies characteristic of sexual and apomictic *Ranunculus*, through the analysis of multiple genotypes (*i.e.* biological replicates). The goals of this study were to (1) identify a list of transcripts showing differential expression between apomictic and sexual ovules, (2) quantify heterochronic gene expression patterns during apomictic and sexual ovule development, and (3) partition transcripts into groups whose expression signatures reflect hybridity, parent or origin effects, or polyploidy.

4.2. Materials and methods

4.2.1. Custom microarray development

The RNAseq approach used to generate the data for microarray design is explained in detail in Pellino *et al*. (2013). In short, flowers were collected from two sexual and two apomictic genotypes of the *Ranunculus auricomus* complex (Table 4.1). For each individual, total RNA was isolated from pooled flowers of five different sizes (one to five mm in length, from an early stage when sepal primordia have enclosed the floral meristem, to the fully developed unopened flower stage) using the Qiagen RNeasy Plant Mini Kit (QIAGEN, www.qiagen.com). After isolation possible DNA contamination was removed using Qiagen RNase-Free DNase, while contamination from the DNase enzyme, polysaccharides, and proteins were removed with a second purification step using the Qiagen RNeasy Mini Kit. In all purification steps the manufacturer's instructions were followed.

To avoid over-representation of highly transcribed genes during subsequent sequencing steps, full length normalized cDNA libraries were produced, whereby each step of the normalization was performed and optimized following the procedures described by Vogel *et al*. (2010) and Vogel & Wheat (2011).

15 μl of normalized cDNA (200ng/ μl) was sent to Fasteris (www.fasteris.com) for RNAseq, whereby a dual-sequencing approach (54-mer single read (SE) and 108-mer pair-end read (PE)) was chosen in order to balance cost and efficiency of a de-novo assembly (see Pellino *et al*., 2013). Both sequencing strategies were conducted using the HiSeqTM 2000 sequencing system (http://www.illumina.com/).

CLC Genomics Workbench (CLC bio version 4.9, www.clcbio.com) was implemented for the sequencing assembly. At first sequences were trimmed for vector

64

contamination, length and quality score using CLC default values. As no reference genome for *Ranunculus* is available, *de novo* and iterated *de novo* approaches were used to assemble the data. For the de novo assembly, all libraries were pooled and assembled using CLC default parameters. In the iterated *de novo* assembly, each library was assembled individually, and the resultant contigs from each individual assembly re-assembled together, including all unassembled reads from each individual assembly. In both strategies the final assembly was trimmed for contigs shorter then 300bp. The two approaches were evaluated based on the total number of matching reads (i.e reads that could be assembled into longer contigs) and N50 values, and the *de novo* assembly was selected for array design. Both contigs and singletons were forwarded to Roche NimbleGen for design and manufacture of a custom 1.4K Microarray.

4.2.2. Transcriptomal profiling of sexual and apomictic ovules

a) Sample selection, ovule microdissection, and RNA extraction

Plants were grown from seedling to pre-flowering stages in outdoor plots at the *Leibniz Institute of Plant Genetics and Crop Plant Research* (IPK), and were then moved into a phytotron for flowering (day: 16 h, 21°C; night: 8 h, 16°C; humidity 70%). Based on the cytological observations of Hojsgaard *et al.* (2014)*,* ovules at four developmental stages were collected at the pre-meiotic, megasporogenesis, tetrad stage/aposporous initial and functional embryo sac stages (I,II,III,IV;) at standardized times (between 7 and 9 AM) over multiple days from both sexual and apomictic *Ranunculus* under a sterile laminar-flow hood with a stereoscopic microscope (100 Stemi; Carl Zeiss). Using a sterile scalpel and forceps to open the flower, carpels were collected and immediately immersed in a sterile 0.55M mannitol solution and placed on ice.

In a second step, microdissection of each single carpel was conducted in a in a sterile laminar air-flow hood under an inverted microscope (Axiovert 200M; Carl Zeiss) using self made sterile glass needles (made with a Narishige PC-10 puller). Individual ovules were collected using an Eppendorf Cell Tram Vario connected to a 150 μm inner diameter glass capillary, immersed in 100 μl RNA-stabilizing buffer (RNA later; Sigma), and immediately frozen in liquid nitrogen. 20 to 40 ovules for each genotype/stage were dissected and stored at -80°C. RNA extraction was carried out using a PicoPure isolation kit (Arcturus Bioscience) and quantification and quality were assessed with RNA Pico chips on an Agilent 2100 Bioanalyzer (Agilent Technologies).

Table 4.1 *Ranunculus* samples used in the experiment

Taxon	Sample	Number of individuals	Ploidy	Reprod.	Locality	Collectors and vouchers
R. carpaticola	REV1	3	2x	Sex	Slovakia, Slovenské rudohorie,Revúca, hill Skalka (forest)	Hörandl, 8483, 01.05.1998 (WU)
R. carpaticola x cassubicifolius	TRE	1	6x	Apo	Slovakia, Strázovské vrchy (near Trençín), between Kubra and Kubrica, close to the bus-stop Kyselka (margin of Carpinus forest and meadow)	Hörandl, Paun, Mládenková, s.n., 30.04.2004 (SAV)
R. carpaticola x *cassubicifolius*	VRU 2	2	6x	Apo	Slovakia, Turçianska kotlina,Vrútky-Piatrová, behind cottage (meadow)	Hörandl, Paun, Mládenková, s.n.,01.05.2004 (SAV)
R. cassubicifolius	YBB 1	1	4x	Sex	Austria, Lower Austria, Wülfachgraben, SE Ybbsitz (forest)	Hörandl, 8472 12.04.1998 (WU)

WU: herbarium of the University of Vienna (Universität Wien), SAV: herbarium of the Institute of Botany, Slovak Academy of science, Bratislava. For ploidy identification see Hörandl & Greilhuber 2002 and Paun *et al.* 2006b.

b) cDNA synthesis and amplification

For each of the samples from the 7 *Ranunculus* individuals (Table 4.1), cDNA synthesis and amplification was conducted using the Sigma TransPlex Complete WTA2 kit (www.sigmaaldrich.com) following the producer's instructions. Amplified cDNA was purified using the GeneEluteTM PCR Cleanup kit following the manufacturer's protocol, and concentration and quality was measured with a NanoDropTM 1000 Spectrophotometer (www.nanodrop.com).

c) Microarray hybridization and data processing

Twenty-eight 1.4M probe custom microarrays were outsourced to Roche NimbleGen (www.NimbleGen.com) for design and manufacture, each of which was used for an independent sample labeling and hybridization reaction for each of the 28 microdissected ovule samples (four stages for seven genotypes; Table 4.1) following the NimbleGen microarray labeling protocol of their One-Color DNA Labeling Kit. The labeled samples where then individually hybridized in random order using the NimbleGen Hybridization System 12 to the custom *Ranunculus* arrays according to the producer's instructions, and scanned using NimbleGen MS 200 Microarray Scanner at 535 nm. Feature intensities were extrapolated using the DEVA software (version 1.1, Roche NimbleGen), and the raw expression data were normalized together using the DEVA implemented robust multiarray average (RMA) algorithm.

The normalized data were analysed using the Qlucore Omics Explorer software (version 2.3; www.Qlucore.com). Principal Component Analysis (PCA) and Qlucore filtering by variance was implemented, whereby contigs with variation (σ / σ max) <0.7 between apomictic and sexual groups were removed from the dataset. This threshold was chosen, using the Qlucore software, such that the differentially expressed apomictic and sexual probe groups could be visually separated on a PCA, while at the same time retaining the maximum number differentially expressed genes. Second, the resultant set of differentially expressed genes were ranked according to p-value (< 0.01) using a paired t-test between the apomictic and sexual groups at each stage separately. In addition to the p-value filter, log2 fold change >2 and adjustment of the p-value for multiple tests using a false discovery rate (FDR) with q-value < 0.05 were applied.

d) Analyses of gene expression through development

In order to calculate significant changes in gene expression patterns throughout ovule development between apomictic and sexual samples, the STEM software (Ernst and Bar-Joseph, 2006) was used. An analysis similar to that made by Sharbel *et al*. (2010) was made, whereby a data set was first constructed by selecting genes showing significant expression differences in at least one stage. Second, for those selected genes the expression over all other stages was added to the data set. Except for the number of permutations, which was set to 1000 to increase accuracy (Sharbel *et al*., 2010), the profiling analysis was performed using default options with Bonferroni correction for multiple testing (Ernst and Bar-Joseph, 2006). For the comparison analysis of different patterns of gene expression across the four developmental stages, the minimum number of intersected genes between sexual and apomictic samples was set to one (Sharbel *et al*., 2010) with maximum uncorrected intersection P value < 0.05.

4.2.3. Analyses for signatures of ploidy, parent of origin effects or hybridization

Considering variable ploidy and evolutionary origins between sexual and hybrid apomictic *Ranunculus*, we sought to classify gene expression patterns into groups reflective of ploidy, parent of origin (*sensu* expression level dominance) or hybridization (*sensu* transgressive) effects. We first selected those genes showing statistically significant different expression (p<0.05, log2 fold change >2, and FDR q-value < 0.05) between hexaploid apomicts and diploid sexuals, the two groups for which three samples each were available (Table 4.1) and hence for which statistical analysis was possible (note that a similar statistical comparison using the single tetraploid sample was not possible, due to sample size). Using this subset of genes defined by the hexaploid apomict-diploid sexual comparison, we then compared gene expression of each group to that of the single tetraploid using a similar statistical approach used by Rapp *et al*. (2009).

a) Parent or origin effect

Genes in the tetraploids which (1) showed no expression differences with those of the hexaploid apomicts (*i.e.* < two standard deviations of mean gene expression), and (2) had expression levels which differed by > two standard deviations of the mean diploid sexual levels, were classified as reflecting a parent of origin /expression dominance effect. Lastly, from the group of all genes which were not initially selected (*i.e.* no

68

significant differences in expression between hexaploid apomicts and diploid sexuals, see above), those in which the tetraploid sample had higher (*i.e.* > two standard deviations of the mean diploid sexual and hexaploid apomict levels) or lower (*i.e.* < two standard deviations of the mean diploid sexual and hexaploid apomict levels) were classified as reflecting a parent of origin /expression dominance pattern.

b) Ploidy effect

On a gene-by-gene basis, the standard deviation of expression was calculated for the hexaploid apomicts and diploid sexuals. Then, genes in the tetraploid whose expression was (1) lower than two standard deviations in the diploid and/or hexaploid, and (2) higher than two standard deviations in the hexaploid and/or diploid, were classified as reflecting a pattern reflective of ploidy/additivity effects (*sensu* Yoo *et al.*, 2013)

c) Transgressive effect

In the opposite case, whereby genes in the tetraploids (1) showed no expression differences with those of the diploid sexuals (*i.e.* < two standard deviations of mean gene expression), and (2) had expression levels which differed by > two standard deviations of the mean hexaploid apomicts levels, these were classified as reflecting a hybrid/transgressive pattern

4.2.4. Microarray validation using qRT-PCR

Ten genes showing differential expression between apomictic and sexual genotypes across the four developmental stages were randomly selected (Table 4.2 After retrieving the sequences from the assembled cDNA database used in the array manufacture, PCR primers were design using the PrimerSelect software (DNASTAR Inc., Madison, WI) and selected, when possible, to overlap the microarray probes with the following parameters: product size <150bp, GC content between 40 and 60%, annealing temperature ca. 60°C. (Table 4.2)

Ranunculus-specific reference genes were developed by identifying sequences which were homologous to a selection of *Arabidopsis thaliana* reference genes (www.TAIR.com) using a blast analysis (blastX 2.2.30+ using the default NCBI parameters). Based on maximum similarity (similarity > 95%, e value <1 e-100) homologous genes to UBQ (gb|ABH08754.1|) and ACTIN 11 (ref|NP_187818.1|) where chosen. Primers were designed following the procedure and parameters described above, and tested for amplification and expected product length in 10 μl PCR reactions

including 25 ng of DNA, 1 μl of PCR Buffer II, 10 pmol for each primer, 0.025 U DNA Taq DNA Polymerase (Sigma-Aldrich), with 3.5 mM of MgCl2 and 4.95 μl of H2O. PCR reactions were performed in a Mastercycler ep384 (Eppendorf, Hamburg, DE) using the following touchdown thermal cycling profile: 94° for 10 min; 9 cycles of 94° for 15 sec, 65° for 15 sec (1 degree decrease in temperature every cycle with a final temperature of 54°), 72° for 30 sec; 35 cycles of 94° for 30 sec, 57° for 15 sec, 68° for 2 min 30 sec; and a final 68° for 15 min.

qRT-PCR reactions, using UBQ and ACT11 (Dyad Id number: 82481 and 151955) as housekeeping genes, were run on a 7900HT FAST RT-PCR machine (Applied Biosystems) using the SYBR Green Master Mix (Applied Biosystems) and the following program: initial denaturation at 90°C for 10 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min. Ct values (PCR cycle number where SYBR Green is detected) were extrapolated and used to infer initial copy number of the genes. Mean expression and standard deviation were calculated between two technical replicates and three biological replicates from apomictic and sexual genotypes using cDNA from the second stage of ovule microdissected tissues (Table 4.2). Relative quantification was calculated using the $\Delta\Delta$Ct methods, and using the genes with higher CT values compared to the calibrator sample.

Table 4.2 list of genes used for qRT-PCR validation and qRT-PCR result according to REST software

Seq id*	Primer 5´- 3´	Primer 3´-5´	Expression	P Value	Regulation
345733[a]	ACCAAAGCGATTAGAAACACCAGT	CCCGAGAAAGCCGATGAAA	0.467	0.089	Up
350109[a]	TCTTGATTCGATTCTTGTGGTTGA	TGTTCCGGTGGCGACTTG	131.622	≤0.001	Up
354538[a]	CAGGCGCAAATGTGAGA	TGGGATCGATAGCAAGAGTC	36.07	0.009	Up
357979[a]	GGCTCTGAGTCAACTTGTCCTAAT	CTTTCGGCTTCCTTTTCTTCTT	4.286	≤0.001	Up
360024[a]	CGACGAGACGGGGTTGAA	ATTGGCGTTGCATCTCTAAGTG	34.088	0.014	Up
366262[a]	ACGTGGTGTTTGGGAATGAAT	TGAACGGATATCCCTGACAAAA	3.772	≤0.001	Up
411537[b]	GAGCCGGAAAGTGCGAGTA	GATGCTGTGATGGAATTGATAGTA	0.018	0.023	Down
505317[b]	GGAAATCCAAGAAACAACC	GAAGATCATTAAGCAAGTGG	0.24	0.057	Down
556851[b]	GATTTTAGGCCATTTGATTGTGC	TTGAGGGTTTGTATGAGATTTGAA	0.006	0.02	Down
345754[b]	TAGCCAACGTACCGATTAGGATAG	AGGTAGTGGCTTGCTCTTTCTGTC	0.033	0.016	Down
405990[b]	ACCCCAGCTGTCTTCCGAGTA	CACAAATGGCATATGAGACACG	0.251	0.031	Down

(a) and (b): transcript up- and downregulated , respectively, according to the microarray analysis

* Dryad entry doi:10.5061/dryad.nk151;

4.2.5. Blastx, tblastx, Gene Ontology

Significantly differentially expressed genes obtained in all apomictic - sexual comparisons were selected for gene ontology (GO) analysis. The Blast2Go (http://www.blast2go.com/b2ghome) software was used for the annotation procedure using a blastx (E-val cut off of E ≤ 1-5) and the default annotation parameters of the program. For genes where no blastx hits was obtained, an additional tblastx analysis was performed (E-val cut off of E ≤ 1-5). Overrepresentation analysis was not possible since only a small fraction of the *Ranunculus* transcriptome could be annotated (see results).

4.3. Results

4.3.1. Microarray development

The assembled RNASeq data 15(62 102 contigs > 300 bp and 400 000 non-assembled singletons; see Pellino *et al.*, 2013) was provided to NimbleGen for custom oligomicroarray development. The NimbleGen selection strategy bioinformatically designed three different probes for each contig, and one to three probes for each singleton, followed by design of a 3 x 1.4 million-spot array. The final array therefore contained multiple technical replicates for each gene expressed during flower development in *Ranunculus*.

4.3.2. Gene expression differences

Raw data were at first normalized using the RMA algorithm implemented in the DEVA software. In order to identify a subset of genes that accounted for the maximum distinction between the apomictic and sexual samples, PCA was applied to the normalized expression values of probes from the original 62 102 contigs, all of which had a minimum length of 300bp. As a result, the apomictic and sexual samples could be distinguished (Figure 4.1). Moreover clear separation with respect to ploidy was found, and the two apomictic *R. carpaticola* x *cassubicifolius* from VRU 2 could also be differentiated (Figure 4.1.)

72
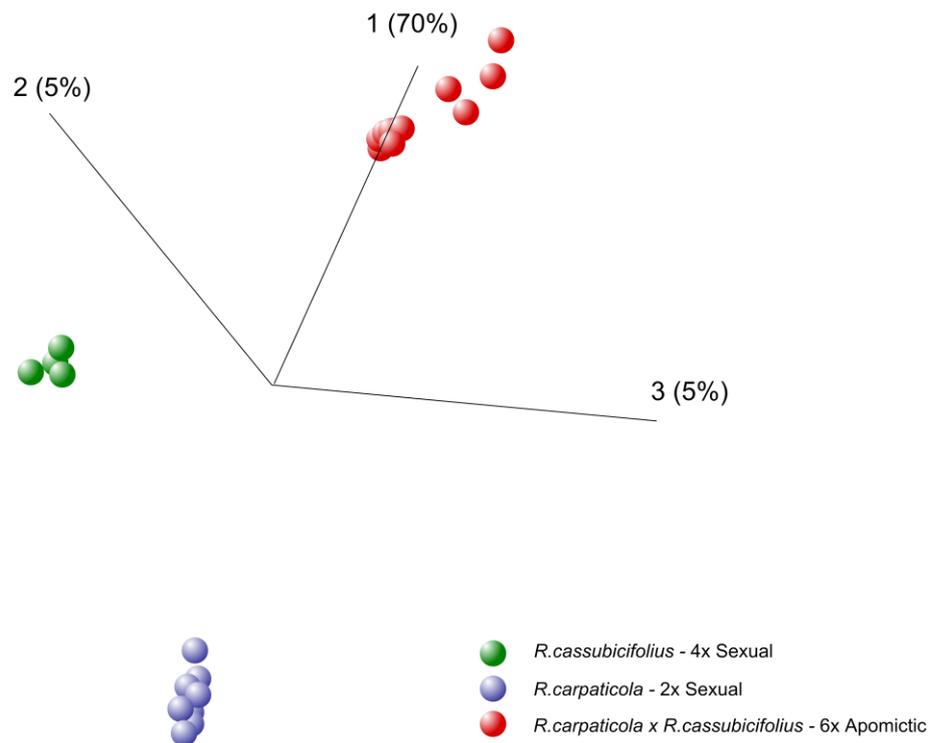
Figure 4.1 Principal component analysis (PCA) on apomictic and sexual *Ranunculus* . PCA applied to normalized microarray data representing four ovule developmental stages from three apomictic (red) and four sexual (blue and green) *Ranunculus* showing ploidy and reproduction specific effects. Each dot represents one ovule stage, and frequencies in parentheses show the percentage of total variation explained by that particular principal component. The numbers on the axes represent the respective component (in the % of variation for each component).

4.3.3. Stage specific differential expression analysis

Transcriptome-wide gene expression variation through ovule development was compared between sexual and apomictic *Ranunculus*, whereby four pre- to post-meiotic ovule stages were compared between seven different genotypes. Overall, across all stages 439 and 339 transcripts were found to be significantly down- or upregulated between apomictic versus sexual genotypes (Figure 4.2. Venn A). The distribution of up- and downregulated transcripts in apomicts differed across the studied developmental stages I to IV (30, 48, 247, and three up regulated, 44, 98, 58, and 27 downregulated for each stage respectively) (Figure 4.2)
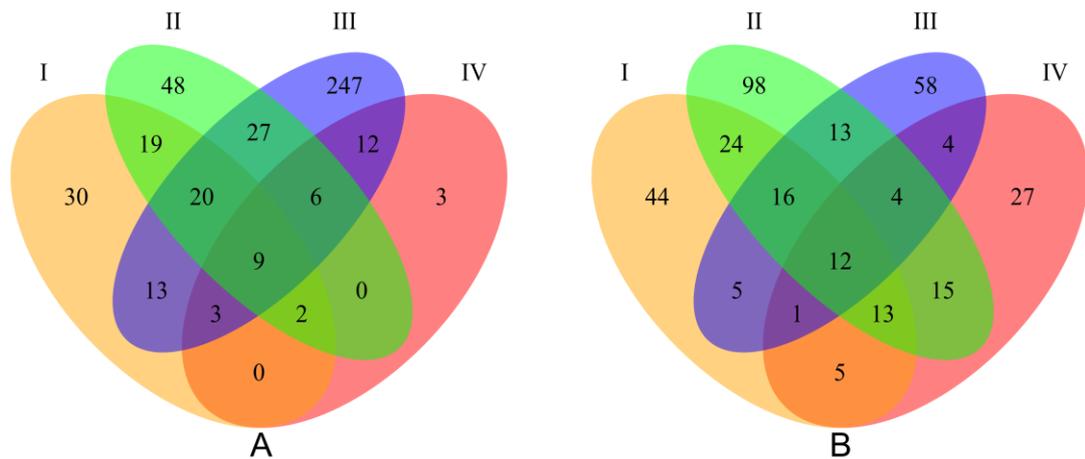
Figure 4.2 Venn diagrams showing number of differentially expressed genes transcripts between apomictic and sexual genotypes at each ovule developmental stage. (A) Genes downregulated in apomictics compare to sexual (B) Genes upregulated in apomictics compare to sexual

Table 4.3 Numbers of differentially expressed transcripts at each ovule stage development.

Number of contigs	Number of contigs > 300bp	Differentially expressed transcripts			Annotated
462102	62102				
		Ovule stage	Upregulated*	Downregulated*	
		I	120 (44)	96 (30)	72
		II	195 (98)	131 (48)	107
		III	113 (58)	337 (247)	156
		IV	81 (3)	35 (3)	30

* In the apomictic individuals compared to the normal sexual expression

In parenthesis the number of stage specific transcripts

Blast2Go analyses could be completed on only a small fraction of the differentially expressed transcripts, the limiting step being the inability to identify significant homologies to nucleotide databases (NCBI nucleotide database February. 2014). Of the number of genes which could be annotated, only 78 and 59 down and up regulated transcripts from apomicts could be assigned a gene ontology (GO) term (Appendix I to VIII). Hence, representation analysis was not possible due to the significant bias introduced by insufficient GO categorization

A clustering-based analysis (Ernst and Bar-Joseph, 2006) was performed to detect significant changes in transcript abundance through ovule development by first grouping transcripts with similar expression patterns through development for the sexual and apomictic array datasets separately. In doing so, 46 and 62 transcripts could be assigned to two particular patterns (p<0.01) in the sexual and apomictic groups respectively. A comparison of patterns between these transcript sets identified eight as

having significant differences (p<0.01) in the corresponding reproductive form, and showed a general trend of expression increase in developmental stage II and sharp drop in stage III in apomicts (Figure 4.3 & 4.4). None of the eight genes could be assigned a GO term, although three had a significant homology to the transposon mutator sub-class protein (XP_006654086.1), salt overly sensitive 1b isoform 1 (EXC05020.1), and annexin like protein (XP_007042996.1)
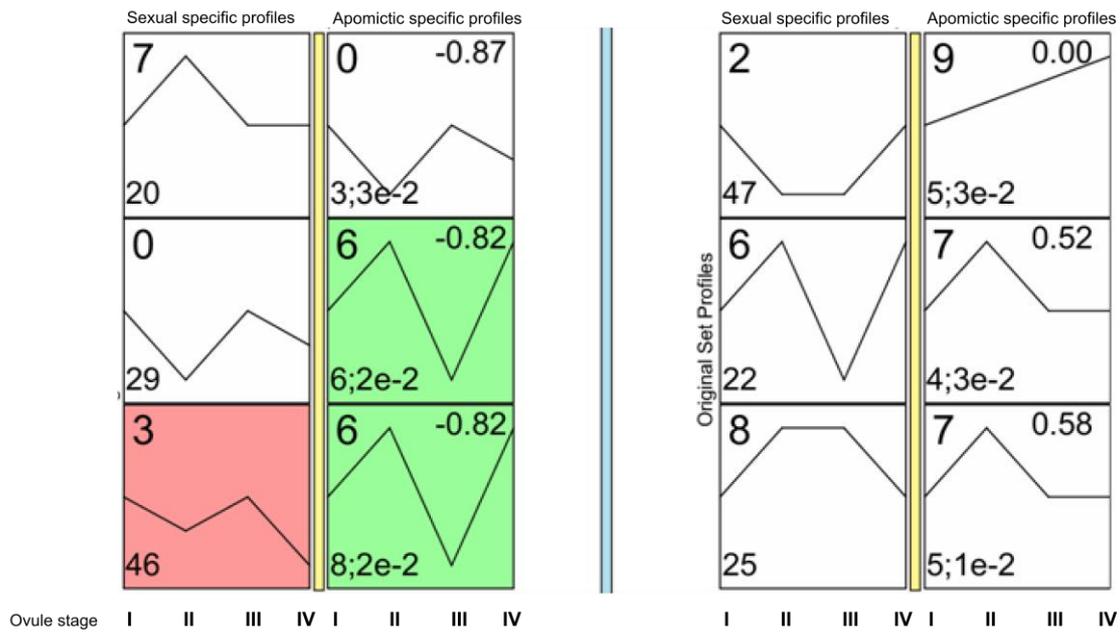
Figure 4.3 Graphs showing the STEM profile analysis of gene expression throughout ovule development. In color profiles with statistically significant association (p<0.01)
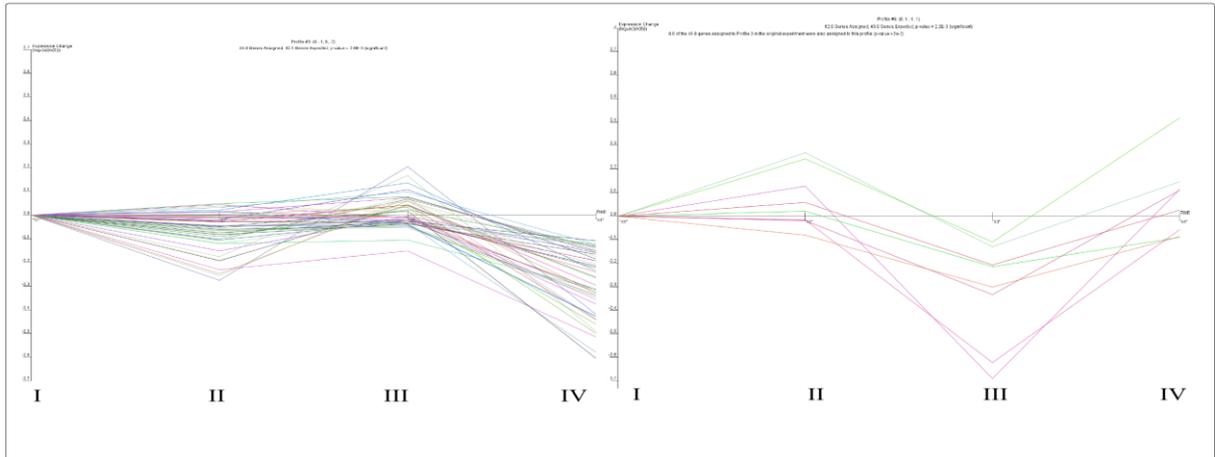
Figure 4.4 Expression pattern of genes in sexual specific ovule development (left) and homologous genes in apomicts with statistically significant heterochronic expression (right); on the x-axis is represented the developmental stage, in the y-axis the log2 relative expression.

4.3.4. Transcriptome wide signatures of hybridization, ploidy variation and parent of origin effects

In order to understand and classify gene expression patterns of the apomictic hybrid (*R. carpaticola* x *R. cassubicifolius*) compared to the ancestral sexual parents (*R. carpaticola and R. cassubicifolius*), 304 differentially expressed transcripts were first identified in at least one developmental stage based on minimum fold change and statistical significance (log2>2, p-value< 0.01, FDR <0.05) between the apomictic hybrids and the diploid sexuals (both of which having three genotypes each for statistical comparison). Secondly, a stage-by-stage comparison was performed for the expression of these 304 transcripts in the second tetraploid sexual parent (*R. cassubicifolius*; for which sample size precluded the first level statistical analysis – see Methods) such that they could be classified into four different expression states (*i.e.* the hybrid relative to each ancestral parent). Genes showing expression bias (*i.e.* not significantly different) with respect to the tetraploid parent were classified as showing a parent of origin effect (4x-parent origin), while those significantly over or under expressed in the apomict compared to both parents were classified as showing transgressive effects. Lastly, genes showing expression level changes in proportion to that of ploidy (*e.g.* increasing expression with increase of ploidy) were classified as showing a ploidy effect

76
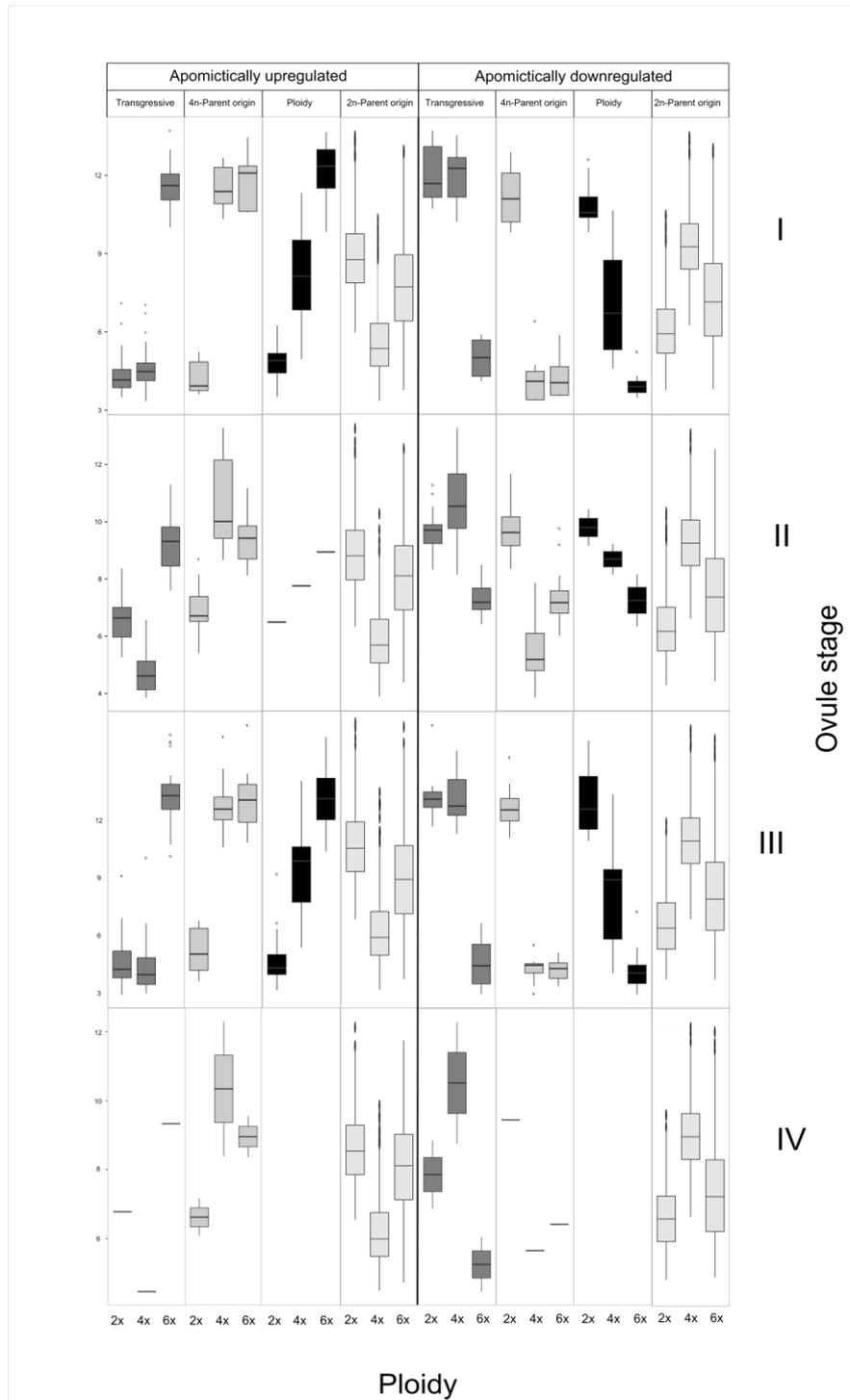
Figure 4.5 Box plots of expression distributions for genes which were significantly up- or downregulated in at least one apomictic ovule stage, as grouped into transgressive, ploidy-mediated and parent of origin patterns.

Table 4.4: Table showing the number of genes classified into transgressive, parent and additivity effects

Expression class*	Stage	Transgressive	Parent of origin	Ploidy	Total
Apomictically upregulated**	I	5 (0.1)	27 (0.49)	23 (0.41)	55
	II	15 (0.58)	10 (0.38)	1 (0.04)	26
	III	16 (0.16)	44 (0.46)	36 (0.38)	96
	IV	2 (0.66)	1 (0.33)	2 (0.66)	5
Apomictic – 2x sexual overexpressed***	I	-	2750	-	2750
	II	-	2874	-	2874
	III	-	2909	-	2909
	IV	-	2551	-	2551
Apomictically downregulated**	I	6 (0.26)	7 (0.3)	10 (0.44)	23
	II	18 (0.32)	37 (0.65)	2 (0.03)	57
	III	9 (0.24)	11 (0.3)	17 (0.46)	37
	IV	2 (0.66)	1 (0.33)	2 (0.66)	5
Apomictic – 2x sexual underexpressed***	I	-	4630	-	4630
	II	-	4887	-	4887
	III	-	5168	-	5168
	IV	-	4395	-	4395

In parenthesis the relative percentage

*Compared to normal sexual diploid expression

** (P<0.001, FDR<0.05, $\log_2$ change >2)

*** Class of genes that showed similar expression between apomictic and sexual diploid but different from the tetraploid

The same procedure was applied to 19116 genes that were not significantly differentially expressed between the apomictic and the sexual diploid in order to classify an additional state of gene expression (2x-parent of origin; see methods). This approach was necessary considering that only a single 4x sample was used, and allowed the comparison of gene expression of this one parent (4x) to the hybrid apomict. This classification was further subdivided with regards to whether expression of the hybrid and the diploid was higher or lower in comparison to the tetraploid. In total 8 different groups of genes were classified, with most being downregulated in apomictic ovules at stage two (Figure 4.3 and Table 4.3),

Overall, most transcripts were expressed in a parent of origin pattern, followed by ploidy and then transgressive patterns (Table 4.3). The strongest effect was apparently due to parent of origin, with a total of 82 upregulated and 138

downregulated s apomictic specific transcripts. Ploidy-mediated patterns of expression were revealed in 62 and 93 apomictic specific upregulated and downregulated transcripts respectively. Transcripts showing transgressive patterns were the least abundant, with 38 and 73 respectively upregulated and downregulated in apomicts (Table 4.3). Of all transcripts showing such patterns, only 83 (49 and 34 up and down-regulated in apomicts respectively) could be annotated (Appendix IX & X).

4.3.5. qRT-PCR validation

Validation of the randomly selected genes showed concordance with the microarray analysis from the five upregulated and five downregulated, respectively, randomly selected genes. According to the REST software all ten genes showed statically significant differential expression when measured with qRT-PCR (Table 4.2)

4.4. Discussion

Whether caused by a single "master gene", or through the interaction of a polygenic complex, apomictic reproduction has long been a dilemma for scientists. Using different experimental approaches, the search for factor(s) underlying the reproductive switch has led to the identification of a number of candidates in different species (Barcaccia and Albertini, 2013) normally involved in sexual reproduction, with support for a shift in gene expression through time (Bicknell and Koltunow, 2004). OMICS approaches have become useful for detecting expression shifts, although functional characterization and screening of the many candidates from such an experiment is a formidable task. This is especially true when dealing with natural systems (as most apomicts are) in which annotated genomic information is not available. Additionally, the effects of polyploidy and hybridization (Comai *et al.*, 2003; Osborn *et al.*, 2003) are highly associated with the apomixis phenotype, adding an extra level of complexity to analyses.

Despite hybridization and polyploidy introducing challenges to experimental approaches, their association with apomixis potentially reflects its mechanism of induction (Carman, 1997; Grossniklaus, 2001). For example, the "genomic shock" (McKlintock, 1950) produced by the union of two different genomes could trigger the cascade of spatial and timing misexpression of sex specific genes, hypothetically leading to the shift to apomixis.

Here we continue our ongoing work (Pellino *et al.*, 2013; Hojsgaard *et al.* 2014b) with wild allopolyploid apomictic *Ranunculus* and related sexuals to shed light upon factors underlying the transition to apomixis. We analyzed genome-wide gene expression through ovule development in both apomictic and sexual genotypes, and partitioned differentially expressed genes into patterns reflective of (a) reproduction specific expression, (b) heterochronic expression across ovule development, and (c) expression of homeologus genes in the apomicts and their phylogenetic parents.

Using the observations of Hojsgaard *et al.* (2014b), ovules at four morphologically defined stages were live-microdissected (see Corral *et al.*, 2013; Pellino *et al.*, 2011) from sexual and apomictic genotypes (Table 4.1). Our sampling strategy, consisting of four and three biological replicates each from sexual and apomictic plants, enabled the identification of differences encompassing both reproductive mode (*e.g.* apomictic versus sexual; Figure 4.1) and genetic background (*e.g.* different sample clusters; Figure 4.1). The sexual group included three diploid and one tetraploid genotype, covariates that furthermore enabled some of the variation to be assigned to ploidy effects (Figure 4.5). Our goal of maximizing biological diversity in the sampling design is reflected in transcriptional variation (Figure 4.1) influenced by the genetic background, polyploidy and hybridity characteristic of *Ranunculus* (Hörandl *et al.*, 2009).

4.4.1. Contrasted sexual *vs.* apomictic development of the ovule reflect transcriptomal variation

Hojsgaard *et al.* (2014b) showed that the first stage of ovule primordium development is relatively undifferentiated between apomictic and sexual *Ranunculus*. In contrast, diverse developmental irregularities accumulate during megasporogenesis, leading to perturbed and non-functional meiosis in apomicts (*i.e.* arrested development and/or altered megaspore selection; Hojsgaard *et al.*, 2014b), followed by enlargement of somatic cells which take on the aposporous initial (AI) role. Deviation between the sexual and apomictic pathways continues during gametogenesis and development of the AI, parallel in timing with normal sexual gametogenesis. The resultant unreduced megagametophyte (arising from the AI) then attains a similar structure to that of the sexuals throughout the final developmental stage (Hojsgaard *et al.*, 2014b). Our tissue sampling design was selected to cover a number of steps leading up to and proceeding throughout AI development (Table 4.1).

A reflection of the phenotypically diverging development pathways, gene expression differences between sexual and apomictic ovules are less pronounced in the first and last stages of sampling, whereas higher numbers of differentially expressed transcripts were sampled during the second and third stages (Table 4.2and Figure 4.1). Stages II and III encompass AI appearance, in parallel with normal meiosis, whereby the AI enlarges to squeeze the forming mega gametophyte and eventually take its place. Despite the weakness of any GO inference (see results), we note that no particular GO classes distinguished differentially expressed genes (Appendix I to VIII) at any stage. Especially in stage II the parallel appearance of meiotic products and aposporous initials (with almost equal frequencies in hexaploids; Hojsgaard *et al*. 2014b) may blur the distinction in gene expression. Later on in stage III, as increasingly more meiotic products have aborted and only unreduced gametophytes develop further, the differences between sexual and apomictic development become most distinct (Figure 4.3). Stage IV might thus be principally similar in gene expression, considering that mature sexual and apomictic embryo sacs do not differ phenotypically.

4.4.2. Homeologous parent of origin, ploidy and transgressive effects in the apomictic hybrid

As complex as the genetic processes underlying apomixis can be, it is believed to be controlled by one or few master regulatory genes (Barcaccia and Albertini, 2013; Eckardt, 2003). In his studies on the genetic control of apomixis in *Ranunculus auricomus*, Nogler (1984a, 1984b) inferred that (a) a single Mendelian factor A$^-$ induced the occurrence of apomixis, and that the frequency (*i.e.* penetrance) of apomixis was dosage dependent and relative to the wild type a+, and (b) the genetically dominant A$^-$ factor behaved as recessive lethal, the implication of which was that A$^-$ could not be transmitted by a diploid parent, as a haploid gamete hosting A$^-$ would be lethal. The implication of this would be that A$^-$ must be transmitted via a polyploid (tetrapolid) (see Table 4.3), although it is unclear whether it would need to be transmitted in the homo- or heterozygous (*i.e.* A$^-$ a$^+$) state, as this would depend on the nature of its lethality (*e.g.* mutational load versus gene dosage effects).

Considering these observations, in addition to the biology of sexual and apomictic *Ranunculus*, we have attempted to partition gene expression patterns into groups reflective of the different influences on the sexual vs. apomictic ovule, including hybridization, variable ploidy and parent of origin effects (Table 4.3 Figure 4.3).

a) Parent of origin effect

Over all stages, 220 genes were found to be similarly expressed between the 6x apomictic *R. carpaticola* x *R. cassubicifolius* and the 4x ancestral parent *R. cassubicifolius*, a pattern reflective of a parent of origin effect. The relatively higher number of genes displaying a parent of origin pattern of expression (in comparison with ploidy and transgressive effect (Table 4.3)) in *Ranunculus* ovules is concordant with similar studies in natural population and synthetic hybrids of cotton (Flagel and Wendel, 2010), maize (Swanson-Wagner *et al*., 2006), rice (He *et al*., 2010) and *Senecio* (Hegarty *et al*., 2006), which showed similarly higher levels of parent of origin expression relative to transgressive effects, a pattern that tended to decrease with the age of the hybrid. These effects appeared in sexual plants without any signs of apomixis, and underline the genomic impact of polyploidization.

Investigations into the causes of parental expression bias point to the effect of *cis-trans* regulation, epigenetic changes and introgression in *Cirsium arvense* (thistle; Bell *et al*., 2013) and cotton (Yoo *et al*., 2013).

If one hypothesizes that Nogler's (1984a, 1984b) apospory-incurring A$^-$ factor is found primarily in the genomes of hybrids, it is plausible that such a factor has been transmitted by one of the parents and possibly (but not necessarily) share the same expression level. If one assumes similar expression level of the A$^-$ between parent and hybrid, then this could be a member of the *Ranunculus* genes demonstrating a "parent of origin" pattern (Table 4.3 Figure 4.3). To begin with, the analysis of tetraploids from the same species used in this study failed to show any sign of AI formation (Hojsgaard *et al*., 2014b), in contrast to what was found in tetraploids of the *R. megacarpus* complex used in Nogler's (1984a, 1984b) work. *R. megacarpus* is, according to its intermediate morphology, most likely a hybrid between a "*R. auricomus*" and "*R. cassubicus*" morphotype (Hörandl and Gutermann 1998). In contrast, *R. cassubicifolius* is a sexual autotetraploid (Hörandl and Greilhuber, 2002). Neither from the widespread diploid cytotype of the species nor from tetraploid natural populations is there any indication of apomixis, as inferred from pollen quality, flower morphology, and population genetic structure (Hörandl *et al*. 1997, Hörandl & Greilhuber 2002, Paun *et al*. 2006a, b; Hojsgaard *et al*. 2014b). Nevertheless, the lethal nature of A$^-$ (or it's linked deleterious mutations) precludes its transmission via haploid gametes (*i.e.* from a diploid parent), therefore A$^-$ must necessarily derive from a polyploid parent.

Considering this, the origin of apomixis in the hexaploids could be explained in two ways. First, apospory could have appeared for the first time spontaneously in natural, triploid hybrids between diploid sexual *R. carpaticola* and tetraploid sexual *R. cassubicifolius* (similar as observed in 2x x 4x crosses in Hojsgaard *et al*. 2014b). A triploid plant is not only capable of producing diploid gametes, it could also produce low levels of functional apomictic seeds (Hojsgaard *et al*. 2014b) and thus transmits the A- factor further by both female and male gametes. Subsequent polyploidization of the triploid could result in the hexaploid that established in nature. Alternatively, A- could have already been present but not phenotypically expressed in tetraploid *R. cassubicifolius*. Based on the observations of Nogler (1984b), A- is associated with AI development but not its timing, and delayed occurrence with respect to meiotic development will not produce any functional unreduced megagametophyte. The occurrence of apomixis is thus associated with both the presence (*i.e.* expression) of A- and its expression time, with the shift in expression timing arising through hybridization or polyploidization (as proposed by Carman, 1997 see below). Alternatively, the level of expression of A- could be influenced by its dosage, as Nogler 1984b) proposed to explain the frequency of AIs (see ploidy below).

Focusing on the diploid parent, the relatively low expression levels of transcripts which characterize the parent of origin effects (with the expression of the hexaploid similar to the tetraploid parent) could be explained by (a) the presence of the A- but not its expression, or (b) the complete absence of A- itself. In order to shed light on these possibilities, we examined high quality single nucleotide polymorphisms (SNPs) mined from a previous study (Pellino *et al*., 2013) involving the same individual plants analysed here. Of the genes showing a parent of origin effect here, only eight were characterized by sufficient DNA sequence read coverage in all individuals, and of those only one (dyad contig id: 368515, Dryad entry doi:10.5061/dryad.nk151;) showed 100% SNP similarity (for 16 SNPs) between the apomictic and the tetraploid genotypes only, implying its potential origin from the tetraploid parent (Appendix XI). More cannot be inferred because of the lack of information on the function (*i.e.* annotation) of these genes, while inference from the SNP dataset is limited considering variability in read coverage between samples, in addition to the fact that the genotypes analysed were not *sensu stricto* those involved in the original hybridization *circa* 80 000 years before present (Pellino *et al*., 2013). Nevertheless we have presented here genes

showing evidence of expression bias and origin from the polyploid parent, data that could support introgression through hybridization.

b) Ploidy effect

In *Ranunculus*, Nogler (Nogler, 1984b) demonstrated that increased dosage of the A- apomixis factor relative to the wild type sexual a+ allele results in higher level of apomixis penetrance, as evidenced by the occurrence of AI cells. Assuming that A- is allelic, it then becomes relevant to understand how dosage could vary, considering that it can only be transferred via > haploid gametes (see Nogler 1984b). Tetraploids can be fully sexual (Hojsgaard *et al*., 2014b), implying either that some tetraploids lack A- altogether and that A- appeared in hybrids only, that the expression of A- occurs at the wrong time of ovule development, or that A- can be present in low dosage with no phenotypic (*i.e.* apomictic) effects.

Multiple copies of A- could be present in the hexaploid apomictic, resulting in increased expression levels compared to the parents of lower ploidy. Assuming that A- is derived exclusively from a polyploid parent, its expression patterns would reflect ploidy, being not present/expressed in the diploid, lowly expressed in the triploid carrier and highly expressed in the apomict.

Together, the occurrence of AI's in the hexaploid, in addition to variability in penetrance of apomixis, are consistent with variable dosage of apomixis factors (*e.g.* A-) as the underlying mechanism. The high variability for facultative sexuality in hexaploids (in the mean 69% of ovules; Hojsgaard *et al*. 2014b) supports the hypothesis that hexaploids are heterozygous for A- and the wild type A+. Thus, how is variable dosage for A- attained?

Increased copy number of the A- factor in hexaploids (to a sufficient level to induce apomixis) could be achieved in a number of ways. A single copy of A- could initially be inherited via a triploid hybrid bridge, autopolyploidization (Paun *et al*., 2006b) and doubling of its copy number. Subsequently, increased dosage and variability in apomixis penetrance could arise via increased selfing and the maintenance of facultative sexuality. For instance, a hexaploid facultative apomictic possessing two A- alleles could potentially increase the copy number up to four in the first progeny. The maximal copy number of A- may be limited by associated deleterious mutations linked to A- (*i.e.* Muller's ratchet; Muller, 1964) or by lethality at homozygous gamete states. The outcomes of this process are two-fold. First, the constraints

associated with the lethality of A⁻ may result in an inability to evolve obligate apomicts which are fully homozygous for A⁻. Second, the combination of dosage effects and residual sexuality may result in variability for apomictic penetrance..

Maintenance of faculative sexuality (means of 29% sexually produced seed) could be further enhanced by selection for occasional recombination around the A-factor, potentially reducing its mutational drag (Hojsgaard and Hörandl, 2015). In support of this, neither fully sexual nor obligate apomictic hexaploid *Ranunculus* individuals have been described (Hojsgaard *et al.*, 2014b; Paun *et al.*, 2006b)

c)   Transgressive effect

Expression patterns of an apomixis factor A⁻ and associated cofactors could additionally be associated with *cis* and *trans* regulatory dynamics that characterize hybrid genomes (Hegarty *et al.*, 2008). Aposporous initials have been described in the tetraploid *R. megacarpus* possessing A⁻ (Nogler, 1984b), while more recently a small percentage of AIs was induced in synthetic diploid and triploid hybrids (Hojsgaard *et al.*, 2014b). A possible explanation could be that aposporous embryo sac formation is indeed controlled by A⁻, although A⁻ does not control the time of induction of the AI (Nogler, 1984b). As the timing of AI induction is a key component for successful apomictic development (Nogler, 1984b), hybridization may play a role considering that shifts in timing of floral genes in hybrids could underlie the transition to apomixis (Carman, 1997; Nogler, 1984b). This scenario could be reflected in (1) genes/transcripts demonstrating transgressive expression patterns (Table 4.3; Figure 4.3), or (2) genes/transcripts for which we could measure heterochronic shifts (Table 4.2; Figure 4.4).

We identified 38 and 35 homologs showing transgressive effects that were respectively down- and upregulated in apomicts (Table 4.3). The relatively low number of genes identified as displaying transgressive expression effects could be a reflection of the estimated 80 000 years of evolutionary age of the *Ranunculus* complex (Paun *et al.*, 2006b; Pellino *et al.*, 2013), as it is speculated that genetic mechanisms and evolutionary forces tend to reduce transgressive effects in natural populations, even after only a few generations subsequent to hybridization and genome doubling (Hegarty *et al.*, 2006).   Genes in this group that could be annotated were involved in megagametogenesis and development, as in the case for the embryo sac developmental arrest protein (EDA), gamete expressed protein (GEX), and a gene of the

argonaute family (AGO) (e-val = 1.43 e-13, 3.28 e-32, 1.61 e-63, respectively and mean similarity of 47%, 70.5% and 84.4%, respectively; Appendix III). Interestingly, *Arabidopsis* EDA mutants show a series of defects during megagametogenesis, resulting in interrupted or abnormal meiotic division (Pagnussat *et al.*, 2005). Furthermore GEX is essential for pollen tube guidance during double fertilization, and its misregulation in *Arabidopsis* resulted in reduced seed set and undeveloped embryos (Ray *et al.*, 1997). Finally, members of the AGO protein family have been associated with the possible epigenetic control of apomixis and gamete formation via a small RNA network, for example in *Arabidopsis* and Maize defective for AGO9 and AGO104 (Olmedo-Monfil *et al.*, 2010; Singh *et al.*, 2011).

Global shifts in gene expression patterns associated with hybridization and polyploidy (Hegarty *et al.*, 2008) have been hypothesized to underlie the switch from sex to apomixis (Carman, 1997; Koltunow, 1993), support for which has been found in apomictic *Boechera* (Sharbel *et al.*, 2010) and between sexual species of *Tripsacum* (Grimanelli *et al.*, 2003). The hypothesis that hexaploid apomictic *Ranunculus* analyzed here have hybrid origins (Hörandl and Greilhuber, 2002; Paun *et al.*, 2006b; Pellino *et al.* 2013) is supported on many levels, and is evidenced in synthetically derived hybrid *Ranunculus* (Hojsgaard *et al.*, 2014b) which showed apomeiotic embryo sac formation, as has also been detected in *Sorgum* and *Antennaria* (Carman, 2007). Interestingly, despite having different forms of apomixis, aposporous *Ranunculus* and diplosporous *Boechera* similarly show a negative spike in differential gene expression during ovule development (compared to sexual ovules), although this occurs at the megaspore mother cell (MMC) stage II in the former (Sharbel *et al.*, 2010), whereas this negative spike occurs later during AI development in the latter (Figure 4.4). Due to limited annotation of the *Ranunculus* transcripts (see results), these data cannot be directly compared to those of *Boechera* (Sharbel *et al.*, 2010), although the negative spike in gene expression occurring at key steps of diplo- and aposporous ovule development is suggestive of conservation in processes leading to apomeiosis.

In this first study on gene expression variation in the apomictic *Ranunculus auricomus* complex we present a list of genes that show differential expression between apomictic and sexual ovules across four developmental stages. Considering the common difficulty in wild species for which genomic sequence information can be readily collected but little to no gene annotation can be made, we have considered the

natural history of the *Ranunculus auricomus* complex in order to classify genes whose expression patterns reflect the evolutionary and genetic mechanisms hypothesized to underlie the switch from sex to apomixis. Our observations based on the number of differentially expressed genes and their expression patterns are consistent with the commonly accepted idea that apomixis is caused by gene deregulation, in addition to being ultimately influenced by hybridization and polyploidization. Here we set the foundations for more specific studies on gene regulation with respect to apomixis expression and penetrance in *Ranunculus*.

FINAL CONCLUSIONS

This thesis and all its components address different questions that arise when dealing with non-model species in the study of apomixis.

The first chapter describes the analysis and selection of stable housekeeping genes in *Boechera*, and represents the first comprehensive analysis of reference genes in a dicot plant with both sexual and apomictic forms (*Boechera*). Most importantly, it is the first report of a housekeeping gene analysis on live-microdissected ovules and anthers in *Boechera*.

Six different housekeeping genes were selected based on homology with *Arabidopsis*, and their expression tested among a range of vegetative and reproductive tissue types of both apomictic and sexual species of *Boechera*. The expression value for each gene was analyzed using two different algorithms, and based on the expression stability (M) the most stable gene was selected specifically for each particular tissue/reproductive mode (Table 2.3). Based on optimized pairwise combinations (V) the best combination of housekeeping genes was obtained for each tissue and reproductive mode (Table 2.3).

In order to validate the selected HKGs, a further test was conducted. Previously studied SuperSAGE tags that had shown reproduction specific expression between two ovule developmental stages in *Boechera* were normalized with the new HKGs list. This validation experiment produced data consistent with the previous work, thereby supporting the validity of the selected HKGs. In addition to that, the HKGs listed in this work are involved in different and independent cellular processes, and thus interaction between them – when used in pairs – is minimized.

Overall these data provide an important tool for transcriptomal analyses of reproductive tissues in *Boechera*, which is a model system for the study of apomixis. The identification of these genes has enabled the precise qRT-PCR validation of two potential candidate genes highly correlated with the presence and origin of apomixis in *Boechera* (Corral *et al.*, 2013; Mau *et al.*, 2013), both of which are presently undergoing functional studies. Identification and optimization of the best HKGs was therefore, a key component for the achievement of an important milestone in deciphering apomixis in *Boechera*.

The next parts of this work focused on the association of apomixis with hybridization, polyploidy and their evolutionary consequences at the transcriptomic level. We took advantage of the known taxonomic information of the *Ranunculus auricomus* group, comprising the diploid *R. carpaticola* and tetraploid *R. cassubicifolius* and their hexaploid apomictic hybrids.

After generating flower-specific RNAseq (Illumina) data for three sexual and two apomictic individuals, we assembled and released the first transcriptome in *Ranunculus*, adding an important piece of knowledge to basal eudicot genomics and enabling, for instance, the custom design of the first *Ranunculus* complex-specific microarray (see below) and the selection of microsatellites (Hojsgaard *et al.* 2014b) that have been used for complementary work. Subsequently, the first genome-wide SNP analysis was carried out on three species of the *R. auricomus* complex, enabling the investigation of the origin and evolution of apomixis in these species and laying the foundation for future studies.

Using high quality SNPs the hybrid origin of the apomictic lineages was confirmed, supporting previous studies on the complex (Paun *et al* 2006a) and more importantly, supporting the general idea of apomixis arising as a consequence of hybridization. Using the SNP dataset to construct a hybridization network, followed by the analysis of post hybridization intra allelic divergence in the apomicts, a single origin event followed by rapid divergence were inferred. An estimation of the time of hybridization was set in this study to the late Pleistocene, ca. 80.000 years ago, strengthening the hypothesis of a relatively young origin of extant apomicts. Signs of mutation accumulation (*i.e.* Muller's ratchet) were detected and, in addition to polyploidy and hybridization, could have led to the observed faster rate of divergence in the apomictic lineages compared to their sexual congeners. The observed rapid divergence of the apomicts could support the idea that apomixis is connected to or is enhanced by rapid selection at particular loci. For instance apomictic *Ranunculus* are geographically more widespread and abundant in number than their sexual parents (Paun *et al*. 2006a; Hörandl *et al*. 2009). Here, divergence could be associated with a heterotic like effect which is evolutionarily advantageous to the hybrids in certain contexts. Alternative explanations for the observed rapid divergence could be found in the higher ploidy level of the apomictic lineages compared to the sexuals, as increase ploidy results in increased number of mutational sites, and the possibility of retaining a

higher number of mutations before being removed by selection (Otto 2000). Lastly, we cannot exclude a hybrid origin of the apomictic lineages, and hence an initial increased heterozigosity level could also have contributed to the observed rapid divergence.

In order to detect signs of selection, pairwise comparisons of dN/dS ratios were conducted, revealing similar distributions between apomicts and sexuals. Interestingly, among the outlier values (*i.e.* genes with elevated dN/dS ratio), enrichment for genes associated with reproduction (including meiosis and gametogenesis) was exclusively found for apomictic vs sexual pairwise comparisons (Figure 3.5 & Table 3.7). This observation can be explained by relaxed selective pressure on genes controlling reproduction in asexual lineages, and/or it could reflect the introgression of divergent alleles of genes controlling reproduction in the hybrid genome. This would be consistent with our findings regarding hybridization and apomixis in the complex, and could potentially be connected to the misregulation of sexual genes to induce apomictic seed production.

Indeed, signs of misregulation and differentially expressed genes between apomicts and sexuals were detected with the analysis of the microarray data, whereby transcriptomic expression levels throughout ovule development were compared between sexuals and apomicts. A total of 555 genes were found to be differentially expressed in at least one developmental stage, and among them eight showed a significant shift in expression patterns throughout development. Correlation of the expression data with cytomorphological observations revealed that most differentially expressed genes, as well as the most significant shift in expression, were found in the second and third developmental stages. Interestingly, these two stages are the key phases of divergence between AI formation vs normal meiotic development of the ovule, while smaller differerences were observed in the first and last stages of development. Across stages two and three the AI appears in parallel with normal meiosis and, by enlarging, it crushes the normal megagametophyte to take its place. Despite the lack of GO information for non-model species and, consequently, a poor characterization of the genes involved, this work clearly connects cytological observations with expression data, reducing the number of possible candidate genes associated with apomixis.

Narrowing down the list of possible candidate genes was also the objective of a further gene partitioning analysis that was based on the pattern of expression among

90

apomictic and sexual individuals, with reference to the work of Nogler (1984b) on the origin of the aposporous controlling factor (A$^-$). Based on the conclusions of Nogler (1984b), gene expression was classified into patterns reflecting transgressive, parent of origin and ploidy effects, with the highest number of differentially expressed genes found in the parent of origin class. While most of the classified genes could not be annotated, among those for which detailed genetic information was available, genes related to gametogenesis and the miRNA pathway potentially involved in meiosis were found, mostly in the transgressive class. The sparse availability of information does not permit further interpretation at this point, however the observed patterns could be an interesting objective to explore in further studies, in particular in combination with the results of Nogler's work.

This work represents the next steps towards a more detailed exploration of apomixis in *Ranunculus*. Overall, these results on the time of origin of the apomictic hybrid is concordant with the general estimation from the *Ranunculus* phylogeny (Emadzade & Hörandl 2011), thus supporting the validity of the methods used and offering a powerful tool for dating phylogenies. In addition, this work corroborates the existing scenario surrounding the origin of the *Raununculus* complex, and adds insights to the evolutionary processes following the initial hybridization event and induction of apomixis. Finally, my work has yielded a broad set of powerful tools that could be implemented for future studies on the system

REFERENCES

Adams, K.L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. J. Heredity. 98, 136–141.

Adams, K.L., and Wendel, J.F. (2005). Allele-Specific, Bidirectional Silencing of an Alcohol Dehydrogenase Gene in Different Organs of Interspecific Diploid Cotton Hybrids. Genetics 171, 2139–2142.

Adams, K.L., Percifield, R., and Wendel, J.F. (2004). Organ-Specific Silencing of Duplicated Genes in a Newly Synthesized Cotton Allotetraploid. Genetics 168, 2217–2226.

Albertin, W., Balliau, T., Brabant, P., Chèvre, A.-M., Eber, F., Malosse, C., and Thiellement, H. (2006). Numerous and Rapid Nonstochastic Modifications of Gene Products in Newly Synthesized Brassica napus Allotetraploids. Genetics 173, 1101–1113.

Aliyu, O.M., Schranz, M.E., and Sharbel, T.F. (2010). Quantitative variation for apomictic reproduction in the genus *Boechera* (Brassicaceae). American Journal of Botany 97, 1719.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Anderson, N.J., Badawi, D.Y., Grossniklaus, H.E., and Stulting, R.D. (2001). Posterior polymorphous membranous dystrophy with overlapping features of iridocorneal endothelial syndrome. Arch. Ophthalmol. *119*, 624–625.

Andersen, C.L., Jensen, J.L., and Orntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research 64, 5245–5250.

Armstrong, S.J., and Jones, G.H. (2003). Meiotic cytology and chromosome behaviour in wild type *Arabidopsis thaliana*. Journal of Experimental Botany 54, 1.

Asker, S., and Jerling, L. (1992). Apomixis in Plants (CRC Press).

Bajgain, P., Richardson, B.A., Price, J.C., Cronn, R.C., and Udall, J.A. (2011). Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (Artemisia tridentata). BMC Genomics 12, 370.

Barbazuk, W.B., Emrich, S., and Schnable, P.S. (2007). SNP Mining from Maize 454 EST Sequences. CSH Protocol 2007, pdb prot4786.

Barcaccia, G., and Albertini, E. (2013). Apomixis in plant reproduction: a novel perspective on an old dilemma. Plant Reproduction 26, 159–179.

Barcaccia, G., Mazzucato, A., Albertini, E., Zethof, J., Gerats, A., Pezzotti, M., and Falcinelli, M. (1998). Inheritance of parthenogenesis in *Poa pratensis* L.: auxin test and AFLP linkage analyses support monogenic control. Theoretical. Applied. Genetics. *97*, 74–82.

Barton, N.H. (2001). The role of hybridization in evolution. Molecular. Ecology. *10*, 551–568.

Barton, N.H., and Charlesworth, B. (1998). Why sex and recombination? Science 281, 1986–1990.

Beilstein, M.A., Al-Shehbaz, I.A., and Kellogg, E.A. (2006). Brassicaceae phylogeny and trichome evolution. American. Journal. Botany. *93*, 607–619

Bell, G. (1982). The Masterpiece of Nature: The Evolution and Genetics of Sexuality (Berkeley: University of California Press).

Bell, G.D.M., Kane, N.C., Rieseberg, L.H., and Adams, K.L. (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.

Bernstein, H., Byerly, H., Hopf, F., and RE, M. (1988). Is meiotic recombination an adaptation for repairing DNA, producing genetic variation, or both? In The Evolution of Sex, L.B. Michod RE, ed. (Sinauer Ass., Massachusetts: U.S.A), p. 139—160.

Bicknell, R.A., and Koltunow, A.M. (2004). Understanding Apomixis: Recent Advances and Remaining Conundrums. Plant Cell 16, S228–S245.

Birky, C.W. (1996). Heterozygosity, Heteromorphy, and Phylogenetic Trees in Asexual Eukaryotes. Genetics *144*, 427–437

Blanc, G., and Wolfe, K.H. (2004). Functional Divergence of Duplicated Genes Formed by Polyploidy during *Arabidopsis* Evolution. Plant Cell Online *16*, 1679–1691

Birky, J., C. W. (2004). Bdelloid rotifers revisited. Proceeding of the National Academy of Science U S A 101, 2651–2652.

Boecher, T. (1951). Cytological and embryological studies in the amphiapomictic *Arabis holboellii* complex. VI 7, 1–57.

Buggs, R.J.A., Chamala, S., Wu, W., Gao, L., May, G.D., Schnable, P.S., Soltis, D.E., Soltis, P.S., and Barbazuk, W.B. (2010). Characterization of duplicate gene evolution in the recent natural allopolyploid Tragopogon miscellus by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Molecular Ecology19, 132–146.

Bustin, S., Benes, V., Nolan, T., and Pfaffl, M. (2005). Quantitative real-time RT-PCR-a perspective. Journal of Molecular Endocrinology 34, 597.

Carman, J.G. (1997). Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biological Journal of the Linnean Society 61, 51–94.

Carman, J.G. (2007). Do duplicate genes cause In: apomixis. Apomixis: Evolution, Mechanisms and Perspectives. E. Hörandl, U. Grossniklaus, T. Sharbel, P. van Dijk (eds.),ARG Gantner Verlag KG, Lichtenstein 169–194.

Caryl, A.P., Armstrong, S.J., Jones, G.H., and Franklin, F.C.H. (2000). A homologue of the yeast HOP1 gene is inactivated in the *Arabidopsis* meiotic mutant asy1. Chromosoma 109, 62–71.

Ceplitis, A. (2003). Coalescence times and the Meselson effect in asexual eukaryotes. Genetical Research 82, 183–190.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6, 836–846.

Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., and Byers, B. (2000). Phenotypic Instability and Rapid Gene Silencing in Newly Formed *Arabidopsis* Allotetraploids. The Plant Cell Online 12, 1551–1567.

Comai, L., Madlung, A., Josefsson, C., and Tyagi, A. (2003). Do the different parental "heteromes" cause genomic shock in newly formed allopolyploids? Philosophical Transactions of the Royal Society B: Biological Sciences 358, 1149–1155.

Cork, J.M., and Purugganan, M.D. (2005). High-diversity genes in the *Arabidopsis* genome. Genetics 170, 1897–1911.

Corral, J.M., Vogel, H., Aliyu, O.M., Hensel, G., Thiel, T., Kumlehn, J., and Sharbel, T.F. (2013). A conserved Apomixis-specific polymorphism is correlated with exclusive exonuclease expression in premeiotic ovules of apomictic *Boechera* species. Plant Physiology. 163, 1660–1672.

Crow, J.F., and Kimura, M. (1965). Evolution in sexual and asexual populations. American Naturalist 439–450.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiology 139, 5–17.

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

Day, R.C., Grossniklaus, U., and Macknight, R.C. (2005). Be more specific! Laser-assisted microdissection of plant cells. Trends in Plant Science 10, 397–406.

Dean, J., Goodwin, P., and Hsiang, T. (2002). Comparison of relative RT-PCR and northern blot analyses to measure expression of 1, 3-glucanase in *Nicotiana benthamiana* infected with *Colltotrichum destructivum*. Plant Molecular Biology Reporter 20, 347–356.

Van Dijk, P.J. van, and Bakx-Schotman, J.M.T. (2004). Formation of unreduced megaspores (diplospory) in apomictic dandelions (*Taraxacum officinale*, s.l.) is controlled by a sex-specific dominant locus. Genetics *166*, 483–492.

Van Dijk, P.J., and Vijverberg, K. (2005). The significance of apomixis in the evolution of the angiosperms: a reappraisal. Plant Species-Level Systematics: New Perspectives on Pattern & Process 143, 101–116.

Ebina, M., Nakagawa, H., Yamamoto, T., Araya, H., Tsuruta, S., Takahara, M., and Nakajima, K. (2005). Co-segregation of AFLP and RAPD markers to apospory in Guineagrass (*Panicum maximum* Jacq.). Grassland. Science. *51*, 71–78.

Eckardt, N.A. (2003). Patterns of gene expression in apomixis. The Plant Cell Online 15, 1499–1501.

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 7, 191.

Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics 78, 737–756.

Fisher, R.A. (1930). The genetical theory of natural selection (Oxford: Clarendon Press).

Flagel, L.E., and Wendel, J.F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytologist 186, 184–193.

Frost, H.B. (1968). Seed reproduction : development of gametes and embryos. Citrus Ind. *2*, 290–324.

Gerstein, A.C., and Otto, S.P. (2009). Ploidy and the causes of genomic evolution. J Hered 100, 571–581.

Ghiselin, M.T. (1974). Economy of nature and the evolution of sex.Systematic Zoology, Vol 23, No 4, 560-562.

Gornall, R. (1999). Population genetic structure in agamospermous plants. Syst. Assoc. Spec. Vol. 57, 118–138.

Grimanelli, D., Leblanc, O., Espinosa, E., Perotti, E., González De León, D., and Savidan, Y. (1998). Mapping diplosporous apomixis in tetraploid *Tripsacum*: one gene or several genes? Heredity 80, 33–39.

Grimanelli, D., Leblanc, O., Perotti, E., and Grossniklaus, U. (2001). Developmental genetics of gametophytic apomixis. Trends in Genetics 17, 597–604.

Grimanelli, D., García, M., Kaszas, E., Perotti, E., and Leblanc, O. (2003). Heterochronic expression of sexual reproductivepPrograms during apomictic development in *Tripsacum*. Genetics 165, 1521–1531.

Grossniklaus, U. (2001). From sexuality to apomixis: Molecular and genetic approaches. In The Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, J.G. Carman, and T. Dresselhaus, eds. (Mexico: CIMMYT, IRD European Commission DG VI (FAIR)), pp. 168–211.

Grossniklaus, U., Nogler, G.A., and van Dijk, P.J. (2001). How to avoid sex: the genetic control of gametophytic apomixis. Plant Cell 13, 1491–1498.

Guo, M., Davis, D., and Birchler, J.A. (1996). Dosage Effects on Gene Expression in a Maize Ploidy Series. Genetics 142, 1349–1355.

Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. JSTOR 147–155.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J., *et al.* (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nature Genetics 39, 207–211.

Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R.N., Mascher, M., Schon, C.C., Taudien, S., Scholz, U., Stein, N., Mayer, K.F.X., *et al.* (2011). From RNA-seq to large-scale genotyping - genomics resources for rye (Secale cereale L.). BMC Plant Biology 11.

He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., *et al.* (2010). Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. Plant Cell 22, 17–33.

Hegarty, M.J., and Hiscock, S.J. (2005). Hybrid speciation in plants: new insights from molecular studies. New Phytologist 165, 411–423.

Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J., and Hiscock, S.J. (2006). Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. Current Biology 16, 1652–1659.

Hegarty, M.J., Barker, G.L., Brennan, A.C., Edwards, K.J., Abbott, R.J., and Hiscock, S.J. (2008). Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 3055–3069.

Van Hiel, M.B., Van Wielendaele, P., Temmerman, L., Van Soest, S., Vuerinckx, K., Huybrechts, R., Broeck, J.V., and Simonet, G. (2009). Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. BMC Molecular Biology 10, 56.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S., Kuroiwa, H., and Kuroiwa, T. (2001). Pollen tube attraction by the synergid cell. Science *293*, 1480–1483.

Hojsgaard, D., Greilhuber, J., Pellino, M., Paun, O., Sharbel, T.F., and Hörandl, E. (2014a). Emergence of apospory and bypass of meiosis via apomixis after sexual hybridisation and polyploidisation. New Phytol 204: 1000-1012.

Hojsgaard, D., Klatt, S., Baier, R., Carman, J.G., and Hörandl, E. (2014b). Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. Critical Reviews in Plant Sciences 33, 414–427.

Hojsgaard D, Hörandl E (2015) A little bit of sex matters for genome evolution in asexual plants. *Front. Plant Sci.* **6**:82. doi: 10.3389/fpls.2015.00082

Hörandl, E**.,** Dobeš, C., Lambrou, M. 1997. Chromosomen- und Pollen-untersuchungen an österreichischen Sippen des *Ranunculus auricomus*-Komplexes. Botanica. Helvetica. 107: 195-209.

Hörandl, E., Gutermann, W. 1998b. Der *Ranunculus auricomus*-Komplex in Österreich. Die *R. cassubicus*-, *R. monophyllus*- und *R. fallax*-Sammelgruppe. Bot. Jahrb. 120: 545-598.

Hörandl E. 2009. A combinational theory for maintenance of sex. Heredity 103: 445-457

Hollingsworth, N.M., and Byers, B. (1989). Hop1 - a Yeast Meiotic Pairing Gene. Genetics 121, 445–462.

Hong, S.Y., Seo, P.J., Yang, M.S., Xiang, F., and Park, C.M. (2008). Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. BMC Plant Biology 8, 112.

Hörandl, E. (1998). Species concepts in Agamic complexes: Applications in the *Ranunculus auricomus* complex and general perspectives. Folia Geobotanica 33, 335–348.

Hörandl, E. (2008). Evolutionary Implications of Self-Compatibility and Reproductive Fitness in the Apomictic *Ranunculus auricomus* Polyploid Complex (Ranunculaceae). International Journal of Plant Sciences 169, 1219–1228.

Hörandl, E. (2009). A combinational theory for maintenance of sex. Heredity 103, 445–457.

Hörandl, E. (2010). The evolution of self-fertility in apomictic plants. Sexual Plant Reproduction 23, 73–86.

Hörandl, M., E., Dobes, C., Lambrou (1997). Chromosomen- und Pollenuntersuchungen an österreichischen Arten des apomiktischen *Ranunculus auricomus*-Komplexes. Botanica Helvetica 107, 195–209.

Hörandl, E., and Greilhuber, J. (2002). Diploid and autotetraploid sexuals and their relationships to apomicts in the *Ranunculus cassubicus* group: insights from DNA content and isozyme variation. Plant Systematic Evololution 234, 85–100.

Hörandl, E., and Hojsgaard, D. (2012). The evolution of apomixis in angiosperms: A reappraisal. Plant Biosystems 146, 681–693.

Hörandl, E., Greilhuber, J., Klimova, K., Paun, O., Temsch, E., Emadzade, K., and Hodalova, I. (2009). Reticulate evolution and taxonomic concepts in the *Ranunculus auricomus* complex (Ranunculaceae): insights from analysis of morphological, karyological and molecular data. Taxon 58, 1194–1215.

Huson, D.H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology Evololution 23, 254–267.

Huson, D.H., Rupp, R., and Scornavacca, C. (2010). Phylogenetic Networks. Concepts, Algorithms and Applicaions (Cambridge Univ. Press).

Ilic, K., SanMiguel, P.J., and Bennetzen, J.L. (2003). A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Proceedings of the National Academy of Sciences. *100*, 12265–12270.

Izmailow, R. (1965). Macrosporogenesis in apomictic species *Ranunculus cassubicus*. Acta Biologica Cracoviensia Series Botanica 8, 183.

Izmailow, R. (1967). Observations in embryo and endosperm development in various chromosomic types of apomictic species *Ranunculus cassubicus* L. Acta Biologica Cracoviensia Series Botanica 10, 99 – 111.

Izmailow, R. (1996). Reproductive strategy in the *Ranunculus auricomus* complex (Ranunculaceae). Acta Societatis Botanicorum Poloniae 65, 167–170.

Judson, O.P., and Normark, B.B. (1996). Ancient asexual scandals. Trends in Ecology. & Evolution. *11*, 41–46.

Karron, J.D., Ivey, C.T., Mitchell, R.J., Whitehead, M.R., Peakall, R., and Case, A.L. (2012). New perspectives on the evolution of plant mating systems. Annual. Botany. *109*, 493–503.

Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., Vries, S. de, and Jong, H. de (2007). Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes. PNAS 104, 14026–14031.

Kearney, M. (2005). Hybridization, glaciation and geographical parthenogenesis. Trends in Ecology.& Evolution. *20*, 495–502.

Koltunow, A. (1993). Apomixis: Embryo Sacs and Embryos Formed without Meiosis or Fertilization in Ovules. Plant Cell 5, 1425–1437.

Koltunow, A.M. (2000). The genetic and molecular analysis of apomixis in the model plant *Hieracium*. Acta Biol. Cracoviensia Ser. Bot. *42*

Koltunow, A.M., and Grossniklaus, U. (2003a). APOMIXIS: A Developmental Perspective. Annual Review of Plant Biology 54, 547–574.

Kondrashov, A.S. (1982). Selection against harmful mutations in large sexual and asexual populations. Genetics Resources 40, 325–332.

Kondrashov, A.S. (1993). Classification of hypotheses on the advantage of amphimixis. J. Heredity. 84, 372–387.

Kotani, Y., Henderson, S.T., Suzuki, G., Johnson, S.D., Okada, T., Siddons, H., Mukai, Y., and Koltunow, A.M.G. (2014). The LOSS OF APOMEIOSIS (LOA) locus in *Hieracium*

praealtum can function independently of the associated large-scale repetitive chromosomal structure. New Phytologist 201, 973–981.

Lewontin, R.C. (1966). Hybridization as a source of variation to new environments. Evolution *20*, 315–336.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, J.W., Schmieder, R., Ward, R.M., Delenick, J., Olivares, E.C., and Mittelman, D. (2012). SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics 28, 1272–1273.

Lin, B.-Y. (1984). Ploidy Barrier to Endosperm Development in Maize. Genetics 107, 103–115.

Liu, B., and Wendel, J.F. (2000). Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43, 874–880.

Lippman, Z.B., and Zamir, D. (2007). Heterosis: revisiting the magic. Trends in Genetic. *23*, 60–66.

Liu, Z.-W., Wang, R.R.-C., and Carman, J.G. (1994). Hybrids and backcross progenies between wheat (*Triticum aestivum* L.) and apomictic Australian wheatgrass [Elymus rectisetus (Nees in Lehm.) A. Löve & Connor]: karyotypic and genomic analyses. Theoretical and. Applied. Genetics. *89*, 599–605.

Lord, E.M., and Russell, S.D. (2002). The Mechanisms of Pollination and Fertilization in Plants. Annual. Review of. Cell Developmental. Biology. *18*, 81–105.

Lovell, J.T., Aliyu, O.M., Mau, M., Schranz, M.E., Koch, M., Kiefer, C., Song, B.-H., Mitchell-Olds, T., and Sharbel, T.F. (2013). On the origin and evolution of apomixis in *Boechera*. Plant Reproduction. *26*, 309–315.

Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. The Quarterly Review of Biology. 257–290.

Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. (1993). The Mutational Meltdown in Asexual Populations. Journal of Heredity 84, 339–344.

Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science *290*, 1151–1155.

Mansfield, S.G., and Briarty, L.G. (1991). Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. Canadian Journal of Botany. *69*, 461–476.

Madlung, A., Tyagi, A.P., Watson, B., Jiang, H., Kagochi, T., Doerge, R.W., Martienssen, R., and Comai, L. (2005). Genomic changes in synthetic *Arabidopsis* polyploids. The Plant Journal 41, 221–230.

Marasek, A., Hasterok, R., Wiejacha, K., and Orlikowska, T. (2004). Determination by GISH and FISH of hybrid status in *Lilium*. Hereditas *140*, 1–7.

Marimuthu, M.P.A., Jolivet, S., Ravi, M., Pereira, L., Davda, J.N., Cromer, L., Wang, L., Nogué, F., Chan, S.W.L., Siddiqi, I., et al. (2011). Synthetic Clonal Reproduction Through Seeds. Science *331*, 876–876

Mark Welch, D., and Meselson, M. (2000). Evidence for the evolution of *bdelloid rotifers* without sexual reproduction or genetic exchange. Science 288, 1211–1215.

Martens, K. (1998). Sex and parthenogenesis: evolutionary ecology of reproductive modes in non-marine ostracods (Backhuys Publishers).

Matzk, F., Prodanovic, S., Bäumlein, H., and Schubert, I. (2005). The Inheritance of Apomixis in Poa pratensis Confirms a Five Locus Model with Differences in Gene Expressivity and Penetrance. Plant Cell Online *17*, 13–24.

Mau, M., Corral, J.M., Vogel, H., Melzer, M., Fuchs, J., Kuhlmann, M., Storme, N. de, Geelen, D., and Sharbel, T.F. (2013). The Conserved Chimeric Transcript UPGRADE2 Is Associated with Unreduced Pollen Formation and Is Exclusively Found in Apomictic *Boechera* Species. Plant Physiology 163, 1640–1659.

Mau, M.,Lovell, J., Corral, J.M., Kiefer, C., Koch, M., Aliyu O.M., Sharbel, T.,F. (2015) Hybrid apomicts trapped in the ecological niches of their sexual ancestors. Proceedings of the National Academy of Sciences. (in press)

Mayr, E. (1963). Animal species and evolution. xiv + 797 pp.

Maynard Smith, J. (1978). The evolution of sex (Cambridge: Cambridge University Press).

McClintock, B. (1950). The Origin and Behavior of Mutable Loci in Maize.Proceedings of the National Academy of Sciences of the United States of America, *36*(6), 344–355.

McMeniman, S., and Lubulwa, G. (1997). Project Development Assessment: An Economic Evaluation of the Potential Benefits of Integrating Apomixis into Hybrid Rice. Australia in International Agricultural Research.

Metcalfe, C.J., Bulazel, K.V., Ferreri, G.C., Schroeder-Reiter, E., Wanner, G., Rens, W., Obergfell, C., Eldridge, M.D.B., and O'Neill, R.J. (2007). Genomic Instability Within Centromeres of Interspecific Marsupial Hybrids. Genetics 177, 2507–2517.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. (2010). Tablet—next generation sequence assembly visualization. Bioinformatics 26, 401–402.

Mittwoch, U. (1978). Parthenogenesis. J Med Genet 15, 165–181.

Morell, V. (1999). Ecology Returns to Speciation Studies. Science *284*, 2106–2108.

Morgan, R., Ozias-Akins, P., and Hanna, W. (1998). Seed set in an apomictic BC 3 pearl millet. International Journal of Plant Sciences 159, 89–97.

Mogie, M. (1992). The evolution of asexual reproduction in plants (London; New York: Chapman & Hall).

Muller, H. (1932). Mussels and Clams: A Seasonal Quarantine-Bicarbonate of Soda as a Factor in the Prevention of Mussel Poisoning. California and Western Medicine *37*, 263–264.

Muller, H.J. (1964). The Relation of Recombination to Mutational Advance. Mutatation Research 1, 2–9.

Nassar, N.M. (2001). The nature of apomixis in cassava (*Manihot esculentum*, Crantz). Hereditas 134, 185–187.

Naumova, T.N., van der Laak, J., Osadtchiy, J., Matzk, F., Kravtchenko, A., Bergervoet, J., Ramulu, K.S., and Boutilier, K. (2001). Reproductive development in apomictic populations of *Arabis holboellii* (Brassicaceae). Sexual Plant Reproduction 14, 195–200.

Nelson, T., Gandotra, N., and Tausta, S.L. (2008). Plant cell types: reporting and sampling with new technologies. Current Opinion in Plant Biology 11, 567–573.

Nogler, G.A. (1984a). Gametophytic apomixis. In Embryology of Angiosperms, B.M. Johri, ed. (New York: Springer Verlag), pp. 475–518.

Nogler, G.A. (1984b). Genetics of apospory in apomictic *Ranunculus auricomus*. V: Conclusion. Botanica Helvetica 94, 411–422.

Novaes, E., Drost, D.R., Farmerie, W.G., Pappas, J., G. J., Grattapaglia, D., Sederoff, R.R., and Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics 9, 312.

Noyes, R.D., and Rieseberg, L.H. (2000). Two Independent Loci Control Agamospermy (Apomixis) in the Triploid Flowering Plant Erigeron annuus. Genetics *155*, 379–390.

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A., and Vielle-Calzada, J.-P. (2010). Control of female gamete formation by a small RNA pathway in *Arabidopsis*. Nature 464, 628–632.

Orsel, M., Krapp, A., and Daniel-Vedele, F. (2002). Analysis of the NRT2 nitrate transporter family in *Arabidopsis*. Structure and gene expression. Plant Physiology 129, 886.

Osborn, T.C., Chris Pires, J., Birchler, J.A., Auger, D.L., Jeffery Chen, Z., Lee, H.-S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., *et al.* (2003). Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19, 141–147.

Otto, S.P., and Whitton, J. (2000). Polyploid Incidence and Evolution. Annual review of genetics. *34*, 401–437.

Ozias-Akins, P. (2006). Apomixis: Developmental Characteristics and Genetics. Critical Reviews in Plant Sciences. *25*, 199–214.

Ozturk, Z.N., Talame, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R., and Bohnert, H.J. (2002). Monitoring large-scale changes in transcript abundance in drought-and salt-stressed barley. Plant Molecular Biology 48, 551–573.

Pagnussat, G.C., Yu, H.-J., Ngo, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.-F., Ye, D., and Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development 132, 603–614.

Paun, O., and Horandl, E. (2006a). Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus* carpaticola: directional bias at dinucleotide loci. Genetics 174, 387–398.

Paun, O., Stuessy, T.F., and Hörandl, E. (2006b). The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus* cassubicus complex. New Phytologist 171, 223–236.

Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H., and Sharbel, T.F. (2013). Asexual genome evolution in the apomictic *Ranunculus auricomus*

complex: examining the effects of hybridization and mutation accumulation. Molecular Ecology 22, 5908–5921.

Pessino, S.C., Evans, C., Ortiz, J.P.A., Armstead, I., Valle, C.B.D., and Hayward, M.D. (1998). A Genetic Map of the Apospory-Region in Brachiaria Hybrids: Identification of two Markers Closely Associated with the Trait. Hereditas *128*, 153–158.

Polegri, L., Calderini, O., Arcioni, S., and Pupilli, F. (2010). Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic *Paspalum* simplex Morong flowers. Journal of Experimental Botany. *61*, 1869–1883.

Pfaffl, M.W., Horgan, G.W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research 30, e36.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnology Letters 26, 509–515.

Pikaard, C.S. (2001). Genomic change and gene silencing in polyploids. Trends in Genetics 17, 675–677.

Ramsey, J., and Schemske, D.W. (2002). Neopolyploidy in Flowering Plants. Annual Review of Ecology and Systematics 33, 589–639.

Rapp, R.A., Udall, J.A., and Wendel, J.F. (2009). Genomic expression dominance in allopolyploids. BMC Biology 7, 18.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends in Genetics : TIG 16, 276–277.

Reiser, L., and Fischer, R. (1993). The Ovule and the Embryo Sac. Plant Cell *5*, 1291–1301.

Richards, A.J. (2003). Apomixis in flowering plants: an overview. Philosophical transactions of the Royal Society of London. 358, 1085–1093.

Rieseberg, L.H. (2009). Evolution: Replacing Genes and Traits through Hybridization. Current Biology 19, R119–R122.

Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J.L., Schwarzbach, A.E., Donovan, L.A., and Lexer, C. (2003). Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. Science 301, 1211–1216.

Rodrigues, J.C.M., Cabral, G.B., Dusi, D.M.A., Mello, L.V. de, Rigden, D.J., and Carneiro, V.T.C. (2003). Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of *Brachiaria brizantha*. Plant Molecular Biology. *53*, 745–757.

Ross, K.J., Fransz, P., and Jones, G.H. (1996). A light microscopic atlas of meiosis in *Arabidopsis thaliana*. Chromosome Research 4, 507–516.

Russell, S.D. (1979). Fine structure of megagametophyte development in *Zea mays*. Canadian. Journa of Botany. *57*, 1093–1110.

Salmona, J., Dussert, S., Descroix, F., de Kochko, A., Bertrand, B., and Joet, T. (2008). Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. Plant Molecular Biology 66, 105–124.

Sánchez-Morán, E., Benavente, E., and Orellana, J. (2001). Analysis of karyotypic stability of homoeologous-pairing (ph) mutants in allopolyploid wheats. Chromosoma 110, 371–377.

Savidan, Y. (1982). Nature and heridity of apomixis in *Panicum maximum*. Trav. Doc. ORSTOM Fr. No 153 159 pp.

Savidan, Y., Carman, J.G., and Dresselhaus, T. (2001). The Flowering of Apomixis: From Mechanisms to Genetic Engineering (CIMMYT).

Schallau, A., Arzenton, F., Johnston, A.J., Hähnel, U., Koszegi, D., Blattner, F.R., Altschmied, L., Haberer, G., Barcaccia, G., and Bäumlein, H. (2010). Identification and genetic analysis of the APOSPORY locus in *Hypericum perforatum* L. Plant J. *62*, 773–784.

Schaefer, I., Domes, K., Heethoff, M., Schneider, K., Schön, I., Norton, R.A., Scheu, S., and Maraun, M. (2006). No evidence for the "Meselson effect" in parthenogenetic oribatid mites (*Oribatida*, Acari). Journal of Evolution Biology 19, 184–193.

Schmid, M.W., Schmidt, A., Klostermeier, U.C., Barann, M., Rosenstiel, P., and Grossniklaus, U. (2012). A Powerful Method for Transcriptional Profiling of Specific Cell Types in Eukaryotes: Laser-Assisted Microdissection and RNA Sequencing. PLoS ONE 7, e29685.

Schmidt, A., Schmid, M.W., Klostermeier, U.C., Qi, W., Guthörl, D., Sailer, C., Waller, M., Rosenstiel, P., and Grossniklaus, U. (2014). Apomictic and Sexual Germline

Development Differ with Respect to Cell Cycle, Transcriptional, Hormonal and Epigenetic Regulation. PLoS Genet 10, e1004476.

Schmittgen, T.D., and Zakrajsek, B.A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. Journal of Biochemical and Biophysical Methods 46, 69–81.

Schön, I., and Martens, K. (2003). No slave to sex. Proceedings of the Royal Society of London B 270, 827–833.

Schranz, M.E., Dobeš, C., Koch, M.A., and Mitchell-Olds, T. (2005). Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). American. Journal of Botany. *92*, 1797–1810.

Schranz, M.E., Windsor, A.J., Song, B., Lawton-Rauh, A., and Mitchell-Olds, T. (2007). Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. Plant Physiology 144, 286.

Schwander, T., and Crespi, B.J. (2009). Multiple Direct Transitions from Sexual Reproduction to Apomictic Parthenogenesis in Timema Stick Insects. Evolution *63*, 84–103

Schwartz, C., Balasubramanian, S., Warthmann, N., Michael, T.P., Lempe, J., Sureshkumar, S., Kobayashi, Y., Maloof, J.N., Borevitz, J.O., Chory, J., *et al.* (2009). Cis-regulatory Changes at FLOWERING LOCUS T Mediate Natural Variation in Flowering Responses of *Arabidopsis thaliana*. Genetics 183, 723–732.

Singh, M., Burson, B.L., and Finlayson, S.A. (2007). Isolation of candidate genes for apomictic development in buffelgrass (*Pennisetum ciliare*). Plant Molecular. Biology. *64*, 673–682.

Sharbel, T.F., Voigt, M.L., Corral, J.M., Thiel, T., Varshney, A., Kumlehn, J., Vogel, H., and Rotter, B. (2009). Molecular signatures of apomictic and sexual ovules in the *Boechera* holboellii complex. The Plant Journal 58, 870–882.

Sharbel, T.F., Voigt, M.L., Corral, J.M., Galla, G., Kumlehn, J., Klukas, C., Schreiber, F., Vogel, H., and Rotter, B. (2010). Apomictic and sexual ovules of *Boechera* display heterochronic global gene expression patterns. The Plant Cell 22, 655–671.

Silveira, E.D., Alves-Ferreira, M., Guimaraes, L.A., da Silva, F.R., and Carneiro, V.T.C. (2009). Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha*. BMC Plant Biology 9, 84.

Silver, N., Cotroneo, E., Proctor, G., Osailan, S., Paterson, K.L., and Carpenter, G.H. (2008). Selection of housekeeping genes for gene expression studies in the adult rat submandibular gland under normal, inflamed, atrophic and regenerative states. BMC Molecular Biology 9, 64.

Singh, M., Goel, S., Meeley, R.B., Dantec, C., Parrinello, H., Michaud, C., Leblanc, O., and Grimanelli, D. (2011). Production of Viable Gametes without Meiosis in Maize Deficient for an ARGONAUTE Protein. Plant Cell 23, 443–458.

Smith, J.M. (1993). The Theory of Evolution (Cambridge University Press).

Smith, M.J. (1978). The Evolution of Sex (Cambridge University Press).

Soltis, D.E., and Soltis, P.S. (1999). Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution 14, 348–352.

Song, Y., Leonard, S.W., Traber, M.G., and Ho, E. (2009). Zinc deficiency affects DNA damage, oxidative stress, antioxidant defenses, and DNA repair in rats. J Nutr 139, 1626–1631.

Spillane, C., Curtis, M.D., and Grossniklaus, U. (2004). Apomixis technology development–virgin births in farmers' fields? Nature Biotechnology 22, 687–691.

Sturzenbaum, S.R., and Kille, P. (2001). Control genes in quantitative molecular biological techniques: the variability of invariance. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 130, 281–289.

Strasburg, J.L., and Kearney, M. (2005). Phylogeography of sexual *Heteronotia binoei* (Gekkonidae) in the Australian arid zone: climatic cycling and repetitive hybridization. Molecular. Ecolology. *14*, 2755–2772.

Suomalainen, E. (1950). Parthenogenesis in animals. Advance. Genetics. 3, 193–253.

Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., and Schnable, P.S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proceedings of the National Academy of Sciences 103, 6805–6810.

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

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. Journal of Biotechnology 75, 291–295.

Thomas, C., Meyer, D., Wolff, M., Himber, C., Alioua, M., and Steinmetz, A. (2003). Molecular characterization and spatial expression of the sunflower ABP1 gene. Plant Molecular Biology 52, 1025–1036.

Trick, M., Long, Y., Meng, J., and Bancroft, I. (2009). Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnology Journal 7, 334–346.

Van Valen, L. (1973). A new evolutionary law. Evolutionary Theory 1, 1–30.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, 34.

Vielle-Calzada, J.-P., Nuccio, M.L., Budiman, M.A., Thomas, T.L., Burson, B.L., Hussey, M.A., and Wing, R.A. (1996). Comparative gene expression in sexual and apomictic ovaries of *Pennisetum ciliare* (L.) Link. Plant Molecular. Biology. *32*, 1085–1092.

Vogel, H., and Wheat, C.W. (2011). Accessing the Transcriptome: How to Normalize mRNA Pools. In Molecular Methods for Evolutionary Genetics, V. Orgogozo, and M.V. Rockman, eds. (Humana Press), pp. 105–128.

Vogel, H., Heidel, A.J., Heckel, D.G., and Groot, A.T. (2010). Transcriptome analysis of the sex pheromone gland of the noctuid moth *Heliothis virescens*. BMC Genomics 11, 29.

Warrington, J.A., Nair, A., Mahadevappa, M., and Tsygankaya, M. (2000). Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiological Genomics 2, 143.

Windsor, A.J., Schranz, M.E., Formanova, N., Gebauer-Jung, S., Bishop, J.G., Schnabelrauch, D., Kroymann, J., and Mitchell-Olds, T. (2006). Partial shotgun sequencing of the *Boechera stricta* genome reveals extensive microsynteny and promoter conservation with *Arabidopsis*. Plant Physiology 140, 1169.

Winkler, H.K.A. (1908). Parthenogenesis und apogamie im pflanzenreiche (G. Fischer).

Wolfe, K.H., Gouy, M.L., Yang, Y.W., Sharp, P.M., and Li, W.H. (1989). Date of the Monocot Dicot Divergence Estimated from Chloroplast DNA-Sequence Data. Proceeding of Nationall Academy of Science U S A 86, 6201–6205.

Yamauchi, A., Hosokawa, A., Nagata, H., and Shimoda, M. (2004). Triploid Bridge and Role of Parthenogenesis in the Evolution of Autopolyploidy. The American Naturalist 164, 101–112.

Yang, C.F., Sun, S., and Guo, Y. (2005). Resource limitation and pollen source (self and outcross) affecting seed production in two louseworts, *Pedicularis siphonantha* and *P. longiflora* (Orobanchaceae). Botanical Journal of the Linnean Society 147, 83–89.

Yoo, M.-J., Szadkowski, E., and Wendel, J.F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110, 171–180.

Zhao, S., and Fernald, R.D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. Journal of Computational Biology 12, 1047–1064.

Acknowledgments

First of all I would like to thank my supervisor and friend Tim Sharbel for his support, patient and his guidance in this long journey. Without his lead I would have never been able to scientifically accomplish what I have been doing so far and most importantly, what I have been learning. I would like to thanks all the member of the apomixis research group with special mention to Samuel Amyteye for supporting me at the very beginning of my work and Jose' Corral for the incredible amount of inputs and discussions. Thomas Thiel for mercilessly destroying any "good" idea I was formulating and by doing that teaching me the power of constructive criticism. Marta Molins for being always available to listen and help me in the moments of sadness. Martin Mau for showing me how a real scientist works. Dorota for the breeze of Evolutionary attitude brought to the office. Lars for being simply a very good soul, a characteristic very difficult to find in these days. And a special thanks to all the guest that have been coming and going in our multicolored group. Diego,Javier, Teresa, Giulio and many more

A mention to all the technicians of the group for being reliable, efficient and friendly, to Heidi Vera and Leane.

To all the friend and people I have meet at the IPK , for all they have been giving to me. Their company made the life in Gatersleben a unique experience that I will always remember. To Nath, Hart, Nadine, Katja, Andre', Takaioshi, Celia, Etienne, Paride, Francesca, and many,many, more.

To my family for never give up on me

Finally to Jana for being the best things it ever happened in my life.

APPENDIX

Appendix I Table showing Blast2Go results for Apomixis specific downregulated transcripts in the first developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
354822	probable histone -like	3.02E-47	98.50%	C:cell wall; C:nucleosome; C:nucleolus; P:DNA mediated transformation; P:response to wounding; P:response to bacterium; F:DNA binding; P:nucleosome assembly
409672	like superfamily protein	2.33E-52	91.75%	P:protein targeting to vacuole; P:response to gibberellin stimulus; C:intracellular; C:plasma membrane; P:N-terminal protein myristoylation; P:regulation of protein localization; F:structural molecule activity
558906	PREDICTED: uncharacterized protein LOC100243721	2.65E-17	88.45%	-
413767	trichome birefringence-like 13 protein	6.85E-50	85.10%	C:Golgi apparatus
405548	dna-dependent atpase snf2h	2.89E-06	84.90%	F:helicase activity; P:chromatin remodeling; F:DNA binding; C:nucleus; F:nucleic acid binding; F:ATP binding; F:nucleosome binding; P:ATP-dependent chromatin remodeling; F:hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides; C:chromatin remodeling complex; F:hydrolase activity, acting on acid anhydrides; F:transferase activity
505317	cysteine-rich rlk (receptor-like protein kinase) 8	9.31E-42	76.20%	F:nucleic acid binding; F:zinc ion binding; P:DNA integration
362712	tyrosine--trna ligase-like	5.47E-06	76.00%	F:ATP binding; C:cytoplasm; F:aminoacyl-tRNA ligase activity; F:nucleotide binding; P:tRNA aminoacylation for protein translation; F:tyrosine-tRNA ligase activity
408581	glutamyl-trna synthetase	8.26E-53	74.10%	C:cytosol; P:proteasomal protein catabolic process; P:glutamyl-tRNA aminoacylation; P:nucleotide biosynthetic process; P:cytoskeleton organization; F:ATP binding; P:gluconeogenesis; F:glutamate-tRNA ligase activity
350516	non-ltr retroelement reverse transcriptase-like protein	5.72E-06	73.06%	-
345754	polyadenylate-binding protein 2-like	8.40E-14	71.25%	F:nucleic acid binding
367772	uncharacterized mitochondrial protein g00810-like	1.14E-39	70.15%	P:metabolic process; F:binding
493073	protein gamete expressed 2-like	2.83E-32	68.95%	P:pollen sperm cell differentiation
410617	uncharacterized partial	1.20E-07	65.25%	-
438526	probable disease resistance protein at5g43740-like	5.47E-04	64.50%	-
557080	en spm-like transposon protein	6.77E-14	61.50%	-
367646	hypothetical protein PRUPE_ppa016504mg, partial	1.49E-22	58.15%	F:transferase activity, transferring phosphorus-containing groups
359667	f-box kelch-repeat protein	6.85E-09	56.75%	-
496594	water chloroplastic-like	4.24E-08	55.40%	P:starch metabolic process; P:phosphorylation; F:kinase activity; C:chloroplast
405690	tnp2-like transposon protein	0	55.35%	P:proteolysis; F:cysteine-type peptidase activity
501235	f-box protein fbw2-like	3.45E-19	55.00%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
345809	hypothetical protein VITISV_027817	0	54.90%	F:binding; F:catalytic activity
408467	myosin-h heavy chain-like	6.33E-29	54.50%	F:binding

366390	protein	5.23E-24	54.05%	F:exonuclease activity; P:RNA processing; F:endonuclease activity; C:intracellular; F:nucleic acid binding; F:zinc ion binding
350311	mitochondrial ubiquitin ligase activator of nfkb 1-like	1.29E-42	53.60%	F:metal ion binding; F:acid-amino acid ligase activity; F:zinc ion binding; F:ligase activity
410564	ribonuclease h protein at1g65750-like	7.71E-19	52.75%	F:nucleic acid binding; F:ribonuclease H activity; F:zinc ion binding; C:intracellular
493755	hypothetical protein VITISV_014784	1.75E-13	52.25%	F:binding; F:catalytic activity
497978	predicted protein	3.12E-15	49.25%	P:proteolysis; F:cysteine-type peptidase activity
497254	PREDICTED: uncharacterized protein LOC101305890	1.33E-17	48.00%	-
405168	bed finger-nbs-lrr resistance protein	1.85E-63	47.10%	P:defense response; F:ADP binding; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity; F:DNA binding
371874	ribonuclease h protein at1g65750-like	1.76E-10	46.95%	F:RNA binding; F:nucleic acid binding; P:RNA-dependent DNA replication; F:ribonuclease H activity; F:RNA-directed DNA polymerase activity; F:zinc ion binding
352336	f-box rni superfamily	3.97E-28	45.75%	F:molecular_function; P:biological_process; C:cellular_component
496232	PREDICTED: uncharacterized protein LOC101305890	6.97E-30	45.50%	-
408815	f-box lrr-repeat protein	8.94E-27	45.45%	F:molecular_function; P:biological_process; C:cellular_component; P:virus induced gene silencing; P:photoperiodism, flowering; P:vegetative phase change
497280	PREDICTED: uncharacterized protein LOC101305890	3.66E-15	45.00%	-
497873	embryo sac development arrest	1.25E-13	44.60%	-
348855	fimbriata-like protein	1.58E-06	44.38%	-
97980	non-ltr retroelement reverse transcriptase	1.37E-06	43.00%	F:nucleic acid binding; P:defense response; F:ADP binding; P:signal transduction; C:intracellular; C:mitochondrion

Appendix II Table showing Blast2Go results for Apomixis specific downregulated transcripts in the second developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
376997	unnamed protein product	5.14E-06	74.50%	-
410617	uncharacterized partial	1.20E-07	65.25%	-
414199	uncharacterized aarf domain-containing protein kinase chloroplastic-like	1.24E-30	86.85%	C:plastoglobule; F:protein kinase activity; P:protein phosphorylation; F:ATP binding; C:mitochondrion
354881	udp-glycosyltransferase 76f1-like	3.33E-40	64.20%	F:transferase activity, transferring glycosyl groups
552674	udp-glycosyltransferase 76f1-like	3.69E-103	67.90%	F:transferase activity, transferring glycosyl groups
413767	trichome birefringence-like 13 protein	6.85E-50	85.10%	C:Golgi apparatus
351009	transcription initiation factor iif subunit alpha-like	9.90E-16	74.60%	P:transcription initiation from RNA polymerase II promoter; F:DNA binding; C:thylakoid; C:membrane part; C:chloroplast; P:positive regulation of transcription, DNA-dependent; F:catalytic activity; C:nucleus
494961	transcription factor myc2-like	7.07E-115	59.65%	P:oxidation-reduction process; F:oxidoreductase activity; F:2-alkenal reductase [NAD(P)] activity
405690	tnp2-like transposon protein	0	55.55%	P:proteolysis; F:cysteine-type peptidase activity
347446	subtilisin-like protease-like	0	73.00%	F:identical protein binding; F:serine-type endopeptidase activity; P:proteolysis; P:negative regulation of catalytic activity
361624	subtilisin-like protease-like	1.63E-63	78.50%	F:identical protein binding; F:serine-type endopeptidase activity; C:cytoplasmic membrane-bounded vesicle; C:extracellular region; P:proteolysis; P:negative regulation of catalytic activity
351887	small heat-shock protein	7.32E-06	54.67%	P:response to stress; C:plastid; C:chloroplast stroma; C:chloroplast; C:mitochondrion
552614	serine threonine-protein kinase ppk4	2.27E-10	43.47%	P:cellular process
350118	ring zinc finger expressed	1.06E-91	71.25%	F:metal ion binding
371874	ribonuclease h protein at1g65750-like	1.76E-10	46.95%	F:RNA binding; F:nucleic acid binding; P:RNA-dependent DNA replication; F:ribonuclease H activity; F:RNA-directed DNA polymerase activity; F:zinc ion binding
410564	ribonuclease h protein at1g65750-like	7.72E-19	52.75%	F:nucleic acid binding; F:ribonuclease H activity; F:zinc ion binding; C:intracellular
496351	protein transparent testa 12-like	2.36E-09	81.30%	P:drug transmembrane transport; C:membrane; F:antiporter activity; F:drug transmembrane transporter activity
500535	protein chromosomal-like	3.05E-06	69.00%	F:ATP binding; C:chloroplast; F:nucleotide binding; F:nucleoside-triphosphatase activity
354822	probable histone -like	3.02E-47	98.50%	C:cell wall; C:nucleosome; C:nucleolus; P:DNA mediated transformation; P:response to wounding; P:response to bacterium; F:DNA binding; P:nucleosome assembly
438526	probable disease resistance protein at5g43740-like	5.47E-04	64.50%	-
497978	predicted protein	3.12E-15	49.25%	P:proteolysis; F:cysteine-type peptidase activity
355574	pentatricopeptide repeat-containing protein mitochondrial-like	2.67E-106	61.70%	-
383751	pentatricopeptide repeat-containing protein mitochondrial-like	3.14E-14	65.35%	-
351751	non-ltr retroelement reverse transcriptase	4.04E-64	61.30%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:endonuclease activity
492093	non-ltr retroelement reverse transcriptase	3.65E-21	48.80%	F:nucleic acid binding; C:mitochondrion; F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:ribonuclease H activity
413184	nadh dehydrogenase complex assembly factor 6-like	5.52E-09	84.83%	P:biosynthetic process; F:transferase activity

350311	mitochondrial ubiquitin ligase activator of nfkb 1-like	1.29E-42	53.60%	F:metal ion binding; F:acid-amino acid ligase activity; F:zinc ion binding; F:ligase activity
366122	lob domain-containing protein 12-like	6.79E-38	75.60%	P:leaf morphogenesis
490287	hxxxd-type acyl-transferase family	5.72E-103	56.85%	F:transferase activity
499052	glucan endo- -beta-glucosidase-like protein 2-like	4.97E-13	85.00%	C:plant-type cell wall; C:anchored to plasma membrane; F:polysaccharide binding; P:callose deposition in cell wall; C:plasmodesma; F:hydrolase activity
413670	gdsl esterase lipase at5g45960-like	2.85E-78	66.40%	F:hydrolase activity, acting on ester bonds
555798	f-box rni-like superfamily protein	1.41E-15	43.75%	F:molecular_function; C:cellular_component
414875	f-box rni superfamily	1.01E-31	47.05%	F:molecular_function; P:biological_process; C:cellular_component
501235	f-box protein fbw2-like	3.45E-19	55.00%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
408255	f-box protein cpr30-like	6.31E-10	47.38%	-
416257	f-box protein	2.16E-06	52.09%	-
497465	en spm-like transposon protein	1.33E-06	55.94%	P:proteolysis; F:cysteine-type peptidase activity
557080	en spm-like transposon protein	6.77E-14	61.50%	-
497873	embryo sac development arrest	1.25E-13	44.60%	-
405548	dna-dependent atpase snf2h	2.89E-06	84.90%	F:helicase activity; P:chromatin remodeling; F:DNA binding; C:nucleus; F:nucleic acid binding; F:ATP binding; F:nucleosome binding; P:ATP-dependent chromatin remodeling; F:hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides; C:chromatin remodeling complex; F:hydrolase activity, acting on acid anhydrides; F:transferase activity
371039	dna replication complex gins protein sld5-like	8.31E-20	81.60%	C:GINS complex; P:DNA replication initiation; P:regulation of cell cycle; P:regulation of DNA replication
496232	PREDICTED: uncharacterized protein LOC101305890	6.97E-30	45.50%	-
497254	PREDICTED: uncharacterized protein LOC101305890	1.33E-17	48.00%	-
497280	PREDICTED: uncharacterized protein LOC101305890	3.66E-15	45.00%	-
558906	PREDICTED: uncharacterized protein LOC100243721	2.64E-17	88.45%	-
409672		2.33E-52	91.75%	P:protein targeting to vacuole; P:response to gibberellin stimulus; C:intracellular; C:plasma membrane; P:N-terminal protein myristoylation; P:regulation of protein localization; F:structural molecule activity

Appendix III Table showing Blast2Go results for Apomixis specific downregulated transcripts in the third developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
490083	60s ribosomal protein l31	7.89E-15	96.75%	C:ribosome; F:structural constituent of ribosome; P:translation
559204	lob domain-containing protein 12-like	2.66E-37	95.00%	-
497568	peptidyl-prolyl cis-trans isomerase cyclophilin-type family protein	1.98E-21	94.70%	P:sugar mediated signaling pathway; P:embryo development ending in seed dormancy; P:mRNA export from nucleus; P:post-translational protein modification; P:root hair cell differentiation; P:protein deubiquitination; P:response to gamma radiation; P:response to freezing; P:stamen development; P:DNA methylation; P:regulation of root meristem growth; P:protein peptidyl-prolyl isomerization; P:RNA processing; P:protein ubiquitination; P:maintenance of meristem identity; P:mitotic recombination; P:telomere maintenance in response to DNA damage; P:leaf shaping; P:cullin deneddylation; P:seed germination; P:trichome morphogenesis; C:Cul4-RING ubiquitin ligase complex; P:meristem initiation; P:leaf formation; P:regulation of telomere maintenance; P:seed dormancy process; P:lipid storage; P:leaf vascular tissue pattern formation; P:protein import into nucleus; P:mitotic cell cycle; F:histone binding; P:actin nucleation; P:sepal formation; P:vegetative to reproductive phase transition of meristem; P:embryo sac egg cell differentiation; F:chromatin binding; P:photomorphogenesis; P:regulation of flower development; P:regulation of histone methylation; P:histone H3-K9 methylation; P:reciprocal meiotic recombination; P:cell adhesion; P:positive regulation of transcription, DNA-dependent; P:gene silencing by RNA; P:petal formation; F:peptidyl-prolyl cis-trans isomerase activity; P:determination of bilateral symmetry; P:carpel development; C:nucleus; P:protein folding; P:cell wall organization
492001	peroxidase	9.42E-20	92.85%	P:response to oxidative stress; P:oxidation-reduction process; P:response to fructose stimulus; P:response to salt stress; F:metal ion binding; F:peroxidase activity; P:water transport; P:cellular cation homeostasis; F:heme binding; P:divalent metal ion transport; P:glucosinolate biosynthetic process; P:cysteine biosynthetic process
409672	like superfamily protein	2.33E-52	91.75%	P:protein targeting to vacuole; C:plasma membrane; P:N-terminal protein myristoylation; C:nucleus; P:regulation of protein localization; F:structural molecule activity
352592	seryl-trna synthetase	2.38E-140	91.05%	P:sulfur amino acid metabolic process; P:regulation of protein dephosphorylation; F:serine-tRNA ligase activity; P:embryo development ending in seed dormancy; P:photosystem II assembly; P:chloroplast relocation; C:chloroplast; P:seryl-tRNA aminoacylation; P:iron-sulfur cluster assembly; P:leaf morphogenesis; P:cellular amino acid biosynthetic process; P:vegetative to reproductive phase transition of meristem; P:serine family amino acid metabolic process; P:cell differentiation; P:mitochondrion organization; P:positive regulation of transcription, DNA-dependent; P:transcription from plastid promoter; P:rRNA processing; F:ATP binding; P:ovule development; P:cell wall modification; P:thylakoid membrane organization; C:mitochondrion
347410	abc transporter b family member 15-like	0	90.20%	P:transmembrane transport; F:xenobiotic-transporting ATPase activity; C:integral to membrane; P:ATP catabolic process; F:ATP binding
375653	probable lrr receptor-like serine threonine-protein kinase at5g45780-like	1.22E-90	90.05%	P:embryo sac egg cell differentiation; P:protein phosphorylation; P:oxidation-reduction process; F:protein serine/threonine kinase activity; C:integral to membrane; P:transmembrane receptor protein tyrosine kinase signaling pathway; F:ATP binding; F:2-alkenal reductase [NAD(P)] activity; C:plasma membrane
416239	monodehydroascorbate reductase	0	89.20%	C:cytoplasm; F:flavin adenine dinucleotide binding; P:cell redox homeostasis; F:monodehydroascorbate reductase (NADH) activity; P:oxidation-reduction process
558906	PREDICTED: uncharacterized protein LOC100243721	2.65E-17	88.45%	-
346432	atp-dependent dna helicase ddm1-like	1.15E-115	87.40%	P:methylation-dependent chromatin silencing; P:chromatin silencing by small RNA; P:RNA interference; P:negative regulation of histone H4 acetylation; P:maintenance of chromatin silencing; P:cell proliferation; F:transferase activity; C:nucleosome; P:DNA mediated transformation; F:DNA binding; P:histone phosphorylation; P:regulation of DNA methylation; F:ATP-dependent DNA helicase activity; P:regulation of gene expression by genetic imprinting; P:ATP catabolic process; P:transposition, RNA-mediated; F:ATP binding; P:positive regulation of histone H3-K9 methylation
350434	aldolase-type tim barrel family protein	0	86.75%	P:oxidation-reduction process; C:peroxisome; P:defense response to bacterium; F:long-chain-(S)-2-hydroxy-long-chain-acid

116

				oxidase activity; F:glycolate oxidase activity; F:medium-chain-(S)-2-hydroxy-acid oxidase activity; P:defense response signaling pathway, resistance gene-independent; F:very-long-chain-(S)-2-hydroxy-acid oxidase activity; P:hydrogen peroxide biosynthetic process; F:FMN binding
388531	probable atp-dependent dna helicase hfm1-like	1.62E-58	86.65%	P:cytokinesis by cell plate formation; F:DNA helicase activity; P:DNA duplex unwinding; P:meiotic chromosome segregation; F:nucleic acid binding; P:regulation of cell cycle process; P:chiasma assembly; P:double-strand break repair; F:ATP-dependent helicase activity; P:mitotic recombination; P:meiotic DNA double-strand break formation; F:ATP binding; P:sister chromatid cohesion; P:mitosis; P:reciprocal meiotic recombination; C:nucleus
413711	kinesin-like calmodulin-binding protein zwichel	1.56E-09	85.20%	F:oxidoreductase activity; P:pollen germination; F:microtubule motor activity; F:microtubule binding; P:oxidation-reduction process; F:ATPase activity; P:ATP catabolic process; P:microtubule-based movement; C:plasma membrane; F:calmodulin binding; C:kinesin complex; F:ATP binding; P:trichome branching
413767	trichome birefringence-like 13 protein	6.85E-50	85.10%	C:chloroplast; C:Golgi apparatus
368755	dna-directed rna polymerases and iii subunit rpabc4-like	1.33E-18	84.95%	F:DNA-directed RNA polymerase activity; C:DNA-directed RNA polymerase IV complex; C:DNA-directed RNA polymerase V complex; C:DNA-directed RNA polymerase II, core complex; F:DNA binding; P:pollen tube growth; P:RNA splicing, via endonucleolytic cleavage and ligation; P:transcription from RNA polymerase II promoter
413184	nadh dehydrogenase complex assembly factor 6-like	5.52E-09	84.83%	P:protein folding; P:protein targeting to mitochondrion; P:response to high light intensity; P:biosynthetic process; P:response to hydrogen peroxide; F:transferase activity; P:response to heat
497996	protein	1.26E-18	83.69%	C:extracellular region; P:pathogenesis
492646	probably inactive leucine-rich repeat receptor-like protein kinase at5g48380-like	1.99E-157	83.15%	P:protein phosphorylation; F:protein kinase activity; P:negative regulation of defense response; C:integral to membrane; C:cytoplasmic membrane-bounded vesicle; F:ATP binding; C:plasma membrane
424479	protein disulfide-isomerase-like	6.04E-09	82.80%	P:protein folding; C:endoplasmic reticulum lumen; F:electron carrier activity; P:cell redox homeostasis; P:glycerol ether metabolic process; F:protein disulfide isomerase activity; F:dolichyl-diphosphooligosaccharide-protein glycotransferase activity; F:protein disulfide oxidoreductase activity
369247	symplekin tight junction protein domain-containing protein	2.71E-10	82.00%	C:cytoplasm
554147	endoribonuclease dicer homolog 2-like	6.28E-72	81.95%	F:ribonuclease III activity; P:cell-cell signaling; F:protein binding; F:ATP-dependent helicase activity; P:covalent chromatin modification; P:vegetative phase change; P:determination of bilateral symmetry; P:chromatin silencing; P:maintenance of DNA methylation; F:DNA binding; P:meristem initiation; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:double-stranded RNA binding; P:positive regulation of transcription, DNA-dependent; P:post-translational protein modification; P:mRNA splicing, via spliceosome; P:meristem maintenance; F:ATP binding; C:nucleus; P:virus induced gene silencing
408414	nc domain-containing protein	1.65E-18	81.60%	P:signal transduction
496351	protein transparent testa 12-like	2.36E-09	81.30%	P:transmembrane transport; F:transmembrane transporter activity
556773	nadph:adrenodoxin mitochondrial-like	1.74E-95	81.25%	P:hydrogen peroxide catabolic process; F:oxidoreductase activity; C:plasma membrane; F:electron carrier activity; C:mitochondrial matrix; P:oxidation-reduction process; F:nucleotide binding
345547	protease 2-like	0	80.75%	C:membrane; P:proteolysis; C:nucleus; F:serine-type endopeptidase activity
561916	pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta	2.79E-14	79.20%	C:6-phosphofructokinase complex; C:plastid; C:membrane; P:embryo development ending in seed dormancy; P:histone H3-K9 methylation; F:diphosphate-fructose-6-phosphate 1-phosphotransferase activity; P:photosynthesis; P:phosphorylation; C:cell wall; P:chromatin silencing; P:glycolysis; P:response to cadmium ion; C:pyrophosphate-dependent phosphofructokinase complex, beta-subunit complex; P:fructose 6-phosphate metabolic process; P:acetyl-CoA biosynthetic process; F:6-phosphofructokinase activity; F:ATP binding
552928	trichome birefringence-like 27	0	78.95%	-
497224	solanesyl diphosphate synthase chloroplastic mitochondrial-like	5.11E-08	78.00%	P:isoprenoid biosynthetic process; F:trans-hexaprenyltranstransferase activity
529192	cytochrome p450 78a3	3.29E-46	77.85%	F:metal ion binding; F:monooxygenase activity; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen

415996	hypothetical protein VITISV_044100	9.28E-09	77.78%	-
433826	at4g24120-like partial	1.70E-08	76.60%	P:transmembrane transport
382767	double-stranded rna-binding protein 4-like	1.40E-06	76.50%	-
354651	pentatricopeptide repeat-containing partial	1.30E-46	76.45%	P:mitochondrial mRNA modification; C:chloroplast; C:mitochondrion
557673	14 kda proline-rich	1.21E-09	76.33%	-
351009	transcription initiation factor iif subunit alpha-like	9.90E-16	74.60%	C:thylakoid; C:membrane part; P:transcription, DNA-dependent
352874	band 7 family protein	1.29E-11	73.94%	C:intracellular membrane-bounded organelle; C:cytoplasmic part; C:intracellular organelle part
552252	eukaryotic translation initiation factor 3 subunit l-like	1.51E-16	73.75%	P:translational initiation; C:cytosol; C:plasmodesma; C:translation preinitiation complex; C:plasma membrane
497779	kinesin heavy	3.40E-28	73.30%	P:microtubule cytoskeleton organization; F:nucleotide binding; P:cytokinesis by cell plate formation; P:histone H3-K9 methylation; P:chromatin silencing; C:cytoplasm; F:microtubule motor activity; C:microtubule associated complex
350516	non-ltr retroelement reverse transcriptase-like protein	5.72E-06	73.06%	F:molecular_function; P:biological_process
351205	1-aminocyclopropane-1-carboxylate oxidase homolog 1-like	8.76E-122	73.05%	F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors
367490	truncated rb	3.14E-20	72.25%	F:purine ribonucleoside binding; F:anion binding; F:adenyl ribonucleotide binding
492243	TPA: hypothetical protein ZEAMMB73_860343	1.75E-05	72.00%	-
554777	eukaryotic translation initiation factor 3 subunit c-like	3.20E-27	71.30%	P:translational initiation; F:RNA binding; C:cytoplasmic part; C:macromolecular complex
500999	eukaryotic translation initiation factor 3 subunit c-like	1.57E-21	71.30%	C:cytosol; C:nucleus; F:protein binding; F:translation initiation factor activity
350118	ring zinc finger expressed	1.06E-91	71.25%	F:metal ion binding; F:zinc ion binding; P:protein ubiquitination; F:ubiquitin-protein ligase activity
345754	polyadenylate-binding protein 2-like	8.39E-14	71.25%	F:organic cyclic compound binding; F:heterocyclic compound binding
355586	thioesterase superfamily	1.37E-35	71.15%	-
372318	eukaryotic translation initiation factor 3 subunit c-like	2.11E-27	70.40%	F:translation initiation factor activity; P:translational initiation; C:cytosol; C:macromolecular complex; F:protein binding; C:nucleus
502939	receptor-like protein kinase hsl1-like	2.27E-05	69.85%	P:root hair elongation; P:root morphogenesis; P:regulation of cell differentiation; P:regulation of meristem growth; P:regulation of hormone levels; F:sequence-specific DNA binding transcription factor activity; P:primary shoot apical meristem specification; P:regulation of transcription, DNA-dependent; P:seed development; C:plasma membrane; P:polysaccharide biosynthetic process; P:cotyledon development; F:protein serine/threonine kinase activity; P:seed dormancy process; P:cell tip growth; P:cell division; P:transmembrane receptor protein tyrosine kinase signaling pathway; P:positive regulation of seed germination; P:post-translational protein modification; P:plant-type cell wall biogenesis; F:ATP binding; P:leaf development; P:pattern specification process; P:anther development; P:growth; F:kinase activity; P:sister chromatid cohesion; P:auxin polar transport; P:embryo development; P:protein phosphorylation; P:regulation of cell size; P:regulation of seed maturation; P:plant-type cell wall organization; P:gravitropism; P:anthocyanin accumulation in tissues in response to UV light; P:positive regulation of transcription, DNA-dependent; P:multidimensional cell growth; F:protein binding; P:regulation of cell cycle process; P:embryonic pattern specification; P:microtubule nucleation; P:cell wall organization; P:seed maturation; C:nucleus
557589	trichome birefringence-like 27	2.93E-08	69.00%	-
500535	protein chromosomal-like	3.05E-06	69.00%	F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity; F:ATPase activity; P:protein metabolic process;

118

				C:chloroplast; P:ATP catabolic process
493073	protein gamete expressed 2-like	2.83E-32	68.95%	P:pollen sperm cell differentiation
362713	retrotransposon unclassified	2.44E-27	68.45%	-
357712	wd40-like beta propeller repeat family expressed	2.09E-34	67.45%	-
490715	plastid protein	8.30E-06	67.00%	C:chloroplast; P:cell differentiation; F:molecular_function; P:protein import into chloroplast stroma; P:thylakoid membrane organization; P:regulation of protein dephosphorylation; P:chloroplast organization; P:plastid organization; P:leaf morphogenesis; P:chloroplast relocation; F:ubiquitin-protein ligase activity; P:endonucleolytic cleavage involved in rRNA processing; P:tRNA metabolic process; C:mitochondrion; P:rRNA processing; P:photosystem II assembly; P:transcription from plastid promoter; P:positive regulation of transcription, DNA-dependent; C:chloroplast envelope; F:zinc ion binding; P:endonucleolytic cleavage of tetracistronic rRNA transcript (SSU-rRNA, LSU-rRNA, 4.5S-rRNA, 5S-rRNA); P:ncRNA metabolic process; C:vacuole; F:hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; P:proton transport; C:vacuolar proton-transporting V-type ATPase complex; C:plasma membrane
347096	protein argonaute 1	0	66.85%	P:embryo development ending in seed dormancy; F:endoribonuclease activity; P:leaf morphogenesis; C:extrinsic to membrane; P:leaf proximal/distal pattern formation; F:siRNA binding; P:RNA interference; C:cytosol; P:virus induced gene silencing; P:chromatin silencing; P:stem cell development; P:production of small RNA involved in gene silencing by RNA; P:meristem initiation; P:response to auxin stimulus; P:response to far red light; P:leaf vascular tissue pattern formation; P:adaxial/abaxial pattern specification; P:covalent chromatin modification; F:protein binding; P:gene silencing by miRNA; F:miRNA binding; P:auxin metabolic process; P:adventitious root development; P:epidermal cell differentiation; C:nucleus
492838	plastid protein	2.74E-10	66.55%	P:transcription, DNA-dependent; P:regulation of primary metabolic process; C:chloroplast envelope; P:endonucleolytic cleavage of tetracistronic rRNA transcript (SSU-rRNA, LSU-rRNA, 4.5S-rRNA, 5S-rRNA); P:regulation of cellular metabolic process; F:ubiquitin-protein ligase activity; P:chloroplast organization; P:protein import into chloroplast stroma; P:developmental process; P:regulation of macromolecule metabolic process; C:mitochondrion
348887	nucleotide-binding site leucine-rich repeat partial	6.92E-76	66.00%	-
513492	ubiquinone biosynthesis monooxygenase coq6	1.33E-04	65.60%	P:oxidation-reduction process; P:ubiquinone biosynthetic process; F:monooxygenase activity; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen; F:flavin adenine dinucleotide binding
451237	u-box domain-containing protein 72-like	1.21E-19	64.30%	P:protein ubiquitination; C:intracellular membrane-bounded organelle; F:ubiquitin-protein ligase activity
354881	udp-glycosyltransferase 76f1-like	3.33E-40	64.20%	F:UDP-glucosyltransferase activity; P:metabolic process
416310	tnp2-like transposon protein	9.79E-165	63.90%	-
117520	non-ltr retroelement reverse transcriptase-like protein	2.74E-28	63.85%	-
354222	zinc finger protein nutcracker-like	1.38E-10	63.83%	-
354720	transposon mutator sub-class	1.74E-07	63.55%	F:nucleic acid binding; F:zinc ion binding; F:DNA binding
410324	mitochondrial transcription termination factor family isoform 1	5.41E-74	63.35%	-
409343	rna polymerase ii c-terminal domain phosphatase-like 3-like	2.06E-56	62.95%	P:negative regulation of abscisic acid mediated signaling pathway; F:phosphoprotein phosphatase activity; P:transcription, DNA-dependent; F:protein C-terminus binding; P:trichome differentiation; P:response to salt stress; P:response to jasmonic acid stimulus; P:response to salicylic acid stimulus; P:root hair cell differentiation; P:vegetative to reproductive phase transition of meristem; C:nucleus
367820	succinate dehydrogenase	2.98E-08	62.80%	C:mitochondrial respiratory chain complex II; P:tricarboxylic acid cycle; P:mitochondrial electron transport, succinate to ubiquinone; F:flavin adenine dinucleotide binding; F:succinate dehydrogenase (ubiquinone) activity
355574	pentatricopeptide repeat-containing	2.87E-106	61.70%	-

protein mitochondrial-like

501019	f-box lrr-repeat protein 23-like	1.60E-09	61.45%	P:ubiquitin-dependent protein catabolic process
351751	non-ltr retroelement reverse transcriptase	4.03E-64	61.30%	-
498691	f-box lrr-repeat protein at3g48880	4.11E-26	60.65%	-
412661	retrotransposon unclassified	4.65E-26	59.85%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
488950	jmjc domain containing expressed	0	59.85%	-
494961	transcription factor myc2-like	7.07E-115	59.65%	P:oxidation-reduction process; F:oxidoreductase activity; F:2-alkenal reductase [NAD(P)] activity; F:protein dimerization activity
554819	unnamed protein product	3.18E-07	59.43%	-
503437	f-box lrr-repeat protein	4.55E-12	59.35%	F:molecular_function; P:virus induced gene silencing; P:photoperiodism, flowering; P:biological_process; P:vegetative phase change; C:nucleus; C:cellular_component
369550	wd repeat-containing protein 76-like	9.13E-13	58.92%	-
368989	f-box protein skip23-like	4.75E-18	58.90%	F:ligase activity
412030	bed zinc family dimerization domain isoform 1	5.41E-102	58.65%	C:plastid; F:nucleic acid binding; F:protein dimerization activity; F:DNA binding
551725	eukaryotic translation initiation factor 3 subunit c-like	3.59E-79	58.60%	P:translation; C:cytoplasmic part
352620	collinsiaxi-like partial	1.46E-17	58.50%	-
352343	ribonuclease h protein at1g65750-like	1.05E-07	57.80%	F:nucleic acid binding; F:NAD binding; F:coenzyme binding; P:glycerol-3-phosphate catabolic process; F:nucleotide binding; C:cytoplasm; P:glycerol-3-phosphate metabolic process; P:oxidation-reduction process; P:systemic acquired resistance; P:carbohydrate metabolic process; F:oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; F:oxidoreductase activity; F:oxidoreductase activity, acting on CH-OH group of donors; F:glycerol-3-phosphate dehydrogenase [NAD+] activity; C:glycerol-3-phosphate dehydrogenase complex
507307	partial	1.93E-20	57.45%	-
347980	non-ltr retroelement reverse transcriptase	3.74E-64	57.45%	-
557932	ribonuclease h protein at1g65750-like	2.15E-04	57.00%	-
410179	annexin d5-like	6.79E-115	56.85%	F:binding
499377	f-box protein skip23-like	6.84E-13	56.65%	F:ligase activity
346646	transposon en spm sub-class	1.91E-98	56.60%	P:organic substance metabolic process; F:chitinase activity
497465	en spm-like transposon protein	1.33E-06	55.94%	-
370598	non-ltr retroelement reverse transcriptase	1.29E-26	55.65%	-
405690	tnp2-like transposon protein	0	55.35%	-
501235	f-box protein fbw2-like	3.45E-19	55.00%	-
39924	protein	2.62E-25	54.30%	-
350311	mitochondrial ubiquitin ligase activator of nfkb 1-like	1.29E-42	53.60%	C:intracellular membrane-bounded organelle; C:cytoplasmic part
354948	ribonuclease h protein at1g65750-like	1.19E-09	52.75%	-

120

504274	ribonuclease h protein at1g65750-like	6.93E-20	51.90%	F:nucleic acid binding; P:protein complex assembly; P:respiratory chain complex IV assembly; F:ribonuclease H activity; C:membrane
347415	hypothetical protein VITISV_012940	9.01E-59	50.65%	-
405926	f-box fbd lrr-repeat protein at4g03220-like	2.98E-07	50.23%	F:lyase activity
418057	f-box kelch-repeat protein at3g06240-like	1.74E-11	49.80%	F:ligase activity
497978	predicted protein	3.12E-15	49.25%	P:regulation of transcription, DNA-dependent
349479	retrotransposon unclassified	1.02E-33	49.00%	-
555936	f-box family protein	1.95E-06	49.00%	F:molecular_function; P:biological_process; F:ligase activity
409213	f-box protein	1.47E-18	48.40%	F:molecular_function; P:biological_process; C:cellular_component; C:nucleus
497254	PREDICTED: uncharacterized protein LOC101305890	1.33E-17	48.00%	-
415293	transposon en spm sub- expressed	1.08E-15	47.70%	-
346877	f-box protein	2.31E-16	47.70%	F:molecular_function; P:biological_process; C:nucleus
362206	ribonuclease h protein at1g65750-like	2.43E-08	47.67%	-
410126	f-box rni-like superfamily protein	1.50E-08	47.50%	-
353918	f-box kelch-repeat protein at1g57790-like	1.01E-14	47.50%	-
408255	f-box protein cpr30-like	6.77E-10	47.38%	-
357738	f-box protein	3.27E-10	46.55%	F:molecular_function; P:biological_process; C:cellular_component; C:nucleus
497319	f-box kelch-repeat protein	7.33E-06	46.00%	F:molecular_function; P:biological_process; C:nucleus
414557	PREDICTED: uncharacterized protein LOC101249853	2.48E-04	46.00%	-
406594	en spm-like transposon protein	2.36E-38	45.85%	-
414880	sodium hydrogen exchanger 7-like	4.26E-12	45.55%	C:integral to membrane; C:membrane; F:solute:hydrogen antiporter activity; P:sodium ion transport; P:transmembrane transport; P:cation transport; P:ion transport; P:transport
497280	PREDICTED: uncharacterized protein LOC101305890	3.66E-15	45.00%	-
497873	embryo sac development arrest	1.25E-13	44.60%	-
348855	fimbriata-like protein	1.58E-06	44.38%	-
552614	serine threonine-protein kinase ppk4	2.27E-10	43.47%	F:protein serine/threonine kinase activity; P:mRNA processing; P:phosphorylation; F:ribonuclease activity; F:protein kinase activity; P:protein phosphorylation; F:ATP binding; F:kinase activity; F:transferase activity, transferring phosphorus-containing groups; C:integral to membrane; P:reproduction; P:nematode larval development; P:lipid storage; P:positive regulation of locomotion; P:response to topologically incorrect protein; P:endoplasmic reticulum unfolded protein response; F:nucleotide binding

121

Appendix IV Table showing Blast2Go results for Apomixis specific downregulated transcripts in the fourth developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
350516	non-ltr retroelement reverse transcriptase-like protein	5.72E-06	73.06%	-
367820	succinate dehydrogenase	2.98E-08	62.80%	C:mitochondrial respiratory chain complex II; P:tricarboxylic acid cycle; P:mitochondrial electron transport, succinate to ubiquinone; F:flavin adenine dinucleotide binding; F:succinate dehydrogenase (ubiquinone) activity
557589	trichome birefringence-like 27	2.93E-08	69.00%	-
492838	plastid protein	2.74E-10	66.55%	P:transcription, DNA-dependent; P:regulation of primary metabolic process; P:endonucleolytic cleavage of tetracistronic rRNA transcript (SSU-rRNA, LSU-rRNA, 4.5S-rRNA, 5S-rRNA); P:regulation of cellular metabolic process; C:chloroplast; P:chloroplast organization; P:developmental process; P:regulation of macromolecule metabolic process; C:mitochondrion
366689	uncharacterized loc101213947	1.46E-10	51.55%	P:proteolysis; F:cysteine-type peptidase activity
418057	f-box kelch-repeat protein at3g06240-like	1.74E-11	49.80%	F:ligase activity
497873	embryo sac development arrest	1.25E-13	44.60%	-
351009	transcription initiation factor iif subunit alpha-like	9.90E-16	74.60%	P:transcription initiation from RNA polymerase II promoter; F:DNA binding; C:thylakoid; C:membrane part; C:chloroplast; P:positive regulation of transcription, DNA-dependent; F:catalytic activity; C:nucleus
558906	PREDICTED: uncharacterized protein LOC100243721	2.65E-17	88.45%	-
501235	f-box protein fbw2-like	3.45E-19	55.00%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
493073	protein gamete expressed 2-like	2.83E-32	68.95%	P:pollen sperm cell differentiation
406594	en spm-like transposon protein	2.36E-38	45.85%	-
350118	ring zinc finger expressed	1.06E-91	71.25%	F:metal ion binding
355574	pentatricopeptide repeat-containing protein mitochondrial-like	2.68E-106	61.70%	-
552928	trichome birefringence-like 27	0	78.95%	-
405690	tnp2-like transposon protein	0	55.35%	P:proteolysis; F:cysteine-type peptidase activity

Appendix V Table showing Blast2Go results for Apomixis specific upregulated transcripts in the first developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
509787	plasma membrane atpase 4-like	1.16E-27	90.10%	F:protein binding; P:regulation of stomatal movement; P:response to water deprivation; F:hydrogen-exporting ATPase activity, phosphorylative mechanism; C:vacuole; P:ATP hydrolysis coupled proton transport; C:integral to membrane; C:Golgi apparatus; P:ATP biosynthetic process; P:protein glycosylation; F:magnesium ion binding; P:response to abscisic acid stimulus; C:plasmodesma; P:ATP catabolic process; C:plasma membrane; C:nucleus; F:ATP binding
407281	kinase superfamily protein isoform 1	2.80E-129	84.95%	P:detection of biotic stimulus; P:response to molecule of bacterial origin; P:response to chitin; P:detection of external stimulus; P:protein phosphorylation; F:ATP binding; F:protein serine/threonine kinase activity; C:plastid
416304	histidine phosphotransfer protein	2.16E-50	83.00%	P:embryo sac development; P:phosphorelay signal transduction system; F:histidine phosphotransfer kinase activity; P:regulation of cytokinin mediated signaling pathway; P:cell growth; F:2-alkenal reductase [NAD(P)] activity; C:nucleus; F:protein histidine kinase binding; P:oxidation-reduction process; P:cell division
556166	elongation factor 1-beta	1.83E-11	83.00%	P:defense response to bacterium; F:translation elongation factor activity; C:eukaryotic translation elongation factor 1 complex; C:apoplast; P:translational elongation
347034	peptide nitrate transporter at2g37900-like	4.10E-94	82.40%	F:sequence-specific DNA binding transcription factor activity; P:oligopeptide transport; F:sequence-specific DNA binding; C:plasma membrane; C:integral to membrane; C:nucleus; P:regulation of transcription, DNA-dependent; F:transporter activity
360492	retrotransposon ty1-copia subclass	3.22E-26	80.15%	F:DNA binding; F:protein dimerization activity; P:cell cycle; F:cyclin-dependent protein serine/threonine kinase regulator activity; C:nucleus; F:zinc ion binding; P:DNA integration; P:regulation of transcription, DNA-dependent
413014	mannan endo- -beta-mannosidase 7-like	4.67E-08	77.21%	P:carbohydrate metabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cation binding
415470	5-methylthioadenosine s-adenosylhomocysteine deaminase-like	1.67E-131	77.00%	F:hydrolase activity
405877	kinesin heavy	1.25E-144	76.90%	P:aerobic respiration; P:microtubule-based movement; F:ATP binding; C:mitochondrion; F:microtubule motor activity; C:microtubule
417794	sgt1	6.78E-19	74.00%	P:hormone-mediated signaling pathway; P:protein catabolic process; P:defense response to fungus
354778	delta 9 desaturase	2.07E-60	72.80%	F:oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; P:phosphatidylglycerol biosynthetic process; F:16:0 monogalactosyldiacylglycerol desaturase activity; P:negative regulation of defense response; P:isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway; P:chlorophyll biosynthetic process; P:carotenoid biosynthetic process; C:thylakoid; P:unsaturated fatty acid biosynthetic process; P:photoinhibition; F:transporter activity; P:defense response to insect; P:long-chain fatty acid metabolic process; P:response to arsenic-containing substance; C:membrane
360599	dna helicase ino80-like	3.91E-09	70.90%	P:gravitropism; P:methylation-dependent chromatin silencing; P:vernalization response; P:cell-cell signaling; P:nuclear-transcribed mRNA catabolic process; F:hydrolase activity, acting on acid anhydrides; P:DNA methylation; P:somatic cell DNA recombination; F:DNA binding; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:ATP binding; P:virus induced gene silencing; P:positive regulation of DNA repair
399090	probable receptor-like protein kinase at1g67000-like	3.71E-11	67.75%	F:binding; F:protein kinase activity; P:cellular metabolic process
354701	zinc finger protein	2.29E-06	66.10%	F:RNA binding; F:nucleic acid binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:nucleotide binding
407555	peroxidase	4.14E-17	64.05%	F:oxidoreductase activity
345945	probable serine threonine-protein kinase abkc-like	4.78E-167	62.70%	F:transferase activity, transferring phosphorus-containing groups; F:kinase activity; P:phosphorylation; C:mitochondrion
498747	disease resistance-responsive (dirigent-like protein) family protein	2.31E-19	62.00%	P:ovule development; F:nucleotide binding; F:aminoacyl-tRNA ligase activity; C:chloroplast stroma; P:tRNA aminoacylation for protein translation; P:cellular component organization
360024	nascent polypeptide-associated complex subunit alpha-like protein	1.18E-07	61.27%	F:hydrolase activity; F:calcium-transporting ATPase activity; C:cytosolic ribosome; P:response to salt stress

296002	auxilin-like protein	9.29E-27	59.30%	F:hydrolase activity
368259	retrotransposon ty1-copia subclass	7.35E-19	59.25%	F:binding
349055	ankyrin repeat-containing protein at3g12360-like	2.38E-13	58.05%	-
553914	f-box protein cpr30	5.56E-07	57.00%	P:negative regulation of protein catabolic process; F:molecular_function; C:cytoplasm; P:negative regulation of defense response; F:protein binding; C:nucleus
377362	PREDICTED: uncharacterized protein LOC101305890	2.50E-19	55.50%	-
406806	protein far1-related sequence 5-like	4.46E-44	53.70%	P:regulation of transcription, DNA-dependent; F:nucleic acid binding; F:sequence-specific DNA binding; F:zinc ion binding; F:sequence-specific DNA binding transcription factor activity
345729	protein	8.19E-28	52.45%	P:proteolysis; F:aspartic-type endopeptidase activity; F:RNA binding; F:nucleic acid binding; P:DNA integration; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
366340	protein	5.59E-23	51.15%	F:nucleic acid binding; F:zinc ion binding; P:biological_process; C:cellular_component
233594	protein	3.09E-32	51.10%	F:nucleic acid binding; F:zinc ion binding; F:DNA binding
364642	ribonuclease h protein at1g65750-like	1.74E-17	50.90%	F:nucleic acid binding; F:ribonuclease H activity; F:zinc ion binding; C:intracellular; F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
493293	f-box kelch-repeat protein at1g57790-like	2.03E-05	50.00%	-
345269	f-box family protein	5.35E-28	47.80%	F:molecular_function; P:biological_process; F:metal ion binding; P:oxidation-reduction process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; F:oxidoreductase activity; F:electron carrier activity; F:iron ion binding; F:monooxygenase activity
399320	hva22-like protein isoform 2	1.96E-05	46.33%	-
356240	hypothetical protein VITISV_000143	1.06E-22	45.55%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:phosphatidylinositol binding; P:cell communication
345733	initiation factor 3g	5.31E-11	45.35%	F:binding; P:primary metabolic process; P:cellular macromolecule metabolic process; C:intracellular
406585	f-box family protein	2.16E-25	44.05%	F:ligase activity; P:glutathione biosynthetic process; F:ATP binding; F:glutathione synthase activity; F:DNA binding; F:molecular_function; P:biological_process; C:cellular_component
406990	epstein-barr virus ebna-1-like protein	2.11E-21	43.35%	F:metal ion binding; P:multicellular organismal development; F:zinc ion binding; P:protein ubiquitination; F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process; C:nucleus
551575	skp1 interacting partner isoform 1	1.44E-04	40.00%	-

Appendix VI Table showing Blast2Go results for Apomixis specific upregulated transcripts in the second developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
527856	obg-like atpase 1-like	1.41E-16	99.40%	C:cytosol; P:protein targeting to vacuole; P:response to salt stress; C:plasmodesma; P:glycolysis; P:gluconeogenesis; P:response to cadmium ion; F:GTP binding
350513	alternative oxidase	6.82E-114	92.95%	F:metal ion binding; P:cellular respiration; P:response to cold; C:mitochondrial inner membrane; P:respiratory gaseous exchange; C:respiratory chain; F:protein binding; F:alternative oxidase activity; C:integral to membrane; P:mitochondria-nucleus signaling pathway; P:electron transport chain
405650	cbl-interacting serine threonine-protein kinase 12-like	1.32E-178	91.60%	F:calmodulin-dependent protein kinase activity; P:calcium-mediated signaling; C:intracellular; P:stomatal movement; F:protein binding; P:protein phosphorylation; F:ATP binding
348444	probable pectinesterase 8-like	2.35E-91	88.75%	F:pectinesterase activity; P:cell wall modification; C:cytoplasmic membrane-bounded vesicle; F:aspartyl esterase activity; C:cell wall
348916	trans-cinnamate 4-hydroxylase	1.92E-159	87.50%	P:pollen development; C:plant-type cell wall; F:trans-cinnamate 4-monooxygenase activity; P:response to wounding; C:plasma membrane; F:electron carrier activity; F:heme binding; P:lignin metabolic process; P:growth; C:endoplasmic reticulum; P:response to light stimulus; P:oxidation-reduction process; P:response to karrikin
410064	probable inorganic phosphate transporter 1-7-like	5.77E-139	86.45%	C:vacuole; C:plasma membrane; F:inorganic phosphate transmembrane transporter activity; P:response to abscisic acid stimulus; P:phosphate ion transmembrane transport; C:integral to membrane; C:nucleus
500476	cysteine-rich receptor-like protein kinase 29-like	4.14E-50	85.95%	C:vacuole; C:plasma membrane; P:protein phosphorylation; F:ATP binding; P:response to abscisic acid stimulus; F:protein serine/threonine kinase activity
369236	mitogen-activated protein kinase	7.25E-25	83.75%	F:MAP kinase activity; P:hypotonic salinity response; P:protein phosphorylation; P:response to fungus; P:response to cold; F:protein tyrosine kinase activity; F:ATP binding; P:jasmonic acid and ethylene-dependent systemic resistance; P:cortical microtubule organization; P:response to abscisic acid stimulus; P:hyperosmotic response; P:response to indolebutyric acid stimulus; C:cytoplasm; C:nucleus; P:MAPK cascade
355920	timeless family isoform partial	1.44E-25	83.20%	C:nucleus; P:DNA-dependent DNA replication; P:regulation of circadian rhythm
409960	amino acid permease 6	0	80.25%	P:aspartate transport; P:amine transport; P:amino acid transmembrane transport; F:neutral amino acid transmembrane transporter activity; F:acidic amino acid transmembrane transporter activity; C:plasma membrane; P:tryptophan transport; C:integral to membrane
71931	phosphate transporter pho1 homolog 3-like	2.34E-28	79.90%	C:cytosol; C:integral to membrane; P:regulation of gene expression; P:positive regulation of cell size; P:endosperm development; P:regulation of seed growth; P:red or far-red light signaling pathway; C:nucleus; C:plasma membrane
351978	polyvinylalcohol dehydrogenase-like	3.42E-52	79.70%	-
410804	udp-glycosyltransferase 74e2-like	6.33E-105	79.05%	P:cellular response to chemical stimulus; P:cellular response to stress; P:response to abiotic stimulus; P:metabolic process; F:indole-3-acetate beta-glucosyltransferase activity
439941	stromal 70 kda heat shock-related chloroplastic-like	1.25E-36	77.20%	P:protein folding; P:response to temperature stimulus; F:ATP binding; F:unfolded protein binding; P:response to stress; C:chloroplast part
358556	pentatricopeptide repeat-containing protein	6.89E-61	76.50%	C:mitochondrion; C:membrane
411302	zinc transporter 1-like	8.40E-68	75.45%	P:zinc ion transmembrane transport; C:cell part; F:2-alkenal reductase [NAD(P)] activity; C:integral to membrane; F:zinc ion transmembrane transporter activity; P:oxidation-reduction process
370966	pinus taeda anonymous locus 0_18758_02 genomic sequence	6.42E-38	74.85%	-
489352	nitrate transporter -like	0	74.75%	P:seed development; P:oligopeptide transport; F:low affinity nitrate transmembrane transporter activity; P:low affinity nitrate transport; P:glucosinolate biosynthetic process; C:plasma membrane
362413	glucan endo- -beta-glucosidase	1.13E-72	73.75%	P:carbohydrate metabolic process; F:hydrolase activity, hydrolyzing O-glycosyl compounds; F:cation binding
557373	elongation factor 1-alpha	6.79E-29	73.20%	P:primary metabolic process; F:nucleic acid binding; F:nucleotide binding; F:isomerase activity; P:cellular metabolic process
349557	methyl-cpg-binding domain protein	1.18E-08	73.10%	C:nucleus; F:DNA binding; F:zinc ion binding; F:cytosine C-5 DNA demethylase activity; F:enzyme binding

125

489011	peroxisome biogenesis protein 1-like	0	70.95%	C:peroxisome; P:protein import into peroxisome matrix; F:ATPase activity, coupled; P:fatty acid beta-oxidation; F:protein binding; F:nucleotide binding; P:response to stress
360599	dna helicase ino80-like	3.91E-09	70.90%	P:gravitropism; P:methylation-dependent chromatin silencing; P:vernalization response; P:cell-cell signaling; P:nuclear-transcribed mRNA catabolic process; F:hydrolase activity, acting on acid anhydrides; P:DNA methylation; P:somatic cell DNA recombination; F:DNA binding; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:ATP binding; P:virus induced gene silencing; P:positive regulation of DNA repair
368181	2-oxoglutarate and fe -dependent oxygenase superfamily isoform 1	6.82E-29	69.55%	F:binding; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors; P:glucosinolate biosynthetic process
359338	n-hydroxycinnamoyl benzoyltransferase	1.21E-59	68.40%	F:transferase activity, transferring acyl groups
353577	af405557_1reverse transcriptase	3.31E-25	68.00%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; C:nucleus
366987	hypothetical protein VITISV_027787	3.12E-54	67.75%	F:binding; P:primary metabolic process; P:cellular macromolecule metabolic process
399090	probable receptor-like protein kinase at1g67000-like	3.72E-11	67.75%	F:binding; F:protein kinase activity; P:cellular metabolic process
491576	myb family transcription factor	1.68E-58	65.50%	F:nucleic acid binding; P:regulation of transcription, DNA-dependent
418224	calcium-dependent protein kinase	6.02E-10	65.35%	F:binding; F:protein kinase activity
407555	peroxidase	4.14E-17	64.05%	F:oxidoreductase activity
352201	abscisic acid-insensitive protein 3	4.35E-08	63.55%	P:regulation of transcription, DNA-dependent; F:DNA binding; P:response to abscisic acid stimulus; P:transcription, DNA-dependent; P:embryo development; C:nucleus; P:abscisic acid mediated signaling pathway; C:cytoplasm; P:multicellular organismal development; F:sequence-specific DNA binding transcription factor activity; P:gibberellic acid mediated signaling pathway; P:positive regulation of transcription, DNA-dependent; P:cell division; P:gibberellin biosynthetic process; P:plastid organization; P:response to auxin stimulus; P:terpenoid biosynthetic process; F:protein binding; C:cytosol; P:cotyledon development; P:mitochondria-nucleus signaling pathway
18740	protein far1-related sequence 5-like	3.25E-36	63.25%	F:transposase activity; P:DNA integration; P:transposition, DNA-mediated; F:DNA binding; F:metal ion binding; F:hydrolase activity
345945	probable serine threonine-protein kinase abkc-like	4.78E-167	62.70%	F:transferase activity, transferring phosphorus-containing groups; F:kinase activity; P:phosphorylation; C:mitochondrion
344982	retroelement pol polyprotein	1.09E-113	62.05%	F:nucleic acid binding; P:DNA integration; F:zinc ion binding
360024	nascent polypeptide-associated complex subunit alpha-like protein	1.18E-07	61.27%	F:hydrolase activity; F:calcium-transporting ATPase activity; C:cytosolic ribosome; P:response to salt stress
296002	auxilin-like protein	9.28E-27	59.30%	F:hydrolase activity
402819	hypothetical protein MTR_7g108140	7.90E-04	58.50%	-
302811	T4.5	9.88E-04	58.00%	-
409514	PREDICTED: uncharacterized protein LOC100261060	9.40E-35	55.15%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; P:DNA repair; F:endonuclease activity; C:intracellular; F:DNA binding; F:ribonuclease T2 activity
413408	eukaryotic translation initiation factor 3 subunit c-like	6.30E-10	54.80%	C:cytoplasm
348656	non-ltr retroelement reverse transcriptase	1.99E-29	54.65%	F:RNA-directed DNA polymerase activity; F:RNA binding; P:RNA-dependent DNA replication; F:endonuclease activity
355701	protein gpr107	1.69E-05	54.00%	C:integral to membrane
406806	protein far1-related sequence 5-like	4.47E-44	53.70%	P:regulation of transcription, DNA-dependent; F:nucleic acid binding; F:sequence-specific DNA binding; F:zinc ion binding; F:sequence-specific DNA binding transcription factor activity
8040	carboxyl-terminal peptidase protein	1.80E-20	53.70%	F:molecular_function; P:biological_process
382954	disease resistance protein rpm1	3.23E-08	53.60%	P:defense response; F:ADP binding; F:ATP binding; F:nucleotide binding; F:nucleoside-triphosphatase activity

126

399746	transposon en spm sub-class	4.66E-20	53.35%	C:mitochondrion
509687	non-ltr retroelement reverse transcriptase	1.54E-06	53.00%	F:RNA binding; F:nucleic acid binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:ribonuclease H activity
345729	protein	8.19E-28	52.45%	P:proteolysis; F:aspartic-type endopeptidase activity; F:RNA binding; F:nucleic acid binding; P:DNA integration; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
413830	fbd-associated f-box protein	1.09E-98	51.95%	F:nucleic acid binding
553167	hypothetical protein VITISV_023635	5.59E-18	51.75%	F:binding; F:catalytic activity
349155	predicted protein	1.72E-13	50.20%	F:molecular_function; P:biological_process
489193	protein	3.83E-55	50.10%	-
493293	f-box kelch-repeat protein at1g57790-like	2.03E-05	50.00%	-
564565	unknown	1.78E-05	48.63%	-
348709	ac069473_8gypsy ty-3 retroelement polyprotein 69905-74404	4.83E-58	48.05%	F:RNA binding; F:nucleic acid binding; P:DNA integration; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; P:proteolysis; F:aspartic-type endopeptidase activity; C:plastid
345269	f-box family protein	5.36E-28	47.80%	F:molecular_function; P:biological_process; F:metal ion binding; P:oxidation-reduction process; F:oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; F:heme binding; F:oxidoreductase activity; F:electron carrier activity; F:iron ion binding; F:monooxygenase activity
345733	initiation factor 3g	5.31E-11	45.35%	F:binding; P:primary metabolic process; P:cellular macromolecule metabolic process; C:intracellular
407937	ankyrin repeat domain protein	2.70E-44	43.65%	C:ubiquitin ligase complex; P:protein ubiquitination; F:ubiquitin-protein ligase activity; P:protein targeting; F:helicase activity; C:membrane; P:protein import; F:nucleic acid binding; F:nucleotide binding; F:ATP binding; P:intracellular protein transport; C:intracellular
368508	ribonuclease h protein at1g65750-like	2.65E-09	42.90%	C:mitochondrion; F:nucleic acid binding; F:ribonuclease H activity
347079	f-box protein at5g07610-like isoform x1	4.80E-14	42.20%	C:cytoplasmic membrane-bounded vesicle
551575	skp1 interacting partner isoform 1	1.44E-04	40.00%	-

Appendix VII Table showing Blast2Go results for Apomixis specific upregulated transcripts in the third developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
350513	alternative oxidase	6.83E-114	92.95%	F:metal ion binding; P:cellular respiration; P:response to cold; C:mitochondrial inner membrane; P:respiratory gaseous exchange; C:respiratory chain; F:protein binding; F:alternative oxidase activity; C:integral to membrane; P:mitochondria-nucleus signaling pathway; P:electron transport chain
348444	probable pectinesterase 8-like	2.35E-91	88.75%	F:pectinesterase activity; P:cell wall modification; C:cytoplasmic membrane-bounded vesicle; F:aspartyl esterase activity; C:cell wall
500476	cysteine-rich receptor-like protein kinase 29-like	4.14E-50	85.95%	C:vacuole; C:plasma membrane; P:protein phosphorylation; F:ATP binding; P:response to abscisic acid stimulus; F:protein serine/threonine kinase activity
556353	cinnamyl-alcohol dehydrogenase cad1	0	85.00%	F:transferase activity, transferring acyl groups other than amino-acyl groups; F:cinnamyl-alcohol dehydrogenase activity; P:lignin biosynthetic process; F:zinc ion binding; C:apoplast; F:nucleotide binding; P:oxidation-reduction process; F:mannitol dehydrogenase activity
489447	cysteine-rich receptor-like protein kinase 2-like	2.49E-139	83.85%	F:transmembrane receptor protein tyrosine kinase activity; C:plasma membrane; P:protein autophosphorylation; P:response to ozone; F:ATP binding; F:protein serine/threonine kinase activity
347034	peptide nitrate transporter at2g37900-like	4.10E-94	82.40%	F:sequence-specific DNA binding transcription factor activity; P:oligopeptide transport; F:sequence-specific DNA binding; C:plasma membrane; C:integral to membrane; C:nucleus; P:regulation of transcription, DNA-dependent; F:transporter activity
360492	retrotransposon ty1-copia subclass	3.22E-26	80.15%	F:DNA binding; F:protein dimerization activity; P:cell cycle; F:cyclin-dependent protein serine/threonine kinase regulator activity; C:nucleus; F:zinc ion binding; P:DNA integration; P:regulation of transcription, DNA-dependent
356464	pentatricopeptide repeat-containing protein at3g29230-like	4.33E-31	78.90%	C:mitochondrion; P:mitochondrial mRNA modification
411199	protein fizzy-related 3-like	1.41E-21	77.25%	C:plastid
489685	abc transporter a family member 7-like	0	76.35%	P:ATP catabolic process; F:ATP binding; F:ATPase activity; C:plasma membrane
95562	polyadenylate-binding protein	2.38E-35	76.25%	F:nucleic acid binding; F:nucleotide binding
406996	hydroxysteroid dehydrogenase 1	6.81E-144	76.00%	P:response to abscisic acid stimulus; F:11-beta-hydroxysteroid dehydrogenase (NADP+) activity; F:17-beta-ketosteroid reductase activity; F:17-beta-hydroxysteroid dehydrogenase (NADP+) activity; P:steroid metabolic process; P:response to brassinosteroid stimulus
370966	pinus taeda anonymous locus 0_18758_02 genomic sequence	6.42E-38	74.85%	-
349557	methyl-cpg-binding domain protein	1.18E-08	73.10%	C:nucleus; F:DNA binding; F:zinc ion binding; F:cytosine C-5 DNA demethylase activity; F:enzyme binding
399397	hypothetical protein VITISV_005743	2.83E-21	73.05%	P:transport; P:cellular biosynthetic process; F:nucleic acid binding; P:thiamine-containing compound metabolic process; P:DNA metabolic process; F:protein kinase activity; F:metal ion binding; F:ion channel activity; F:nucleotide binding; F:hydrolase activity; C:membrane
157798	predicted protein	2.48E-22	72.25%	P:gibberellic acid mediated signaling pathway; F:molecular_function; P:seed dormancy process; P:biological_process
555506	pentatricopeptide repeat-containing protein mitochondrial-like	5.48E-05	71.00%	-
360599	dna helicase ino80-like	3.91E-09	70.90%	P:gravitropism; P:methylation-dependent chromatin silencing; P:vernalization response; P:cell-cell signaling; P:nuclear-transcribed mRNA catabolic process; F:hydrolase activity, acting on acid anhydrides; P:DNA methylation; P:somatic cell DNA recombination; F:DNA binding; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:ATP binding; P:virus induced gene silencing; P:positive regulation of DNA repair
383226	upf0481 protein at3g47200-like	8.44E-15	69.15%	-
362546	metal ion binding	6.84E-15	68.70%	F:metal ion binding
399090	probable receptor-like protein kinase at1g67000-like	3.72E-11	67.75%	F:binding; F:protein kinase activity; P:cellular metabolic process

128

416856	exordium like 3	5.20E-91	65.35%	C:plant-type cell wall; P:response to karrikin
407555	peroxidase	4.13E-17	64.05%	F:oxidoreductase activity
345945	probable serine threonine-protein kinase abkc-like	4.78E-167	62.70%	F:transferase activity, transferring phosphorus-containing groups; F:kinase activity; P:phosphorylation; C:mitochondrion
360024	nascent polypeptide-associated complex subunit alpha-like protein	1.18E-07	61.27%	F:hydrolase activity; F:calcium-transporting ATPase activity; C:cytosolic ribosome; P:response to salt stress
368259	retrotransposon ty1-copia subclass	7.35E-19	59.25%	F:binding
302811	T4.5	9.88E-04	58.00%	-
37970	ring fyve phd zinc finger-containing protein	2.66E-07	55.92%	F:metal ion binding; F:zinc ion binding; P:regulation of transcription, DNA-dependent; C:cellular_component; F:DNA binding; P:biological_process
377362	PREDICTED: uncharacterized protein LOC101305890	2.50E-19	55.50%	-
499953	uncharacterized loc101212789	1.61E-22	55.00%	C:extracellular region; P:pathogenesis
345729	protein	8.19E-28	52.45%	P:proteolysis; F:aspartic-type endopeptidase activity; F:RNA binding; F:nucleic acid binding; P:DNA integration; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity
553167	hypothetical protein VITISV_023635	5.58E-18	51.75%	F:binding; F:catalytic activity
33310	non-ltr retroelement reverse transcriptase	5.79E-17	51.00%	F:RNA binding; P:RNA-dependent DNA replication; F:endonuclease activity; F:RNA-directed DNA polymerase activity; F:nucleic acid binding; F:zinc ion binding
497511	uncharacterized loc101212067	1.81E-45	50.35%	C:extracellular region; P:pathogenesis
407525	hypothetical protein VITISV_025096	1.93E-14	45.10%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; P:DNA repair; F:endonuclease activity; C:intracellular; F:DNA binding
406990	epstein-barr virus ebna-1-like protein	2.11E-21	43.35%	F:metal ion binding; P:multicellular organismal development; F:zinc ion binding; P:protein ubiquitination; F:ubiquitin-protein ligase activity; P:ubiquitin-dependent protein catabolic process; C:nucleus

Appendix VIII Table showing Blast2Go results for Apomixis specific upregulated transcripts in the fourth developmental stage

Seq. Name	Seq. Description	min. eValue	mean Similarity	GOs
412420	momilactone a synthase-like	2.55E-87	83.85%	F:oxidoreductase activity; P:diterpene phytoalexin precursor biosynthetic process pathway; C:plastid; P:oxidation-reduction process; F:nucleotide binding
413477	condensin complex subunit 2-like	1.02E-09	80.65%	P:mitosis; P:chromosome condensation; P:cell division
500029	udp-glycosyltransferase 85a1	6.68E-67	79.45%	F:transferase activity, transferring hexosyl groups; P:metabolic process
360599	dna helicase ino80-like	3.91E-09	70.90%	P:gravitropism; P:methylation-dependent chromatin silencing; P:vernalization response; P:cell-cell signaling; P:nuclear-transcribed mRNA catabolic process; F:hydrolase activity, acting on acid anhydrides; P:DNA methylation; P:somatic cell DNA recombination; F:DNA binding; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:ATP binding; P:virus induced gene silencing; P:positive regulation of DNA repair
399090	probable receptor-like protein kinase at1g67000-like	3.72E-11	67.75%	F:binding; F:protein kinase activity; P:cellular metabolic process
354701	zinc finger protein	2.29E-06	66.10%	F:RNA binding; F:nucleic acid binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:nucleotide binding
350344	dna polymerase zeta catalytic	3.64E-66	65.65%	-
351620	ubiquitin carboxyl-terminal	2.15E-58	62.00%	C:plasmodesma; F:hydrolase activity; C:cytosol
368259	retrotransposon ty1-copia subclass	7.34E-19	59.25%	F:binding
353516	disease resistance rpp13-like protein 1-like	6.50E-51	56.55%	P:defense response; F:ADP binding; F:hydrolase activity; F:phosphoprotein phosphatase activity; C:cytoplasmic membrane-bounded vesicle
377362	PREDICTED: uncharacterized protein LOC101305890	2.50E-19	55.50%	-
399746	transposon en spm sub-class	4.67E-20	53.35%	C:mitochondrion
553167	hypothetical protein VITISV_023635	5.58E-18	51.75%	F:binding; F:catalytic activity
349155	predicted protein	1.72E-13	50.20%	F:molecular_function; P:biological_process
489193	protein	3.83E-55	50.10%	-
493293	f-box kelch-repeat protein at1g57790-like	2.03E-05	50.00%	-
489339	protein binding	8.81E-10	48.72%	P:photoinhibition; F:serine-type endopeptidase activity; C:chloroplast; P:proteolysis

Appendix IX Table showing Blast2Go results for Apomixis specific upregulated transcripts in and their homeologous classification (H.effect : Hybrid effect)

Gene ID	Seq.Description	min.eValue	mean.Similarity	GOs.	grp	Stage
414015	3297496 protein	7.29E-05	68.00%	-	H.effect	II
350118	5310787 ccp	7.21E-93	71.20%	F:metal ion binding	H.effect	III
366078	argonaute family protein isoform partial	1.61E-63	84.45%	P:translational initiation; F:translation initiation factor activity	H.effect	II
497873	embryo sac development arrest	1.45E-13	46.83%	P:megagametogenesis; F:molecular_function	H.effect	I
551725	eukaryotic translation initiation factor 3 subunit c-like	4.16E-79	58.60%	P:translation; C:cytoplasmic part	H.effect	III
408848	flotillin-like protein	1.43E-11	69.55%	C:apical plasma membrane; P:root development; P:induction of tumor, nodule, or growth in other organism involved in symbiotic interaction	H.effect	II
345754	polyadenylate-binding protein 2-like	9.75E-14	71.25%	F:nucleic acid binding	H.effect	I
558906	PREDICTED: uncharacterized protein LOC100243721	3.07E-17	88.25%	-	H.effect	III
493073	protein gamete expressed 2-like	3.28E-32	70.45%	P:pollen sperm cell differentiation	H.effect	III
347415	ribonuclease h protein at1g65750-like	1.04E-58	50.35%	F:RNA binding; P:RNA-dependent DNA replication; F:RNA-directed DNA polymerase activity; F:unfolded protein binding; F:ATP binding; P:response to stress; P:protein folding; F:hydrolase activity	H.effect	III
367900	serine threonine-protein phosphatase 7 long form homolog	7.79E-28	49.85%	F:molecular_function; P:biological_process; C:nucleus	H.effect	II
405690	tnp2-like transposon protein	0	55.55%	P:proteolysis; F:cysteine-type peptidase activity	H.effect	I
409672	vesicle-associated protein 4-2-like	2.70E-52	91.40%	P:response to gibberellin stimulus; P:protein targeting to vacuole; P:regulation of protein localization; P:N-terminal protein myristoylation; F:structural molecule activity; C:nucleus; C:plasma membrane	H.effect	III

Appendix X Table showing Blast2Go results for Apomixis specific downregulated transcripts in and their homeologous classification (H.effect : Hybrid effect, P2.effect: Parent of origin, Ploidy.effect: Ploidy effect)

Gene ID	Seq.Description	min.eValue	mean.Similarity	GOs.	grp	Stage
413767	1110748 at2g14530	6.80E-51	86.50%	C:chloroplast; C:Golgi apparatus	H.effect	II
414542	1833194 os03g0257000	1.10E-64	88.55%	C:nucleus	H.effect	I
405548	19988529 dna-dependent atpase snf2h	3.35E-06	84.90%	P:embryo sac development; F:helicase activity; F:DNA binding; P:cell growth; P:ATP-dependent chromatin remodeling; F:ATP binding; C:nucleus; F:nucleosome binding	H.effect	II
350118	5310787 ccp	7.21E-93	71.20%	F:metal ion binding	H.effect	II
490083	60s ribosomal protein l31	9.15E-15	96.60%	F:structural constituent of ribosome; C:plasma membrane; C:cell wall; C:cytosolic large ribosomal subunit; C:chloroplast; P:translation	P2.effect	II
404837	60s ribosomal protein l3-like	7.12E-11	70.55%	P:primary metabolic process; P:macromolecule metabolic process; C:ribonucleoprotein complex; P:cellular metabolic process	P2.effect	II
360024	6812847 klth0d17842p	1.37E-07	61.33%	F:hydrolase activity; F:calcium-transporting ATPase activity; C:cytosolic ribosome; P:response to salt stress	Ploidy.effect	III
407555	878199 peroxidase	4.15E-21	64.70%	F:oxidoreductase activity	Ploidy.effect	III
350513	alternative oxidase	7.92E-114	93.00%	F:metal ion binding; P:cellular respiration; P:response to cold; C:mitochondrial inner membrane; C:respiratory chain; F:protein binding; F:alternative oxidase activity; C:integral to membrane; P:mitochondria-nucleus signaling pathway; P:electron transport chain	Ploidy.effect	III
366582	atpase plasma membrane-type-like	3.06E-113	83.30%	F:cation-transporting ATPase activity; P:vacuole organization; P:proanthocyanidin biosynthetic process; C:integral to membrane; C:chloroplast; P:vacuolar acidification; F:metal ion binding; P:ATP biosynthetic process; P:cation transport; C:plasmodesma; P:ATP catabolic process; C:plasma membrane; F:ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism; F:ATP binding	H.effect	III
349370	cell wall vacuolar inhibitor of fructosidase 2-like	6.94E-41	75.05%	F:pectinesterase activity; P:response to karrikin; C:extracellular region; P:negative regulation of catalytic activity; F:pectinesterase inhibitor activity	P2.effect	II
360599	dna helicase ino80-like	4.54E-09	71.00%	P:gravitropism; P:methylation-dependent chromatin silencing; P:vernalization response; P:cell-cell signaling; P:nuclear-transcribed mRNA catabolic process; F:hydrolase activity, acting on acid anhydrides; P:DNA methylation; P:somatic cell DNA recombination; F:DNA binding; P:production of miRNAs involved in gene silencing by miRNA; P:production of ta-siRNAs involved in RNA interference; F:ATP binding; C:nucleus; P:virus induced gene silencing; P:positive regulation of DNA repair	Ploidy.effect	III
497873	embryo sac development arrest	1.45E-13	46.83%	P:megagametogenesis; F:molecular_function	H.effect	IV
497465	en spm-like transposon protein	1.54E-06	55.88%	P:proteolysis; F:cysteine-type peptidase activity	H.effect	II
408848	flotillin-like protein	1.43E-11	69.55%	C:apical plasma membrane; P:root development; P:induction of tumor, nodule, or growth in other organism involved in symbiotic interaction	H.effect	I
416304	histidine-containing phosphotransfer protein 1 isoform 1	2.51E-50	82.85%	P:phosphorelay signal transduction system; P:regulation of seed germination; F:histidine phosphotransfer kinase activity; P:regulation of shoot system development; P:oxidation-reduction process; P:cell growth; P:embryo sac development; P:regulation of cytokinin mediated signaling pathway; P:transition metal ion transport; P:regulation of anthocyanin metabolic process; P:cell division; F:protein histidine kinase binding; P:inorganic anion transport; F:2-alkenal reductase [NAD(P)] activity; C:cytoplasm; C:nucleus	Ploidy.effect	I
553167	hypothetical protein VITISV_023635	6.48E-18	51.70%	F:heterocyclic compound binding; F:organic cyclic compound binding; F:catalytic activity	Ploidy.effect	III
492201	peroxidase 44-like	1.88E-55	73.25%	C:cell wall; P:trichoblast differentiation; F:binding; P:transition metal ion transport; F:oxidoreductase activity; P:response to stress	Ploidy.effect	I
367265	PREDICTED: uncharacterized protein	0.000338	42.00%	-	P2.effect	II

132

	LOC101305583					
496232	PREDICTED: uncharacterized protein LOC101305890	8.10E-30	45.50%	-	H.effect	II
497280	PREDICTED: uncharacterized protein LOC101305890	4.25E-15	45.00%	-	H.effect	II
377362	PREDICTED: uncharacterized protein LOC101305890	2.90E-19	55.50%	-	P2.effect	III
399090	probable receptor-like protein kinase at1g67000-like	4.31E-11	68.80%	P:phosphorylation; F:protein kinase activity; F:nucleotide binding	P2.effect	III
489493	proteasome subunit beta type-5 precursor	1.39E-15	53.30%	F:molecular_function; P:biological_process; C:cellular_component; P:proteolysis involved in cellular protein catabolic process; F:endopeptidase activity; F:hydrolase activity; C:nucleus; C:cytoplasm; F:threonine-type endopeptidase activity; C:proteasome core complex; C:proteasome complex; F:peptidase activity; P:proteolysis	P2.effect	II
424479	protein disulfide-isomerase-like	5.87E-09	83.80%	P:glycerol ether metabolic process; P:seed development; P:response to endoplasmic reticulum stress; P:response to salt stress; C:lytic vacuole within protein storage vacuole; P:cell redox homeostasis; F:electron carrier activity; C:plant-type cell wall; C:thylakoid; P:regulation of programmed cell death; F:isomerase activity; P:response to zinc ion; C:chloroplast; P:response to cadmium ion; C:endoplasmic reticulum; F:protein disulfide oxidoreductase activity; C:plasma membrane; P:embryo development	P2.effect	II
507600	retrotransposon unclassified	1.30E-08	69.40%	P:DNA metabolic process; F:nucleic acid binding; F:catalytic activity	P2.effect	II
552614	serine threonine kinase ire1	2.63E-10	43.69%	P:single-organism cellular process; P:cellular macromolecule metabolic process; P:positive regulation of biological process; P:response to topologically incorrect protein; P:primary metabolic process; P:regulation of cellular process; F:catalytic activity	H.effect	II
367900	serine threonine-protein phosphatase 7 long form homolog	7.79E-28	49.85%	F:molecular_function; P:biological_process; C:nucleus	Ploidy.effect	I
417794	sgt1	7.87E-19	74.00%	P:defense response to bacterium; P:hormone-mediated signaling pathway; F:protein binding; P:protein catabolic process; C:cytoplasm; P:defense response to fungus; C:nucleus	Ploidy.effect	III
405690	tnp2-like transposon protein	0	55.55%	P:proteolysis; F:cysteine-type peptidase activity	H.effect	II
412853	transcription factor fer-like iron deficiency-induced transcription factor-like	1.58E-29	60.25%	P:cellular response to iron ion; P:cellular response to ethylene stimulus; P:regulation of iron ion transport; F:protein binding; P:cellular response to nitric oxide	H.effect	III
410617	uncharacterized partial	1.40E-07	65.25%	-	H.effect	II
386226	unconventional myosin	1.56E-06	83.40%	F:RNA binding; F:Rab GTPase binding; P:root hair elongation; P:RNA-dependent DNA replication; C:peroxisome; P:peroxisome localization; C:myosin complex; F:RNA-directed DNA polymerase activity; F:ATP binding; F:GTP-dependent protein binding; F:motor activity; F:actin binding; P:mitochondrion localization; P:Golgi localization	Ploidy.effect	II

Appendix XI Table of the SNPs profile of one gene showing parent of origin effect

Contig ID	Position	apo 1	apo2	carp	cass
368515	148	A	A	C	A
	153	A	A	T	A
	155	G	G	C	G
	159	T	T	C	T
	162	A	A	G	A
	163	G	G	A	G
	169	T	T	A	T
	170	G	G	T,A	G
	175	A	A	C	A
	176	T	T	A	T
	200	G	G	C	G
	203	A	A	G	A
	308	C	C	T	C
	314	T	T	A	T
	316	A	A	T	A
	326	G		A	G