From a Phenotype to Transcriptomics

Apomixis Initiation in the genus *Boechera*

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

Diplom-Biologe Marie-Luise Voigt
Born in: Jena, Germany
Oral-examination:

From a Phenotype to Transcriptomics

Apomixis initiation in the genus *Boechera*

Referees:
Prof. Dr. Marcus Koch
Dr. Timothy F. Sharbel

The research presented in this thesis was carried out at the research group Apomixis at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany.

Gatersleben, 2009

Table of contents

01. Summary

The North American genus *Boechera* provides a highly polymorphic source of natural variation for studies on the regulation of apomictic development (asexual reproduction via seed). The aim of the research conducted was to qualify and quantify naturally-occurring reproductive variation in apomictic and sexual accessions (genotypes), and to further elucidate molecular factors responsible for the initiation of the first step in the apomictic pathway, *apomeiosis*. Sexual reproduction is characterized by double fertilization (i.e. a pair of sperm nuclei fertilize both the egg and central cell), whereby apomictic plants produce seeds without fertilization of the egg cell (yielding maternal genetic clones). In apomictic *Boechera* accessions the sexual pathway is altered for three traits: i) formation of an unreduced embryo sac, e.g. through meiotically-unreduced megaspore formation (*apomeiosis*), ii) development of an embryo from an unfertilized and unreduced egg cell (*parthenogenesis*), and iii) formation of functional endosperm (embryo nourishing tissue), e.g. fertilization of the binucleate central cell (*pseudogamy*). This type of apomixis is called "*diplospory*".

In general, apomixis is correlated with polyploidy. The genus *Boechera* contains one of the rare cases of diploid apomixis, and thus provides an ideal model whereby both diploid and polyploid apomicts can be compared. We began by examining both pollen and seed formation in a number of ecotypes in order to identify variation in the apomictic phenotype. Both apomicts showed flexibility with regards to combinations of the apomixis components, but diploid apomicts were characterized by higher flexibility to variant ploidy ratios in embryo : endosperm, which can affect seed development. In performing a comprehensive comparative study between reproduction traits of both apomictic karyotypes, I show that most traits exhibit lineage-specificity rather than correlations with ploidy. This could reflect the consequences of natural hybridization (we tested independently-evolved apomictic lineages), where natural selection acts upon novel variation in several traits to allow their establishment in specific niches, thus obscuring the effects of ploidy on reproductive success.

The phenotypic data were used to select highly expressive diploid apomicts for a deep transcriptomic comparison between microdissected live sexual and apomeiotic ovules. This approach was taken to elucidate the first step in apomixis, *apomeiosis*, which is proposed to be the key factor for stable apomixis expression in *Boechera*. Approximately, 4 000 differentially expressed mRNAs were identified between sexual and apomictic ovules at the megaspore mother cell stage (MMC), the hypothesised stage of apomeiosis initiation, indicating : i) heterochronic expression of genes, ii) differential gene expression, and iii) a parent-of origin effect. In a following approach I preselected *A. thaliana* genes and performed a sequence homology search between *Arabidopsis thaliana* and *Boechera*. Identified sequences in *Boechera* were analysed with respect to their expression profiles generated from a SuperSAGE (serial analysis of gene expression) experiments. This study demonstrated that apomixis in *Boechera* might be influenced by chromatin remodelling, which could suppress or enhance gene transcription, and the cause for chromatin remodelling could be the heterozygous (i.e. hybrid) state of apomictic *Boechera*.

This study yielded a set of promising apomixis initiation candidates, which could be used as a first subset for confirmation and functional studies. These data furthermore give deeper insight into the apomixis pathway and its complexity in the genus *Boechera*.

0.2 Zusammenfassung

Die nordamerikanische Gattung *Boechera* pflanzt sich entweder sexuell oder asexuell fort. Das natürliche Auftreten unterschiedlicher Fortpflanzungswege macht sie zu einem besonderem Model, um die Regulation des asexuellen Weges (bezeichnet als „Apomixis“) zu studieren. Die sexuelle Fortpflanzung ist durch den Prozess der Doppelbefruchtung (Eizelle und Zentralzelle werden jeweils durch ein Spermium befruchtet) charakterisiert. Hingegen produzieren apomiktische Pflanzen den Samen ohne die vorausgegangene Befruchtung der Eizelle (die Nachkommen sind genetische Klone). In apomiktischen *Boechera* Pflanzen wurde der sexuelle Prozess in drei Komponenten verändert: 1) die Herausbildung eines unreduzierten Embryosacks, durch die Umgehung der Meiose I während der Megasporenausbildung (*apomeiosis*), 2) die parthenogenetische Entwicklung der Eizelle zum Embryo (*parthenogenesis*) und 3) die Ausbildung von dazugehörigem Endospermgewebe durch Befruchtung der Zentralzelle (*pseudogamy*). Diese Form der Apomixis wird als *Diplospory* bezeichnet.

In der Natur wird eine auffällige Korrelation von Apomixis mit Polyploidie vorgefunden. Die Gattung *Boechera* beherbergt den seltenen Umstand des gleichzeitigen Vorkommens von diploiden und triploiden Apomikten, das sie zu einem idealen Vergleichsmodel macht. Infolgedessen wurde die Pollen- und Samenausbildung von unterschiedlichen *Boechera* Akzessionen untersucht, um Variationen im apomiktischen Verlauf zu identifizieren. In beiden Karyotypen wurde Flexibilität in der Ausprägung von den drei Apomixiskomponenten festgestellt. Aber charakteristisch für diploide Apomikten war die höhere Anpassungsfähigkeit in Veränderungen der Ploidiebeziehung von Embryo und Endosperm. In einer umfassenden Studie, in der Reproduktionsmerkmale verglichen wurden, wurde vornehmlich eine *Boechera* Akzessionen-spezifische Ausprägung von Merkmalen herausgefunden. Dies läst auf keinen Vorteil von Apomixis im Zusammenhang mit Polyploidie schließen. Ein diskutierter Aspekt könnte die verfälschte Ausprägung der Merkmale auf Grund von Hybridisierung und der einhergehenden natürlichen Selektion auf neuartige Ausprägungen sein, die die Etablierung in Nischen ermöglicht.

Die phänotypischen Daten wurden genutzt um, für eine weitreichende Transkriptomanalyse, stabile diploide apomiktische Pflanzen auszusuchen. Es wurden von zwei sexuell diploiden und zwei apomiktisch diploiden Pflanzen Samenanlagen herauspräpariert. Diese befanden sich am Eintritt zum Meioseprozess. Der Versuch wurde angelegt um in einer Gen – Expressionsanalyse (SuperSAGE; Serial Analyse of Gene Expression) molekulare Faktoren zu identifizieren, die bei der Initiierung von Apomixis (*apomeiosis)* von Bedeutung sind. In diesem Versuch wurden 4.000 differentiell exprimierte mRNAs zwischen den sexuellen und apomiktischen Samenanlagen identifiziert. Diese weisen hin auf i) Heterochronische Expression von Genen, ii) signifikant unterschiedliche Expression von Genen, und iii) ein *Parent-of origin* Effekt. Weiterhin wurde ein Sequenzähnlichkeitstest zwischen Gensequenzen von *Arabidopsis thaliana* und *Boechera* durchgeführt. Identifizierte Gene wurden in ihrem Expressionsverhalten studiert. Dabei wurden Hinweise gefunden, dass Apomixis in *Boechera* eventuell auf Umstrukturierungen des Chromatins zurückzuführen ist, welche durch den heterozygoten Zustand der Apomikten (Alloploide Hybriden) ausgelöst wurde.

Mit diesem Versuch konnte eine Liste mit vielversprechenden Apomixis-Kandidaten aufgestellt werden, die für nachfolgende detaillierte Funktionsanalysen zur Verfügung stehen. Die gesammelten Daten geben einen tieferen Einblick in den Apomixis-Prozess von *Boechera*.

GENERAL INTRODUCTION

1. *Sexual and asexual reproduction in angiosperms.*

Reproduction is the most fundamental aspect in the life history of any organism. Sexual reproduction in angiosperms follows a life-cycle with an alteration between diploid (sporophytic) and haploid (gametophytic) generation (Figure 1). To reproduce sexually two steps are necessary, *reduction* and *fusion* (*syngamy*). To form progenies the mature diploid plant develops reproductive structures (flowers), which contain the reproductive organs stamen (male) and pistil (female). The stamen consists of a filament and an anther head (microsporangia), and in the microsporangia pollen mother cells (PMC) will undergo a reduction in genome content (meiosis) leading to the formation of haploid male gametophytes (pollen grains). The pistil consists of stigma (where pollen attaches), style ("neck") and ovary (where progenies develop). The inner cells of the ovary initiate ovule development (Hill and Lord, 1994), and these ovules (megasporangia) form a defined megaspore mother cell (MMC) which will undergo a reduction in genome content (meiosis) to develop the female gametophyte (embryo sac). The male pollen grain and female embryo sac contain the haploid gametes, sperm and egg cell, respectively. The reductional process is necessary to ensure the normal somatic genome constitution after fusion (syngamy). Fusion in angiosperms is interestingly a double fertilization event (see below; Nawaschin, 1898), whereas in animals or gymnosperms a single fertilization is the norm.

In the microsporangia the pollen mother cell (PMC) goes through meiosis to generate four haploid microspores, which are released after cytokinesis. Each microspore contains one haploid nucleus, which undergoes two mitosis steps. Of the two nuclei after the first division, one is called the vegetative nucleus and the second a germ nucleus. Depending on the species, the time point where the germ nucleus divides by a second mitosis into two sperm nuclei can be before pollen tube development or within the pollen tube. In the megasporangia the megaspore mother cell (MMC) goes through meiosis, and of the four haploid meiotic products (megaspores) only one proceeds, whereby the other three megaspores degenerate. The surviving megaspore undergoes three mitosis cycles, resulting in a mature eight-nucleated seven celled embryo sac (Figure 1). When the pollen grain dehisces from the anther head and makes contact with the style, it develops a pollen tube which delivers the two sperm nuclei to the ovule. One fuses with the egg cell and the other with the binucleated central cell (Figure 1) of the embryo sac (i.e. double fertilization). Double fertilization is highly constrained within flowering plants and ensures viable seed development, as the union of the binucleated central cell with a sperm nucleus results mainly in triploid endosperm tissue which nourishes the embryo (Berger et al. 2006; 2008). In contrast, in gymnosperms the female gametophyte is a large

multicellular organ (the archegonia), which harbours the egg cells. After fertilization of the egg cell the remaining female gametophyte tissue nourishes the developing embryo (Baroux et al., 2002).

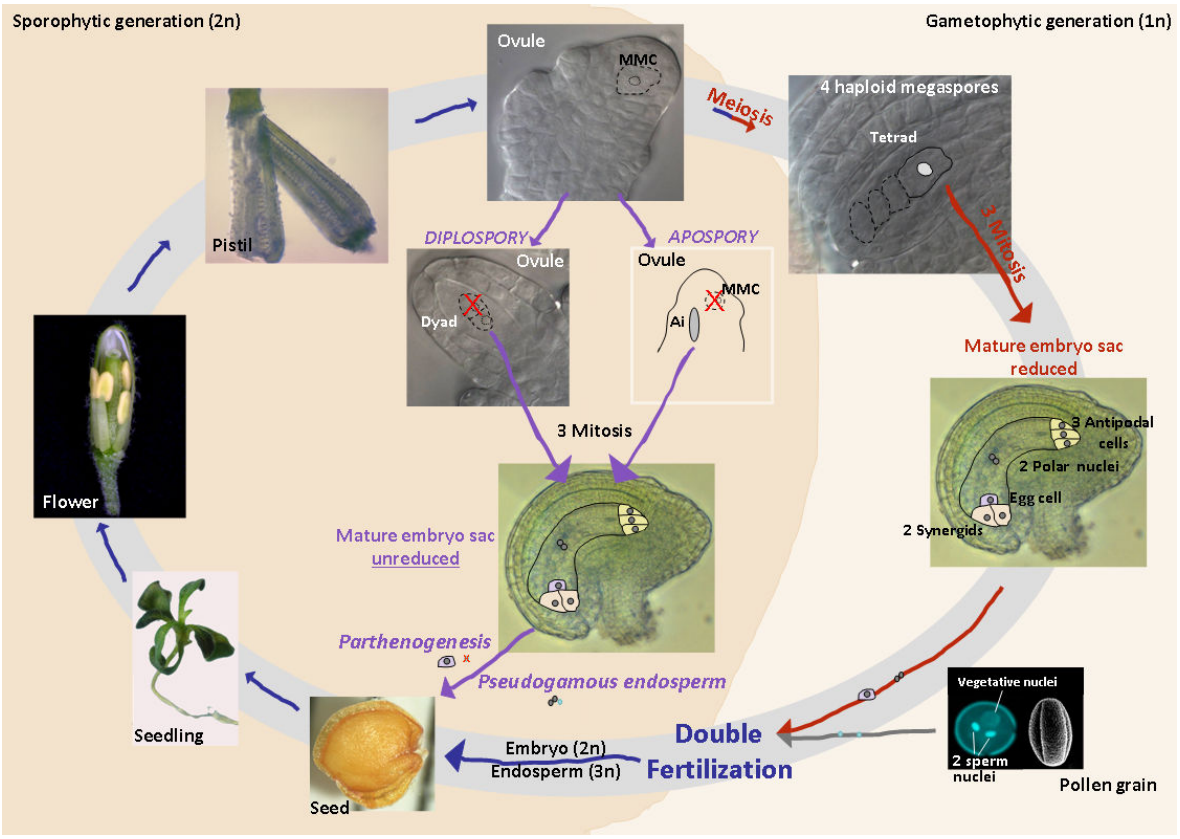


Figure 1. The life-cycle in flowering plants follows an alternation of sporophytic and gametophytic generations. In sexual plants, the megaspore mother cell (MMC) in the ovule goes through meiosis and results in four haploid megaspores. One of these proceeds with 3 mitosis steps to form a seven celled-eight nuclear embryo sac (mature embryo sac). Both the egg cell and the two polar nuclei in the central cell fuse with a sperm nucleus to form embryo and endosperm (double fertilization). In asexual plants, no meiotic reduction occurs, which is achieved either through *diplospory* (failing to enter meiosis I, resulting in a dyad instead of a tetrad, where one cell proceeds with three mitosis steps to form an unreduced embryo sac), or *apospory* in which an alternative cell (Ai, aposporous initial) of the embryo sac takes over the fate of the MMC. In both types, the egg cell develops without fertilization by a sperm nucleus into an embryo (parthenogenesis). The endosperm can be the result of autonomous formation (no fertilization) or by a fertilization event between polar nuclei and a sperm nucleus (pseudogamy). All photos are taken from *Boechera*.

The consequences of sexual reproduction are highly variable and unpredictable with respect to genotype and phenotype, as the offspring are a mixture of merged genomes. In contrast, asexual reproduction is characterized by major key differences. Asexual reproductive pathways are very different among plants, and can be generally subdivided into vegetative, or reproduction with clonal seed formation. Vegetative asexuality occurs when somatic cells grow additional structures which will become independent new individuals, for example in flowering plants these structures can be stolons (*e.g.* strawberry), rhizomes (*e.g.* grass), bulbs (*e.g.* onion), tubes (*e.g.* potato) or plantlets (*e.g.* duckweed). Vegetative propagation is employed in agriculture, taking

advantage of the predictable genotypic and phenotypic constitution of the derived offspring. Asexual reproduction though seed requires the formation of gametes, and the offspring is genotypically predictable and essentially maternal clones (although see *Cupressus*, Pichot et al., 2001).

Asexual seed formation in plants is referred to as apomixis (agamospermy, Asker and Jerling, 1992). Apomixis has frequently evolved from sexuality in plants (Carman, 1997), a switch which requires a number of adaptations. For example, sexual reproduction requires a reductional phase (during meiosis I) to ensure the normal somatic genome constitution after fusion (syngamy), whereas this phase is avoided in apomixis (*apomeiosis*). A second trait is fertilization-independent development of the unfertilized egg cell (parthenogenesis), and additionally in plants, the capacity to develop *autonomou*s (*Taraxacum*, van Dijk et al., 1999) or *pseudogamous* endosperm tissue (reviewed in Grimanelli et al., 2001; Koltunow and Grossniklaus, 2003; Spielman et al., 2003; Nogler, 1984a; Grossniklaus et al., 2001a).

2. *Apomixis*.

History. As early as 1841, the first report made by J. Smith (1841) illustrates the observation of seed production in the absence of male plants of *Alchornea ilicifolia*, although Strasburger (1878) found that the seeds of *Alchornea* can arise from adventitious buds (additionally developed) in the tissue of the nucellus and not from unfertilized egg cell. Increasing numbers of observations were made showing seed production without fertilization, including in ferns (Druery, 1886) and algae (Klebs, 1896). Improvements in chromosome staining and cell preparation yielded very detailed cytological and embryological descriptions of sexual *versus* apomictic reproduction. Strasburger (1905; 1909) noticed that many apomictic genera of plants are characterized by polyploidy, and he assumed that sexual species carried a hidden potential for apomixis which is expressed in their hybrids. In the first chapter of the “Historical Survey” by Gustafssons (1946), it is interesting to discover that, a hundred years ago, Strasburger and Ernst had disputed about the causes of apomixis: hybridization, as pointed out by Ernst, or hybridization only in conjunction with an asexual propensity in sexual parent species, as claimed Strasburger. At present, the question is still unresolved, and it seems that more conditions than these two are necessary to express apomixis. The complexity of naturally occurring apomixis could explain why many questions about apomixis are still unanswered. Some examples of apomictic plants, many of which were reported at the beginning of the twentieth century and are still under investigation to elucidate apomixis, include *Hieracium* (Rosenberg, 1907), *Taraxacum* (Murbeck, 1904), *Erigeron* and

Eupatorium (Holmgren, 1919), *Potentilla* and *Poa* (Müntzing, 1928), *Crepis* (Babcock et al., 1938) and *Ranunculus* (Nogler, 1984b).

Process. Apomixis is the formation of seeds which are genetically identical to the maternal plant (Koltunow and Grossniklaus, 2003; exception *Cupressus*, see Pichot et al., 2001). Apomixis has been found in over 400 plant species (Nogler, 1984a), with the majority of apomictic species occurring within Poacea *(Poa)*, Rosaceae *(Rubus, Sorbus)* and Asteraceae *(Achillea, Crepis, Hieracium, Taraxacum).* With very few exceptions apomicts are mainly polyploid and highly polymorphic on the morphological level, much of which can be attributed to hybridisation (Otto and Whitton, 2000; Comai, 2005). The association between apomixis and polyploidy has been the subject of debate in understanding the origin and evolution of asexuality. Apomixis can be divided into two forms: gametophytic apomixis and adventitious embryony.

In flowering plants, female gametogenesis occurs within ovules, a specialized reproductive organ (Drews et al. 1998). The formation of an embryo from a somatic cell of the nucellus or inner integument of the mature ovule, which is described in detail in *Citrus,* is called *adventitious embryony* (Koltunow, 1993). Gametophytic apomixis can be subdivided into *diplospory* and *apospory* (Nogler, 1984a; Asker and Jerling, 1992; Figure 1). *Diplospory* refers to the process in which the megaspore mother cell (MMC), a cell that is committed to the sexual pathway, fails to enter or to complete the reductional phase of meiosis I (*apomeiosis*), yielding an unreduced embryo sac. In *apospory*, an alternative cell (AI or aposporous initial) of the nucellus proceeds with mitosis to develop into an unreduced embryo sac (Nogler, 1984a) while the nearby MMC degenerates.

In all forms of apomixis, the formation of an unreduced female gametophyte and its unfertilized development into an embryo (*parthenogenesis*) are the differences between sex and apomixis. Endosperm formation can be similar between sexual and apomictic plants, as most apomicts need fertilization of the central cell for proper endosperm formation (Haig and Westoby, 1991; Quarin, 1999; Spielman et al., 2003; Grossniklaus et al., 2001a). It is known that for normal embryo development stable endosperm formation is necessary. In *Arabidopsis thaliana*, four different signalling pathways between embryo and endosperm during sexual seed development have been postulated (Ungru et al., 2008): a first immediate signal from fertilized embryo to endosperm to stimulate endosperm fate; a second signal from endosperm to embryo; a third signal from endosperm to embryo again to stimulate development beyond the globular stage; and, a last signal from the embryo which is responsible for seed survival. Nowack et al. (2006) have shown, using a *cdka*-1 pollen mutant that contains only one sperm nucleus, that the fusion of the egg cell with this nucleus is enough for the triggering of endosperm proliferation which later arrests (3DAP), and eventually aborts (9DAP;

before globular stage). In apomictic plants, no fertilization of the egg cell is necessary, which means that the first signal (immediate) from embryo to endosperm is non-existent or modified.

Regulation. How many genes control apomixis expression? One has to consider whether apomixis is a novel breeding system, or an alteration of the sexual one in which the traits apomeiosis, parthenogenesis and autonomous endosperm formation were derived. Over time, different models have been postulated, from the single locus regulation hypothesis (Bicknell et al., 2000; Nogel, 1948b; Grossniklaus et al., 2001a; Grimanelli et al., 2001) with a master gene (Mogie, 1992), to more recent works which point to independent regulation of apomixis components: e.g. *Taraxacum officinale* (van Dijk et al., 1999; 2003), *Hieracium* (Bicknell and Koltunow, 2004), *Poa pratensis* (Albertini et al., 2001), *Ranunculus megacarpus* (Nogler, 1984b) and *Erigeron annuus* (Noyes, 2000). As diverse as the apomictic forms are, the occurrence of facultative apomicts with sexual and apomictic seed formation in one plant, makes it tempting to consider apomixis as an alteration of the sexual pathway (Koltunow and Grossniklaus, 2003) and not controlled by a single locus. Moreover, it is becoming apparent that apomixis expression is a consequence of asynchronous and local changes in gene expression, which may be induced by hybridization and subject to the complex control of gene regulation by epigenetic and abiotic factors (Carman, 1997; Grimanelli et al., 2001; Sharbel et al., 2009).

From its discovery over a hundred years ago, researchers have been fascinated by apomixis and the potential it holds for agricultural improvements and support (Spillane et al., 2004). Due to the nature of apomixis (aside from rare exceptions like *Cupressus*, (Pichot et al., 2001), offspring are genetically identical to the mother plant. Ordinarily, recombination during meiosis in sexual plants leads to new allelic combinations in offspring, and can break apart favourable gene combinations. The absence of recombination in apomicts circumvents this process, and all favourable or introduced gene combinations are inherited as a single linkage group in the next generation. Fully understanding apomixis would greatly facilitate the ability of plant breeders to fixate and propagate genetic heterozygosity, and associated hybrid vigor in crop plants (Spillane et al., 2004).

The transfer of apomixis into sexual species through introgression has not yet been successful. Introgression of genes from the apomict *Tripsacum dactyloides* to maize (Sokolov et al., 1998a/b) was very laborious, and so far has not lead to strains for commercial use (Sokolov et al., 2000) due to low seed set and male sterility. Furthermore, a suspected barrier seems to be dysfunctional endosperm formation, resulting from deviation from the normal angiosperm 2:1 maternal:paternal (2m:1p) genome balance. Deviations from this ratio in sexual plants can have lethal impacts on development of the

embryo (Scott et al., 1998). Some apomictic species produce unreduced sperm nuclei as well as eggs (*Boechera,* Voigt et al., 2007), but it has yet to be confirmed whether apomeiosis *sensu stricto* occurs on the male side, or if it is a secondary adaptation that maintains the 2m:1p ratio. Autonomous endosperm formation is very rare, and how it evolved is unknown. Noyes et al. (2007) proposed a pleiotropic factor 'F' which signals to the plant that fertilization has occurred, initiating autonomous embryo and endosperm development. In *Arabidopsis*, a mutation in the *FIS* gene induces autonomous endosperm formation, but the resultant seeds do not reach maturity. This may be due to missing signals between embryo and endosperm during development (Ungru et al., 2008).

Engineering of apomixis through "misexpression" of apomixis candidate genes in sexual plants has not been successful either. A problem is that among agricultural important crop plants no natural apomicts exist. Furthermore, knowledge regarding which genes could be deregulated is scant, let alone thoughts on how the deregulation could be performed. In recent years a number of mutations in *Arabidopsis* have been shown to mimic traits of apomixis. In 2008, Ravi et al. reported a mutation in *Arabidosis* which led to the formation of an unreduced egg cell (2n; albeit at very low frequencies). A mutation in the gene *SWI*, called DYAD (Agashe et al., 2002), causes synapsis to fail in female meiosis, but this effect depends on the presents of two DYAD copies. The egg cell of DYAD mutants can be fertilized by haploid sperm to generate triploid offspring. Autonomous endosperm formation is promoted by a mutation in *FIS* (Ohad et al., 1996), and parthenogenesis by a mutation in *MSI* (Guitton and Berger, 2005). Intriguingly, such single mutations disrupted established sexual pathways, and led to traits reflective of apomixis. But the low penetrance of these mutations and the difficulties in coordinating the findings to establish apomictic *Arabidopsis* indicates that global changes influencing expression regulatory networks likely need to be fine balanced for stable apomixis expression. A strategy in comparing global expression profiles between sexual and apomictic tissues using network analyses would yield more complex, but fruitful knowledge.

Species of the genus *Boechera* (formerly *Arabis*) represent an optimal model system, as many comprise both sexual and apomictic forms. Importantly, both diploid and triploid apomicts are known, and hence diploid apomicts can be compared to diploid sexual without the added influences of polyploidy. Furthermore, its close relationship to *Arabidopsis* (Koch et al., 2003) facilitates the use of powerful genomics and molecular tools of *Arabidopsis* for comparative genomic studies of unique traits.

3. *The genus Boechera (Brassicaceae) and Apomixis.*

Boechera comprises 109 North American taxa (Al-Shehbaz et al., 2006; 2007a/b), wherein 71 are sexual diploid and 38 are thought to be hybrid in origin, and mainly triploid and apomict. Based on morphological characters, *Boechera* was originally included within the genus *Arabis* (Rollins, 1993; Mulligan, 1995), but chromosome number (x = 7) and molecular data (Koch et al., 1999; 2000) led to the consideration of *Boechera* as an independent genus. It furthermore showed closer relationships to other North American genera like *Pennellia* and *Haliomolobus* rather than to European *Arabis* (Bailey et al., 2006). The highly polymorphic nature of the genus *Boechera*, species of which can be characterized by simple or branched stems and trichomes, siliques upwards, downwards or sideways, flower color from white to purple, open flowers or closed flowers, makes the taxonomic relationships of this group complex (Al-Shehbaz et al., 2006; 2007a/b).

Al-Shahbaz (2007b) identified 38 hybrids, the most prominent and best studied of which is *Boechera divaricarpa*. It is hypothesized that *B. stricta* and *B. holboellii* were the ancestral parental species which crossed to generate *B. divaricarpa* (Koch et al., 2003; Dobeš et al., 2004a/b; Song et al., 2006). *Boechera stricta* has been shown to be predominantly diploid and sexual, while *B. holboellii* and *B. divaricarpa* are facultative apomicts, and highly variable with respect to ploidy, morphology, and genetic polymorphism (Roy, 1995; Sharbel et al., 2001; Koch et al., 2003; Dobeš et al., 2004a; Sharbel et al., 2004; Schranz et al., 2005; Sharbel et al., 2005; Kantama et al., 2007). The basic chromosome number (x) is 7, and polyploidy (mainly triploidy) and aneuploidy are common (Böcher, 1951; Sharbel and Mitchell-Olds, 2001; Sharbel et al., 2004; Sharbel et al., 2005). Polyploidy and aneuploidy are widely distributed, and have originated multiple times in different geographical regions (Sharbel et al., 2001; Sharbel et al., 2005).

Boechera is distributed throughout North America and its radiation was significantly influenced by Pleistocene glaciations (Dobeš et al., 2004a/b), which lead to separation and isolation of species groups, followed by retreating glaciers and complex recolonization patterns. The study of population structure in *B. stricta* showed similar high levels of genetic diversity for northern (Idaho, Montatna) and southern population (Utah, Colorado; Song et al. 2006), suggesting that several glacial refugia (in the North and South) during glacial history existed and that recolonization proceeded from there (Song et al., 2006). Furthermore, the study revealed recent genetic admixture between northern and southern groups, as well as with other species like *B. holboellii* (Song et al., 2006).

An analysis of geographic patterns and population structure in *B. holboellii* showed, based on chloroplast DNA sequence variation, an allocation of ancestral haplotypes in northern areas (northern Rocky Mountains of the US, Idaho, Montana) and their derivates in southern areas (Sothern Rocky Mountains, and Sierra Nevada, Utah,

Colorado) (Dobeš et al. 2004b). In a microsatellite study a decrease in microsatellite diversity was shown in northern compared to southern populations (Dobeš et al. 2004a), suggesting that postglacial recolonization from southern refugia towards north resulted in a decline in intraspecific diversity (Soltis et al., 1997). The complex recolonization pathways of *B. stricta and B. holboelli* are apparent in frequent hybridization and allopolyploid lineages (Boecher, 1951; Rollins, 1993; Mulligan, 1995; Sharbel and Mitchell-Olds, 2001; Dobeš et al., 2004b; Kantama et al., 2007), which indicates incomplete reproduction isolation between species.

Boechera was originally studied by the Danish researcher Tyge Böcher (1947; 1951; 1954) who characterized numerous taxa using cytological, embryological and morphological traits. He observed irregular male and female gametophyte development in some accessions, indicating apomixis. He documented disturbed pollen development which was the source of unreduced pollen grains, and furthermore he concluded that there were differences in the formation of unreduced male and female gametes. In an emasculation experiment he found that pollen is necessary for apomictic seed development. Chromosome counts of embryo and endosperm showed that both were autonomously derived, and he concluded that pollen arrival was necessary to trigger autonomous development.

Two of his observations have subsequently been modified. First, using the flow cytrometric seed screen (FCSS), a method whereby embryo and endosperm ploidy from single seeds can be measured (Matzk, 2000), it has been shown that the central cell must occasionally fuse with a sperm nucleus, and hence pseudogamous endosperm formation also occurs (Voigt et al., 2007; Aliyu unpublished data; Chapter II). Secondly, in measuring seeds from apomictic plants, we found only in very rare cases sexually produced seeds from facultative apomictic accessions (reduced female gametes, below 1%, Chapter II; see also Schranz et al., 2006), which is in contrast to Böcher's observations of apomictic seed development in lower and middle parts of the shoot, and sexual reproduction in newly developing flowers (Böcher, 1951).

Apomictic *Boechera* are diplosporous. Cytological studies on *Boechera* (Böcher, 1951; Naumova et al., 2001) have shown that embryo sac formation in sexual plants is of the *Polygonum*-type, whereby the apomictic plants follow the *Taraxacum*-type. The MMC of apomictic plants fails to enter the reductional phase of meiosis (meiosis I) and variable degrees of chromosome pairing (univalent, bivalent and trivalent) have been observed in polyploids (Böcher, 1951). This cell proceeds with the second phase (meiosis II), leading to the formation of an unreduced dyad instead of a reduced tetrad. One of the dyad cells degenerates, while the other continues development, as in the case of the sexual pathway, with three mitosis steps. The mature embryo sac is morphologically

indistinguishable from the sexual or apomictically-derived pathway (Naumova et al., 2001). Pseudogamous endosperm formation has for the most part been observed, but rare cases of autonomous endosperm formation have also been reported (Naumova et al. 2001; Voigt et al., 2007).

Microsporogenesis is disturbed in apomictic *Boechera*, and leads to the formation of reduced, unreduced and aneuploid pollen grains (Böcher, 1951; Dobeš et al., 2004b; Schranz et al., 2005; Sharbel et al., 2005; Voigt et al., 2007). Morphologically, the pollen grains from sexual and apomictic plants are very distinct (Voigt et al. 2007). In sexual plants the pollen grains are regular and ellipsoid, while irregular sizes and shapes of the pollen grains characterize apomictic plants (Voigt et al., 2007). In several studies, pollen size has been used to distinguish between sexual and apomictic plants (Al-shebaz et al., 2006; Schranz et al., 2006; Koch et al., 2003; Dobeš et al., 2004 a/b), but this method should be taken with caution in *Boechera*. The regular shape is a good in indicator of breeding system, but is less reliable with respect to ploidy level (Voigt et al., 2007).

B chromosome or not? B chromosomes are supernumerary chromosomes to the normal autosomes, are not essential for the organism, and can vary in number within and between populations. They are mainly heterochromatic, do not follow Mendelian laws of inheritance and have been found in all major groups of animals and plants (Camacho et al., 2000). The aneuploid chromosome of *Boechera* has been hypothesized to be a B chromosome (Sharbel et al., 2004; 2005), since it is variable in recombination potential and is heterochromatic (Camacho et al., 2000; Sharbel et al., 2004). An attractive possibility is that aneuploidy (in the form of a B chromosome) may somehow be involved in the expression of apomixis (Roche et al., 2001). Recent cytogenetic studies identified two aberrant chromosomes for *Boechera* that were previously defined as B chromosomes (Kantama et al., 2007). One of the chromosomes is small in appearance, but highly heterochromatized (*Het*) in comparison to its complement, whereas the second aberrant chromosome is the shortest chromosome of all (*Del*). Thus far, the *Het* chromosome has been only identified in apomictic plants, and the Del chromosome only in apomicts with 15 chromosomes, but their role in apomixis is unknown.

4. *AIMS of the Dissertation.*

Two significant PhD theses, both focusing on *Boechera* and apomixis from different perspectives, have contributed to our knowledge of this species group. Laksama Kantama performed chromosome studies and genetic analyses (PhD thesis, University Wageningen, 2005), whereas Christian Kiefer focused on evolution and phylogeography of *Boechera* and the evolution of apomixis (PhD thesis, University of Heidelberg, 2008). These studies have brought new insights for the evolution apomixis in the genus *Boechera*, especially with respect to the identification of a *Het* and a *Del* chromosome, and the very complex structure of the evolution of the genus.

The aim of my thesis was the characterization of apomixis in *Boechera* and the elucidation of gene expression differences between apomictic and sexual plants, followed by identification of apomixis initiation candidate genes and underlying gene regulation networks.

I) Phenotype characterization

I.I) Characterisation of the apomictic pathway *versus* sexual pathway in *Boechera* plants.

I.II) Comparing diploid apomicts with triploid apomicts to test how ploidy affects apomixis establishment and maintenance.

The genus *Boechera* provides one of the rare opportunities to study differences between both diploid and triploid apomicts on a phenotypic level, but moreover the diploid apomicts can be compared to diploid sexual without the added influences of polyploidy in a transcriptome study.

II) Gene expression

II.I) Elucidation of gene expression differences between diploid apomicts and diploid sexual lines of isolated ovules at meiosis entrance in performing a serial analysis of gene expression (SuperSAGE, Matsumura et al., 2005).

II.II) Combining two independent transcriptome data sets to identify apomixis initiation candidate genes and to gain insight into apomixis initiating processes.

5. *General information for thesis.*

A) The geographical information for all *Boechera* lines used in the course of the thesis are summarized in Table 1, and illustrated in Figure 2. The plants were initially identified by Rollins (1993) and Dorn (1984) and classified as *Boechera stricta* and *Boechera holboellii* (exemplary photos of lines Figure 3A-H). One of the most distinctive species is *B.*

stricta, as it is diploid sexual and monophyletic with the widest distribution (Al-Shehbaz, 2007b; Schranz et al., 2005), but *B. holboellii* represents a polyphyletic group of diploids and triploids which can be sexual and/or apomictic. With the new nomenclature by Al-Shehbaz (2006; 2007a/b), hybrids and apomicts were separated from sexual diploid taxa, and strictly speaking *B. holboellii* are now plants originating from Greenland (Al-Shehbaz et al., 2006). Unfortunately, our plants have been yet not been reviewed, therefore I will remain with the initial identification and focus rather on reproductive mode and genotype/line than on taxonomy.

B) In Chapter I, diploid apomicts were referred to as 2n+B (B = B chromosome), whereas in all other Chapters, 2n+x. Recent findings convinced me to keep designations more general, since it is unclear whether all aneuploid chromosomes can be considered B's.

C) Chromosome count data are indicated by "n", and thus plants with 14 chromosomes are titled "2n", plants with 15 chromosomes with "2n+x" and with 21 chromosomes "3n". Genome size measurements via flow cytometry are indicated by "C". Comparing the genome content value with a defined standard measurement, values are interpreted as diploid = 2C; triploid = 3C; tetraploid = 4C; hexaploid = 6C and so on.

Table 1. Geographic distribution of plants used in different experiments, reproductive mode is indicated by sex = sexual reproduction and apo = apomixis.

Species	ID	sex/apo	Population	Map Nr.
B. holboellii	28	apo	Birch Creek, Ravalli Co, MT	1
	61	sex	Wallowa Mountains, OR	2
	67	apo	Ranch Creek, Granit Co, MT	3
	78	apo	Ranch Creek, Granit Co, MT	3
	105	sex	Bandy Ranch, Missoula Co, MT	4
	120	apo	Mule Ranch, Beaverhead Co, MT	5
	148	apo	Challis, Lemhi Co, ID	6
	195	apo	Highwood Mountains, MT	7
	205	apo	Charlies Gulch, Ravalli Co, MT	8
	209	apo	Charlies Gulch, Ravalli Co, MT	8
	218	apo	Highwood Mountains, MT	7
	290	sex	Sonora Pass, Toiyabe NF, CA	9
	300	apo	Birch Creek, Ravalli Co, MT	1
B. stricta	132	sex	Canyon Creek, Beaverhead Co., MT	10
	ES 910	sex	Parker Meadow, ID	11
	ES 503	sex	Gold creek, CO	12
	ES 596	sex	Blue Lakes Road, CA	13

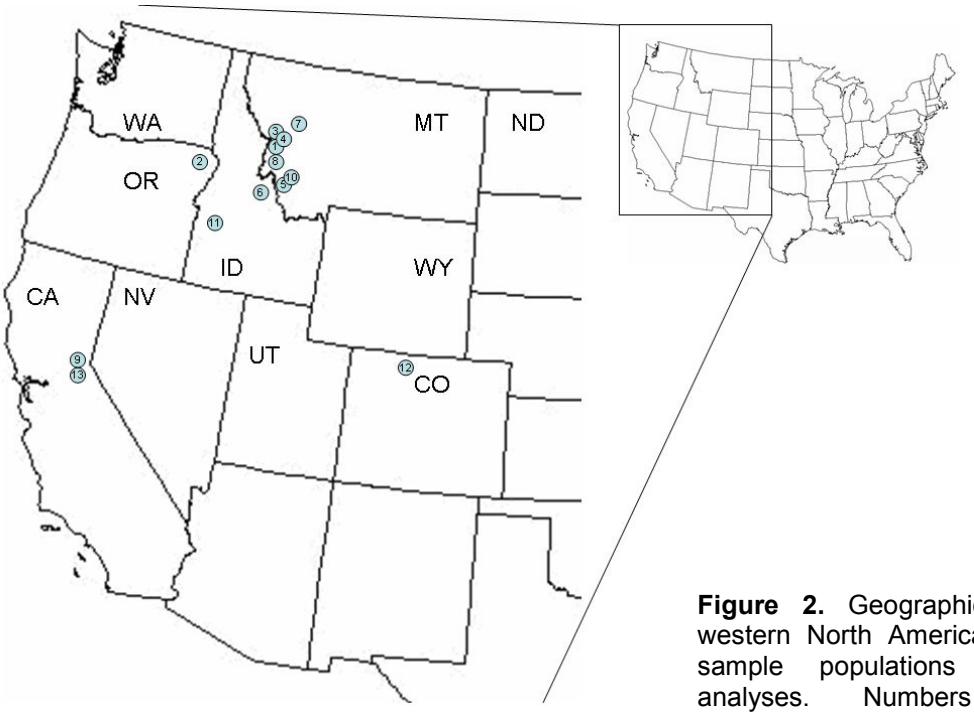


Figure 2. Geographic map of western North America, showing sample populations used in analyses. Numbers in circle correspond to numbers in Table 1.

Figure 3. Exemplary photos of *Boechera* lines used in conducted work.

(A) ID 132 *B. stricta* sexual diploid

(B) ID 105 *B. holboellii* sexual diploid

(C) ID 120 *B. holboellii* apomict diploid

(D) ID 28 *B. holboellii*, apomict diploid

(E) ID 300 *B. holboellii*, apomict diploid

(F) ID 148 *B. holboellii*, apomict triploid

(G) ID 195 *B. holboellii*, apomict triploid

(H) ID218 *B. holboellii*, apomict triploid

6. Literature cited.

Agashe B, Prasad CK and Siddiqi I. 2002. Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in *Arabidopsis*. Development 129: 3935-3943.

Albertini E, Porceddu A, Ferranti F, Reale L, Barcaccia G, Romano B and Falcinelli M. 2001. Apospory and parthenogenesis may be uncoupled in Poa pratensis: a cytological Investigation. Sexual Plant Reproduction 14: 213–217.

Al-Shehbaz IA and Windham MD. 2006. New and noteworthy Species of *Boechera* (Brassicaceae) I: Sexual diploids. Harvard Papers in Botany 11: 61–88.

Al-Shehbaz IA and Windham MD. 2007a. New and noteworthy Species of *Boechera* (BRASSICACEAE) II: Apomictic hybrids. Harvard Papers in Botany 11: 257–274.

Al-Shehabz IA and Windham MD. 2007b. New and noteworthy Species of *Boechera* (BRASSICACEAE) III: Additional Sexual diploids and Apomictic hybrids. Harvard Papers in Botany 12: 235–257.

Asker S and Jerling L. 1992. Apomixis in Plants. Boca Raton, FL. CRC Press.

Babcock EB and Stebbins GL. 1938. The American species of Crepis. Publ. Carn. Inst. Wash 504: 1-99. (Cited after Gustafsson, 1946)

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick Jr SI, Windham MD and Al-Shehbaz IA. 2006. Toward a global phylogeny of the Brassicaceae. Molecular Biology and Evolution 23: 2142-2160.

Baroux C, Spillane C and Grossniklaus U. 2002. Evolutionary origins of the endosperm in flowering plants. Genome Biology 3: Reviews 1026.

Berger F, Grini PE and Schnittger A. 2006. Endosperm: an integrator of seed growth and development. Current Opinion in Plant Biology 9: 664-670.

Berger F. 2008. Double-fertilization, from myths to reality. Sexual Plant Reproduction 21:3-5.

Bertolla F and Simonet P. 1999. Horizontal gene transfers in the environment: natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. Research in Microbiology 150: 375-384.

Bicknell RA, Borst NK and Koltunow AM .2000. Monogenic inheritance of apomixis in two *Hieracium* species with distinct developmental mechanisms. Heredity 84: 228–237.

Böcher TW 1947. Cytologocal studies of *Arabis holboellii*. Hereditas 33.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Det Kongelige Danske Videnskabernes Selskab 6: 1-59.

Böcher TW. 1954. Experimental taxonomical studies in the *Arabis holboellii* complex. Svensk Botanisk Tidskrift 48: 31-44.

Camacho JPM, Sharbel TF and Beukeboom LW. 2000. B chromosome evolution. Philosophical Transactions of the Royal Society B 355: 163-178.

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

Comai L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews 6: 836-846.

Dobeš C, Mitchell-Olds T and Koch MA. 2004a. Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera xdivaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers--an integrative approach. American Journal of Botany. 91: 2087-2101.

Dobeš C, Mitchell-Olds T and Koch MA. 2004b. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* x*divericarpa*, and *A. holboellii* (Brassicaceae). Molecular Ecology. 13: 349-370.

Dorn RD. 1984. Vascular Plants of Montana. Mountain West Publishing, Cheyenne.

Drews GN, Lee D and Christensen CA. 1998. Genetic analysis of female gametophyte development and function. The Plant Cell 10: 5–18. (Cited after Gustafsson, 1946)

Druery CT. 1886. Observations on a singular mode of development in the Lady-Fern (Athyrium Filix-foemina). Journal of the Linnean Society of London 21: 354-360.

Ernst A. 1913. Embryobildung bei Balanophora. Flora 106: 129-159. (cited after Gustafsson, 1946-1947)

Grimanelli D, Leblanc O, Perotti E and Grossniklaus U. 2001. Developmental genetics of gametophytic apomixis. Trends in Genetic 17: 597-604.

Grossniklaus U, Nogler A and van Dijk PJ. 2001a. How to avoid sex: the genetic control of gametophytic apomixis. The Plant Cell 13: 1491–1497.

Grossniklaus U**,** Spillane C, Page DR and Köhler C. 2001b. Genomic imprinting and seed development: Endosperm formation with and without sex. Current Opinion in Plant Science 4: 21–27.

Guitton AE and Berger F. 2005. Loss of function of MULTICOPY SUPPRESSOR OF IRA 1 produces nonviable parthenogenetic embryos in Arabidopsis. Current Biology 15: 750-754.

Gustafsson Å. 1946-1947. Apomixis in higher plants. I-III Lunds Univ. Årsskr. N. F. Avd. 2:42

Haig D and Westoby M. 1991. Genomic imprinting in endosperm: its effects on seed development in crosses between different ploidies of the same species, and its implications for the evolution of apomixis. Philosophical Transactions of the Royal Society B 333: 1–13.

Hill JL and Lord DR. 1994. Wild-type flower development:gynoecial initiation in Arabidopsis: An Atlas of Morphology and Development (Bwmann J., ed.). New York: Springer Verlag, pp. 158-159.

Holmgren I. 1919. Zytologische Studien über die Fortpflanzung bei den Gattungen Erigeron und Eupatorium. Kgl. Sv. Vet. Ak. Handl. 59. No. 7: 1-118. (Cited after Gustafsson, 1946).

Kantama L. 2005. Chromosome studies and genetic analyses of natural and synthetic apomictic model species. Wageningen University, ISBN 90-8504-321-2

Kantama L, Sharbel TF, Schranz ME, Mitchell-Olds T, de Vries S and de Jong H. 2007. Diploid apomicts of the *Boechera holboelli* complex display large scale chromosome substitutions and aberrant chromosomes. Proceedings of the National Academy of Sciences of the United States of America 104: 14026-14031.

Kiefer C. 2008. Evolution and Phylogeography of the North American genus Boechera (Brassicaceae) and the Evolution of Apomixis. University of Heidelberg.

Klebs G. 1896. Die Bedingungen der Fortpflanzung bei einigen Algen und Pilze. Jena 543 pp. (Cited after Gustafsson, 1946)

Koch M, Bishop J and Mitchell-Olds T. 1999. Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biology. 1: 529-537.

Koch M, Haubold B and Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Molecular Biology and Evolution 17: 1483-1498.

Koch M, Dobeš C and Mitchell-Olds T. 2003. Multiple Hybrid Formation in Natural Populations: Concerted Evolution of the Internal Transcribed Spacer of Nuclear Ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molecular Biology and Evolution 20: 338-350.

Koltunow AM and Grossniklaus U. 2003. Apomixis: a developmental perspective. Annual Review of Plant Biology 54: 547-574.

Koltunow AM. 1993. Apomixis: Embryo Sacs and Embryos Formed without Meiosis or Fertilization in Ovules. The Plant Cell 5: 1425-1437.

Nawaschin SG. 1898. Resultate einer Revision der Befruchtungsvorgaenge bei *Lilium martagon* und *Fritillaria tenella*. Bull Acad Imp Sci St Petersburg 9: 377-382.

Matzk F, Meister A, and Schubert I. 2000. An efficient screen for reproductive pathways using mature seeds of monocots and dicots. The Plant Journal 21: 97-108.

Mogie M. 1992. The Evolution of Asexual Reproduction in Plants. Chapman & Hall, London.

Mulligan GA. 1995. Synopsis of the genus *Arabis* (Brassicaceae) in Canada, Alaska and Greenland. Rhodora 97: 109-163.

Müntzing A. 1928. Pseudogamie in der Gattung *Pofentillu.* Hereditas 11: 267-283.

Murbeck S. 1904. Parthenogenese bei den gattungen Taraxacum und Hieracium. Bot. Not. (Lund): 285-296. (Cited after Gustafsson, 1946)

Naumova TN, 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.

Nawaschin SG. 1898. Resultate einer Revision der Befruchtungsvorgänge bei *Lilium martagon* und *Fritillaria tenella*. Bull. Acad. Imp. Des Sci. St. Petersberg 9: 377-382.

Nogler GA. 1984a. Gametophytic apomixis. Pp. 475-518 *in* B. M. Johri, ed. Embryology of Angiosperms. Springer Verlag, New York.

Nogler GA. 1984b. Genetics of apospory in apomictic *Ranunculus auricomus*.V. Conclusion. Bot Helv 94: 411–422.

Nogler GA. 1995. Genetics of apomixis *in Ranunculus auricomus*. VI Epilogue. Bot Helv 105: 111–115.

Nowack MK, Grini PE, Jakoby MJ, Lafos M, Koncz C and Schnittger A. 2006. A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nature Genetics 38: 63-67.

Noyes RD. 2000. Diplospory and Parthenogenesis in Sexual x Agamospermous (Apomictic) *Erigeron* (Asteraceae) Hybrids. International Journal of Plant Sciences 161: 1-12.

Noyes RD, Baker R and Mai B. 2007. Mendelian segregation for two-factor apomixis in *Erigeron annuus* (Asteraceae). Heredity 98: 92–98.

Ohad N, Margossian L, Hsu YC, Williams C, Repetti P and Fischer RL. 1996. A mutation that allows endosperm development without fertilization. Proceedings of the National Academy of Sciences of the United States of America 93: 5319-5324.

Otto SP and Whitton J. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34: 401-437.

Pichot C, El Maataoui M, Raddi S and Raddi P. 2001. Surrogate mother for endangered *Cupressus*. Nature 412: 39.

Quarin CL. 1999. Effects of pollen source and pollen ploidy on endosperm formation and seed set in pseudogamous apomictic *Paspalum notatum*. Sexual Plant Reproduction 11: 331–335.

Ravi M, Marimuthu MPA and Siddiqi I. 2008. Gamete formation without meiosis in *Arabidopsis*. Nature. 451: 1121-1124.

Roy BA. 1995. The breeding systems of six species of *Arabis* (Brassicaceae). American Journal of Botany 82: 869-877.

Roche D, Hanna WW and Ozias-Akins P. 2001. Is supernumerary chromatin involved in gametophytic apomixis of polyploid plants? Sexual Plant Reproduction 13: 343-349.

Rollins RC. 1993. The Cruciferae of Continental North America. Stanford, California: Stanford University Press.

Rosenberg O. 1907. Cytological studies on the apogamy in Hieracium. Bot. Tidskr. (Köbenhayn) 28: 143-170. (Cited after Gustafsson, 1946)

Schranz ME, Dobeš C, Koch MA 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, Kantama L, de Jong H, and Mitchell-Olds T. 2006. Asexual reproduction in a close relative of Arabidopsis: a genetic investigation of apomixis in *Boechera* (Brassicaceae). The New Phytologist 171: 425-438.

Scott RJ, Spielman M, Bailey J and Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125: 3329-3341.

Smith J. 1841. Notice of a plant which produces seeds without any apparent action of pollen. Transactions of the Linnean Society of London 18. (Cited after Gustafsson, 1946)

Sokolov VA, Kindinger B and Khatypova IV. 1998a. Apomictically reproducing 39-chromosome maize-Tripsacum hybrids. Genetika 34: 499-506.

Sokolov VA, Lukina LA and Khatypova IV. 1998b. Perspectives of developing apomixis in maize. Maize News Letter 72: 67-68.

Solokov VA and Khatypova IV. 2000. The Development of apomictic maize: Update, problems and perspective. Genetika Vol. 32, 3: 331-353.

Soltis DE, Gitzendanner MA, Strenge DD and Soltis PS. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution 206: 353–373.

Song BH, Clauss MJ, Pepper A and Mitchell-Olds T. 2006. Geographic patterns of microsatellite variation in Boechera stricta, a close relative of *Arabido*psis. Molecular Ecology 15: 357-369.

Spielman M, Vinkenoog R and Scott RJ. 2003. Genetic mechanisms of apomixis. Philosophical Transactions of the Royal Society B 358: 1095-1103.

Spillane C, Curtis MD, and Grossniklaus U. 2004. Apomixis technology development - virgin births in farmers' fields? Nature Biotechnology 22: 687.

Sharbel TF and Mitchell-Olds T. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity 87: 59-68.

Sharbel TF, Voigt ML, Mitchell-Olds T, Kantama L and de Jong H. 2004. Is the aneuploid chromosome in an apomictic *Boechera holboellii* a genuine B chromosome? Cytogenetic and Genome Research 106: 173-183.

Sharbel TF, Mitchell-Olds T, Dobeš C, Kantama L, and de Jong H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic and Genome Research 109: 283-292.

Sharbel TF, Voigt ML, Corral JM, 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 in press.

Strasburger ED. 1878. Über Polyembryonie. Jen. Zschr. Naturw. 12: 647-670. (Cited after Gustafsson, 1946)

Strasburger ED. 1905. Die Apogamie in Eualchemillen und allgemeine Gesichtspunkte, die sich aus ihr ergeben. Jahrb. Wiss. Bot. 41: 88-164. (Cited after Gustafsson, 1946)

Strasburger ED. 1909. Zeitpunkt der Bestimmung des Geschlechtes, Apogamie, Parthenogenesis und Reducktionsteilung. Hist. Beit. 7: 1-124. (Cited after Gustafsson, 1946)

Ungru A, Nowack MK, Reymond M, Shirzadi R, Kumar M, Biewers S, Grini PE and Schnittger A. 2008. Natural variation in the degree of autonomous endosperm formation reveals independence and constraints of embryo growth during seed development in *Arabidopsis thaliana*. Genetics 179: 829-841.

Van Dijk P, Tas ICQ, Falque M and Bakx-Schotman T. 1999. Crosses between sexual and apomictic dandelions (*Taraxacum*). II. The breakdown of apomixis. Heredity 83: 715–721.

Van Dijk P, van Baarlen P and de Jong JH. 2003. The occurrence of phenotypically complementary apomixis-recombinants in crosses between sexual and apomictic dandelions (*Taraxacum officinale*). Sexual Plant Reproduction 16: 71–76.

Voigt ML, Melzer M, Rutten T, Mitchell-Olds T and Sharbel TF. 2007. Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. Vienna: A.R.G. Gantner Verlag K.G.

Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective

Marie-Luise Voigt, Michael Melzer, Twan Rutten, Thomas Mitchell-Olds & Timothy F. Sharbel

Published in the Book
Apomixis: Evolution, mechanisms and perspectives
Editores: E. Hörandl, U. Grossniklaus, P. Van Dijk, T. F. Sharbel
A.R.G. Gantner Verlag K.G.
2007, Chapter XIII p. 236-257

Abstract

The highly polymorphic *Boechera holboellii* complex provides a source of natural variation for genetic studies on the regulation of apomictic development. Sexuals are diploid, whereas apomictic individuals can be diploid or triploid, with or without B chromosomes. By using the flow cytometric seed screen (FCSS) this work has demonstrated high levels of intra- and interindividual variability in terms of embryo and endosperm ploidy, which points to disturbed microsporogenesis. A multifaceted analysis of pollen formation has demonstrated that pollen grain size, morphology and ploidy level are very different between sexual and apomictic accessions, and in addition that pollen grain size and ploidy can not be correlated in apomictic accessions. Diploid apomicts produce reduced as well as unreduced pollen grains, whereas triploid apomicts produce unreduced pollen grains. Fertilization of apomeiotically-derived egg cells and autonomous endosperm formation are nonetheless within the reproductive potential of apomictic individuals. Natural selection acts at various stages in the apomictic pathway to reduce the potential variability generated by an apomictic plant, and it appears that both mutation accumulation and epigenetic phenomena lead to gamete inviability and seed abortion.

KEYWORDS: apomixis, *Boechera holboellii*, pollen, polyploid, sperm.

INTRODUCTION

The evolutionary dynamics of a species or population are fundamentally dependent upon their genetic composition and how this may vary through space and time. Factors affecting genetic variability include effective population size, population structure, and historical contingency. In addition, the type of mating system determines the relative rate at which such variability is accumulated and vertically transferred, and may have implications for how members of a gene pool can adapt to change.

The transfer of DNA from one generation to the next can involve mechanisms which allow the exchange of genetic information between cells or organisms. Meiosis and syngamy are the characteristic mechanisms through which genetic variability is introduced into sexually reproducing (amphimictic) organisms (Birdsell and Wills, 1996). Alternatively, asexual (apomictic) reproduction involves no genetic exchange between separate individuals, but may nonetheless be accompanied by cellular mechanisms which generate limited variability (e.g., automixis). The implications of amphimixis relative to apomixis for the individual (populations, species) are multifaceted, and include both costs (i.e., meiosis, mate finding, etc.) and benefits (i.e. accelerated evolution, decreasing mutational load) which undoubtedly work in an antagonistic fashion to generate an overall positive or negative effect (Butlin, 2002). For the multitude of taxa in which amphimixis has repeatedly evolved, the sum total of the costs and benefits has weighed towards positive effects, thus pointing to significant incurred advantages which allow the maintenance of this form of reproduction.

Parthenogenesis (in animals) and apomixis (in plants) are, for the most part, phenomena associated with females ("gynogenesis"), and thus it is interesting that males are often maintained in asexual populations (D'Souza et al., 2004). A rather rare, but intriguing case of male parthenogenesis and apomixis is "androgenesis", in which the offspring carries only the paternal genome (McKone and Halpern, 2003). In androgenesis female egg cell development is necessary for zygote growth, but the maternal genome is later eliminated. This has been identified in a number of taxa in the genera *Bacillus* (Mantovani and Scali, 1992) and *Corbicula* (Komaru et al., 1998), and in *Cupressus dupreziana* (Pichot et al., 2001).

Males are important insomuch that they are required for sexual reproduction and its associated benefits. But they are also the main factor in the so called "twofold cost" of sex, which leads to a 50% fitness decrease in sexual females when compared to a population composed only of asexual females. This cost can be minimized in populations where sexual and facultative asexual individuals co-exist. The production of males by asexuals provides potential mates for sexual females and may lead to an increase in the

genetic diversity of sexual and asexual offspring. It is intriguing that males (or the male function in hermaphrodites) are maintained within pure asexual populations, since this represents a potential energetic cost with no immediate benefit. The large majority of studies dealing with the costs and benefits of sex has taken the animal point of view, and has only rarely been addressed from the plant perspective (Joshi and Moody, 1995). Plants are a challenging system since most angiosperms (flowering plants) are hermaphroditic (i.e., no female choice of male) and furthermore the phenomenon of double fertilization (i.e., pollen are required for the fertilization of the egg and central cells) means that male gametes have a secondary evolutionary "value". Most apomictic plants and parthenogenetic animals are pseudogamous, although in plants this is due to the requirement of producing a balanced endosperm, while in animals fertilization is required to trigger embryo development.

The Boechera (formerly Arabis) holboellii complex is composed of B. holboellii, B. stricta (formerly *B. drummondii*), and *B. × divaricarpa*, the latter being a hybrid between the first two species (Koch et al., 2003; Dobeš et al., 2004a, b). The basic chromosome number (x) of *Boechera holboellii* is 7, and polyploidy (typically triploid) and aneuploidy are common (Böcher, 1951, 1954). The breeding system of this complex is variable, consisting of both sexual and facultative apomictic forms (Böcher, 1951; Roy, 1995; Naumova et al., 2001). Compounding this variability is the wide distribution of polyploidy and aneuploidy (Böcher, 1951), the former of which has originated multiple times in geographically and genetically distinct populations (Sharbel and Mitchell-Olds, 2001; Sharbel et al., 2005). As is typical for amphi-apomictic complexes, there is a general correlation between polyploidy and apomixis, but this is not always the case as indications of diploid apomixis have been found in certain populations (Böcher, 1951; Roy, 1995). Fixed heterozygosity for allozyme markers in putative diploid and polyploidy apomicts (Roy, 1995) imply that apomictic lineages result from hybridization between divergent genotypes.

As with many apomictic taxa, macro- and microsporogenesis are typically disturbed in polyploid individuals. In *B. holboellii* normal reduced (functional), unreduced and aneuploid pollen can be found within and between different individuals (Böcher, 1951; Dobeš et al., 2004a; Sharbel et al., 2005). Oogenesis proceeds by nuclear restitution after meiosis I, and variable degrees of chromosome pairing (univalent, bivalent and trivalent) have been observed in polyploids (Böcher, 1951). Both pseudogamous and autonomous endosperm formation have been identified (Naumova et al., 2001), the former of which may represent a form of selective pressure for maintenance of viable pollen formation, at least with respect to pollen-tube formation (Böcher, 1951).

We have been studying seed formation in a number of sexual and apomictic accessions of *Boechera* using the flow cytometric seed screen (Matzk et al., 2000). This work has demonstrated high levels of intra- and interindividual variability in terms of embryo and endosperm ploidy (unpublished data), and has hence raised questions regarding pollen development and production. Our observed endosperm ploidy measurements (e.g., 3C, 5C, 6C, 7C, 8C, 9C, 10C) point to disturbed microsporogenesis as the source of this variability. We have thus launched a comprehensive analysis of pollen formation in *Boechera* with the intention of elucidating how pollen production differs between sexual and apomictic accessions.

MATERIALS AND METHODS

Plant material. Seeds were collected from 14 individual plants (accessions) of known genotypes from *Boechera holboellii* and *B. stricta* sampled from throughout North America (Table 1). Plants were identified using keys following (Dorn, 1984) and (Rollins, 1981), and in addition, species classifications were compared to known herbarium specimens. Thirty seeds from each collection were placed onto moist filter paper in Petri plates and cold treated at 4° C for 3 weeks in the dark. Germinating seedlings were then transferred to pots (11 × 11 × 13 cm) containing sterilized soil and grown under natural light in greenhouses at a mean temperature of 26° C. Up to three offspring from each family (in total 26 individuals) were randomly selected for further investigation.

Table 1. General information of *Boechera* mother plants from which offspring were analyzed.

Species	ID	Karyotype[a]	Population[b]
B. holboellii	28	2C+B	5
	61	2C	26
	67	2C+B	21
	78	2C+B	21
	105	2C	7
	120	2C	19
	148	3C	20
	195	3C	4
	205	2C+B	6
	209	2C+B	6
	218	3C	4
	290	2C	31
	300	2C+B	21
B. stricta	132	2C	10

[a] Based on flow cytometric analyses of genome size (Sharbel & Mitchell-Olds, 2001; Sharbel & al., 2005)

[b] Numbers correspond to map locations in Fig. 1 from Sharbel & al. (2005)

Mode of reproduction and fecundity. Twenty to 70 single seeds per individual plant were analyzed using the flow cytometric seed screen (Matzk et al., 2000). In addition,

immature seeds were analyzed by separation of the embryo from endosperm/testa, which were prepared separately to confirm ploidy of embryo to endosperm/testa. Immature seed were chosen in order to identify seeds with unusual ploidy levels in embryo or endosperm which may have been otherwise undetected due to seed abortion during development. Mature seeds were chosen to confirm viable embryo and endosperm ploidy levels. The numbers of non-aborted and aborted seeds were counted from opened single siliques (3 to 14 per plant) for fecundity comparisons between sexual and apomictic plants.

Pollen analyses. In order to define and analyze similar developmental stages of flowers from sexual and apomictic plants, anther head color (yellow) and flower size was used. Preliminary analyses demonstrated that mature pollen grains were found in closed flowers of apomictic plants and in open flowers of sexual plants, and thus anthers from 2 to 3 similarly-sized open (sexual) and similarly-sized unopened (apomictic) flowers per plant were selected for analysis. Anthers were isolated using a dissecting microscope (Stemi 200C, Zeiss), and a Zeiss Axioplan 2 imaging light microscope connected to a FUJIX Digital Camera (HC 300Z) was employed to capture photos. The anthers were used in the following analyses:

Area of pollen grains. It has been shown that there is a strong correlation between ploidy level and pollen grain size (Koch et al., 2003), and thus this indirect procedure was used to estimate the ploidy level of *Boechera* pollen. Forceps were used to homogenize two anther heads per flower on separate microscope slides, followed by squashing in an Aceto-carmine (45 ml acetic acid, 55 ml water, 0.4 g carmine) staining solution under a cover slip. The staining was used to achieve better contrast between pollen grain and background for area measurements. Samples were examined under a light microscope and digital images captured under 20× magnification. To measure area (µm²) of pollen grains the Scion image software for Windows (beta 4.0.2 version 2000) was used (http://www.scioncorp.com/).

Pollen fecundity and viability. Pollen viability was estimated by staining pollen grains with Alexander's staining solution (Alexander, 1969), which yields either green (aborted) or red to deep red (non-aborted) pollen staining (Johnson-Brousseau and McCormick, 2004). Four anther heads per flower were separately mounted directly in a drop of the stain and then covered with cover slip, and pollen grains were released by putting pressure on the cover slip. After 15 min, the microscope slides were heated over a small flame and then examined using normal light microscopy. Total fecundity estimates were performed by one person, and each squashed anther head slide was initially viewed under 5× magnification to identify regions characterized by 8 differentiable frequency classes of pollen. Levels of non-aborted pollen per flower were then counted under 20×

magnification, taking 3 to 4 photos and averaging the numbers of non-aborted pollen to a particular frequency class, using two to six flowers per plant and 35–726 pollen grains per flower.

Fluorescence microscopy. Anther heads were fixed in ethanol : acetic acid glacial (3 : 1) overnight at room temperature and stored in 70% EtOH. Fixed anther heads were immersed in a drop of 4',6-Diamidino-2-phenylindole (DAPI) 1mg/ml in Vectashield staining solution (Vector, Peterborough, U.K.), and then squashed under a cover slip to release the pollen. After one hour incubation at room temperature the released pollen grains were viewed with Zeiss Axioplat 2 microscope under 100× magnification and using the appropriate filter combination for DAPI staining.

Scanning electron microscopy. Anthers were collected at the time of anthesis and carefully stamped onto carbon coated aluminum sample blocks to release the mature pollen. After gold coating in an Edwards S150B sputter coater (Edwards High Vacuum Inc., Crowley, West Sussex, U.K.), pollen was examined in a Hitachi S4100 SEM (Hisco Europe, Ratingen, Germany) at 10 kV acceleration voltage. Digital recordings were made and saved as tif-files.

Fixation, substitution and embedding for light microscopy. For the primary fixation anthers sampled at the time of anthesis were kept for 6h at room temperature in 50 mM cacodylate buffer (pH 7.2) containing 0.5% (v/v) glutaraldehyde and 2.0% (v/v) formaldehyde, followed by one wash with buffer and two washes with distilled water. For the secondary fixation samples were transferred to a solution of 1% (w/v) OsO4. After 1h samples were washed three times with distilled water. Dehydration of the samples was done stepwise by increasing the concentration of ethanol in the following order: 30% (v/v), 50% (v/v), 60% (v/v), 75% (v/v), 90% (v/v) and 100% (v/v) ethanol for 1h each. After 1h dehydration with propylene oxide the samples were infiltrated with Spurr (Plano GmbH, Marburg, Germany) as follows: 33% (v/v) Spurr resin in propylene oxide for 10h, 50% (v/v) and 66% (v/v) for 4h each and then 100% (v/v) Spurr overnight. Samples were then transferred into embedding moulds where they were stored for 6h in fresh resin followed by polymerization at 70° C for 24h. For light microscopical examination using a Zeiss Axiovert 135 (Carl Zeiss AG, Jena, Germany) microscope, sections with a thickness of approximately 3 µm were mounted onto slides and stained for 2 minutes in a mixture of (1 : 1) 1% Azur II and 1% methylene blue in 1% aqueous borax at 60° C. Images were taken using a Zeiss AxioCam HRc CCD camera system.

RESULTS

Reproductive mode. The flow cytometric seed screen (FCSS) of 2C individuals confirmed that they were sexual, as their seeds were produced by the union of reduced

egg and central cells (containing two polar nuclei) with reduced male gametes (Table 2). No deviations from the 2C : 3C embryo:endosperm ratio were measured from either the immature or mature seeds.

Table 2. Reproductive mode classification summarized for different ploidy classes, based upon flow cytometric seed screen analysis of immature and mature seeds (Matzk et al., 2000).

	FCSS						
No. of plants	Embryo: Endosperm	Immature No.[a]	Mature No.[a]	Reproduction	Embryo development	Endosperm formation	Sperm nucleus ploidy[c]
Diploid (2C) accessions							
5	2C:3C	57	94	obligate sexual	fertilization		1C
Diploid (2C+B) accessions							
12	2C:3C	-	17	sexual	fertilization		1C
	2C:4C	4	3	apomictic	parthenogenesis	autonomous	-
	2C:5C	-	9	apomictic	parthenogenesis	pseudogamy	1C
	2C:6C	99	108	apomictic	parthenogenesis	pseudogamy	2C
	2C:7C	-	1	apomictic	parthenogenesis	pseudogamy	3C
	2C:8C	5	21	apomictic	parthenogenesis	pseudogamy	4C
	2C:10C	1	1	apomictic	parthenogenesis	pseudogamy	6C
	4C:6C	9	3	apomictic	fertilization	pseudogamy	2C + 2C
	4C:8C	2	3	apomictic	fertilization	pseudogamy	2C+ 4C
Triploid (3C) accessions							
5	3C:6C	49	4	apomictic	parthenogenesis	autonomous	-
	3C:9C	89	75	apomictic	parthenogenesis	pseudogamy	3C
	6C:6C	3	-	apomictic	fertilization	autonomous	3C
	6C:9C	15	2	apomictic	fertilization	pseudogamy	3C + 3C
	9C:6C	4	-	apomictic	fertilization	autonomous	6C
	9C:(9C)	-	1	apomictic	fertilization	?[b]	6C

[a] Number of analyzed single seeds
[b] Only one peak identified
[c] Inferred sperm ploidy

The 2C+B and 3C plants produced seed apomictically, as is evidenced by the frequent occurrence of apomeiosis and pseudogamous endosperm development (Table 2). A single 2C+B plant (209.4) was found to sexually produce seeds with a 2C : 3C embryo : endosperm ratio. While the mother of this individual was characterized by a B chromosome, it is unclear whether this particular offspring (209.4) also contained the B. Evidence for fertilization of the apomeiotically-generated egg cell was found mainly in immature seeds, with fewer incidences of this phenomenon in screens of the mature seeds. Autonomous endosperm development (no fertilization of the central cell by male gamete) was found in immature seeds more often than mature ones (Table 2).

Diploid apomicts exhibited a wide range of endosperm ploidy, which can only be explained by pollination with male gametes of differing ploidy. Based upon the FCSS

(Table 2), it is evident that diploid apomicts could be fertilized with at least five different kinds of pollen, both reduced (1C) and unreduced (2C, 3C, 4C and 6C). Surprisingly only unreduced (3C) pollen were responsible for endosperm formation within triploid individuals, although the generation of 9C embryos would required an unreduced hexaploid pollen sperm nucleus (see discussion).

Seed set and average number of ovules also differ between individuals. 2C, 3C and 2C+B individuals produce on average 93, 63 and 110 ovules per silique, respectively. Closer examination revealed that the number of ovules produced per silique can differ within a single individual (e.g., 300.9, 78.9, 195.3), although this appeared to be characteristic of apomictic plants only (Fig. 1). Certain plants were characterized by very poor seed set (means of viable seeds per silique for 2C, 2C+B and 3C plants respectively = 0.81, 0.47, and 0.47), especially in apomictic plants (e.g., 290.2, 120.6, 28.18, 28.8 and 218.2) in which over 80% of ovules or developing seeds were aborted (data not shown).

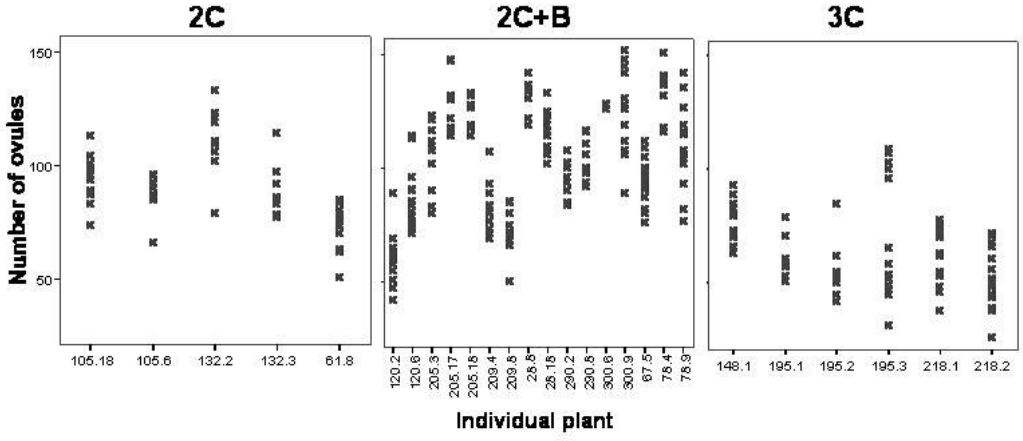


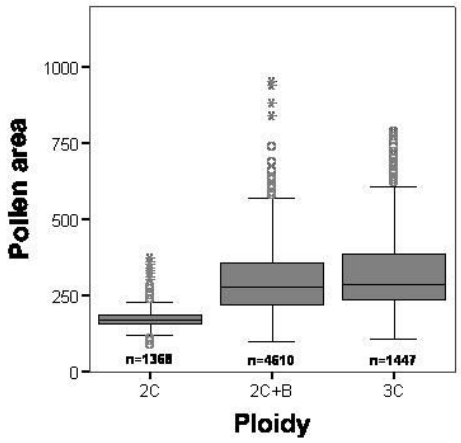
Figure 1. Scatter-plot of counted ovules per silique for individual diploid (2C), apomictic diploid (2C+B) and triploid (3C) plants (x = represent amount per silique). 2C, 2C+B and 3C individuals produced on average 93, 110 and 63 ovules per silique respectively.

Pollen size. The pollen area measurements from grouped 2C, 2C+B and 3C individuals demonstrated a wide range of values (Fig. 2A). Diploid (2C) sexual individuals (four plants and 1368 pollen area measurements) were characterized by the smallest pollen sizes and showed a tight box-plot with some outliers above and below the main distribution. The upper outliers and extreme values of the 2C pollen corresponded to the main distributions of pollen from 2C+B and 3C apomictic individuals. The pollen size distributions of both 2C+B (16 plants and 4610 pollen area measurements) and 3C individuals (5 plants and 1447 pollen area measurements) were similar to one another, both in terms of the main distribution and outlier and extreme values (Fig. 2A).

The pollen size distributions of individual 2C plants overlapped for the most part, and were characterized by small numbers of outliers (Fig. 2B). The 2C+B individuals

exhibited the greatest variability in pollen sizes which ranged from the lower to upper distribution bounds of the 2C and 3C individuals, respectively.

A)



B)

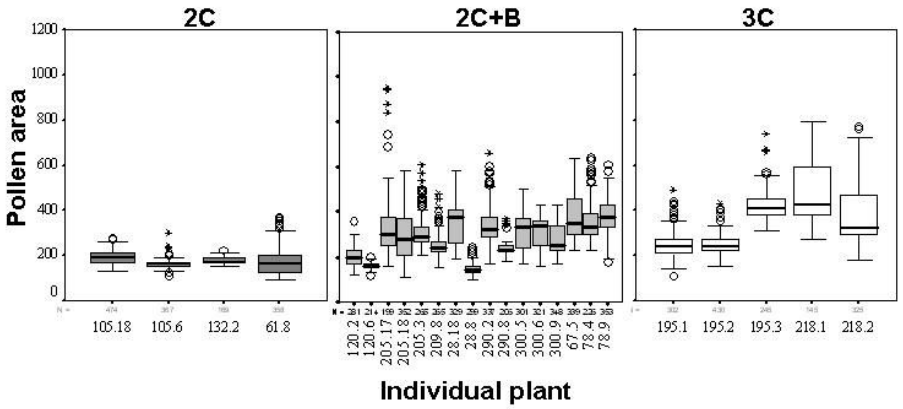


Figure 2. Box-plots of pollen area (μm^2) measurements for (A) clustered and (B) diploid (2C), apomictic diploid (2C+B) and triploid (3C) individuals. Bars within Box-plot represents median (n=number of pollen grains analyzed, circles and stars represent outliers and extreme values respectively).

Two 2C+B individuals (120.6 and 28.8) showed tight distributions equal or smaller to 2C individual pollen sizes. Two of the 3C individuals (195.1 and 195.2) were characterized by small and relatively tight pollen size distributions in the lower range of 2C+B individuals, and had upper outlier values corresponding to the main distributions of the three other 3C individuals (Fig. 2B). A third 3C individual originating from the same mother plant (195.3), was characterized by a larger pollen size distribution relative to its siblings (195.1 and 195.2; Fig. 2B).

Closer examination of single flowers in individuals having broad pollen size distributions (Fig. 2B) revealed that almost half (11 plants) were characterized by pollen size variation between flowers and/or the occurrence of different pollen grain sizes in one flower (Supplementary data-A). The pollen size distributions of all flowers of 2C

individuals were clearly unimodal, while B chromosome carriers were highly variable (Supplementary data-A), eight of them had a clearly bimodal pollen size distribution.

Fluorescent microscopy furthermore demonstrated that bimodal distribution not only reflects variation in size, but also variation in numbers of contained nuclei. Small pollen grains contain the expected three nuclei (1 vegetative + 2 sperm), but in comparison large pollen grains contain either the expected 3 or sometimes 6 (2 vegetative + 4 sperm) nuclei (Fig. 3). Only a single 3C individual demonstrated a bimodal size distribution, while other individuals were mainly unimodal with several outliers (Supplementary data - A).

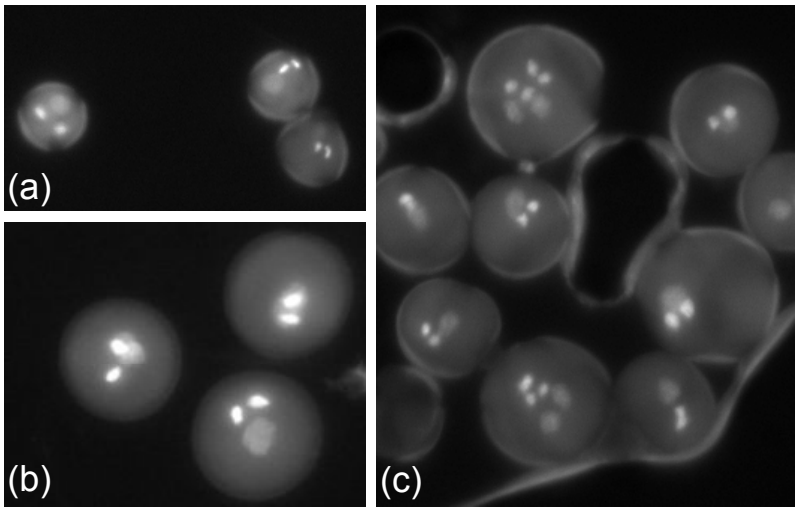


Figure 3. Fluorescence microscopy of pollen grains stained with DAPI under 20× magnification showing exemplary: A, one vegetative nucleus and two sperm nuclei (strongly stained) produced by diploids (2C); B, one vegetative nucleus and two sperm nuclei produced by triploids (3C); C, three types of pollen grains produced by apomictic diploids (2C+B). All pictures are at the same scale.

Considering only the mode values of the different size distributions, four general size clusters could be differentiated (Fig. 4). The first cluster (<200 µm²) included all diploid plants and three B chromosome carriers (120.2, 120.6, 28.8).

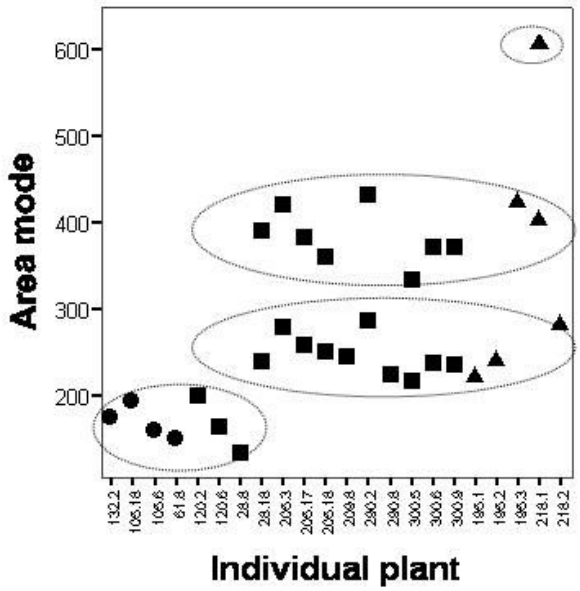


Figure 4. Plot of size distribution mode (µm^2) for 2C (circle), 2C+B (square) and 3C (triangle) individuals (shading reflects different size class groups).

A second cluster (230–290 µm²) included the remaining B chromosome carriers and three triploids (195.1, 195.2, 218.2), and a third cluster (340 µm² to 440 µm²) contained a number of B chromosome carriers and 2 triploids (195.3 and 218.1). Individual 218.1 (3C) was found in the largest forth "cluster" (~600 µm²; Fig. 4).

Pollen morphology. The appearance of variable and broad pollen size distributions within single flowers (data not shown) led us to consider pollen grain asymmetry and irregular surface morphology as possible sources of this variation. Scanning electron microscopy (SEM) analysis demonstrated that pollen grains from diploid plants displayed uniformity in shape and size, whereas pollen grains from B chromosome carriers and triploid individuals exhibited highly variable size, shape, and surface morphology (Fig. 5). The asymmetry, size variation and surface deformation could have lead to problems in pollen size estimations.

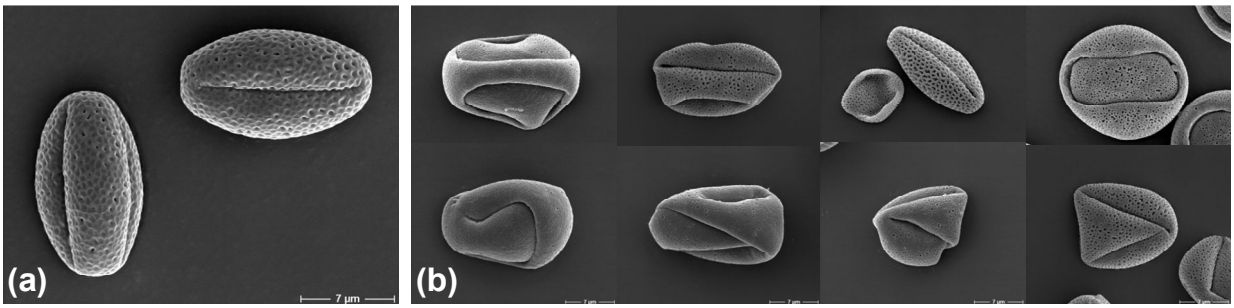


Figure 5. Scanning electron microscopy pictures showing pollen grains from (a) 2C sexual and (b) 2C+B and 3C apomictic plants.

Pollen fecundity and viability. The frequency of pollen per sampled anther head differed between ploidy classes (Supplementary - B). The majority of anther heads produced by diploid individuals contained between 250–300 pollen grains, although one plant (61.8, Supplementary data - B) produced significantly lower levels of pollen. B chromosome carriers demonstrated variable fecundity, with the majority of anther heads producing 200 to 300 pollen grains, while triploids typically produced 100 to 250 pollen grains. All diploid, four B chromosome carriers (205.18; 28.18; 300.9; 78.4) and one triploid (218.1) were characterized by stable pollen production (i.e., no difference between anther heads). In contrast, most B chromosome carriers and triploid individuals demonstrated differences in pollen production between anther heads (Supplementary - B).

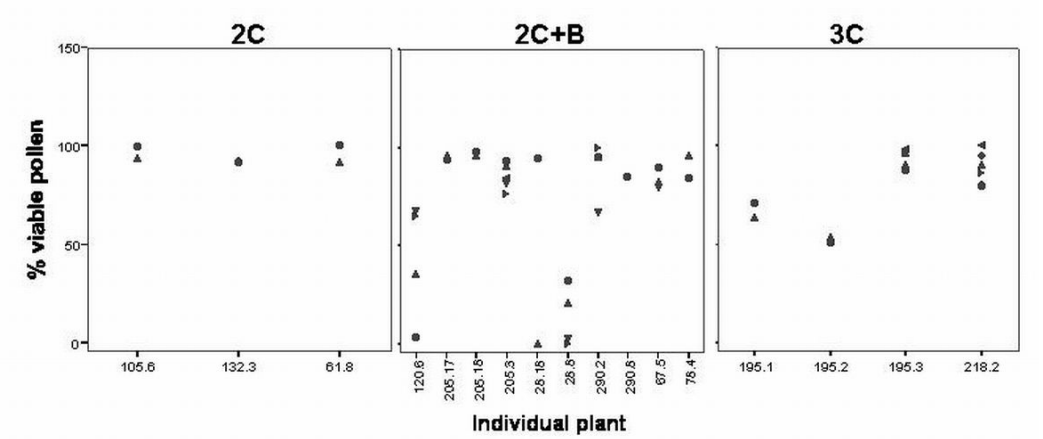


Figure 6. Non-aborted pollen frequency per flower 2C, 2C+B and 3C individuals, as measured using histochemical staining. Different shapes (circles and triangles) within a single individual represent different flowers.

Histochemical staining revealed that the production of viable pollen per flower in 2C sexual plants was stable at approximately 90% (Fig. 6). Most 2C+B and all 3C individuals displayed stable viable pollen production, although 4B chromosome carriers (120.6, 28.18, 28.8, and 290.2) showed distinct variation between flowers (Fig. 6). These data were supported by Semi-Thin Sections of anther heads, which confirmed that these plants have high levels of pollen abortion (Fig. 7).

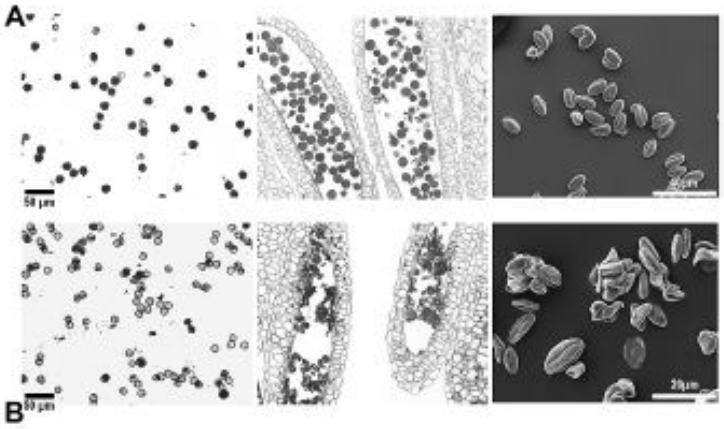


Figure 7. From left to right: pollen grains stained with Alexander's stain, Semi-thin sections of anther heads, and scanning electron microscopy picture. (A) 2C Individual 105.6 and (B) 2C+B Individual 120.6.

DISSCUSION

Pollen.

Our comparisons of pollen morphology, size and ploidy between 2C sexual and 2C+B and 3C apomictic *Boechera* accessions have demonstrated a great deal of variability between the asexual clones, and are concordant with pollen size data previously collected from apomictic *Boechera* (Böcher, 1951, 1954; Koch et al., 2003; Dobeš et al., 2004a; Sharbel et al., 2005). Pollen size differences have generally reflected relative genome size differences between accessions (i.e., poly- ploids have larger pollen than diploids), and hence pollen size has been interpreted as an accurate correlate of pollen ploidy (Dobeš et al., 2004a; Sharbel et al., 2005). The data presented here demonstrate that the determination of pollen ploidy using size criteria is more complicated than previously expected, at least with respect to apomictic plants.

It is clear that some of the pollen size variability measured in this study could be attributed to irregular pollen symmetry (Figs. 6 and 7) or differences in pollen development between samples (McCormick, 1993; Chen and McCormick, 1996). Böcher (1951) additionally indicated the position of the flower (bottom or top) as a correlate of pollen size differences. As anther dehiscence appears to occur earlier in apomictic accessions relative to those of sexuals, it was difficult to choose identical pollen developmental stages for comparative purposes. Nonetheless, the overall trends in pollen size variability appear to be correlated with reproductive mode.

Pollen size (area) distributions per individual were either uni- or bimodal, or so variable that no majority pollen size could be identified (Supplementary data - A). We have thus considered modal pollen size values only as being characteristic for a particular individual, and using this criterion we can identify three major classes of pollen size (Fig. 4). The diploid sexual plants produce the smallest modal pollen grain size, with some variation between their modal values (Fig. 4). If one assumes that the variation between 2C sexual modal values is reflective of measurement error due to developmental stage, then the differences between the three major modal size clusters appear to be significant (Fig. 4).

Interestingly, 2C+B individuals produce pollen grains distributed through all three size clusters, while 3Cs produce pollen grains distributed among two of the clusters. The smallest cluster (130–90 μm^2) includes 2C sexual accessions, as identified by flow cytometric seed screen analysis (Table 2), and thus this cluster is reflective of reduced (1C) pollen grains. Considering this, it appears that three B chromosome carriers (120.2, 120.6, and 28.8) produce reduced (1C) pollen grains, although the FCSS revealed for all three a clear 2C : 6C embryo : endosperm ratio, which required fertilization from 2C sperm cells.

Moreover, it has been shown for 28.8 and 120.6 (Fig. 6) that these plants produce a high number of aborted pollen grains, and thus we suspect that the mode value here represents aborted/collapsed pollen grains (Fig. 6) which fall within the same size class as reduced 1C pollen grains (Schranz et al., 2005).

Both 2C+B and 3C accessions produced pollen sizes which fell within the second (230–290 µm²) and third (340–440 µm²) modal clusters (Fig. 4). Assuming that pollen size and ploidy are tightly correlated, these data imply that 2C+B and 3C individuals produce pollen of the same ploidy, which is clearly incorrect when one considers the flow cytometric seed analyses of embryo and endosperm genome size (Table 2).

The flow cytometric seed screen revealed that the majority 2C+B individuals produce seeds containing 6C or 8C endosperm (Table 2), thus demonstrating that they have been generated through fertilization by either 2C or 4C pollen sperm cells. In contrast, 3C plants form seeds with 9C endosperm, and have thus arisen through fertilization by 3C pollen sperm cells. Considering the 2C+B data only, it is thus probable that the second and third modal pollen size clusters (Fig. 4) represent 2C and 4C pollen, respectively. In the same light, one would infer that the second and third modal pollen size clusters represent 3C and 6C pollen in 3C apomictic individuals (Fig. 4) (Böcher, 1951, Kravtchenko et al., 2001). These latter data are puzzling since seed screening demonstrated 3Cs to be characterized by 9C and never 12C endosperm (Table 2), the latter of which would be expected in the case of 6C pollen. In only a single case was a 9C embryo identified (Table 2), which either requires a 6C pollen grain or was the result of fertilization with two 3C sperm cells (i.e., polyspermy; Spielman et al., 2003).

While these data are inconclusive with regards to variability in pollen ploidy in the apomictic accessions, they do clearly demonstrate that pollen size and ploidy are not correlated. In addition, fluorescence microscopy (Fig. 3c) illustrate that larger pollen grains (Fig. 4, cluster III) can contain different numbers of nuclei (3 nucleate = 1 vegetative + 2 sperm nuclei, or 6 nucleate = 2 vegetative + 4 sperm nuclei; Böcher, 1951). Furthermore, it appears that the nuclear fluorescence intensity of large 6 nucleate pollen grains is equal to that of small 3-nucleate pollen, with large 3-nucleate pollen having the brightest intensity (Spielman et al., 1997; González-Melendi et al., 2005). This suggests higher ploidy of nuclei within large 3 nucleate pollen grains and could explain results from FCSS, but it is unclear whether these grains differ in maturity or are characterized by mitotic disturbance.

Variability in embryo and endosperm ploidy in the apomictic accessions demonstrates that fertilization can occur with pollen of differing ploidy (Table 2). There is even evidence for double fertilization or fertilization with multinucleate pollen cells (Fig.

3), as 4C : 8C embryo : endosperm ratios have been found in low levels in B chromosome carriers (Table 2). To express a well defined explanation is rather difficult, it could have resulted from polyspermy by two sperm nuclei of different ploidy (Spielman et al., 2003) or the occurrence of three polar nuclei is the cause (F. Matzk, pers. comm.). It is unclear what mechanism could lead to polyploidy pollen formation (e.g., endoreduplication versus disturbance in cytokinesis), although the pollen morphology in some of our accessions (Fig. 5) is similar to an *Arabidopsis* mutant characterized by disturbed cytokinesis (Hulskamp et al., 1997; Spielman et al., 1997).

Embryo and endosperm variability.

The flow cytometric analyses of both immature and ripe seeds have shown that at least some of the variability we have characterized at the level of pollen can lead to successful fertilization. The comparison of immature green and ripe brown seeds demonstrates that natural selection may be acting upon the seeds during their development. Diploid *B. stricta* and *B. holboellii* produce exclusively 2C : 3C embryo to endosperm ploidies (Table 2), and considering that these individuals also produce reduced pollen cells (Fig. 4), it is clear that they produce seed sexually.

B chromosome carriers are more flexible with respect to embryo and endosperm formation. Seeds are characterized by fertilized sexual, and both fertilized and unfertilized apomeiotic embryos (Table 2). The flow cytometric seed screen revealed that certain plants only produce hexaploid endosperm, whereas others produce hexa- and octaploid endosperm (Table 2). One plant (120.6) furthermore demonstrated extremely variable endosperm production (e.g., 3C, 4C, 5C, 6C, 7C). Variable endosperm ploidies can arise through both autonomous and pseudogamous development in addition to both self- and cross-fertilization (data taken from isolated flowers are not included here) with pollen of differing ploidy. We could identify no significant difference between levels of autonomous and pseudogamous development between immature and mature seeds (χ^2= 0.4, P = 0.516).

As found by Naumova et al. (2001), fertilization of the apomeiotically-derived egg cell is relatively rare, and based upon the data presented there is no statistically significant difference in the numbers of fertilized egg cells between measurements made on immature and mature seeds (χ^2= 3.8, P = 0.051). The low levels of detected autonomous endosperm development (Naumova et al., 2001) imply that an imbalanced maternal to paternal genome ratio do not completely limit embryo development, at least in certain apomictic clonal lineages.

It has been suggested that developmental abnormalities are associated with haploid gametes, or that natural selection acts against haploid gamete formation in 2C

apomictic (with or without B chromosomes) *Boechera* (Naumova et al., 2001). The data presented here give no clear proof of the ability of 2C apomicts to produce haploid gametes, only rare cases in which haploid sperm gametes merged with an apomeiotically derived central cell, but never with apomeiotic egg cell. These data could be interpreted in a number of ways: (1) 2C apomicts could produce haploid gametes upon which there is no negative selection acting per se, but rather associated effects (e.g., mutation accumulation, developmental timing, etc.) may influence the ability of haploid gametes to fertilize and generate viable seed; (2) fertilization could have occurred with sexual individuals growing in the same greenhouse, although this is unlikely since it appears that fertilization occurs within the closed flowers; (3) haploid gametes are inviable due to mutation accumulation and/or the effects of hybridity; and (4) selection for fertilization which maintains the 2m : 1p endosperm ratio. The latter explanation seems most likely; the production of unreduced gametes within the 2C apomicts ensures balanced endosperm formation (Scott et al., 1998; Adams et al., 2000).

Triploid *Boechera* demonstrate the same range of reproductive variation as do B chromosome carriers, although a number of differences between the two cytotypes do exist. There is a significantly higher incidence of fertilized apomeiotic egg cells in immature versus mature seeds (χ^2= 5.96, P = 0.014*), which implies that 6C and 9C embryos suffer deleterious consequences of elevated ploidy and abort after fertilization (Table 2). In addition, autonomous endosperm formation is significantly more apparent in immature seeds (χ^2= 23.82, P < 0.001***), thus pointing to insufficient endosperm nourishment for the developing embryos which contributes to their abortion (Table 2).

Finally, there is a significantly higher incidence of autonomous endosperm formation in the immature seeds of 3C relative to 2C+B individuals (χ^2= 37.74, P < 0.001***), although this difference disappears in comparisons of mature seeds. On the one hand the differences in immature seeds could be genotype specific (similar effects due to common ancestry), or alternatively the differences could have arisen through secondary effects associated with ploidy. We are unable to differentiate between these scenarios based upon the data presented here.

Fertilization disturbances and gene flow.

Experimental crosses have demonstrated that both diploid and triploid apomictic *Boechera* are facultative, and hence two extremely rare conditions (diploid apomixis and partial triploid sexuality) among the flowering plants can be found in this complex (Schranz et al., 2005). Both types of facultative apomict can furthermore produce sexual seed either as a maternal or paternal parent (Schranz et al., 2005). It is thus clear that both pollen and egg cells produced by apomictic individuals are, at least occasionally,

compatible with sexual fertilization, and this is corroborated by the seed screen data collected here (Table 2).

Apomictic accessions are nonetheless characterized by decreased levels of fertility (Koch et al., 2003; Sharbel et al., 2005). Decreased fitness in asexual taxa is expected to result from mutation accumulation (Kondrashov, 1994) and chromosome imbalance associated with hybridity and/or polyploidy. The different apomictic accessions studied here demonstrate characteristics which are typical for asexual organisms in general (e.g., embryo abortion, pollen production problems, etc.), although significant differences between the accessions nonetheless exist. Distinct differences with respect to the three apomixis components (apomeiosis, parthenogenesis and pseudogamy) characterize 2C+B and 3C individuals, and disturbances to all three can be classified according to their relative time of action.

Pre-fertilization disturbances.

Among the apomictic accessions, 2C+B individuals were characterized by higher numbers of ovules per silique and pollen grains per anther head relative to 3Cs. In contrast, it is evident that 2C+B individuals produce a greater range of pollen size and a greater rage of viable pollen grain within and between accessions. Perturbed pollen production in 3C individuals likely results from the combined effects of irregular chromosome segregation in addition to mutational load, while in 2C+B individuals pollen production is expected to be affected by mutational load alone (assuming the diploid genome is less affected by irregular chromosome segregation). While the 3C individuals analyzed here do show greater pollen size variation relative to 2C sexuals, it is unclear how much of this variation can be attributed to differing chromosome numbers between pollen grains produced by a single individual, since we have also shown that pollen size and ploidy are not strictly correlated (see above). Evidence for bimodal size classes in certain 3C accessions (Böcher, 1951) may demonstrate whole genome segregation (i.e., C and 2C gametes) as has been documented in different animal taxa (Uzzell et al., 1980), although both experimental crosses the data presented here give no evidence for the production of such gametes.

Assuming that 2C+B apomicts do not suffer from the deleterious effects of irregular chromosome segregation, it is interesting to note that we can distinguish two classes of 2C+B apomicts. Certain plants (Fig. 6; e.g., 120.6; 28.18; 28.8) have a lower viability of their gametes relative to the other 2C+B and 3Cs, which can be explained in two ways. Firstly, these 2C+B accessions could be suffering from a higher mutational load than the others, and assuming a constant genomic mutation rate, this would imply that these clonal lineages (120.6; 28.18; 28.8) are evolutionarily older. The age-related mutational load hypothesis not only applies for whole genome comparisons, but could

also result from a relatively old genomic region which has accumulated mutation and undergone segregation into various genetic backgrounds through occasional sex (i.e., the B chromosome). An alternative hypothesis is that the 2C+Bs may have a similar mutational load compared to the 3Cs, but the 2C+B lineages suffer the effects of mutations more readily since there are fewer copies of each gene in comparison to a 3C (i.e., a mutated allele in a 3C still leaves 2 unmutated allele copies, while only a single unmutated copy is left in a 2C). This latter explanation would support hypotheses which have been proposed to explain the evolutionary advantage (and ubiquity) of polyploid asexuality (Otto and Whitton, 2000).

Post-fertilization disturbances.

The variable pollen types produced by apomictic *Boechera* are characterized by differing viability, and natural selection represents a microsporogenesis bottleneck which restricts the types of pollen which may participate in successful fertilization. Our flow cytometric analyses of seeds at different developmental stages additionally show that a subset of this surviving pollen variation can lead to fertilization, but that natural selection subsequently acts during seed development to further reduce the reproductive variation initially seen in the immature seeds.

It appears that naturally selection acts differently on the developing seeds of 2C+B versus 3C apomicts. A striking difference is the higher level of autonomous endosperm formation in the immature seeds of triploid compared to B carriers. It is unclear why autonomous endosperm formation is more prevalent initially in 3Cs, although abortion of these seeds leads to comparable surviving levels of autonomous endosperm seeds in both apomictic cytotypes. It is likely that deviation from the 2 : 1 maternal to paternal genome ratio (Scott et al., 1998; Adams et al., 2000) contributes to the abortion of seeds with autonomous endosperm formation during their development. Fertilization of the apomeiotically-derived egg cell was also more prevalent in immature relative to ripe seeds, which points to developmental disturbances.

While these data do not give any hint as to what mechanism could lead to the initially high levels of fertilization and autonomous endosperm formation, they do show that the apomictic pathway is flexible, and that natural selection acts at different stages of the pathway to successively remove variants from the ripening seed set. It is evident that natural selection continues to act upon the germinating seeds, since many of the ploidy variants measured in the ripe seeds have never been found in wild populations (Sharbel and Mitchell-Olds, 2001; Dobeš et al., 2004a; Sharbel et al., 2005).

Conclusions.

The *Boechera holboellii* complex is characterized by multiple evolutionary origins of polyploidy, and a B chromosome which is found in genetically and geographically diverse populations (Sharbel and Mitchell-Olds, 2001; Sharbel et al., 2004, 2005). It is furthermore hypothesized that apomixis has been repeatedly expressed from a sexual background (Sharbel and Mitchell-Olds, 2001; Sharbel et al., 2004, 2005). One explanation for this hypothesis is that apomictic individuals can occasionally backcross with sexual individuals, and that this could help spread "apomixis factors" throughout a population. Experimental crosses using *Boechera* of differing ploidy and reproductive mode have demonstrated that, while backcrossing is theoretically possible, it typically leads to ploidy jumps in offspring which are not reflected in natural populations (Schranz et al., 2005).

We have identified no evidence for the ability of apomictic accessions to fertilize (i.e., backcross) sexuals with reduced pollen cells, since the apomicts studied here produced unreduced pollen. Pollen fertilization is nonetheless a requirement for both sexual and apomictic forms, and consequently the "two-fold cost" of sex could be applied for apomicts as well in this species complex.

Finally, our data demonstrate the plasticity of gametogenesis and fertilization in apomictic *Boechera*, the extent of which is also correlated with differences in ploidy and B chromosome presence. Natural selection acts at various stages in the apomictic pathway to reduce the potential variability generated by an apomictic plant, and it appears that both mutation accumulation and epigenetic phenomena (e.g., maternal – paternal genome imbalance and endosperm formation) lead to gamete inviability and seed abortion. The fertilization of apomeiotically-derived egg cells and autonomous endosperm formation are nonetheless within the reproductive potential of apomictic individuals, and this information is being used to focus in on the apomixis phenotype in order to decipher its mechanism in the genus *Boechera*.

ACKNOWLEDGEMENTS

We would like to thank the technical assistance of H. Block, V. Katzorke and L. Börner. Thank you also to Dr. J. Kumlehn and Dr. R. Scott for enlightening discussions regarding data interpretation and future avenues of research. Special thanks to M. Zielinski for his unconditional support and patience.

LITERATURE CITED

Adams S, Vinkenoog R, Spielman M, Dickinson HG and Scott RJ. 2000. Parent of- origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. Development 127: 2493–2502.

Alexander MP. 1969. Differential staining of aborted and nonaborted pollen. Stain Technology 44: 117–122.

Birdsell J and Wills C. 1996. Significant competitive advantage conferred by meiosis and syngamy in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA. 93: 908–912.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Kongel. Danske Videns. Selsk. Biol. Skr. 6: 1–59.

Böcher TW. 1954. Experimental taxonomical studies in the *Arabis holboellii* complex. Svensk. Bot. Tidskr. 48: 31–44.

Butlin R. 2002. Opinion–evolution of sex: The costs and benefits of sex: new insights from old asexual lineages. Nat. Rev. Genet. 3: 311–317.

Chen YC and McCormick S. 1996. Sidecar pollen, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development. Development 122: 3243–3253.

D'Souza TG, Storhas M, Schulenburg H, Beukeboom L W and Michiels NK. 2004. Occasional sex in an 'asexual' polyploid hermaphrodite. Proc. Biol. Sci. 271: 1001–1007.

Dobeš C, Mitchell-Olds T and Koch MA. 2004a. Intraspecific diversification in North *American Boechera stricta* (= *Arabis drummondii*), *Boechera × divaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. Amer. J. Bot. 91: 2087–2101.

Dobeš C, Mitchell-Olds T and Koch M. 2004b. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. × divericarpa*, and *A. holboellii* (Brassicaceae). Molec. Ecol. 13: 349–370.

Dorn RD. 1984. Vascular Plants of Montana. Mountain West Publishing, Cheyenne. González-Melendi P, Ramiret C, Testillano PS, Kumlehn J, Risueño MC. 2005. Three dimensional confocal and electron microscopy imaging define the dynamics and mechanisms of diploidisation at early stages of barley microspore-derived embryogenesis. Planta 222: 47–57

Hulskamp M, Parekh NS, Grini P, Schneitz K, Zimmermann I, Lolle SJ and Pruitt RE. 1997. The STUD gene is required for male-specific cytokinesis after telophase II of meiosis in *Arabidopsis thaliana*. Developm. Biol. 187: 114–124.

Johnson-Brousseau SA and McCormick S. 2004. A compendium of methods useful for characterizing Arabidopsis pollen mutants and gametophytically-expressed genes. Plant J. 39: 761–775.

Joshi A and Moody ME. 1995. Male gamete output of asexuals and the dynamics of populations polymorphic for reproductive mode. J. Theor. Biol. 174: 189–197.

Koch MA, Dobeš C and Mitchell-Olds T. 2003. Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molec. Biol. Evol. 20: 338–350.

Komaru A, Kawagishi T, and Konishi K. 1998. Cytological evidence of spontaneous androgenesis in the freshwater clam *Corbicula leana* Prime. Developm. Genes Evol. 208: 46–50.

Kondrashov AS. 1994. Muller's ratchet under epistatic selection. Genetics 136: 1469–1473.

Kravtchenko AY, Naumova TN and van der Laak J. 2001. Formation of unreduced pollen in the natural apomictic species *Arabis holboellii*. Presented at: 2nd International Apomixis Conference APO2001, April 24–28 2001, Como, Italy.

Mantovani B and Scali V. 1992. Hybridogenesis and androgenesis in the stick-insect *Bacillus rossius-grandii benazzii*. Evolution 46: 783–796.

Matzk F, Meister A and Schubert I. 2000. An efficient screen for reproductive pathways using mature seeds of monocots and dicots. Plant J. 21: 97–108.

McCormick S. 1993. Male gametophyte development. Plant Cell 5: 1265–1275.

McKone MJ and Halpern SL. 2003. The evolution of androgenesis. Amer. Naturalist 161: 641–656.

Naumova TN, van der Laak J, Osadtchiy J, Matzk F, Kravtchenko A, Bergervoet J, Ramulu KS and Boutilier K. 2001. Reproductive development in apomictic populations of *Arabis holboellii* (Brassicaceae). Sexual Pl. Reprod. 14: 195–200.

Otto SP and Whitton J. 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34: 401–437.

Pichot C, El Maataoui M, Raddi S and Raddi P. 2001. Surrogate mother for endangered Cupressus. Nature 412: 39.

Rollins RC. 1981. Studies of *Arabis* (Cruciferae) of western North America. Syst. Bot. 6: 55–64.

Roy BA. 1995. The breeding systems of six species of *Arabis* (Brassicaceae). Amer. J. Bot. 82: 869–877.

Schranz ME, Dobeš C, Koch MA and Mitchell-Olds T. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). Amer. J. Bot. 92: 1797–1810.

Scott RJ, Spielman M, Bailey J and Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125: 3329–3341.

Sharbel TF and Mitchell-Olds T. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity 87: 59–68.

Sharbel TF, Mitchell-Olds T, Dobeš C, Kantama L and de Jong H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogen. Genome Res. 109: 283–292.

Sharbel TF, Voigt ML, Mitchell-Olds T, Kantama L and de Jong H. 2004. Is the aneuploid chromosome in an apomictic *Boechera holboellii* a genuine B chromosome? Cytogen. Genome Res. 106: 173–183.

Spielman M, Preuss D, Li FL, Browne WE, Scott RJ and Dickinson HG. 1997. TETRASPORE is required for male meiotic cytokinesis in *Arabidopsis thaliana*. Development 124: 2645–2657.

Spielman M, Vinkenoog R and Scott RJ. 2003. Genetic mechanisms of apomixes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358: 1095–1103.

Uzzell T, Hotz H and Berger l. 1980. Genome exclusion in gametogenesis by an I nterspecific Rana hybrid: evidence from electrophoresis of individual oocytes. J. Exp. Zool. 214: 251–259.

Supplementary data

Supplementary data – Table A. Overview of pollen size variation measured from 25 *Boechera* accessions. Individual numbers refer to accession and individual (before and after dot respectively).

Supplementary data – Table B. Frequency of individual pollen fecundity classes per anther head (the most frequent class per individual is in bold to facilitate overview).

Diploid *versus* Triploid apomicts

Marie-Luise Voigt, Marcin Piwczynski, Timothy F. Sharbel

Abstract

Despite the wide range of developmental pathways leading to genetically identical offspring, naturally occurring apomicts share some common features (Asker & Jerling, 1992). For example, with very few exceptions apomicts are mainly polyploid and highly polymorphic on the morphological level, much of which can be attributed to hybridisation. The association between apomixis and polyploidy has been the subject of debate in understanding the origin and evolution of asexuality. This study approaches one aspect of this debate, the question of whether polyploidy buffers the effects of mutation in the asexual genome. The *Boechera holboellii* complex contains one of the rare cases of diploid apomixis, and thus provides an ideal comparative model whereby both diploid and polyploid apomicts can be compared.

In the following study we examined various fitness traits over two generations of apomictic plants and their genetically clonal offspring, an analysis which encompassed a wide range of life stages. We first examined fitness traits correlated with reproduction, including flower and ovule number, fruit and seed set, and seed size. Naturally occurring seed size variation was classified into small, medium and large size groups, and we studied both the source of this variation and correlations with fitness traits including germination and growth rate in the second generation. Surprisingly, most fitness traits exhibit lineage-specificity rather than correlations with ploidy, which could reflect the consequences of natural hybridisation, where natural selection acts upon novel variation in several traits to allow their establishment in specific niches.

KEYWORDS: diploid, polyploid, apomixis, fitness, life history traits, mutational load, seed size variation, clone, *Boechera*

INTRODUCTION

Apomixis is asexual reproduction in plants whereby seeds are produced clonally (Asker and Jerling, 1992). Many similarities between asexual (apomictic) plants and (parthenogenetic) animals (Beukeboom and Vrijenhoek, 1998) exist, although double fertilization (i.e. a pair of sperm nuclei fertilize both the egg and central cell to generate an embryo and functional endosperm) is a secondary adaptive trait which is not found in animals and gymnosperms. Apomixis in flowering plants (angiosperms) is commonly divided into three components: production of an unreduced egg cell (*apomeiosis*); *parthenogenetic* development of the unfertilized egg cell into an embryo; and fertilization of the central cell for proper endosperm formation (i.e. *pseudogamy*; Asker and Jerling, 1992). Apomixis has been described in over 500 species (Savidan, 2007), although it is only expressed in a limited number of families (~ 40 different families). This could imply that certain families may have a genetic predisposition for apomixis expression due to a common ancestry. However, the variety of apomictic pathways observed among the different families makes common ancestry questionable.

It is generally accepted that apomixis is derived from an aberration of the sexual pathway (Koltunow and Grossniklaus, 2003), which could have been triggered in several ways. One idea of apomixis origin is that it is triggered by hybridization between genetically distinct sexual species (Schultz, 1969), resulting in asynchronous gene expression between the "allo-genomes", with concomitant changes in cellular architecture and regulatory pathways (Carman, 1997). An alternative idea is that apomixis somehow arises via polyploidization (autopolyploidy) in a diploid sexual mother, with an apomictic offspring arising by spontaneous genome doubling or non-reduction during gametophyte development to result in increased genome content (Grimanelli et al. 2001, Quarin et al., 2001). Autopolyploidy can have major effects on reproduction (formation of multivalents during meiosis I, where the dissociation is more difficult; Santos et al., 2003), metabolism efficiency, regulatory changes in gene expression and epigenetic instability (reviewed in Comai, 2005). Which exact factors or pathways have to be changed in a sexual individual to induce apomixis remains elusive, but upon its origin, apomixis allows an organism to fix and stabilize the consequences of hybridization and autopolyploidy.

One trait that most apomictic species share is polyploidy (albeit polyploidy does not always mean asexuality; Asker and Jerling, 1992). Mogie (et al., 2007) referred to the correlation of asexuality and polyploidy as *joint causal determination*, whereby one condition supports the other in establishment and maintenance. In other words polyploidy is a highly adaptive trait for apomicts, but is not a factor leading to its origin. In tackling the question why apomixis and polyploidy are highly correlated, the question of

precedence arises: i) did polyploidization (autopolyploid or hybridization *i.e*. between tetraploid and diploid plants; Grimanelli et al., 2001; Quarin et al., 2001) and its consequences induce apomixis, or ii) did apomixis (as a consequence of alloploidy, i.e. asynchrony of gene expression; Carman, 1997) lead to formation of non-reduced gametes which led to polyploidy. Regardless of which was first, the advantageous coexistence of polyploidy and apomixis has been hypothesized to occur on various levels.

What mechanisms lead to the formation of polyploidy in plants? It is estimated that between 30 and 80 % of all plants are polyploid in nature (Mahony and Robinson, 1980), and it has been shown that polyploidy has multiple and dynamic origins within species complexes (Soltis and Soltis, 1993; 1999). Polyploids are subdivided into two classes depending on their chromosomal composition, autopolyploid (nonhybrid, chromosome from the same species) and allopolyploid (hybrid, chromosomes from different species). Autopolyploidy is a consequence of disturbed mitosis (somatic doubling) or meiosis (to produce non-reduced gametes), the latter of which can have an affect on both male and/or female gamete formation which results in triploid or tetraploid offspring (Comai, 2005). Allopolyploidyzation results from hybridization between genetically different species, and difficulties in passing mitosis and meiosis could result in spontaneous production of unreduced gametes (Carman, 1997; Ramsey and Schemske, 1998; 2002; Otto and Whitton, 2000; Comai, 2005).

The genetic consequences of polyploidy include an overall increase in the frequency of deleterious mutations (Bicknell et al., 2000). Elevated heterozygosity in alloploids could additionally compensate for deleterious mutations, and enable polyploids (relative to diploids) to adapt to a broader array of ecological conditions through increased metabolic flexibility (Otto and Whitton, 2000). Apomictic organisms accumulate mutation (*i.e.* leading to a *mutational load*) over time due to loss of recombination (Maynard-Smith, 1978), and upon the origin of an apomictic lineage it can have (1) the same mutational load as its mother plant (if it was a product of non-reduction); or, (2) divergent mutational load if it was a product of hybridization. In either case, the independent accumulation of deleterious and beneficial mutations, as a consequence of the loss of recombination (Maynard Smith, 1978), will slowly lead to genetic divergence from its sexual ancestor (Halkett et al., 2005), as well as from its asexual origin (Paun et al., 2006). It is predicted that the fate of asexual lineages is early extinction (Maynard Smith, 1978), but it has been shown that certain ancient asexual organisms do not follow this paradigm (Schön and Martens, 1998; Belshaw and Quicke, 2003; D'Souza et al., 2004; Khakhlova and Bock, 2006).

Does the higher mutational load of apomicts (Haag and Roze, 2007) explain the latter's association with polyploidy, leading to the hypothesis that apomictic polyploid

plants can cope with higher mutational loads than hypothetical diploid apomicts? Very few studies have examined these issues, which are complicated by the difficulty in disentangling the effects of polyploidy, hybridization and reproductive mode. Studies focusing on diploid sexual parents and their polyploid sexual offspring lead to different conclusions depending on whether natural or artificial hybrids were used (Levin, 1983; Kirk, 2005). Artificial hybrids, as agriculture products, are known to exhibit heterosis, although the differing rates of hybrid-breakdown for complex phenotypic traits in F2 and subsequent generations lead to complications in data interpretation. Conversely, studying natural processes affecting population dynamics in the early generations of hybrids can be missed (Kirk, 2005).

To complicate the situation even more, most apomictic species are polyploid, whereas their diploid ancestors are sexual, and thus tackling questions of why sexual reproduction is predominant in plants and animals, despite the advantages of asexual reproduction, is very difficult (Barton and Charlesworth, 1998). It is assumed that sexual reproduction has a two-fold cost which can be interpreted in two ways. The two-fold cost of providing for and supporting two genders (Maynard Smith, 1978) is shared with most apomicts species, because male gametes are necessary to trigger up-stream signal pathways or fusion with the central cell to form proper endosperm (Nogler, 1984a; Grossniklaus et al., 2001a). Secondly, the cost of meiosis in sexual species with equal ratio of male and female offspring is not applicable in apomicts, because all offspring are females and therefore it is hypothesized that they produce twice as many females at same time as a sexual (Williams, 1975). Much theoretical research has been done to elucidate the relative benefits of sex in contrast to its costs (Kondrashov, 1988; West et al., 1999), but few studies have actually measured the costs directly.

Van Dijk (2007) compared sexual and apomictic plants of hermaphrodite *Taraxacum officinale* to measure cost of sex. Apomictic *Taraxacum* plants are autonomous of males, and even though they produce pollen grains, no fertilization of the egg or central cell is essential (Tas, 1999). In sex-allocation theory a trade-off between male and female function is predicted (Charnov, 1982), which in apomicts could predict that degenerate pollen accumulation through the relaxation of selection pressure on traits associated with pollination, pollen tube growth and fertilization, may occur (Maynard Smith, 1978). Van Dijk and co-workers looked at several fitness traits such as flower number, seed set and seed weight between triploid apomicts and diploid sexuals to gain insight into sex and resource allocation. Apomictic pollen production was disturbed, but had no influence on female sex allocation or seed size. The only critical limitation in sexual plants was determined to be self-incompatibility, as pollination is not always assured. In summary it was concluded that the cost of sex in realty is very complex

(Lewis, 1987; van Dijk, 2007), and that species specific traits are influenced by breeding system, self-compatibility, outcrossing, pollination, frequent back crossing (pollen from apomicts fertilizing sexual ovules), competition between males and the establishment of neo-apomictic lineages (Verduijn, 2004).

Coming back to the intriguing question of why most apomicts are polyploid, a conundrum which has been investigated theoretically (Williams, 1975; Maynard Smith, 1978; Bell, 1982; Kondrashov, 1993; Otto and Gerstein, 2006) and thoroughly reviewed (Otto and Whitton, 2000; Comai, 2005), but has been virtually impossible to test in natural asexual systems (see example in Janko, 2007). It would be interesting to measure true effects of ploidy on fitness traits to test whether polyploidy buffers higher mutational load (requirement for stable apomixis expression for long-term), or whether it required for apomixis expression (although not all polyploids are apomictic; Asker and Jerling, 1992).

A good model system to test these ideas is the genus *Boechera*. It contains one of the rare cases of both diploid and polyploid apomicts (Böcher 1951; Roy 1995; Naumova et al., 2001). The sexual diploid plants have 2n = 2x = 14 chromosomes, whereas diploid apomicts can have as well 2n or 2n + x (aneuploid) chromosomes (Laksama et al., 2007), and polyploid apomicts are mainly triploid (2n = 3x = 21). Sexual plants exhibit high levels of homozygosity due to self-fertilization (Song et al., 2006) while apomictic plants exhibit high heterozygosity, which has its origins in hybridization (Koch, 2003; Dobes et al., 2004a/b; Sharbel et al., 2001; Schranz et al., 2005; Kantama et al., 2007) and mutation accumulation over time (Corral et al., 2009). It is hypothesized that the diploid sexual and monophyletic *Boechera stricta*, both (Dobes et al., 2004b; Schranz et al., 2005) hybridized with another species of the genus *Boechera* to form the apomict *B. divaricarpa* (Al-shebaz, 2007b). *Boechera divaricarpa* is polymorphic and can be diploid (2n), diploid aneuploid (2n+x) and polyploid.

We have performed a comprehensive study of life-history fitness traits (flower number, fruit set and length, ovule number, seed set, and seed size) in three independent lineages of both diploid and triploid apomictic *Boechera,* in order to investigate how their variation is influence by ploidy. We furthermore collected the data with respect to inflorescence position, as Böcher (1951) described a positional effect in which apomictic seeds occurred in the first developed flowers and sexual seeds in latest, a phenomenon which has also been shown in other species in which flowers are grouped into inflorescences (Wolfe and Denton 2001; Medrano et al., 2000).

Moreover, we have analysed seed size variation in order to answer the following questions: i) Is there a consistent seed size difference between diploid and triploid

apomicts, as expected due to higher genome content in triploids? ii) Is variation in the embryo: endosperm ploidy ratio a source of seed size variation? It has been proposed that genomic imprinting and parental conflict (Haig and Westoboy, 1991) in endosperm influence seed size. Another influence could be the maternal to paternal genome dosage effects on endosperm (Haig and Westoby, 1991; Scott et al., 1998; Dilkes and Comai, 2004). Two maternal to one paternal genome contribution is necessary for normal endosperm formation, and varying this ratio leads to disturbed endosperm formation and often seed abortion. It was shown that paternal excess in endosperm leads to promoted growth and large seeds, whereby maternal excess to repressed growth and small seeds (Scott et al., 1998; Haig and Westoby, 1991; Adams et al., 2000). III) Is variation in the embryo:endosperm ploidy ratio a characteristic of apomixis stability? And finally, IV) are seed size and/or ploidy correlated with germination and growth rate, assuming that bigger seeds have more resources?

MATERIAL AND METHODS

A. Plant System

The genus *Boechera* is composed of 71 sexual diploid species, and 38 hybrid taxa with variable ploidy, as described in new nomenclature (Al-Shehbaz and Windham 2006; 2007a, b). For our study we used plants that were initially identified by Rollins (1993) and Dorn (1984) and classified as *B. holboellii*.

B. Samples

From 6 *B. holboellii* accessions (genotypes) (Supplementary data – A), 30 seeds were placed onto moist filter paper in Petri plates and cold treated at 4° C for 3 weeks in the dark. Germinating seedlings were then transferred to pots (11 × 11 × 13 cm) containing sterilized soil and grown under long day conditions in greenhouses at a mean temperature of 23° C. Up to six offspring from each accession were randomly selected for further investigation. All results are listed in Table 2.

Karyotyping. Root tips were collected from two plants of each accession between 8 and 9 am, placed in ice water and stored at 4°C. After 24 h the root tips were fixed in 3:1 EtOH: glacial acetic acid and stored at 4°C. Fixed root tips were rinsed 3 x in distilled water, then rinsed in 10 mM sodium citrate buffer (pH 4.5) and digested for 1.5 h at 37°C in a cellulose (0.5% w/v), pectolyase (0.5% w/v) and cytohelicase (0.5% w/v) mixture (Sigma). After digestion, root tips were rinsed in water and placed into a drop of 45% acetic acid on a clean glass slide. With the help of needles the tissue was spread out and the slide was transferred onto a hot plate (45°C) for 2 min, during which continuous

stirring of the tissue was maintained. When the fixative had evaporated the slide was rinsed in 3:1 EtOH: glacial acetic acid, dried again and a drop of 98% ethanol was added to the dried spot. After the ethanol had evaporated, 10 µl of a DAPI (2 µg/ml) in Vectashield (Vector, Peterborough, U.K.) staining solution was added and the slide was incubated for 1 h at RT in the dark. The samples were viewed with a Zeiss Axioplan 2 microscope under 100X magnification using the appropriate filter combination for DAPI staining.

Genotyping. To rule out contamination between seed pouches, the mother plants and all analyzed offspring were genotyped using nine microsatellites which have been described elsewhere (Clauss et al. 2002; Dobeš et al. 2004; Song et al. 2006; Table 1). Leaf tissue was collected from each plant and genomic DNA was extracted using a Macherey-Nagel NucleoSpin 96 Plant-Kit according to manufacturer's instructions. All DNA was quantified by running 5 µl on a 0.8% agarose gel.

Table 1. Microsatellite loci used for genotyping.

Locus[a]	primer 5' to 3'
ATTSO392	F-TGTAAAACGACGGCCAGTTTTGGAGTTAGACACGGATCTG
	R-GTTGATCGCAGCTTGATAAGC
BF18	F-TGTAAAACGACGGCCAGTAACCTCCCAAGATTCGCTTC
	R-TTCGCCATTGTTGTGATTTG
MBK21B3	F-TGTAAAACGACGGCCAGTATTGCTTCCGTTTTGCTCAT
	R-TTTCAATTCTCTCTGCCTCTCA
BF20	F-TGTAAAACGACGGCCAGTTTCTCGGGAAAGTAATGAGGAG
	R-GCAAATCTGACCAATGCAAG
Bdru266	F-TGTAAAACGACGGCCAGTTTTAATTTGTGCGTTTGATCC
	R-CAAAATCGCAGAATGAGAGG
R3.25	F-TGTAAAACGACGGCCAGTAAAGCCCGTTTCAGATGTTG
	R-TAAAGCCCTTCTTGCTTCGT
Bdru878	F-TGTAAAACGACGGCCAGTGGGAACCTTCATGTCCAAAG
	R-TGCCTTTTCCGTTTTTCTAATC
R3.12	F-TGTAAAACGACGGCCAGTACGTTTGAAGCCCAATACGA
	R-AATCGGAAACCCAGAAATCC
D3	F-TGTAAAACGACGGCCAGTGGTTATGTGAGAGTTAAG
	R-ATTGTTGAATGCAACAGG

[a] Microsatellite primers taken from various sources (Clauss et al. 2002; Dobeš et al. 2004; Song et al. 2006)

PCR amplification was performed using the following conditions: 1x PCR buffer (Bioline), 1.5 mM $MgCl_2$, 0.15 mM dNTPs, 0.2 µM M13 tailed forward primer, 0.4 µM reverse primer, 0.36 µM WellRed labeled M13(-18) primer (Proligo), 0.25 U Biotaq DNA polymerase (Bioline) and 2 µl extracted DNA template. Amplification was carried out using Mastercycler Epgradient thermal cycler (Eppendorf) with the following profile: i) denaturation for 4 min at 94°C; ii) with every third cycle, increase annealing temperature

by 1°C starting with 94°C for 30 s, 50°C for 15 s, 72°C for 20 s and ending with 54°C; iii) 25 cycles with 94°C for 30 s, 59°C for 15 s and 72°C for 20 s; iv) final extension of 10 min at 72°C. PCR products from differently-labeled reactions were pooled, and purified using magnetic beads (Agencourt Ampure, Beckman Coulter). DNA fragments were separated on CEQ8000 Fragment analyzer system, and genotypic data collected using the system fragment analysis software (Beckman Coulter TM).

Reproductive mode. To confirm earlier studies (Voigt et al., 2007), 10 seeds per lineage were additionally tested. Single seed measurements were performed on a PA II flow cytometer (PARTEC, GmbH, Münster, Germany) using the flow cytometry seed screen protocol from Matzk et al. (2000), which allows the measurement of embryo and endosperm ploidy. For the sample preparation a single seed was placed in a Petri dish and chopped with a razor blade to release nuclei in 1ml staining solution (0.107 g $MgCl_2$ x $6H_2O$, 0.5 g NaCl, 1.211 g Tris, 0.10 ml Triton x-100, 0.9 g Sodiumcitrate-Dihydrate, 1.0 ml DAPI [4',6-Diamidin-2'-phenylindoldihydrochlorid = 1mg/ml], pH = 7.0). The solution was filtered through 30µm mesh (PARTEC staining solution, Münster, Germany) and incubated for 15 min on ice before measurement. The same protocol was applied for leaf tissue.

C. Reproduction traits – 1st Generation

Five replicates (plants) from each accession were analyzed (Table 2; Figure 1). All statistical analyses were performed with SPSS (version 11.5) and R (R Development Core Team 2007) using standard libraries and the "ade4" package for Mantel tests (Chassel et. al., 2004).

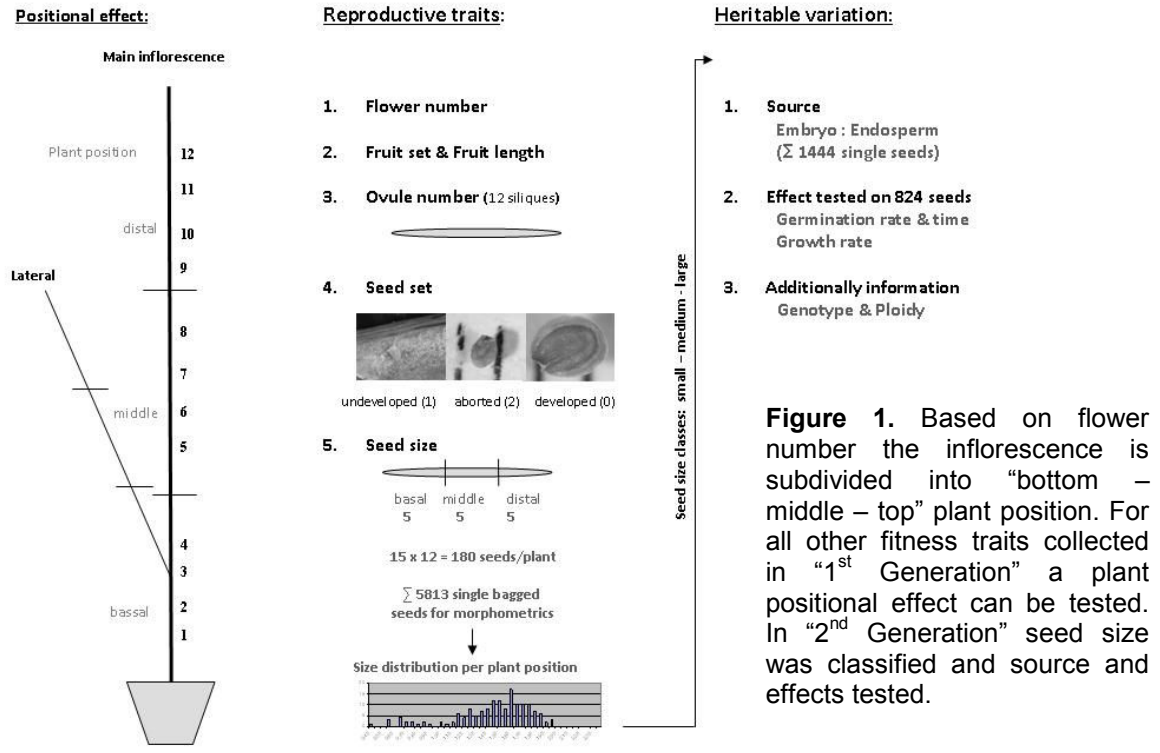


Figure 1. Based on flower number the inflorescence is subdivided into "bottom – middle – top" plant position. For all other fitness traits collected in "1st Generation" a plant positional effect can be tested. In "2nd Generation" seed size was classified and source and effects tested.

Size of reproductive structures. For each plant, one or two inflorescences were picked and identified as either main or lateral inflorescence. As was often the case for diploid apomictic plants, where no main inflorescence was available, two lateral inflorescences were examined. The flower number was counted along the inflorescence, and the count divided by three subdivided the florescence into three positions (bottom - middle – top) in order to test for positional effects. At each plant position four siliques were randomly chosen and their length measured. Each silique was divided into 3 sections (basal - middle - distal), and five (developed) seeds were randomly chosen and individually bagged. In total for each plant (if possible) 180 seeds were individually bagged (12 siliques x 15 seeds) and each seed was labeled with a specific code containing information about plant ID, plant position, silique section, and silique number. All seeds were photographed using a Zeiss Stemi 2000-C in connection with a Fujix Digital Camera HC-300Z. Seed area measurements were performed using DigiShape (Cortex Nova, 2005) software.

Fecundity and fertility. The four siliques per plant position (per plant) were evaluated for i) number of ovules; and, ii) seed set, whereby seeds were coded along the silique (basal to distal end) as developed (0), undeveloped (1), or aborted (2) (Figure 1). The seed set was calculated as the number of developed seeds per silique. The silique is a two valve capsule, and cautious opening of one side did not prevent spontaneous opening of the second valve, which released seeds prior to coding, making only one valve of the silique countable. All correlations of seed set with positional effects were made for each silique based upon the ovule number divided into three sections (basal - middle - distal).

Data analysis. A *paired sample t-test* was used to test whether trait values along the main shoot were different from trait values along the lateral shoot. All data were first tested using the *Kolmogorov-Smirnov* test for normal distribution of variables and the *Levene test* for homogeneity of variance. If there was no significant difference between values of main and lateral shoot, data points were combined.

The variability of each genotype was estimated using the coefficient of variation (CV). It was calculates as standard deviation divided by the mean. We selected a relative measure of variation to ensure comparability among different phenotypic traits with different units of measurement. We tested the phenotypic trait i) ovule number, ii) seed set and iii) seed size. CV values were calculated for each replicate and then combined among genotypes, and results were illustrated in a graph.

Differences in traits between triploid and diploid apomicts were evaluated using nested analyses of variance, with inflorescence position within plants within genotype within ploidy treated as random effects. A particularly serious concern in the analysis of variance is the issue of unbalanced design (unequal amount of data points collected per

trait between plants), for in this case the significance tests lack theoretical justification (Shaw and Mitchell-Olds, 1993). Considering lack of balance in our design, a randomization approach was used to confirm results received from nested ANOVA. An R script was written to perform 5000 randomizations on trait data, and the resultant factor having the highest variation in the response variable was chosen and analyzed using one-way ANOVA followed by a *Tukey's HSD* test to obtain detailed information about differences between the factor's levels.

D. Heritable variation 2nd Generation

Based upon the mean seed size distribution per individual plant, single seeds were classified as small, medium or large. Eight seeds of each seed size class per plant and plant position were selected. The seeds were collected for two different experiments: one to explore the source of observed seed size variation; and in a second, to test whether seed size plays a role in fitness. For the first experiment and for each replicate we choose the five smallest, five middle and five largest seeds per plant position (bottom, middle, top; in total = 45 seeds) and performed a flow cytometric seed screen to measure the ratio of embryo to endosperm ploidy. All statistical calculations were performed with SPSS (version 11.5) or excel (Microsoft ® Office Excel 2003).

1. Source: *Embryo to endosperm ratio*. For the three positions on each plant, five seeds per size class were chosen for single seed measurements of embryo and endosperm ploidy via flow cytometry seed screening (FCSS, in total 45 seeds per plant/ 15 seeds per plant part; Matzk, 2000; see section B.). To compare seed size with ploidy ratios the *Pearson chi*^{2} test was applied, and to calculate over-representations of certain embryo:endosperm ratios between seed size groups the *Fisher's exact* test (One-tailed) was applied taking total seed number within ploidy as reference.

2. Effects: *Germination rate and time*. Additionally three seeds per size class of each plant position for one inflorescence were chosen (in total 27 seeds per plant) and placed onto moist filter paper in Petri plates, cold treated at 4° C for 2 weeks in the dark, and then transferred into light and RT. When the seed coat broke and root tips were visible, that time point was noted.

Growth rate. Germinated seedlings were placed in pots (11 × 11 × 13 cm) containing sterilized soil and grown under long day conditions in the greenhouse at a mean temperature of 26° C. After 34 d the measurement of rosette diameter (distance between two largest leaves) was conducted. To correlate seed size with germination time and rate, the *Person chi-square test* was applied.

3. *Effect of common ancestry on among-genotype similarities*. From all surviving plants, leaf tissue was collected for ploidy determination via flow cytometry (see section B) and

for genomic DNA extraction. The software POPDIST (Guldbrandtsen et al., 2000) (http://genetics.agrsci.dk/~bernt/popgen) was used to create a genetic distance matrix using the algorithm of Tomiuk and Loeschcke (1991; 1996) which enables the comparison of populations of differing ploidy. The software QuickTree (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form= quicktree) transformed the distance matrix into New Hampshire format which then could be visualized by TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) (Page, 1996).

A phenotypic distance matrix between genotypes was built on the basis of 4 traits (seed size, fruit size, seed set, fruit set) using Euclidean distance. A Mantel-test (Mantel, 1967) was used to compare the phenotypic and genetic distance matrices, which tests whether the architecture of any given matrix is predicted by a hypothesis matrix by randomizing the entries in one matrix, recalculating the matrix correlation coefficient, and determining an empirical distribution of the matrix correlation coefficient. The actual value of the inter-matrix correlation is then compared with an empirical distribution to determine the true probability of rejection of the null hypothesis of no correlation (Cheverud et. al., 1989). In our case, the empirical distribution was based on 10000 replicates.

RESULTS

B.1 *Confirmation of previous results and classification*

Karyotype. The analyses confirmed that the tested plants were sexual diploid (2n = 14), diploid apomict (2n+x = 15), and triploid apomict (3n = 21; Figure 2). For all diploid apomict accessions an aneuploid chromosome was counted (Table 2).

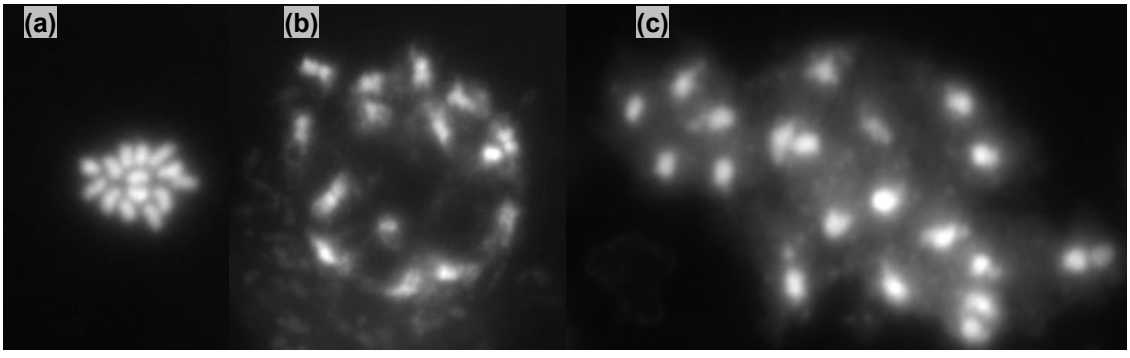


Figure 2. Chromosome complements of (a) sexual diploid, (b) apomictic diploid (2n+x = 15 chromosomes) and (c) apomict triploid (3n = 21 chromosomes).

Reproductive mode. Ten mature seeds from each accession were analyzed. Sexually produced seed are expected to have a diploid embryo (maternal = 1n + paternal = 1n) and a triploid endosperm (maternal = [1n + 1n]+ paternal = 1n). Apomictically-derived seeds from diploid plants are expected to have a non-reduced diploid embryo (maternal = 2n) having the same ploidy of mother plant (e.g. diploid) and a hexaploid endosperm (maternal = [2n + 2n] + paternal = 2n), while triploid apomicts have non-reduced triploid embryo (maternal = 3n) and nonaploid endosperm (maternal = [3n + 3n] + paternal = 3n) (Supplementary data – B). The reproductive mode of all accessions is indicated in Table 2 and confirmed previous results.

Table 2. *Boechera* lines which were used in this study.

Line/ genotype	Species	Repro. mode	Ploidy	Karyotype	Population	Location US
132	*B. stricta*	sex	2n	14	CACR 5	Canyon Creek/ Vipond Park, MT
120	*B. holboellii*	apo	2n+x	15	MURA 2	Mule Ranch, Beaverhead Co, MT
300	*B. holboellii*	apo	2n+x	15	BC 12	Birch Creek, Ravalli Co, MT
28	*B. holboellii*	apo	2n+x	15	BC13	Birch Creek, Ravalli Co, MT
148	*B. holboellii*	apo	3n	21	MT 2TWRO	Challis, Lemhi Co, ID
195	*B. holboellii*	apo	3n	21	MC 5	Highwood Mtns, MT
218	*B. holboellii*	apo	3n	21	MC 4	Highwood Mtns, MT

Genotype analysis. Nine microsatellite loci were successfully amplified for all individuals, and accession-specific alleles and consequently accession-specific genotypes (i.e. amplicon fragment sizes) were identified (Table 3). The sexual diploid line 132 was heterozygous at only one locus, while all apomictic lineages are highly heterozygous (from 5 to 9 of all loci). All apomictic genotypes were identical to their respective mother plant, thus confirming their apomictic nature. Based upon the nine loci, lines 28 and 300 have identical genotypes, although they are known to differ at other loci (Unpublished data).

Table 3. Multilocus genotypes of 7 *Boechera* lines over 9 microsatellite loci (amplicon fragment sizes in bp are presented).

Line	No. $^{a}/_{b}$	Loci								
		BF18	ATTSO392	MBK21B3	BF-20	Bdru266	R3.25	Bdru-878	R3.12	D3
132	6/ 91	*138/ 138*	*160/160*	-	*184/ 232*	*149/149*	*202/ 202*	*138/138*	*103/ 103*	*186/ 186*
28	6/ 101	*135/ 137*	*157/160*	*206/206*	*225/ 237*	*147/147*	*190/ 192*	*126/136*	*181/ 181*	*186/ 186*

Table3. continued

Line	No. [a]/[b]	Loci								
		BF18	ATTSO392	MBK21B3	BF-20	Bdru266	R3.25	Bdru-878	R3.12	D3
300	7/ 129	*135/ 137*	*157/160*	*206/206*	*225/ 237*	*147/147*	*190/ 192*	*126/136*	*181/ 181*	*186/ 186*
120	5/ 38	*135/ 137*	*157/160*	*193/205*	*226/ 226*	*148/152*	*186/ 192*	*138/142*	*103/ 104/ 179*	*186/ 186*
148	6/ 93	*135/ 137/ 138*	*157/160*	*193/205*	*184/ 232*	*147/149*	*192/ 196*	*136/138*	*110/ 179*	*186/ 215*
195	7/ 143	*135/ 137/ 141*	*154/157/ 160*	*192/204*	*226/ 223*	*147/149 / 151*	*189/ 191/ 201*	*134/138*	*175/ 181/ 183*	*186/ 192*
218	6/85	*139/ 141*	*157/160*	*192/200*	*225/ 231*	*148/152*	*190/ 201*	*126/150*	*175/ 179/ 181*	*192*

[a] Number of offspring per line tested, which are involved in 1st generation fitness experiment.
[b] Number of surviving offspring in 2nd generation experiment which were genotyped.

C.1 *Plant positional effect on traits.*

Nested analysis of variation for fruit and seed allocation indicated that inflorescence is not the main source for variance. As seen in Figure 3, we found positional effects in all traits and lines, but this effect is not predictable and changed with respect to accession and trait. The variation in traits was mainly effected by the genotype (Table 4, Figure 3, Supplementary data – C, randomization test).

Table 4. Nested analysis of variance for fruit and seed allocation among ploidy levels in three diploid and three triploid genotypes of *Boechera* (the highest variance attributed to a given factor and significant F values are bold faced).

Reproductive component	*F*	*p*	Variance components		
			Genotypes within ploidy	Plants within genotype within ploidy	Residuals = inflorescence positions
Fruit length	0.00	0.99	**5.257**	0.430	0.711
Seed size	0.55	0.50	**1.043**	0.263	0.056
Fruit set	9.91	**0.04**	1.048	92.37*	-
Seed set	0.89	0.40	**695.27**	60.30	340
Flower number	1.00	0.38	**675.26**	219.2*	-
Ovule number	4.64	0.10	**365.92**	23.65	55.8

Note: The highest variance attributed to a given factor and significant F values are bold faced.
*In the case of fruit set and flower number, inflorescence positions are not taken into account.

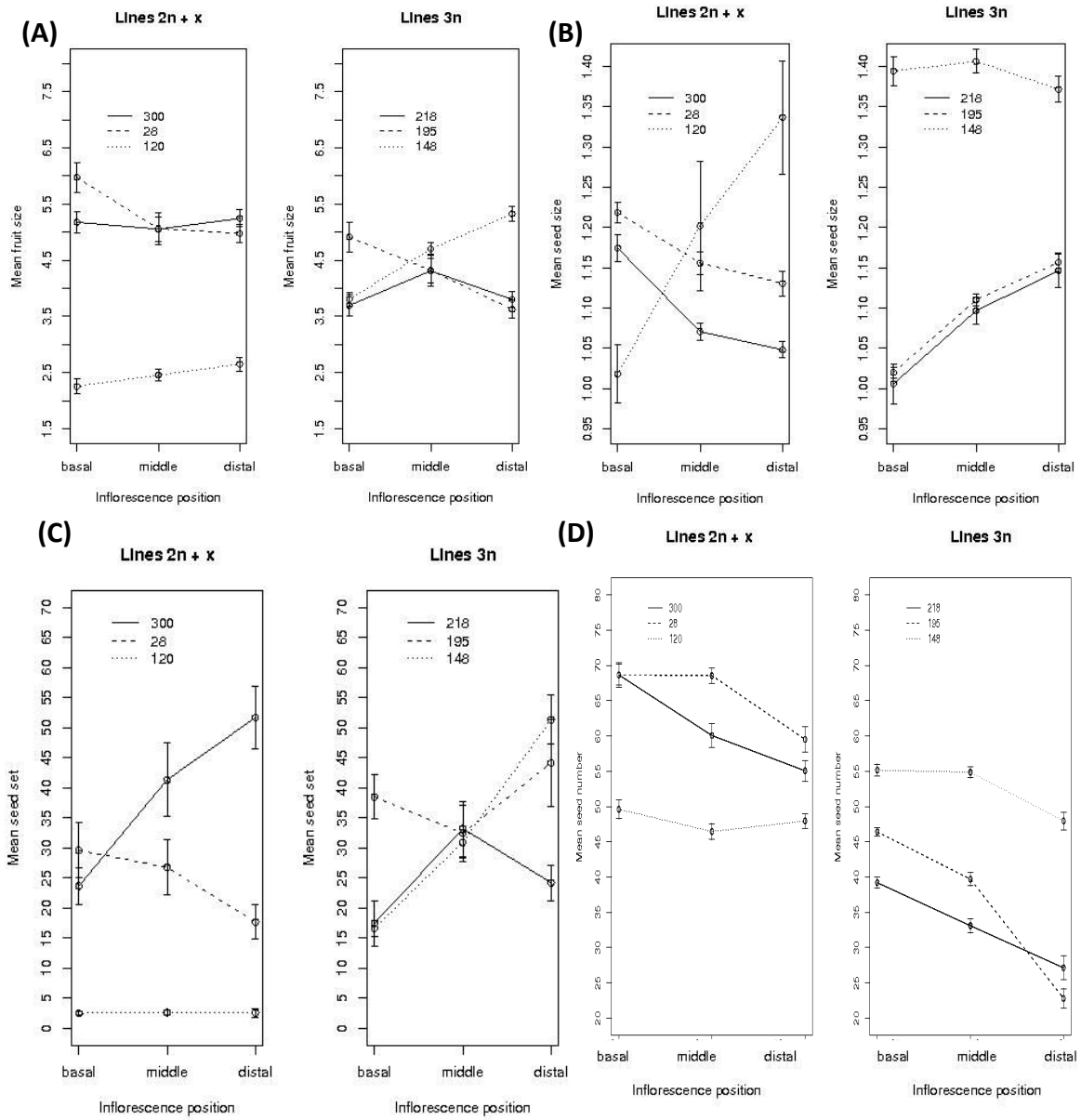


Figure 3. Line plots of (A) mean fruit length, (B) mean seed size, (C) mean seed set and (D) mean ovule number per lineage (different line styles refer to specific lineages (Table 3) with standard error bars at each plant position) for diploid (2n + x) and triploid (3n) apomicts.

The correspondence between genetic distance and phenoypic variation of four reproductive traits was estimated (Mantel test, Supplementary data - D) and indicated furthermore that morphology and life-history traits were independent from genetic relatedness (Supplementary data – D).

Size of reproduction structures. As seen in Figure 4A, line 120 produced the shortest siliques (mean = 2.4 cm), and *Tukey test* indicated its significant difference to all other lines ($P < 0.001$). The *Tukey test* furthermore found no significant differences between line 195 + 218; 195 + 148; 148 + 300 and 28 + 300. Although seeds of diploid apomicts were smaller in mean (mean = 1.12 mm^2) than triploid apomicts (mean = 1.21 mm^2), no

significant interaction between size of seeds and ploidy (diploid or triploid) were observed (Table 4).

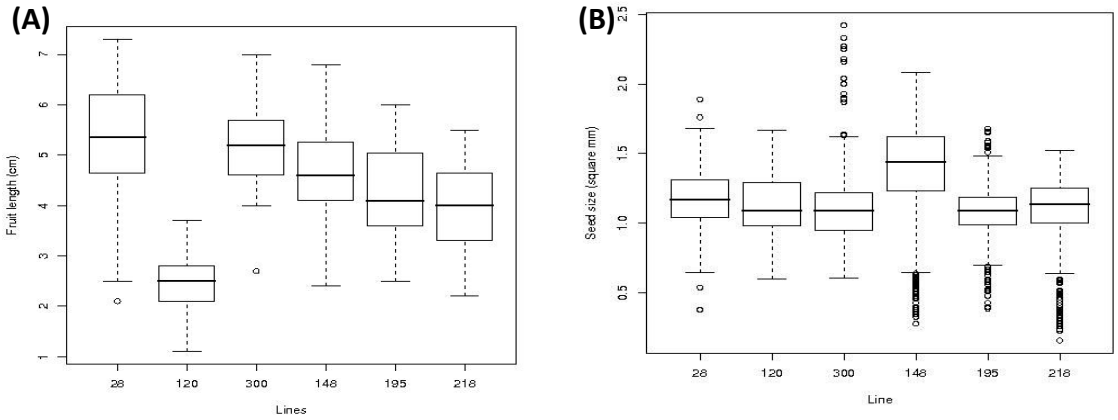


Figure 4. Box-plots illustrating mean values ± SE per lineage of (A) fruit length in cm and (B) seed size in mm^2. Outliers are indicated by circles.

Seed size varied significantly between line 148 (mean = 1.45 mm^2) and all other lines ($P < 0.001$). Furthermore, significant differences between line 28 and all other lines (except against line 120) were also observed (*Tukey test*, $P < 0.001$). The triploid lines displayed a range of small outliers (Figure 4B), whereby lineage 300 was characterized by large outliers (Supplementary data – E).

C.2 *Fecundity and fertility.*

Analyzing the reproductive effort of plants we counted how many flowers were formed in one generation and how many of those developed a fruit (Figure 5). For two lines (148 & 195) comparison between flower numbers (Figure 5A) of main and lateral shoots were calculated and in both cases significant differences were observed (ID 148, $t = 6.125$, $P = 0.026$ and ID 195, $t = 21.92$, $P = 0.002$), while all other traits were not significant (Supplementary data - F).

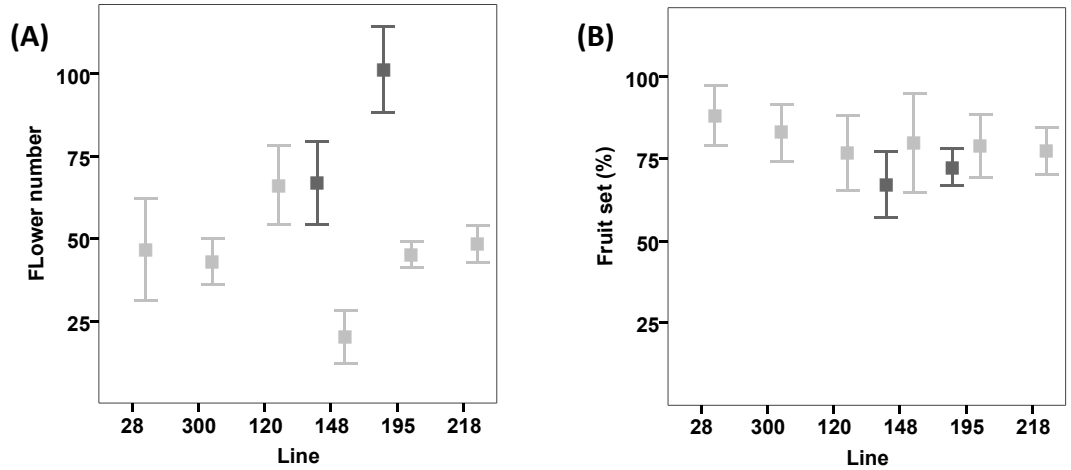


Figure 5. Error bars of mean values ± SE of flower number (A) and fruit set (B) per accession. Grey illustrates mean value of main shoots, and light grey that of lateral shoots.

Interestingly, fruit set (calculated as how many flowers developed a fruit after fertilization) was similar between accessions (only around 67% to 89% of flowers developed a fruit) regardless of differences in flower number (Table 4, Figure 5B, Supplementary data - G).

Seeds in the two valve capsule (silique) are arranged in a zipper-like formation. Siliques were opened, and seed set data were determined from basal to distal end and calculated with regards to overall ovule number in one half of the silique only (as only one half could be counted due to the automatic opening of the second half and subsequent loss of positional information). The nested analysis of variance indicated no significant difference between diploid and triploid apomicts even though the mean value was higher in the former (diploid: mean = 56.88; triploid: mean = 40.11; Table 4, Figure 6A, Supplementary data -G), with genotype rather than ploidy being the main source of variance.

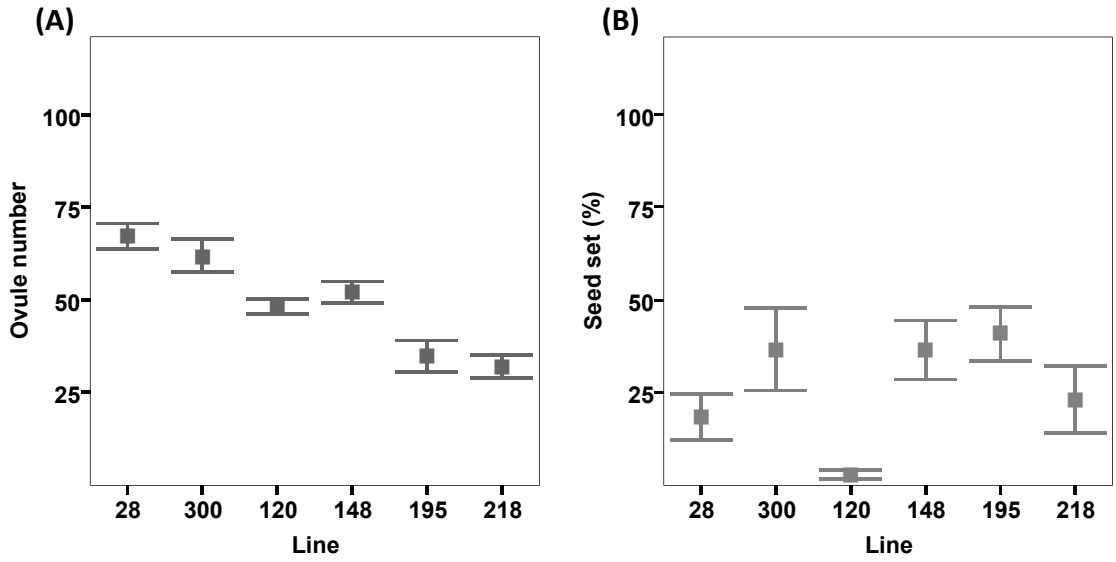


Figure 6. Mean ± SE of ovule number across lines (A) and mean ± SE of seed set (%) within lines (B).

Interestingly, regardless of ovule number (Figure 6A), the mean seed set (% of developed seeds per silique) across all apomictic lines was below 40% (Figure 6B). Mean seed set among diploid apomicts was line-specific (ID 120 mean = 2.65; ID 28: mean = 24.7 and ID 300: mean = 38.47; Table 4), and among triploid apomicts, two subgroups emerged (*Turkey test*: ID 218: mean = 24.1; ID 195: mean = 39.5, ID 148: mean = 37.17; Figure 6B, Supplementary data - C).

We additionally coded the seed set information from basal to distal end of the silique. Seeds which contained a visibly large embryo were considered developed; those which contained no embryo, a very small embryo or one with a shriveled appearance were classified as aborted; and seeds which were dried out and beige in color but still

having the size of an ovule were coded as undeveloped (Figure 1). The number of aborted seeds was different between diploid and triploid apomicts, with diploid accessions having more aborted seeds (2n+x: mean = 5.5; 3n: mean = 0.4), especially in lines 28 and 300. Aborted seeds were rarely found among the triploid lines (Figure 7A). The nested analysis of variance indicated a significant correlation between ploidy and amount of aborted seeds (Table 5). In diploid apomictic accessions, aborted seeds were found more often at the basal and distal ends of a silique (Figure 7B) compared to the middle section. The nested analysis of variance for the amount of undeveloped seeds and ploidy indicated that the source of variation was mainly the genotype (Table 5, Supplementary data - C).

Table 5. Nested analysis of variance for seed size in three diploid and three triploid genotypes of *Boechera* (the highest variance attributed to a given factor and significant F values are bold faced).

Reproductive component	*F*	*p*	Variance components		
			Genotypes within ploidy	Plants within genotype	Residuals = inflorescence positions
Aborted seeds	11.7	**0.03**	0.437	0.475	0.441
Undeveloped seeds	0.0003	0.98	**0.254**	0.053	0.150

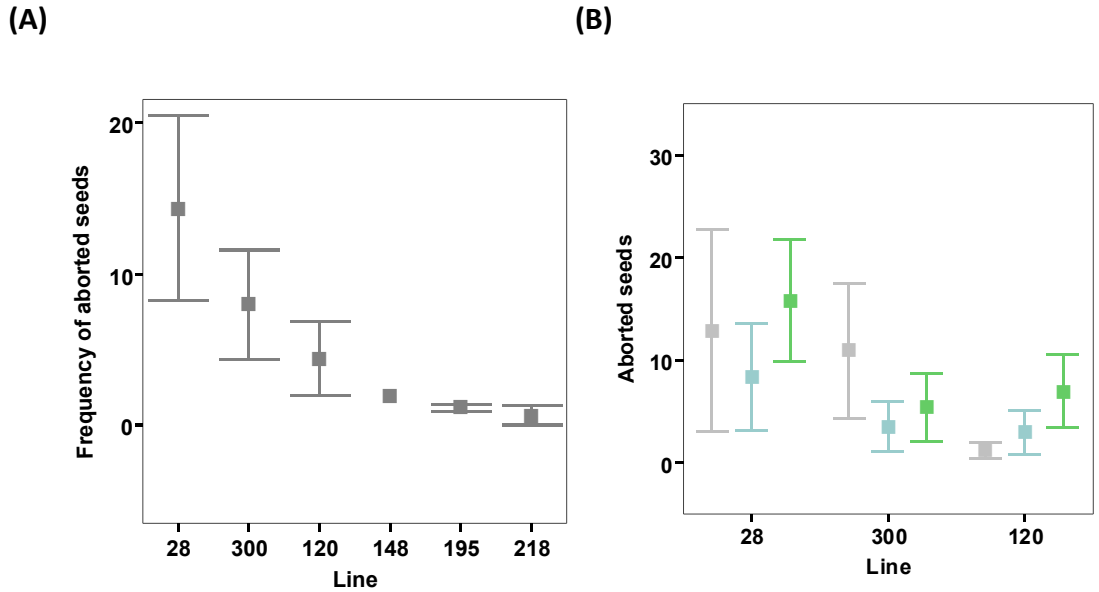


Figure 7. Aborted seed frequency (mean ± SE across lines; A), and (B) frequency of aborted seed (mean ± SE per silique section; grey = basal, blue = middle, green = distal) in more detail for diploid apomictic lines.

In triploids, the number of developed seeds per inflorescence was different between main and lateral shoots, whereby the main shoot produced twice as many developed seeds as the lateral shoot (Table 6). Furthermore, in diploid apomicts, which

have no main shoot, the amount of developed seeds per lateral shoot were similar to those from main shoot of triploids (Table 6, Figure 8).

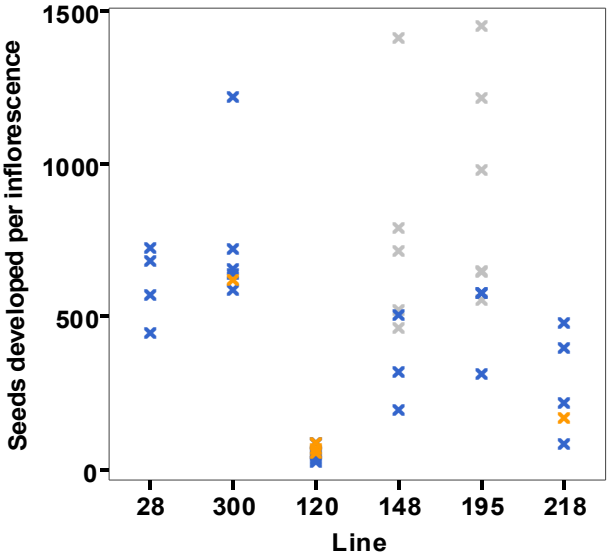


Table 6. Mean and SD values of developed seeds per line (l = lateral shoot, m = main shoot), n = number of shoots included in estimation.

Line	shoot	n	$\tilde{x}$	SD
28	l	4	607.2	123.3
300	l	6	740.9	238.5
120	l	8	58.2	24.05
148	m	5	779.8	376.6
	l	3	340.6	157.3
195	m	6	916.3	360.9
	l	3	491.1	150.8
218	l	5	269.9	165.6

Figure 8. Scatter-plot of developed seed output per inflorescence and line (grey = main shoot; blue = 1st lateral shoot; orange = 2nd lateral shoot). The calculation is a function of mean value of developed seeds per fruit ($\tilde{x}$) times amount of fruits (fruit N) per inflorescence.

Variation within lines. The variation of the phenotype within lines was measured between line replicates (clonal offspring of one mother).

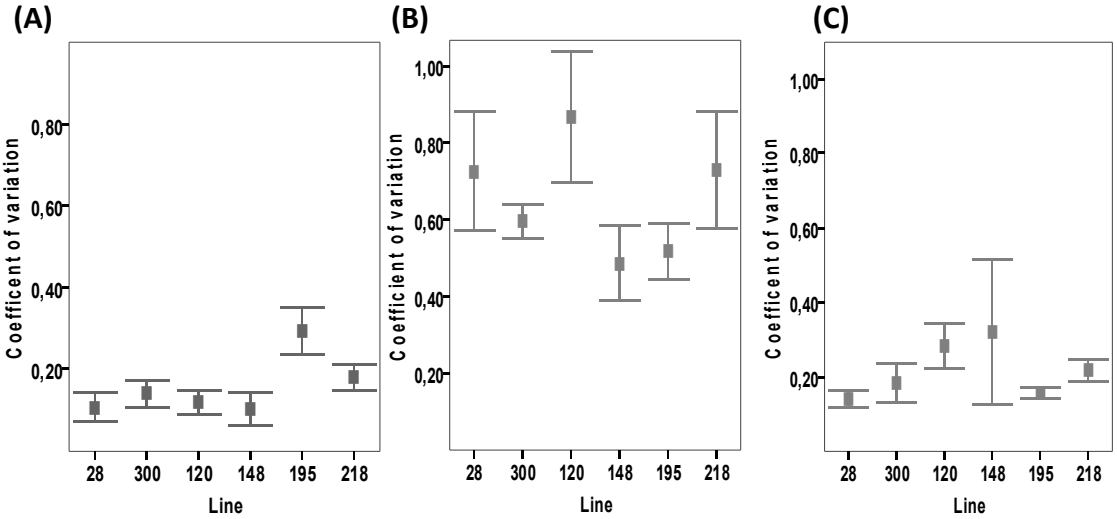


Figure 9. Coefficient of variation ± SE for (A) ovule number, (B) of seed set (%), and(C) seed size between lines.

The coefficient of variation (CV) among replicates across all lines was lower for ovule number in comparison to CV estimation of seed set and seed size.

D.1 Heritable variation in offspring.

In total 1212 single seeds were measured, 488 seeds from the three diploid apomictic lines and 724 seeds from the three triploid apomict lines. The ratios were classified as: i) expected; ii) unexpected, and iii) questionable (Tables 7, 8). Expected ratios were considered on the basis of Chapter I (Table. 2), where for diploid apomictic lineages the embryo : endosperm ratio 2C:6C were the most prominent, and for triploid apomicts 3C:9C. As seen in Chapter I, one diploid apomictic lineage (line 120) produces haploid sperm nuclei, which is reflected in their characteristic embryo:endosperm ploidy ratio of 2C:5C. Aberrations from these ratios were classified as *unexpected*, and those in which either embryo or endosperm ploidy could not be estimated were classified as *questionable* (Table 7, 8).

Table 7. Number of seed measured and classification into groups, as described in text.

Ratio	2n+x	3n	Total
no result	12	70	82
Expected	400	539	939
Unexpected	41	26	67
Questionable	35	89	124
Total	488	724	1212

In total, 939 (77.4 %) embryo:endosperm ratios were as expected, 67 (5.5 %) were unexpected, 124 (10.2 %) cases were questionable and for 82 cases (6.7%) the measurement failed (Table 7). In comparison of seed results from diploid versus triploid plants, 90% (441 cases) of seeds gave complete embryo:endosperm ratios in the former, whereas only 78% (565 cases) of seeds yielded a complete measurement in triploids. Questionable results were mainly obtained from triploid apomicts, and were characterized by unsuccessful estimation of the endosperm ploidy (i.e 3C:?; Table 8). Seeds for which estimations of embryo and endosperm ploidy were not successful were typically from the smallest size group ($X^2 = 20.9$; $P < 0.001$).

Table 8. Number of seeds showing different embyo:endosperm ploidy ratios identified among diploid and triploid *Boechera*.
[1] unexpected, [2] questionable, [3] expected, [4] autonomous endosperm

Ratio	2n+x			Ratio	3n		
	small	medium	large *		small	medium	large
2C.4C [1,4]	3	1		2C:? [2]	1	2	1
2C:7C [1]	1			3C:6C [1, 4]	5		
2C:8C [1]	4	4	16	3C:12C [1]	2	2	3
2C:10C [1]	1			3C:? [2]	34	21	13
2C:? [2]	7	4	12	6C:? [2]	4	2	2
2C-3C:? [2]	1			6C:9C [1]	4	4	4
2C-3C:5C [2]		1		6C:12C [1]			1
2C:4C-5C [2]	1			?:9C [2]	8		1
4C:6C-7C [2]	1			9C:12C [1]			1
3C:6C [1]			1	No result	58	6	6
3C:? [2]			1				
4C:? [2]	4	3					
4C:6C [1]			7				
4C:8C [1]	1						
4C:10C [1]	2						
No result	10	1	1				
TOTAL	36	14	38		116	37	32
2C:6C (5C)[3]	127	152	121	3C:9C[3]	151	202	186
TOTAL	163	166	159		267	239	218

*Classification of small/medium/large size group based upon line specific seed size distribution.

The groups having *unexpected* and *questionable* ratios were of most interest. Within the diploid apomicts (Table 8) more variation was visible, with 17 different ratios, compared to triploid apomicts which had 10 different ratios. Within the small seed size group of diploid apomicts 6 unexpected ratios were identified, while among the medium and large seed size groups, 2 and 3 unexpected ratios were found respectively. Within the triploids in the medium seed size class 2 unexpected ratios were identified, whereby among the small and large seed size groups 3 and 4 different unexpected ratios were found respectively.

To study the effects of egg cell fertilization on seed size ratio, expected and unexpected ratios were compared. In diploid offspring, 2C:6C, 4C:6C and 3C:6C seeds were all characterized by similar endosperm formation, but in two cases fertilization of the egg cell took place, and interestingly those seeds were larger (Table 9). Interestingly, this effect was not found in the offspring of triploids, as 3C:9C and 6C:9C ratios were equally distributed among seed size classes. The effects of paternal or maternal genome excess (dosage effects) during endosperm formation and embryo development were likely

apparent in offspring of triploids, as these showed autonomous endosperm formation exclusively in small seeds. Autonomous endosperm formation in diploid apomicts was identified in the small and medium seed size groups.

Paternal excess can be studied in seeds having the following ratios: 2C:6C [2m: 4m+2p; m – maternal, p – paternal] and 2C:8C or 2C:10C. A significant difference was calculated for 2C:8C [2m: 4m+4p], being mainly found in the large seed size class (Fisher's exact, $P = 0.024$), while the one seed showing 2C:10C [2m: 4m+6p] belonged to small seed size class (Table 8). Such effects were not visible among the offspring of triploids, as paternal excess without egg cell fertilization (3C:12C) [3m: 6m+6p] was equally identified across seed size classes (Table 8). Egg cell fertilization and paternal excess in endosperm was found rarely in both karyotypes (2n+x = 3 seeds; 3n = 2 seeds; Table 9). The three seeds from the diploids were identified in line 28, and all of them were small. For triploids egg cell fertilization or paternal excess alone had no seed size effect, while in combination all seeds were large.

Table 9. Numbers of seeds showing different embryo:endosperm ploidy ratios per line, equivalent ratios between 2n+x and 3n were put in the same row.

			2n+x				**3n**		
Ratio	28	300	120	Ratio	148	195	218	Event	
2C:4C		1	3	3C:6C	4		1	maternal excess	
2C:5C			57					expected	
2C:6C	171	172		3C:9C	168	231	140	expected	
2C:7C			1					paternal excess	
2C:8C	13	10	1	3C:12C		3	4	paternal excess	
2C:10C		1						paternal excess	
4C:6C	4	3		6C:9C	4	6	2	fertilization of egg cell	
4C:8C	1			6C:12C		1		double fertilization	
4C:10C	2							double fertilization	
3C:6C		1						double fertilization	
				9C:12C		1		double fertilization	
2C:?	2	19	1	3C:?	20	23	25		
3C:?		1		2C:?	4				
4C:?	3	5		6C:?	4	2	2		
				?:9C	8	1			
2C-3C:?		1							
2C-3C:5C		1							
2C:4C-5C			1						
4C:6C-7C	1								
No result	5	6	1	No result	29	8	33		

Effects of embryo and endosperm. Taking all measurements together in both apomictic karyotypes, we found unfertilized apomeiotically-derived embryos (95.9 %), fertilized apomeiotically-derived embryos (3.8 %) and reduced unfertilized embryos (0.3 %) (Figure 10). Seed size and embryo ploidy are independent from each other (*Pearson chi square* 2n+x: x^2= 0.14, 3n: x^2 = 0.84), although a tendency in the diploid apomicts was visible towards fertilized egg cells in small and large seeds (small = 7; medium = 3; large = 9). Such a tendency was not seen within triploid apomicts.

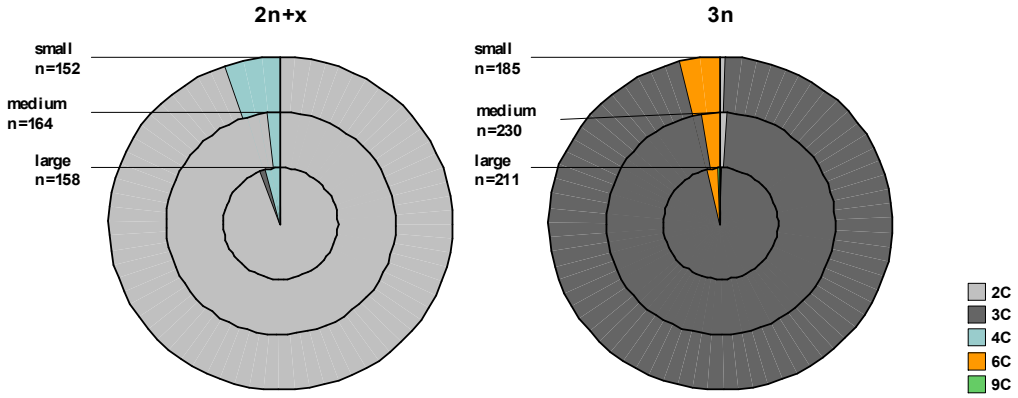


Figure 10. Pie-charts showing different embryo ploidies, and clustered by seed size group (inner circle = large, middle = medium and outer = small) for each apomictic ploidy class. Color scale on side indicates ploidy of embryo.

The rare occurrence of reduced-unfertilized diploid embryos from triploid mother plants was found exclusively in line 148 across all seed size groups. The occurrence of apomeiotically-derived and fertilized egg cells is different between lines. A nonaploid embryo from a triploid mother plant was found exclusively in line 195. The triploid line 218 showed an increased frequency of hexaploid embryos (3C egg cell + 3C sperm nucleus) in the small seed size group. In the diploid lines 28 and 300 tetraploid embryos (2C egg cell + 2C sperm nucleus) were found in higher frequency among both small and large seed size groups (Table 9, Figure 11).

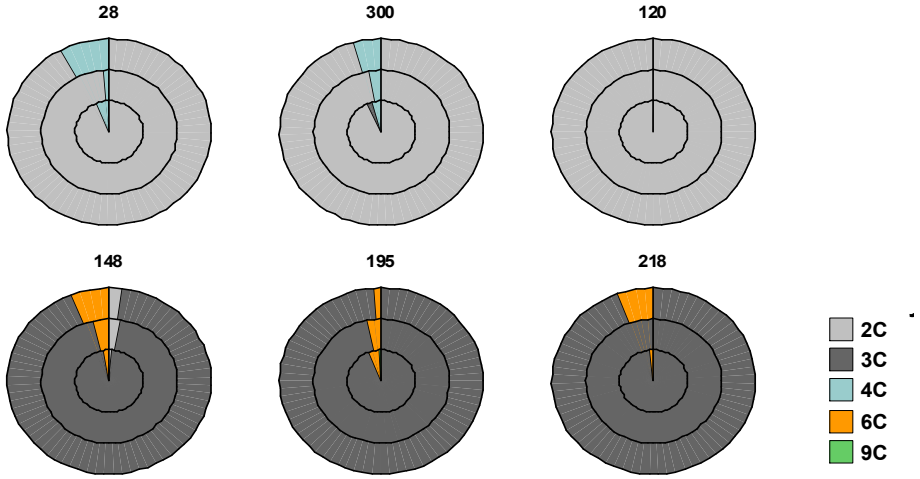


Figure 11. Pie-charts showing different embryo ploidies, and clustered by seed size group (inner circle = large, middle = medium and outer = small) for different *Boechera* lines. Color scale on side indicates ploidy of embryo.

In line 300 we found even a rare cases of triploid embryos (2C egg cell + 1n sperm nucleus), indicating fertilization between an apomeiotically derived egg cell and a haploid sperm nucleus (Figure 11).

Endosperm was produced both autonomously and pseudogamously (Table 8). Autonomous endosperm (no paternal genome contribution) was found with low frequency (2n + x: 0.9%, 3n: 0.9%), whereby pseudogamous endosperm formation occurred most frequently, with different maternal and paternal contributions. Among the diploid apomicts, we observed more endosperm ploidies (n = 7) than among the triploid apomicts (n = 3; Figure 12), although there are line-specific differences.

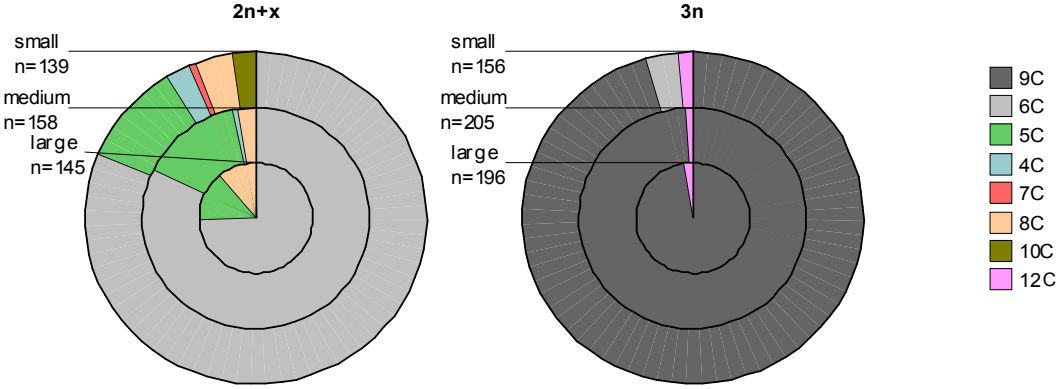


Figure 12. Pie-charts showing different endosperm ploidies, and clustered by seed size group (inner circle = large, middle = medium and outer = small) for different apomictic ploidy classes. Color scale on side indicates ploidy of endosperm.

Seed size and endosperm ploidy varied independently from each other within lines, except in lineage 148, where the small seed size group showed elevated levels of autonomous endosperm (*Pearson chi square* 148: x^2 = 0.004). Considering all identified endosperm ploidies in diploid lines, an over-representation of 8C endosperm in the large seed size group (P < 0.001) was found. Accession 120 had mainly 5C endosperm, the product of the fusion between non-reduced polar nuclei and a reduced sperm nucleus (Figure 13).

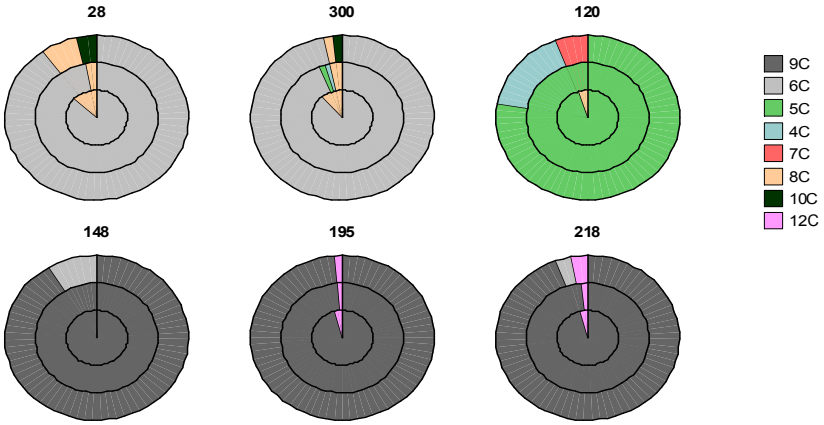


Figure 13. Pie-charts showing different endosperm ploidies, and clustered by seed size group (inner circle = large, middle = medium and outer = small) for different *Boechera* lines. Color scale on side indicates ploidy of endosperm.

Germination. We tested for a correlation between seed size and offspring fitness, and therefore three small, three medium and three large seeds per plant position (bottom, middle, top; in total 27 seeds) were chosen for each plant for germination (Supplementary data – H). In Figure 14 the number of germinated and non-germinated seeds per seed size group and line are illustrated. In line 148 (chi^2 $P < 0.001$) and 218 (chi^2 $P = 0.025$) a correlation between seed size and germination rate could be found (Figure 14, Supplementary data H,I), it indicated that almost all seeds of the large seed size group germinated whereby a significant amount of small seeds did not germinate.

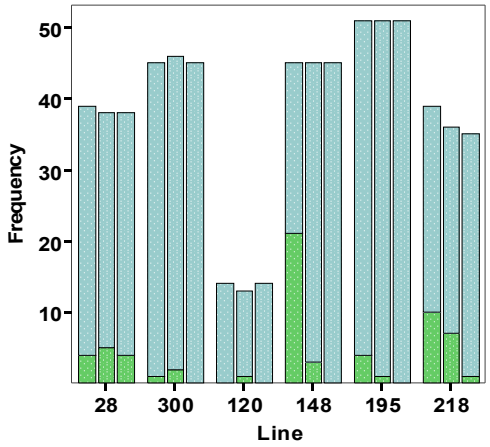


Figure 14. Numbers of germinated (blue) and (green) non-germinated seeds per accession (the three bars illustrate results of small, medium and large seed size classes)..

The seeds of triploid apomictic plants germinated earlier than diploid apomictic plants, sometimes even before they were put into room temperature and light conditions. Upon the onset of germination, all seeds of the diploid lines had completed germination within four days, whereas the seeds of the triploid lines needed up to 15 d (Figure 15).

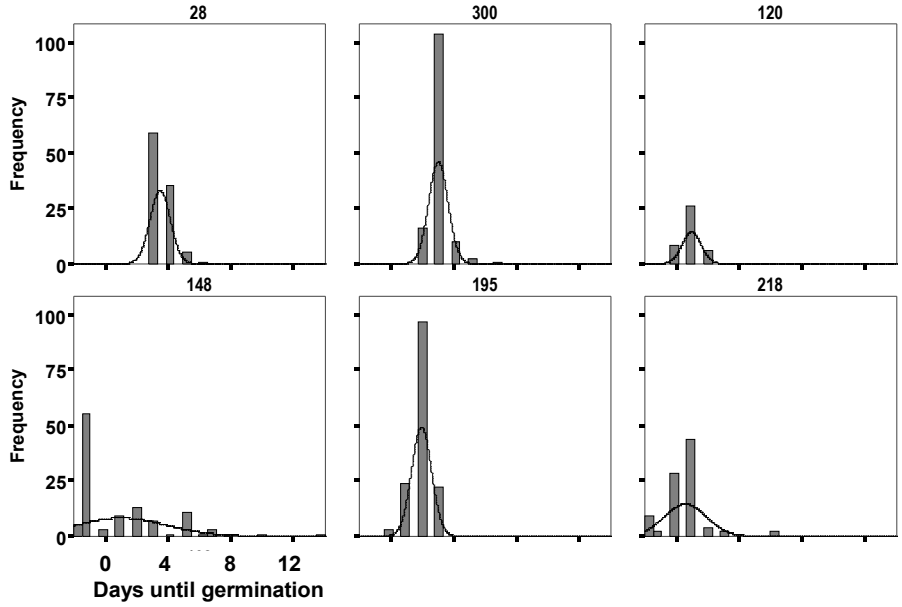


Figure 15. Histograms (with incorporated normal curve) showing the number of days until seed germination per line.

Growth rate, ploidy and genotype. After 34 d of growth, rosette diameters were measured and a growth rate per day was calculated. The accessions have distinct growth rates (Figure 16A) with the slowest growth rate shown by accession 218 (0.13 cm/d) and the fastest seen in accession 120 (0.25 cm/d). A *Turkey test* identified five subgroups, with line 300 and 148 being clustered together, while the remaining lines formed independent groups. From each surviving plant, ploidy measurements were taken and by combining growth rate and ploidy level data it could be shown that an increase of ploidy led to significant decrease of growth rate (Figure 16B). Tetraploid offspring from a diploid mother and hexaploid offspring from triploid mothers always showed a decrease in growth rate (Figure 16B).

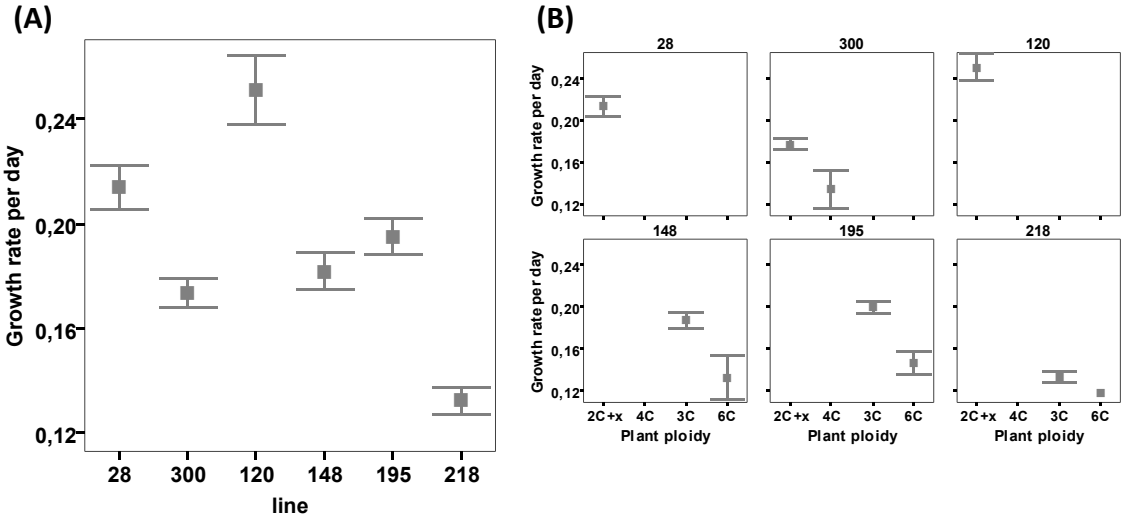


Figure 16. Mean growth rate per accession (A; Error bar mean +/- SE), and (B) offspring growth rate per line and plant ploidy

Furthermore all identified naturally derived polyploidized plants had the same genotype as the mother plants (Table 3, Supplementary data - J), which indicated self-fertilization. All apomictic lines were strictly heterozygous, and no deviations from mother genotype were found. Naturally arisen polyploid plants showed no correlation with seed size group (Supplementary I), confirming what was found earlier during flow cytometric seed screen, which revealed no correlation of embryo ploidy with seed size.

DISCUSSION

In measuring ploidy of surviving plants and calculating their corresponding growth rates, a strong correlation between ploidy and growth rate could be made. Naturally arisen autopolyploid plants grow slower than their parents, illustrating a lack of heterosis, inbreeding depression (Birchler, 2003) and dramatic consequences arising from increased genome content (Arnold and Hodges, 1995; Arnold, 1997; Campbell and Waser, 2001;

Johnston et al., 2001). The *Boechera* autopolyploids identified here were derived from selfing and have the identical genome content as their mother plant, although they differ in allele dosage. Apomictic mother plants are highly heterozygous, a state which can mask recessive deleterious alleles, but after genomic duplication deleterious alleles become homozygous and can potentially influence fitness. This could be a reason for the rare occurrence of tetraploids in nature, because in addition to slower growth, seed output was not as high as in mother plants (personal observation). In other traits, like number of days until germination or seed size, no differences between mother plant and polyploidized offspring were observed.

The capacity of apomictic plants to undergo egg cell fertilization was furthermore identified in measuring embryo and endosperm ploidy of seeds. Our initial assumption had been to find a correlation between seed size and embryo:endosperm ploidy ratio. It was previously shown for *Boechera* that deviations from the expected embryo:endosperm ratio occur (Voigt et al., 2007), and we predicted that seed size would be influenced by differing maternal to paternal genome ratios in endosperm development.

Seed size.

Seeds consist of a seed coat, embryo and endosperm (nourishing tissue). In apomictic seeds, seed coat and embryo represent maternal tissue, whereby the endosperm is the result of maternal and paternal contributions. Seed development in *Arabidopsis thaliana* can be subdivided into two phases i) active proliferation and elongation of endosperm, and ii) embryo growth (Garcia et al., 2005). The final seed size is mainly achieved during the first phase. Endosperm is the most important factor for successful seed development (Dumas and Rogowsky, 2008), and deviations from the two maternal to one paternal genome contribution lead to disturbed endosperm development (Haig and Westoby, 1991; Scott et al., 1998; Dilkes and Comai, 2004). It has been shown that paternal excess in endosperm leads to promoted growth and large seeds, whereby maternal excess leads to repressed growth and small seeds (Haig and Westoby, 1991; Scott et al., 1998; Adams et al., 2000). If the deviation from the balanced endosperm ratio (EBN; Johnstone et al., 1980) between maternal and paternal contributions is too large it has a lethal effect on the embryo (Scott et al., 1998).

An alternative explanation for seed size variation is genomic imprinting and parental conflict (Haig and Westoboy, 1991) during endosperm development. The prediction is that genes allocating resources from mother to offspring, sired by different fathers, will have different expression depending on the parental origin of alleles (maternal to paternal, parent of origin effect). For endosperm repression mainly maternal (m) alleles will be expressed (thereby insuring equal resource allocation to all seeds),

whereby for endosperm activation paternal (p) alleles should act to increase resource allocation to seeds related to oneself (Xiao et al. 2006). We have not tested this hypothesis by performing crosses because *Boechera* propagate mainly by selfing and we wanted to focus in this study on naturally seed formation.

For both karyotypes we identified that the majority of seeds (77.4%) resulted from the expected apomictic pathway involving unfertilized apomeiotically-derived embryos (apomeiosis, parthenogenesis) and fertilization of central cell (pseudogamy). This pathway was equally distributed across seed size classes, indicating no correlation between embryo:endosperm ratios and seed size. Just 15.7% (191 of 1212) of seeds showed an aberrant ratio, and of these 124 cases (10.2%) had incomplete ratios. Together we found more variant ratios within the diploid than triploid apomicts, and in both karyotypes more variant ratios were identified among the small and large seed size groups rather than in medium sized seeds. This points to effects of embryo:endosperm ratios on seed development, but the very low frequencies of unexpected ratios (5.5%) were not sufficient to investigate dosage effects, and can therefore only be used to show tendencies.

To study effects of egg cell fertilization on seed size we compared seeds with 2C:6C and 4C:6C ratios among the offspring of diploids, and those seeds generated via fertilization of egg cell were large (Table 8). Interestingly, this effect can not be found in the offspring of triploids, as both 3C:9C and 6C:9C ratios showed equal distribution among seed size classes (Table 8). Dosage effects, or maternal or paternal genome excess during endosperm formation were identified in triploids, which had autonomous endosperm exclusively in small seeds (Scott et al., 1998). The very low frequency and difficulties in identifying autonomous endosperm formation from the FCSS output profiles make the identification of this phenomenon problematic, as growing embryos by definition have active cells, many of which are in the duplicated 6C stage. Thus distinguishing between 6C embryo and 6C autonomous endosperm cells is challenging, although this can be inferred if endosperm cells are also active, as indicated by a 12C peak (Supplementary data – B). This can explain the high levels of questionable results within the small seed size group of triploid lines in comparison to diploid lines. As with the triploids, autonomous endosperm formation within diploid lines was found among the small and medium seed size, and hence both groups shared the lack of autonomous endosperm formation in large seeds.

Paternal excess was studied in comparing ratios whereby only changes in endosperm ploidy took place (for example between 2C:6C [2m: 4m+2p] and 2C:8C, 2C:10C, or 3C:12C with 3C:9C). The expected ratio 2C:6C and 3C:9C reflected the balanced 2maternal:1paternal endosperm ratio, whereby other ratios leaded to a paternal excess. For diploids paternal excess was characterized by equal contribution of 1maternal :

1paternal (2C:8C [2m: 4m+4p] genomes, which was mainly found in large seeds (Fisher's exact, *P* = 0.024), and a unbalanced contribution of 2m:3p maternal to paternal genomes (2C:10C [2m: 4m+6p(2m:3p)] which was found in a small seeds. This could indicate a ploidy effect as described by Scott at al (1998), but this was not visible among the offspring of triploids, as paternal excess without fertilization of the egg cell (3C:12C) [3m: 6m+6p] was equally identified across seed size classes. If both fertilization of the egg cell and paternal excess in endosperm was found combined, it showed genotype-specificity (Table 9) and indicated opposing effects between diploid and triploid apomicts. In diploids the combination of both were found in small seeds, whereby the combination of both phenomena in triploids were found in large seeds. The very low frequency of this combination makes it difficult to determine the reasons for the observed seed size effects or to make predictions regarding dosage effect on endosperm and embryo development in *Boechera*.

The majority of seeds classified as small, medium or large showed identical ploidy ratios (77.4%) and thus the contribution by the male in endosperm formation is not a critical force for seed size variation. Additionally sources for seed size variation, such as position along inflorescence or silique (fruit) could not be confirmed. Possible consequences of different seed sizes, including bigger seeds providing more reserves for seedlings, and faster germination or more rapid growth rates were not identified, although we can not exclude an advantage in later life stages (de Jong and Klinkhamer, 2005). Seed size itself in *Boechera* is not an indicator for ploidy, as seeds from diploid apomicts can have similar size as those from triploid apomicts.

Apomixis pathway.

The unexpected embryo:endosperm ratios were those which told us the most about the potential of apomictic plants, as they showed that the three components of apomixis in *Boechera* can be flexible in their combination. In triploids we saw rare events of what appeared to be meiotic reduction (ratio 2C:?), but unfortunately the estimation of endosperm ploidy was inconclusive and therefore we cannot differentiate between a haploid egg cell being fertilized by a haploid sperm, or a diploid egg cell which developed parthenogenetically into an embryo. In either case reduction during meiosis occurred, which furthermore points to the possibility that apomicts are still capable of meiotic reduction. It would be interesting to measure how autopolyploid plants (4C or 6C) form seeds, as they potentially could have normal reduction during meiosis I. Pairing (synapsis) and bivalent formation would be possible, whereby in the allo-ploid parents synapsis is disturbed (Kantama et al., 2007). If an autopolyploid (for example, *ABAB*) shows normal reduction during meiosis I leading to reduced egg cell formation (*AB*), it could imply that disturbed meiosis I in apomicts is a consequence of inadequate chromosome pairing

(syngamy). The rare occurrence of tetraploids in nature could support this assumption. Rare tetraploid plants exhibit sexual reproduction (Böcher, 1969; Dobeš et al., 2006), which gives rise to new diploid lines, although F1 tetraploids (but of allotetraploid constitution) were shown to have mixed fertility and were facultative apomictic or sexual (Schranz et al., 2005).

The flexibility of apomixis in *Boechera,* as is evident from embryo:endosperm ploidies furthermore demonstrates that the 3 apomixis components (apomeiosis, parthenogenesis and endosperm fertilization) are unlinked, although one would predict a coupling of apomeiosis and parthenogenesis. Apomeiosis followed by fertilization would generate offspring with increased ploidy which would increase in a stepwise manner each generation, whereby parthenogenesis and normal meiosis would lead to haploid individuals. The unreduced egg cell likely waits for fertilization within a certain developmental time window, and when no fertilization occurs it has the capability for parthenogenetic development, although if it is occasionally fertilized then it behaves like in the sexual pathway. Autonomous endosperm formation was very rarely identified, and more common pseudogamous endosperm formation for the most part involved unreduced pollen sperm cells. Böcher (1951) described, in his cytological studies of pollen development in apomicts, disturbed microsporogenesis which yielded unreduced sperm nuclei. Our data also support his observations, as we found seeds characterized by fertilized embryos and central cells by two diploid sperm nuclei: with a 4C (2m+2p):6C (2m+2m+2p) embryo : endosperm ratio. Böcher (1951) furthermore concluded that the formation of unreduced female gametes is different from the formation of unreduced male gametes, as is supported by the diploid line 120 whose unreduced egg cell and central cell were fertilized by haploid sperm (ratio: 2C:5C [2m: 4m+1p]).

Fertilization of the central cell with both sperm nuclei was also apparently possible. The second highest frequency of embryo:endosperm ratio within diploid apomicts was 2C:8C (24 seeds), and within triploid apomicts 3C:12C (7 seeds). These ratios could be explained by both unreduced sperm nuclei merging with the central cell (2m+2m+2p+2p), indicating a weak barrier for multiple fertilization of the central cell (Spielman et al., 2003), or alternatively by increased sperm ploidy (e.g. 4C or 6C; Voigt et al., 2007 and personal observation). Both diploid and triploid apomicts showed flexibility with regards to combinations of apomixis components, but diploid apomicts were characterized by higher flexibility to variant ploidy ratios. Is this an indication of triploids being more stable and less variable?

Apomixis and Polyploidy.

The genus *Boechera* exhibits one of the rare naturally-occurring cases whereby two different karyotypes express stable asexual (apomictic) reproduction. This system is valuable for the study of ploidy effects on reproduction success and fitness in natural populations. With *Boechera* we have the opportunity to compare diploid versus triploid apomicts, and to furthermore test for effects of mutation accumulation in order to shed light on the intriguing correlation between apomixis and polyploidy. With the increased genome content (i.e. 3 alleles per gene) in triploids we predicted the masking of mutations, as would be manifested by less reproductive variation in triploid *versus* diploid apomicts. The variance in measured traits illustrated for only one trait (*number of aborted seeds*) a difference between diploid and triploids. In all other traits the variance seen is mainly explained by genotype-specificity and is independent of ploidy.

One variance component was inflorescence position, which in other taxa can be correlated with lower fruit set and /or seed set per flower in distal/late-opening flowers compared to basal/early-opening flowers (Medrano et al., 2000). This variation can be attributed to three hypotheses: i) “resource competition” (Stephenson et al., 1981; Lee et al., 1988) whereby the first developed flowers have unlimited resources while later developed flowers have only remaining resources; ii) architectural effect, whereby position determines trait performance (see review Diggle, 2002); and iii) pollen quality and quantity (Vaughton and Ramsey, 1995, 1997; Brunet and Charlesworth, 1995).

In our study we examined four traits (fruit size; ovule number; seed set and seed size) for a positional effect, and found no clear pattern across any line. For fruit set, seed set and seed size, no clear pattern was observed and each line behaved differently (with some intra-lineage variability as well), with some showing position effects while others not (Supplementary data – L). For ovule number we observed a tendency of reduction towards plant distal end across all lines, but since all lines were so different we hypothesize that line-specific responses to environmental variation (i.e. green house conditions) rather than true position effects on the basis of resource competition or architectural effect were the cause. Furthermore, we have also studied pollen quality and quantity (unpublished data) and no correlation with position effect could be made. The only curious observation made among the diploids was the identification of unreduced egg-cell fertilization in 95% (17 of 18 cases) of seeds from basal and middle inflorescence positions, while only one such seed was found in distal position (Supplementary data – k).

Genotype-specificity was identified as major cause of variation for all traits, indicating genotype-specific use of resources. We assumed that all plants had equal amounts of resources and would produce offspring as long as resources were available.

Even though flower number is line-specific (Table 4), we observed no line-specificity in fruit set (Table 4). All lines produced approximately the same percentage of fruit, which is intriguing. An explanation could be if the aborted flowers appeared at the same time due to the absence of necessary resources for a limited time, but this is difficult to measure because the subdivision of the shoot into basal – middle - distal was based upon flower number and not time. Pollination and fertilization of the pistil triggers fruit development, and in our case it seems that fruit development occurs independently of seed development (e.g. the development of seedless fruits is referred to *parthenocarpy, Gustafson,* 1947). In line 120 and through emasculation experiments (data not shown) we observed that fruit/silique elongation started without fertilization of the pistil by pollen. This trait could be very interesting for the production of seedless fruits in agriculture (*i.e.* bananas, tomato, watermelon, grapes and cucumber).

The coefficient of variation (CV) was used to test phenotype variation within a genotype, as we tested five replicates per genotype (clones of mother plant). The offspring of an apomictic plant are expected to be genetical clones, and hence the same genotype should lead to the same phenotype in the same environment. We tested for three traits (Figure 9) and expected lower within genotype variation for ovule number, because they are maternally identical and are exposed to the same environmental conditions. For the traits seed set and seed size we expected to see higher variation, because even though the replicates are genetically identical within a genotype, the trait's phenotype depends on male contribution, which could be influenced by pollen quality and quantity (irregular pollen formation, Böcher, 1951; Voigt et al., 2007).

The mean number of ovules was different between diploids and triploids, although the variation is mainly explained by line-specificity. Diploids had more ovules within a silique (fruit) than triploids, although one triploid line (ID 148) contained as many ovules as the diploid line 120. Lower ovule number in triploids was found by earlier studies as well (Voigt et al., 2007). Ovules are female reproductive organs, containing the megaspore mother cell, which is determined to form the female gametophyte. Pistils of diploids contain one third more ovules, inferring an advantage over triploids in offspring output.

Interestingly, seed set was below 40% in all apomictic lines. In most apomictic lines of *Boechera,* pollen fertilization is necessary for proper endosperm formation (Voigt et al., 2007) and therefore seed development. In coding the seeds into i) developed, ii) aborted and iii) undeveloped ovules, we found a higher rate of aborted seeds among the diploid apomicts ($P < 0.03$). This can be interpreted in two different ways: i) either those ovules started autonomous development without prior fertilization by a sperm nucleus, which later resulted in disturbed endosperm formation and missing signals between the

developing embryo and endosperm (Ungru et al., 2008); or ii) a sperm nucleus of inadequate ploidy fertilized the central cell which yielded in a shift in the necessary maternal to paternal genome contribution (Scott et al., 1998) resulting in seed abortion. At this stage we can not distinguish between either scenario, but we have shown that apomictic *Boechera* produce pollen of different sperm ploidy (Voigt et al., 2007).

The aborted seeds in line ID 28 and ID 300 were more frequent in the distal and basal parts of siliques. This pattern was independent of plant position, and thus likely represents a shift between female and male gametophyte development. Within a pistil, basally-developed ovules are advanced in development in comparison to distal ones (Christensen et al., 1997). When mature pollen is released it fertilizes mature ovules, with a prevalence for the most easily reachable ovules (Hulskamp et al., 1995). This could mean that pollen tubes prefer closer ovules, and thus basally-developed ovules start autonomous development which leads to abortion. Furthermore, distal ovules which matured later may be less successfully fertilized due to lower pollen availability. In our calculations of seed output per inflorescence, differences between lines were visible, and were correlated with seed abortion (line 28 and 300), seed size (line 148) and the amount of lateral inflorescences (bushier appearance of triploids, Supplementary data – A).

Considering selected reproductive traits across one generation for diploid and triploid plants, a significant difference between both karyotypes was found only for *numbers of aborted seeds*. All other traits we examined showed variance which was both genotype-specific and independent from ploidy. The Mantel test indicated independence with respect to morphology and genetic relatedness, which supports the idea that independent evolution of different clonal lineages, their different genetic backgrounds, mutation accumulation and environmental conditions are stronger forces acting upon phenotype modulation than ploidy. It is conceivable that apomicts which have established themselves in a population display variation reflective of natural selection acting continuously on several traits (Bretagnolle and Lumaret, 1995).

Apomixis, Polyploidy and Hybridization.

That rarely any trait showed dependency on ploidy, or that we did not identify less variation among triploids was surprising, and leaves the original question unanswered: why we find the high correlation of apomixis with polyploidy? If ploidy does not matter, why do we not find more cases of diploid apomixis? Perhaps *Boechera* is unique with respect to most other apomictic taxa. For example, the predominance of triploid apomicts in *Boechera* is interesting. Triploids are considered to play a roll in tetraploid evolution (i.e. triploid bridge), but are considered to be evolutionary dead ends (Bretagnolle and Thompson, 1996; Ramsey and Schemske, 1998). One has to distinguish between newly

arisen triploids and unstable meiosis, which can function as intermediates to produce tetraploids (Ramsey and Schemske, 1998), and established triploids with either asexual reproduction or with specialized meiosis producing diploid sperm and haploid egg (Normann and Quarin, 1987) or haploid sperm and diploid egg (Stock et al., 2002). The triploids of *Boechera* are reproductively stable due to apomixis, which allows bypassing of the triploid block (van Dijk et al., 2004; Verduijin et al., 2004; Saura et al., 1993). The predominance of diploid apomicts in *Boechera* is curious, as diploid organisms typically reproduce sexually. In *Boechera* diploid sexual plants are selfers, which is reflected in high homozygosity (Song et al., 2006), whereas diploid and triploid apomicts displayed high heterozygousity.

Elevated heterozygosity in apomicts is a consequence of hybridization and accumulation of mutation. Both diploids and triploids are hybrid in origin (Böcher, 1951; Rollins, 1983; Sharbel and Mitchell-Olds, 2001; Koch et al., 2003; Dobeš et al., 2004a, b, Kantama et al., 2007), and thus both are allo-ploid. The best studied hybrid is *B. divaricarpa*, which includes plants where *B. stricta* is one of the parents (Al-Shehbaz, 2007b). The occurrence of hybridization is apparent in diploid apomicts, as their chromosome set is a combination of *B. stricta* - and *B. holboellii* - like chromosomes (Kantama et al., 2007). Furthermore, the expected balance of seven *B. stricta* and seven *B. holboellii*-like chromosomes (as expected for an F1 hybrid) is disturbed, showing plants for example with 4 *B. stricta* and 10 *B. holboellii*-like chromosomes, thus indicating gene flow between lines or meiotic chromosome segregation in apomictic lineages. Triploid apomicts are hypothesized to have arisen multiple times from diploid ancestors through hybridization (Böcher, 1951; Rollins, 1983; Sharbel and Mitchell-Olds, 2001; Koch et al., 2003; Dobeš et al., 2004a, b). Kantama et al. (2007) concluded, based on their observation of *B. stricta* generating low frequencies of male 2n gametes, that this could be the source of allotriploids or allotetraploids. They proposed that the increase in genome content and the allo-ploid chromosome constitution resulted in genomic stress, in which modified gene expression gave rise to apomixis which may subsequently take several generations before stabilization (Kantama et al., 2007).

In our study we expected that the use of lines with different genetic backgrounds would enable the identification of differences based upon ploidy in order to test the correlation between polyploidy and asexuality, but we instead identified genotype-specificity. We suspect that heterosis (i.e. hybrids exhibit characteristics that lie outside the range of parents; Shull, 1908) likely obscured the phenotypic changes involved with the switch from sexual to asexual reproduction. Considering that hybridization between two distinct sexual species, as seen between *B. stricta* (A) and *B. holboellii* (B) to generate the apomict *B. divaricarpa* (AB), associated meiotic disturbances could have somehow led

to a positive fitness advantage of apomeiosis. Parthenogenesis and endosperm formation are flexible in diploids (Tables 9) and thus the critical factor for stable apomixis expression is likely the establishment of apomeiosis. For example in *Arabidopsis thaliana* one mutation has been shown to result in dyad formation during egg cell development (2n; albeit at very low frequencies; Ravi et al., 2008). Furthermore, a mutation in the gene *SWI*, called DYAD (Agashe et al., 2002), causes synapsis to fail in female meiosis, but this effect depends on the presence of two DYAD copies. The egg cell of DYAD mutants can be fertilized by haploid sperm to generate triploid offspring, although the reproductive mode of the F1 triploids was not reported.

Boechera and Apomixis.

Assuming that apomixis has evolved repeatedly from different sexual ancestors in *Boechera* (Sharbel and Mitchell-Olds, Sharbel et al 2004, 2005), and that the three apomixis components result from expression changes in different genes, it is difficult to imagine a single mutation (i.e. apomixis factor) inducing this trait on a broad geographic scale. The genus *Boechera* went through a complex history of hybridization and radiation in the last 1.8 mya (Boecher, 1951; Rollins, 1993; Mulligan, 1995; Sharbel and Mitchell-Olds, 2001; Dobeš et al., 2004b; Roy, 1995), and thus one possibility is that a neutral mutation from a *Boechera* ancestor, or a "factor" which is either recessive or expressed without obvious effect on phenotype, could somehow lead to apomixis expression in hybrids.

When two *Boechera* species hybridize (e.g. *Boechera stricta* can be fertilized by pollen from many species; Schranz at al., 2005), such a "factor" would find itself in an allo-ploid surrounding, which somehow could affect meiosis. How could this happen? The alloploid constitution could result in two different initiating effects for apomeiosis, one mechanism effect and one effecting gene regulation. The mechanism effect could be that the physical pairing of homolog chromosomes is essential for normal gene expression patterns (e.g. *transvection* - pairing dependent gene expression, Grant-Downton and Dicknson, 2004). In apomictic *Boechera* chromosome condensation takes place (leptotene, first stage of prophase I of meiosis I) which indicates signalling pathways for meiosis are triggered, although the following stages have not yet been identified (Naumova, 2001). The unsuccessful formation of bivalents and disturbed homolog chromosome pairing could, as a consequence, disrupt normal gene expression and gene regulation processes and lead to unreduced dyad formation during meiosis. In *Drosophila* it was shown that chromosomal rearrangements can disrupt pairing at a locus to disturb its normal transcription (Wu and Morris 1999; Duncan 2002).

Alternatively, it could be that the "apomixis factor" is a normally-expressed gene in the sexuals parents, and which becomes asynchronously expressed or misregulated in allo-ploids to induce apomixis (Carman, 1997). Deregulation or asynchrony can be influenced by cis-trans interactions (Landry et al., 2005) and/or epigenetic change through changes in DNA methylation and chromatin structure (Koltunow and Grossniklaus, 2003). Recent transcriptomic comparisons between apomictic and sexual *Boechera* have identified heterochronic expression of genes, differential gene expression, and a parent-of origin effect (Sharbel et al., 2009). The identified parent of origin effect was interesting, as it was shown that *B. holboelli* specific-alleles were up and down regulated in the apomicts, whereas *B. stricta* specific-alleles were mainly down-regulated in the apomicts (Sharbel et al., 2009), thus indicating *trans* acting elements as potential factors associated with apomeiosis. Both *trans* acting elements and somatic chromosome pairing can have overlapping effects. It has been shown in nematodes that imprinting of X chromosomes depends on pairing; if no pairing partner is available unpaired DNA is targeted for methylation of histone H3 at Lys9 (H3-Lys9) and can be silenced (Bean et al., 2003). It would be interesting for further studies of apomixis expression to keep both processes, although difficult to separate, in mind.

Origin and establishment of triploids.

For the origin of the triploid *Boechera* several hypothesis are proposed: i) a sexual diploid *Boechera* forming 2n (*AA*) gametes which fused with haploid gamete (*B*) (Schranz et al., 2005; Kantama et al., 2007), ii) sexual autotetraploids *Boechera* (originated from diploid apomict) produce reduced 2n (*AB*) gametes which fuse with 1n gamete (C) (Schranz et al., 2005; Dobeš et al., 2006), and iii) unreduced eggs cells from diploid apomicts (*AB*) were fertilized by haploid sperm (*C*) (Schranz et al., 2005). Hypothesis 1 is possible, but the limitation is that both parental taxa are sexual and the resulting triploid (*AAB*) is most likely sterile (Schranz 2005). In this scenario the developed F1 triploid has to overcome the constraints of hybridization (allo-ploidy) and polyploidization (difficulties during meiosis, triploid block, Bretagnolle and Thompson, 1996; Ramsey and Schemske, 1998). While possible, the second and third scenarios seem more likely for developing triploids, although the sexual reduction in autotetraploids needs to be confirmed. In these cases, the proposed triploid (*ABC*) offspring has the apomeiosis "factor" transmitted from one parent and therefore the triploid block can be circumvent (Schranz et al., 2005), and influences of allo-ploidy could be less effective, because one genome contribution (*AB*) has already overcome the constrains of hybridization.

The analysed lines were chosen from different populations and geographic areas, and thus correlations between ploidy and apomixis may be obscured by the independent accumulation of mutation, different ages of apomictic lineages, and differing genetic

backgrounds of the original parents during hybridization (multiple hybidization events, Boecher, 1951; Rollins, 1993; Mulligan, 1995; Sharbel and Mitchell-Olds, 2001; Dobeš et al., 2004b; Roy, 1995). To test for effects of ploidy on apomictic reproduction, diploid and triploids of one population should in the future be studied, as natural selection acting upon novel variation in several traits could allow their establishment in specific niches and fainted true effects of ploidy on reproduction success.

ACKNOWLEDGEMENTS

We would like to thank H. Block, V. Katzorke and L. Börner for technical assistance and Tom Mitchell-Olds for providing seed material.

Contribution: Ideas for project and performance by MLV and MP; Mantel test and nested ANOVA was performed mainly by MP; writing of the manuscript by MLV, with critical comments by TFS and MP.

LITERATURE CITED

Adams S, Vinkenoog R, Spielman M, Dickinson HG, and Scott RJ. 2000. Parent-of-origin effects on seed development in Arabidopsis thaliana require DNA methylation. Development 127: 2493-2502.

Agashe B, Prasad CK, and Siddiqi I. 2002. Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in *Arabidopsis*. Development 129: 3935-3943.

Al-Shehbaz IA, and Windham MD. 2006. New and noteworthy Species of Boechera (Brassicaceae) I: Sexual diploids. Harvard Papers in Botany 11: 61–88.

Al-Shehbaz IA, and Windham MD. 2007a. New and noteworthy Species of BOECHERA (BRASSICACEAE) II: Apomictic hybrids. Harvard Papers in Botany 11: 257–274.

Al-Shehbaz IA, and Windham MD. 2007b. New and Nothworthy species of BOECHERA (BRASSICACEAE) III: Additional sexual diploids and apomictic hybrids. Harvard Papers in Botany 12: 235–257.

Arnold ML, and Hodges SA. 1995. Are natural hybrids fit or unfit relative to their parents? Trends in Ecology and Evolution 10 pp. 67–71.

Asker S, and Jerling L. 1992. Apomixis in Plants. CRC Press, Boca Raton, FL.

Barton NH, and Charlesworth B. 1998. Why sex and recombination? Science (New York, N.Y 281: 1986-1990.

Bean CJ, Schaner CE, and Kelly WG. 2004. Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. Nature Genetics 36: 100-105.

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

Belshaw R, and Quicke DL. 2003. The cytogenetics of thelytoky in a predominantly asexual parasitoid wasp with covert sex. Genome 46: 170-173.

Beukeboom LW, and Vrijenhoek RC. 1998. Evolutionary genetics and ecology of sperm dependent parthenogenesis. Jornal of evolutionary Biology 11: 755-782.

Bicknell RA, Borst NK, and Koltunow AM. 2000. Monogenic inheritance of apomixis in two Hieracium species with distinct developmental mechanisms. Heredity 84 (Pt 2): 228-237.

Birchler JA, Bhadra U, Bhadra MP, and Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Developmental Biology 234: 275-288.

Bleeker W, and Matthies A. 2005. Hybrid zones between invasive Rorippa austriaca and native R. sylvestris (Brassicaceae) in Germany: ploidy levels and patterns of fitness in the field. Heredity 94: 664-670.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Det Kongelige Danske Videnskabernes Selskab 6: 1-59.

Böcher TW. 1954. Experimental taxonomical studies in the *Arabis holboellii* complex. Svensk Botanisk Tidskrift 48: 31-44.

Böcher TW. 1969. Further studies in Arabis holboellii and allied species. Sacrtryk Bot. Tidsskr. 64: 141-161.

Bretagnolle F, and Lumaret R. 1995. Bilateral polyploidization in *Dactylis glomerata*, L. ssp. *Lusitanica*: occurence, morphological and genetic characteristics of first polyploids. Euphytica 84: 197–207.

Brunet J, and Charlesworth D. 1995. Floral sex allocation in sequentially blooming plants. Evolution 49: 70–79.

Campbell DR, and Waser NM. 2001. Genotype by environment interaction and the fitness of plant hybrids in the wild. Evolution 55: 669-676.

Carman JG. 1997. Asynchromous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biological Journal of the Linnean Society 61: 51-94.

Carputo D. 1999. Post-zygotic gametic selection due to endosperm balance number explains unusual chromosome numbers of 3× × 2× progeny in Solanum. Sexual Plant Reproduction 12: 27-31.

Carputo D, Frusciante L, and Peloquin SJ. 2003. The role of 2n gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing Solanums. Genetics 163: 287-294.

Carputo D, Monti L, Werner JE, and Fursciante L. 1999. Uses and usefulness of endosperm balance number. Theoretical Applied Genetics 98: 478-484.

Charnov EL. 1982. The theory of sex allocation. Monographs in Population Biology 18:1-355.

Chessel D, Dufour AB and Thioulouse J. 2004. The ade4 package-I- One-table methods. R News 4: 5-10.

Cheverud JM, Wagner GP, and Dow MM. 1989. Methods for the comparative analysis of variaion patterns. Systematic Zoology 38: 201-213.

Clauss MJ, Cobban H, and Mitchell-Olds T. 2002. Cross-species microsatellite markers for elucidating population genetic structure in Arabidopsis and Arabis (Brassicaeae). Moleculare Ecology 11: 591-601.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews 6: 836-846.

Corral JM, Piwczynski M, and Sharbel TF. 2009. Allelic sequence divergence in the apomictic *Boechera holboellii* complex. In Lost Sex, the Evolutionary Biology of Parthenogenetic Reproduction. (Schön I, Martens K and Van Dijk PJ. eds). Berlin: Springer Verlag, p. In press.

Christensen CA, King EJ, Jordan JR, and Drews GN. 1997. Megagametogenesis in *Arabidopsis* wild type and the *Gf* mutant. Sexual Plant Reproduction 10: 49-64.

De Jong T, and Klinkhamer P. 2005. Evolutionary Ecology of plant reproduction strategies.Campride University Press. Pp 110 – 131.

Diggle PK. 1995. Architectural effects and the interpretation of patterns of fruit and seed development. Annual Review of Ecology and Systematics. 26: 531–552.

Dilkes BP, and Comai L. 2004. A differential dosage hypothesis for parental effects in seed development. The Plant Cell 16: 3174-3180.

Dobeš C, Mitchell-Olds T, and Koch MA. 2004a. Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera xdivaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers--an integrative approach. American Journal of Botany 91: 2087-2101.

Dobeš C, Mitchell-Olds T, and Koch MA. 2004b. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* x*divericarpa*, and *A. holboellii* (Brassicaceae). Molecular ecology. 13: 349-370.

Dobeš C, Koch MA, Sharbel T. 2006. Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Ann. Miss. Bot. Gard. 93: 517–533.

Dorn RD. 1984. Vascular Plants of Montana. Mountain West Publishing, Cheyenne.

D'Souza TG, Storhas M, Schulenburg H, Beukeboom LW, and Michiels NK. 2004. Occasional sex in an 'asexual' polyploid hermaphrodite. Proc Biol Sci. 271: 1001-1007.

Ducan IW. 2002. Transvection effects in Drosophila. Annual Review of Genetics 36: 521-556.

Dumas C, and Rogowsky P. 2008. Fertilization and early seed formation. Comptes rendus biologies 331: 715-725.

Garcia D, Fitz Gerald JN, and Berger F. 2005. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. The Plant Cell 17: 52-60.

Grant-Downton RT, and Dickinson HG. 2004. Plants, pairing and phenotypes--two's company? Trends in Genetics 20: 188-195.

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

Grimanelli D, Garcia M, Kaszas E, Perotti E, and Leblanc O. 2003. Heterochronic Expression of Sexual Reproductive Programs During Apomictic Development in Tripsacum. Genetics 165: 1521-1531.

Grossniklaus U, Nogler A, and van Dijk PJ. 2001a. How to avoid sex: the genetic control of gametophytic apomixis. The Plant Cell 13: 1491–1497.

Guldbrandtsen B, Tomiuk J, and Loeschcke V. 2000. POPDIST, version 1.1.1: a program to calculate population genetic distance and identity measures. Journal of Heredity 91:178-179.

Haag CR, and Roze D. 2007. Genetic load in sexual and asexual diploids: segregation, dominance and genetic drift. Genetics 176: 1663-1678.

Haig D, and Westoby M. 1991. Genomic imprinting in endosperm: its effects on seed development in crosses between different ploidies of the same species, and its implications for the evolution of apomixis. Philosophical Transactions of the Royal Society B 333: 1–13.

Halkett F, Simon JC, and Balloux F. 2005. Tackling the population genetics of clonal and partially clonal organisms. Trends in Ecology & Evolution 20: 194.

Hulskamp M, Schneitz K, and Pruitt RE. 1995. Genetic Evidence for a Long-Range Activity That Directs Pollen Tube Guidance in Arabidopsis. The Plant Cell 7: 57-64.

Janko K, Bohlen J, Lamatsch D, Flajshans M, Epplen JT, Rab P, Kotlik P, and Slechtova V. 2007. The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis: Teleostei), and their ability to establish successful clonal lineages--on the evolution of polyploidy in asexual vertebrates. Genetica 131: 185-194.

Johnston SA, den Nijs TPM, Peloquin, SJ, and Hanneman Jr. RE. 1980. The significance of genic balance to endosperm development in inter-specific crosses. Theoretical. Applied Genetics. 56, pp. 293–297.

Kantama L, Sharbel TF, Schranz ME, Mitchell-Olds T, de Vries S, and de Jong H. 2007. Diploid apomicts of the *Boechera holboelli* complex display large scale chromosome substitutions and aberrant chromosomes. Proceedings of the National Academy of Sciences of the United States of America 104: 14026-14031.

Khakhlova O, and Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. The Plant Journal 46: 85-94.

Kirk H, Vrieling K, and Klinkhamer PG. 2005. Maternal effects and heterosis influence the fitness of plant hybrids. The New Phytologist 166: 685-694.

Koch MA, Dobeš C, and Mitchell-Olds T. 2003. Multiple Hybrid Formation in Natural Populations: Concerted Evolution of the Internal Transcribed Spacer of Nuclear Ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molecular Biology and Evolution 20: 338-350.

Koltunow AM, and Grossniklaus U. 2003. Apomixis: a developmental perspective. Annual Review of Plant Biology 54: 547-574.

Kondrashov AS. 1988. Deleterious mutation and the evolution of sexual reproduction. Nature 336.

Kondrashov AS. 1993. Classification of hypotheses on the advantage of amphimixis. The Journal of Heredity 84:372-386.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, and Hartl DL. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics 171: 1813-1822.

Levin DA. 1983. Polyploidy and novelty in flowering plants. American Naturalist 122:1–25.

Lewis WM. 1987. The cost of sex. Pp 33-57 in: Stearns, S. C. (ed), The Evolution of Sex and its Consequences. Birkhäuser Verlag, Basel.

Mahony MJ, and Robinson ES. 1980. Polyploidy in the Australian leptodactylid frog genus *Neobatrachus*. Chromosoma 81: 199–212

Mantel NA .1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209–220.

Matzk F, Meister A, and Schubert I. 2000. An efficient screen for reproductive pathways using mature seeds of monocots and dicots. The Plant Journal 21: 97-108.

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

Medrano M, Guitian P, and Guitian J. 2000. Patterns of fruit and seed set within inflorescence of Pancratium maritimum (Amaryllidaceae): nonuniform pollination, resource limitation, or architectural effects? American Journal of Botany 87(4): 493-501.

Mogie M, Britton NF, and Stewart-Cox JA. 2007. Asexuality, polyploidy and the male function.A. R. G. Ganter Verlag K.G., Vienna.

Mulligan GA. 1995. Synopsis of the genus *Arabis* (Brassicaceae) in Canada, Alaska and Greenland. Rhodora 97: 109-163.

Naumova TN, van der Laak J, Osadtchiy J, Matzk F, Kravtchenko A, Bergervoet J, Ramulu KS, and Boutilier K. 2001. Reproductive development in apomictic populations of Arabis holboellii (Brassicaceae). Sexual Plant Reproduction 14: 195-200.

Nogler GA. 1984a. Gametophytic apomixis. Pp. 475-518 *in* B. M. Johri, ed. Embryology of Angiosperms. Springer Verlag, New York.

Normann GA, and Quarin CL. 1987. Permanent odd polyploidy in a grass (*Andropogon ternatus*). Genome 29: 340-344.

Otto SP, and Whitton J. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34: 401-437.

Otto SP, and Gerstein AC. 2006. Why have sex? The population genetics of sex and recombination. Biochemical Society transactions 34: 519-522.

Page RDM. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358.

Paun O, Stuessy TF, and Horandl E. 2006. The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic Ranunculus cassubicus complex. The New Phytologist 171: 223-236.

Quarin CL. 1999. Effects of pollen source and pollen ploidy on endosperm formation and seed set in pseudogamous apomictic *Paspalum notatum*. Sexual Plant Reproduction 11: 331–335.

Quintanilla LG, and Escudero A. 2006. Spore fitness components do not differ between diploid and allotetraploid species of Dryopteris (Dryopteridaceae). Annals of Botany 98: 609-618.

R Development Core Team 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Ramsey J, and Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29: 467-501.

Ramsey J, and Schemske DW. 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33: 589-639.

Ravi M, Marimuthu MP, and Siddiqi I. 2008. Gamete formation without meiosis in Arabidopsis. Nature 451: 1121-1124.

Roy BA. 1995. The breeding systems of six species of Arabis (Brassicaceae). American Journal of Botany 82: 869-877.

Rollins RC. 1981. Studies on Arabis (Cruciferae) of Western North America. Systematic Botany 6: 55-64.

Rollins RC. 1993. The Cruciferae of continental North America. Stanford University Press. Stanford, California. P. 976.

Santos JL, Alfaro D, Sanchez-Moran E, Armstrong SJ, Franklin FC, and Jones GH. 2003. Partial diploidization of meiosis in autotetraploid Arabidopsis thaliana. Genetics 165: 1533-1540.

Savidan Y. 2007. Apomixis in higher plants. A.R.G. Gantner Verlag K.G., Vienna.

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

Schulz RJ. 1969. Hybridization, unisexuality and polyploidy in the teleost *Poeciliopsis* (Poeciliidae) and other vertebrates. Am. Nat**.** 103: 605-619.

Schön I, and Martens K. 1998. DNA repair in an ancient asexual: a new solution for an old problem? Journal of Natural History 32: 943-948.

Scott RJ, Spielman M, Bailey J and Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development 125: 3329-3341.

Sharbel TF, and Mitchell-Olds T. 2001. Recurrent polyploid origins and chloroplast phylogeography in the Arabis holboellii complex (Brassicaceae). Heredity 87: 59-68.

Sharbel TF, Mitchell-Olds T, Dobeš C, Kantama L, and de Jong H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic and Genome Research 109: 283-292.

Sharbel TF, Voigt ML, Corral JM, 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 in press.

Shaw RG, and Mitchell-Olds T. 1993. ANOVA for unbalanced data: an overview. Ecology 74: 1638-1645.

Shull GH. 1908. The composition of a field of maize. Am. Breed. Assoc. Rep. 4: 296–3 01.

Soltis DE, and Soltis PS. 1993. Molecular data and the dynamic nature of polyploidy. Critical Reviews in Plant Sciences 12: 243-273.

Soltis DE, and Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology and Evolution 14:348-352.

Song BH, Clauss MJ, Pepper A, and Mitchell-Olds T. 2006. Geographic patterns of microsatellite variation in Boechera stricta, a close relative of Arabidopsis. Moleculare Ecology 15: 357-369.

Spielman M, Vinkenoog R, and Scott R J. 2003. Genetic mechanisms of apomixis. Philosophical Transactions of the Royal Society B 358: 1095-1103.

Stephenson AG. 1981. Flower and fruit abortion: proximate causes and ultimate functios. Annual Review of Ecology and Systematics 12: 253-279.

Stock M, Lamatsch DK, Steinlein C, Epplen JT, Grosse WR, Hock R, Klapperstuck T, Lampert KP, Scheer U, Schmid M, and Schartl M..2002. A bisexually reproducing all-triploid vertebrate. Nature Genetics 30: 325-328.

Tas ICQ, and van Dijk PJ. 1999. Crosses between sexuel and apomictic dandelions (*Taraxacum*). I. The inheritance of apomixis. Heredity 83: 707-714.

Tomiuk J, and Loeschcke V. 1991. A new measure of genetic identity between populations of sexual and asexual species. Evolution; International Journal of Organic Evolution 45: 1685-1694.

Tomiuk J, and Loeschcke V. 1996. A maximum-likelihood estimator of the genetic identity between polyploid species. Journal of Theoretical Biology 179: 51-54.

Ungru A, Nowack MK, Reymond M, Shirzadi R, Kumar M, Biewers S, Grini PE, and Schnittger A. 2008. Natural variation in the degree of autonomous endosperm formation reveals independence and constraints of embryo growth during seed development in Arabidopsis thaliana. Genetics 179: 829-841.

Vaughton G, and Ramsey M. 1995 Pollinators and seed production. pp. 475-490. In: J. Kigel and G. Galili (eds), Seed development and germination. Marcel Dekker, NY.

Vaughton G, and Ramsey M. 1997 Seed mass variation in the shrub *Banksia spinulosa* (Proteaceae): resource constraints and pollination effects. International Journal of Plant Sciences 158: 424-431.

Verduijn MH, Van Dijk PJ, and Van Damme JM. 2004. The role of tetraploids in the sexual-asexual cycl e in dandelions (Taraxacum). Heredity 93: 390-398.

Van Dijk PJ. 2007. Potential and realized costs of sex in dandelions, *Taraxacum officinale* s.l., A.R.G. Gantner Verlag K.G., Vienna.

van Dijk PJ, and Bakx-Schotman JMT. 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.

Voigt ML, Melzer M, Rutten T, Mitchell-Olds T, Sharbel TF. 2007. Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. A.R.G. Gantner Verlag K.G., Vienna.

West SA, Lively CM and Read AF. 1999. A pluralistic approach to sex and recombination. J. Evol. Biol. 12: 1003-1012.

Williams E. (1975): Meiotic chromosome pairing in interspecific hybrids of *Nicotiana* I. Comparative pairing in interspecific hybrids of section Suaveolentes. N. Z. J. Bot. 13: 601–609.

Wolfe LM, and Denton W. 2001. Morphological constraints on fruit size in *Linaria canadensis.* International Journal of Plant Sciences 162: 335-342.

Wu CT, and Morris JR. 1999. Transvection and other homology effects. Current Opinion on Gentetics Dev. 9: 237-246.

Xiao W, Custard KD, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, and Fischera RL. 2006. DNA Methylation Is Critical for Arabidopsis Embryogenesis and Seed Viability. The Plant Cell 18: 805-814.

Supplementary data

A) Plant pictures.

B) FCSS profiles.

C) Nested ANOVA, model and randomization test.

D) Mantel test.

E) Summary of collected data for reproduction structure size per line.

F) Paired sample t-test, comparison main/lateral shoot.

G) Summary of collected data for fecundity and fertility.

H) Summary of collected data for germination and plant position under seed size groups.

I) Seed size group and performance trait (Chi square test).

J) Plant ploidy of survived plants.

K) Summary of collected data in FCSS measurements and plant position.

L) Summary of behavior of single plants/replicates per line and trait.

Molecular signatures of apomictic and sexual ovules in the *Boechera holboellii* complex

Timothy F. Sharbel, **Marie-Luise Voigt**, José María Corral, Thomas Thiel, Alok Varshney, Jochen Kumlehn, Heiko Vogel & Björn Rotter

In press: The Plant Journal
http://www3.interscience.wiley.com/cgi-bin/fulltext/122193058/PDFSTART

Abstract

Apomixis, a natural form of asexual seed production in plants, has evolved independently in various taxa and represents an important potential technology for agriculture. The switch to apomixis is based on de-regulation of developmental pathways originally leading to sexual seed formation. Hybridization and polyploidy, both typical characteristics of asexual plants and animals, are mechanisms that could trigger de-regulation. Here we show that up-regulation of alleles in apomeiotic ovules is mediated by genomic duplication, heterochrony and the residual effects of ancient hybridization in diploid apomicts of the *Boechera holboellii* complex. Using SuperSAGE, we have identified over 4000 differentially expressed mRNA tags between the micro-dissected ovules from 2 diploid sexual (*Boechera stricta* and *B. holboellii*) and two diploid apomictic (*Boechera divaricarpa*) accessions. Pairwise sequence comparisons between tags enabled identification of allelic variants of the same loci. Up-regulated candidate apomeiosis alleles consistently have more than three related alleles, thus demonstrating transcription from duplicated loci. A further 543 alleles were heterochronically expressed between sexual and apomeiotic ovules at developmental stages 2-II to 2-IV. Intriguingly, 69 *B. holboellii* specific alleles were preferentially up-regulated in apomeiotic ovules, thus showing a remnant 'parent of origin' effect stemming from Pleistocene origin of the hybrid *B. divaricarpa* from taxa related to *B. holboellii* and *B. stricta*. These data implicate polyploid gene dosage in the expression of asexual seed formation, and support hypotheses of de-regulation of the sexual pathway. The observed 'parent of origin' effect suggests that the genomic memory of hybridization has somehow been maintained after hundreds, if not thousands, of asexual generations.

KEY WORDS: apomixis, *Boechera*, hybridization, gene duplication, gene expression, parent of origin effect

INTRODUCTION

The origin and persistence of asexual reproduction, despite mutation accumulation and limited genetic variability, remain one of the most challenging phenomena in evolutionary biology (Bell, 1982). A comparison of asexual animals and plants demonstrates that, despite widely differing reproductive mechanisms, a number of characteristics are frequently found in common, the most obvious of which include hybrid origin and polyploidy (Richards, 2003). Thus, while asexuality across these large phylogenetic distances has been attained independently, the natural selective forces which have shaped its evolution have led convergently to similar variation on the nuclear level.

Asexual animals (parthenogenetic) and plants (apomictic) have arisen repeatedly from sexual ancestors, and it remains unclear how this switch in reproductive mode occurs at the molecular genetic level, or how polyploidy and hybridity contribute to asexual lineage origin and its subsequent evolution in the absence of normal meiosis. For example, polyploidy could both initiate and maintain asexuality through gene dosage effects (Grimanelli et al., 2001), and may additionally provide lineage stability by buffering deleterious mutations in gametes and somatic cells (Richards, 2003). Moreover, hybridization could change regulatory gene expression patterns to induce asexuality (Carman, 1997), or it may have positive fitness effects through fixed heterosis in asexual lineages (Kearney, 2005). Clearly both hybridization and polyploidy could induce the global regulatory changes needed to initiate asexuality (Comai et al., 2003; Osborn et al., 2003), but disentangling their relative contributions has been difficult.

Among plants, gametophytic apomixis is a naturally-occurring form of asexual seed formation producing progeny that are genetically identical to the maternal genotype from meiotically unreduced embryo sacs (Koltunow and Grossniklaus, 2003). The majority of gametophytic apomictic species is found in the *Asteraceae*, *Rosaceae* and *Poaceae*, where they have arisen independently and recurrently (Grossniklaus et al., 2001). Nevertheless, polyploidy, facultative apomixis (both sexual and apomictic seed production within one individual), and faster development of the apomeiotic ovule relative to the sexual one (Savidan, 2007) are traits that are shared among many of these taxa. Compared to parthenogenesis in animals, apomixis has been the subject of deeper analyses into its developmental mechanisms, because of its potential importance as an enabling technology for agriculture. Harnessing apomixis would greatly facilitate and accelerate the ability of plant breeders to fix and faithfully propagate genetic heterozygosity and associated hybrid vigor in crop plants (Spillane et al., 2004).

Although exceptions do exist, three independent developmental steps must be acquired by an amphimictic (i.e. sexual) plant to produce seeds apomictically: formation of an unreduced embryo sac, e.g. through meiotically unreduced megaspore formation (apomeiosis), development of an embryo from an unfertilized and unreduced egg cell (parthenogenesis), and formation of functional endosperm, e.g. fertilization of the binucleate central cell (pseudogamy). The apomeiotically derived embryo thus receives its entire genome through the female line (although paternal apomixis does occur in cypress; Pichot et al., 2001). As these components are under separate genetic control, it is highly unlikely that all three could evolve in unison in a sexual ancestor through randomly occurring mutations, as expression of any single step would decrease the fitness of its sexual carrier (but see van Dijk and Vijverberg, 2005). An alternative hypothesis is that apomixis arises through de-regulation of the sexual developmental pathway (Koltunow, 1993; Grossniklaus, 2001), which would be manifested at multiple loci simultaneously. In wild apomictic taxa, this coordinated de-regulation could be influenced by global regulatory changes resulting from hybridization and/or polyploidy (Grossniklaus, 2001). Both naturally occurring and induced mutants showing each component separately have been identified (Curtis and Grossniklaus, 2007; Ravi et al., 2008), implying that many taxa have the potential to express apomixis-like traits, in addition to corroborating the hypothesized independent regulation of each component.

The *Boechera* (formerly *Arabis*) *holboellii* complex comprises *B. holboellii, B*. *stricta* (formerly *B. drummondii*), and their hybrid *B. ×divaricarpa* (Koch et al., 2003; Dobeš et al., 2004b). The breeding system of this complex is variable, consisting of both sexual and facultative apomictic forms (Böcher, 1951; Roy, 1995; Naumova et al., 2001). Compounding this variability is the wide distribution of polyploidy (mostly 2n = 3x) and aneuploidy (2n = 2x+1 or 2n = 3x+1; (Böcher, 1951), with the former originated multiple times in geographically and genetically distinct populations (Sharbel and Mitchell-Olds, 2001; Sharbel et al., 2005). The aneuploid chromosome is variable in recombination potential, heterochromatic, and degenerate at the morphological and molecular genetic levels, and has thus been classified as a B chromosome (Camacho et al., 2000; Sharbel et al., 2004). Generally speaking, *B. stricta* has been shown to be predominantly diploid and sexual, while *B. holboellii* and *B. ×divaricarpa* are facultative apomicts and highly variable with respect to ploidy, morphology, and genetic polymorphism (Roy and Rieseberg, 1989; Roy, 1995; Sharbel and Mitchell-Olds, 2001; Koch et al., 2003; Dobeš et al., 2004a; Sharbel et al., 2004; Schranz et al., 2005; Sharbel et al., 2005; Kantama et al., 2007).

Apomictic *Boechera* are characterized by Taraxacum-type diplospory whereby the megaspore mother cell undergoes meiosis I without completing the reductional phase (apomeiosis), followed by meiosis II leading to a nucleus which has the same ploidy as the

mother plant (Böcher, 1951; Naumova et al., 2001). As with many asexual taxa, microsporogenesis is typically disturbed in apomictic individuals, and variation in levels of normally reduced, non-reduced and aneuploid pollen exist within and between accessions (Böcher, 1951; Dobeš et al., 2004b; Schranz et al., 2005; Sharbel et al., 2005; Voigt, 2007). Analyses of meiosis in pollen cells have furthermore demonstrated variable degrees of chromosomal synapsis (univalent and multivalent) in apomictic accessions (Böcher, 1951), and both pseudogamous and autonomous endosperm formation have also been identified (Naumova et al., 2001; Voigt et al., 2007).

Here we have attempted to elucidate the first step in apomixis, apomeiosis, by means of a deep transcriptomic comparison between microdissected live sexual and apomeiotic ovules. Importantly, we have exploited the rare phenomenon of diploid gametophytic apomixis (see also Siena et al., 2008) in the *Boechera holboellii* complex (Böcher, 1951) in order to overcome the confounding effects of polyploidy. Our data show that, in comparison to sexual ovule formation, apomeiosis is characterized by shifts in gene expression (heterochrony), "parent of origin" effects stemming from ancestral hybridization, and allele expression from duplicated loci. In screening the transcriptomes of both sexual and apomictic ovules using combined SuperSAGE and 454 sequencing analyses, we have laid the groundwork for independent validation experiments using additional *Boechera* accessions and tissues.

MATERIAL and METHODS

Sample selection and ovule microdissection. We used the flow cytometric seed screen (Matzk et al., 2000) to analyse reproductive variability in four *Boechera* accessions (Table 1).

Table 1. *Boechera* accessions used in SuperSAGE and RT-PCR analyses.

Species[a]	*ID*	*analysis*	*Collection locality*
B. stricta	B06-485	SuperSAGE	Gold Creek, Colorado
B. holboellii	105.6.7	SuperSAGE	Bandy Ranch, Missoula Co, Montana
B. divaricarpa3	300.6.1	SuperSAGE	Birch Creek, Ravalli Co, Montana
B. divaricarpa8	205.3.4	SuperSAGE	Charlies Gulch, Ravalli Co, Montana
B. holboellii	105.6.7-2	RT-PCR	Bandy Ranch, Missoula Co, Montana
B. divaricarpa	300.6.1-1K	RT-PCR	Birch Creek, Ravalli Co, Montana

[a]Species identifications were based upon silique orientation and trichome morphology.

Early flower and embryo sac development was compared to that of *Arabidopsis* (Smyth et al., 1990; Schneitz et al., 1995), and we selected ovules at megasporogenesis

(between stages 2-II to 2-IV (Schneitz et al., 1995); differentiated megaspore mother cell, inner and outer integument initiated) in order to examine changes in gene expression associated with apomeiosis (Fig. 1).

The gynoecia of sexual and apomictic *Boechera* (Table 1) were dissected out from non-pollinated flowers at the stage at the megasporogenesis stage in 0.55 M sterile mannitol solution, at a standardized time (between 8 AM and 9 AM) over multiple days. Micro-dissections were performed in a sterile laminar air flow cabinet using a stereoscopic Microscope (1000 Stemi; Carl Zeiss, http://www.zeiss.com/) under 2X magnification. The gynoecium was held with forceps, and a sterile scalpel was used to cut longitudinally such that the halves of the silique together with the ovules were immediately exposed to the mannitol. Individual live ovules were subsequently collected under an inverted microscope (Axiovert 200M, Carl Zeiss) under sterile conditions, using sterile glass needles (self-made using a Narishige PC-10 puller (Narishige Group, http://www.narishige-group.com), and bent to an angle of approximately 100° to isolate the ovules from placental tissue. Using a glass capillary (with an opening of 150 µm interior diameter) interfaced to an Eppendorf Cell Tram Vario (http://www.eppendorf.de), the ovules were collected in sterile Eppendorf tubes containing 20 µl of RNA stabilizing buffer (RNAlater, Sigma, http://www.sigmaaldrich.com/). Ten ovules per tube and two samples per accession were collected in this way, frozen directly in liquid nitrogen and stored at -80 °C.

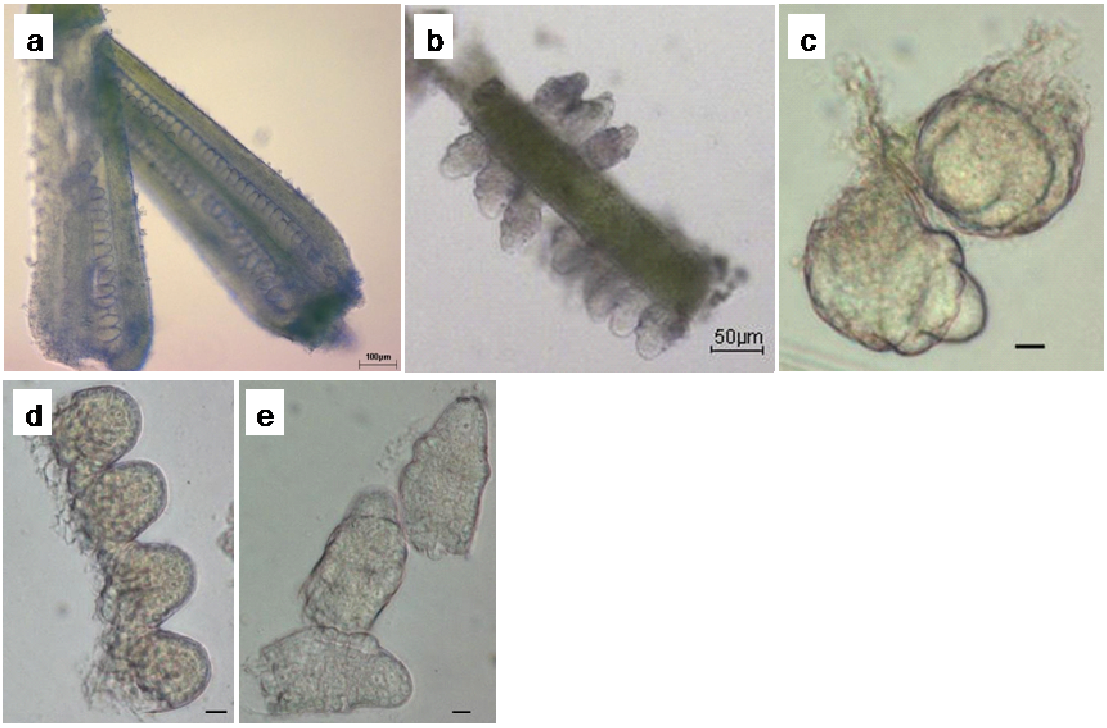


Figure 1. Photographs showing immature *Boechera* silique material (a, b) from which live ovules (c) between stages 2-II to 2-IV (Schneitz et al., 1995) were microdissected for SuperSAGE analysis, and stages 1I-1II (d) and 2II-2IV (e; Schneitz et al., 1995) from which ovules were collected for RT-PCR analyses (unless otherwise indicated, all scale bars = 10µm).

RNA isolation, amplification and cDNA generation. RNA from micro-dissected ovule material was isolated using a PicoPure RNA isolation kit (Arcturus Bioscience, http://www.arctur.com) with several modifications. The standard lysis buffer was supplemented with 1% (v/v) of NucleoGuard stock solution and 2µl N-Carrier (AmpTec, http://www.amp-tec.com). An additional treatment with Turbo DNAse (Ambion, http://www.ambion.com) treatment was included to eliminate any contaminating DNA. A second purification step was performed with RNeasy columns (Qiagen, http://www.qiagen.com/) to eliminate contaminating polysaccharides, proteins and the DNAse enzyme. RNA integrity and quantity was verified on an Agilent 2100 Bioanalyzer using the RNA Pico chips (Agilent Technologies, http://www.agilent.com).

Linear mRNA amplification was achieved using the ExpressArt mRNA amplification kit (AmpTec) with several modifications. Approximately 1 ng of total RNA was used in two independent reactions per sample to eliminate any random methodological effects. The two independent reactions were then pooled prior to cDNA conversion. RNA was converted to cDNA using an anchored oligo (dT)-T7-promoter primer with a mix of the AmpTec reverse transcriptase enzyme and the ArrayScript reverse transcriptase (Ambion) and double-stranded cDNA was generated using the box-random-trinucleotide primer included in the kit. The resulting RNA after the first amplification round was purified with RNeasy MinElute columns (Qiagen), followed by a second and third amplification round. RNA integrity and quantity were verified on an Agilent 2100 Bioanalyzer using RNA Nano chips (Agilent). RNA quantity was determined using a Nanodrop ND-1000 spectrophotometer. Subsequently, 5 µg of DNA-free amplified mRNA was converted into double-stranded cDNA using a 5´-biotinylated primer and subsequently used for SuperSAGE generation.

SuperSAGE libraries were produced by GenXPro GmbH, Frankfurt am Main, Germany), essentially as described by Matsumura et al. (2006) with the following adaptations: instead of concatemerization and cloning of the ditags, these were directly sequenced using a GS 20 sequencer (454 Life Sciences; Roche, http://www.roche.com). All ditags consisting of the same tag combination were eliminated using the GenXProgram software, which also sorts and counts the 26 bp-tags. All tags containing “Ns” were removed from the analysis, which has been shown to significantly increase the accuracy of GS-20 reads (Huse et al., 2007). The likelihood for the differential expression of each SuperSAGE tag was calculated using normalized tag numbers with a correction for multiple tests (Audic and Claverie, 1997), as implemented in Robertson et al. (2007).

Preparation of normalized dscDNA for 454 transcriptome sequencing. RNA was isolated from pooled flower stages 1 - 12 (Smyth et al., 1990) from three diploid sexual plants (two *B. holboellii* and one *B. stricta*, accessions ES910-2-2K, 105.6-1K, B07261) and three

apomictic plants (two diploid *B. divaricarpa* and one triploid *B. divaricarpa*, accessions 67.5-K, 300.6.1-1K, 218.2-2K) with the TRIzol Reagent (Invitrogen, http://www.invitrogen.com/) according to the manufacturer's protocol. Any remaining genomic DNA contamination was removed by TURBO DNAse (Ambion). The DNAse enzyme was removed and the RNA was further purified using a RNeasy MinElute clean-up Kit (Qiagen) according to the manufacturer's protocol, and eluted in 20 µl of RNA storage solution (Ambion). Poly(A)+ mRNA was isolated using a Poly(A)Purist mRNA purification kit according to the manufacturer's protocol (Ambion).

Two normalized, full-length-enriched cDNA libraries were generated using a SMART cDNA library construction kit (BD Clontech, http://www.clontech.com) and the Trimmer Direct cDNA normalization kit (Evrogen, http://evrogen.com), generally following the manufacturer's protocol but with several modifications. In brief, 1 µg of poly(A)+ mRNA was used for each cDNA library generated. Reverse transcription was performed with a mixture of several reverse transcription enzymes for 1h at 42°C and 90 minutes at 50°C. Each step of the normalization procedure was carefully monitored to avoid generation of artefacts e.g. due to overcycling. The resulting full-length-enriched, normalized cDNAs were linearly amplified, resulting in a total of 15µg double-stranded cDNA. The double-stranded cDNA was digested using SfiI enzyme to remove most of the adaptor sequences for 454 sequencing, column-purified, concentrated, and sequenced using 454 technology (GS FLX, Agencourt Genomics Services, http://www.agencourt.com). Sequences were analysed using a standard Newbler assembly (454 Life Sciences).

RT-PCR. 8 SuperSAGE tags which were differentially expressed between the apomictic and sexual accessions were selected. The tags were blasted against the 454 sequences of both apomictic and sexual transcriptome libraries to obtain corresponding gene sequences that were analysed using GeneMark (Besemer and Borodovsky, 2005) and GENSCAN (Burge and Karlin, 1997) and compared with *Arabidopsis thaliana* for the prediction of the coding regions (Table S1). Predicted coding regions were used for PCR primer design using the Genome Lab™ GeXP system (Beckman Coulter®, http://www.beckmancoulter.com) to generate a multiplex group corresponding to eight transcripts, plus three pairs of primers for known housekeeping genes: eEF1Balpha2 (NM_121956), ACT2 (NM_112764) and ACT11 (NM_112046)(Sturzenbaum and Kille, 2001).

Eleven sets of 20 ovules were isolated by tissue micro-dissection from two different accessions (one sexual and one apomictic; Table 1) at two different development stages (1I-1II, 2II-2IV, Schneitz et al., 1995) for RNA extraction. Following RNA extraction using a PicoPure RNA isolation kit (Arcturus Bioscience) and quantification on an Agilent 2100 Bioanalyzer using RNA Pico chips, multiplex cDNA synthesis and PCR reactions were

performed using the Genome Lab™ GeXP system following the manufacturer's instructions (Beckman Coulter). The multiplexed PCR products were analyzed in a CEQ8000 sequencer using the GenomeLab™ GeXP genetic analysis System (Beckman Coulter). Relative peak area, as measured by the GeXP system, was compared between all genes. Of the housekeeping genes, ACT2 exhibited the least amount of variation across all samples and hence was used to standardize expression levels of the 8 chosen genes.

Bioinformatics analyses. All SuperSAGE tag sequences of a given subset (four *Boechera* libraries, in silico SuperSAGE set of *Arabidopsis*) were compared, the number of positional nucleotide differences was counted, and the ratio of observed versus expected frequencies was plotted in a histogram for each data set. The expected random distribution of *k* bp differences in the data was calculated using the density function of the binomial distribution $P(X = k) = B(k|p, n)$ with n = 22 (tag size minus size of the common restriction site) and $p = 1-(p_A^2 + p_C^2 + p_G^2 + p_T^2)$ the probability of a nucleotide mismatch given the nucleotide frequency (*B. divaricarpa*3: A=0.343, C=0.169, G=0.184, T=0.304; *B. divaricarpa*8: A=0.348, C=0.165, G=0.177, T=0.310; *B. stricta*: A=0.340, C=0.154, G=0.176, T=0.330; *B. holboellii*: A=0.346, C=0.160, G=0.174, T=0.320; *A. thaliana*: A=0.295, C=0.178, G=0.215, T=0.312) and thus $p = 0.728$, $p = 0.724$, $p = 0.721$, $p = 0.722$ and $p = 0.738$ for *B. divaricarpa* 3, *B. divaricarpa* 8, *B. stricta*, *B. holboellii* and *Arabidopsis*, respectively. In order to obtain virtual *A. thaliana* SuperSAGE tags, the predicted transcripts data set (TAIR7_cdna_20070425) was downloaded from the TAIR ftp site (ftp://ftp.arabidopsis.org/). The tags were generated using a modified version of the Perl script published on the TAIR ftp site, which allows the extraction of most downstream tags of 26 bp size with a 5' recognition sequence CATG in a virtual SuperSAGE experiment.

All SuperSAGE tags were blasted to the 454 cDNA database obtained from sexual and apomictic *Boechera* using the following parameters (blastall -p blastn -m 8 -e 1 -W 7 -r 1 -q -1 –i). *Boechera* sequences representing significant hits were annotated by Blast2GO using default parameters and the ANNEX annotation augmentation function (version 2.3.1, Conesa and Gotz, 2008). The combined graph function of Blast2Go was used to generate pie charts of the functional annotation of groups of sequences (representing different SuperSAGE tag classes, Table 4) based on gene ontology (GO) categorization, and a Fisher's exact test (Conesa and Gotz, 2008) was used to analyse *Boechera* 454 sequence groups corresponding to each SuperSAGE tag class (Table 4) for significant enrichment of particular GO classes. Due to the large numbers of sequences in the sex- and apomeiosis-specific data sets (Table 4), a 5% sequence filter value was used as a cut-off to generate combined graphs (Conesa and Gotz, 2008). No cut-off value was used for any other groups of sequences. Sequences were evaluated for their predicted involvement in

biological processes, molecular functions, and cellular localization, and all data are presented at level 3 GO categorization (Table S2 and Figure S2).

RESULTS

Using flow cytometric seed screening (Matzk et al., 2000) to assess reproductive mode, we chose 2 diploid sexual (*B. stricta* and *B. holboellii*) and two highly expressive diploid apomictic individuals (*B. divaricarpa*3 and *B. divaricarpa*8) from which 10 live ovules each were microdissected for linear cDNA amplification and SuperSAGE analysis (Figure 1, Tables 1 and 2). Based on the flow cytometric seed screen (Matzk et al., 2000), the 2 chosen sexual individuals exclusively comprised 2C embryo:3C endosperm, which, as expected, reflected an embryo composition of C maternal (Cm) genomes + C paternal (Cp) = 2C genomes, and an endosperm composition of 2Cm + Cp = 3C. In contrast, the two diploid apomicts were both characterized by > 95% apomeiotically derived embryos (unfertilized 2Cm) and > 95% fertilized endosperm (e.g. 4Cm + 2Cp = 6C; Table 2).

Table 2. Flow cytometric seed screen results showing number of seeds for each *Boechera* accession characterized by the embryo:endosperm[a] ploidy values.

Species (ID)	2:3[a]	2:6	2:8	2:10	4:6	4:8
B. stricta[b] (B06-485)	16	-	-	-	-	-
B. holboellii (105.6.7)	84	-	-	-	-	-
B. divaricarpa3 (300.6.1)	-	51	9	-	-	2
B. divaricarpa8 (205.3.4)	-	213	20	1	3	2

[b]The sexuality of this accession (SAD4) was also confirmed by crosses and isozyme analyses.

SuperSAGE analysis of the four ovule-specific cDNA libraries yielded a total of 452 516 sequenced mRNA tags (26 bp), 339 888 of which were found in two or more copies in all libraries together, and, from these, 43 289 different tag sequences were identified (Table 3). An additional 112 628 tags occurred as singletons. Both sex flower-specific and apomictic flower-specific cDNA libraries were generated using 454 (FLX, 454 Life Siences, Roche, http://www.roche.com) technology. The sexual flower-specific library had a total of 12 365 737 high quality sequenced base pairs (mean sequence length 409 bp, median sequence length 244 bp), and Newbler assembly (454 Life Sciences) generated 30 237 assembled contigs. The apomictic flower-specific library had a total of 14 622 601 high-quality sequenced base pairs (mean sequence length 414 bp, median sequence length 246 bp), and Newbler assembly generated 35 362 assembled contigs.

Table 3. Number of shared (above diagonal) and differentially expressed[a] SuperSAGE tags (below diagonal) between *Boechera* ovule-specific libraries, in addition to the total number of sequenced and singleton tags per library.

	Sexual		Apomictic	
	stricta	*holboellii*	*divaricarpa3*	*divaricarpa8*
B. stricta	-	28271	24298	33353
B. holboellii	923 / 2014	-	29432	34864
B. divaricarpa3	801 / 1612	1344 / 1015	-	33699
B. divaricarpa8	781 / 3424	894 / 2975	622 / 2227	-
Tags sequenced per library	67433	77713	61692	133040
Number of singletons	19331	29481	24142	39674

[a] $P < 0.05$ (Audic and Claverie, 1997); For each relative comparison (up / down-regulated), "up-regulated" refers to the first number in the column (*e.g.* in *holboellii* vs *stricta*, 923 and 2014 tags were up- and down-regulated, respectively, in *holboellii*).

Allelic variation. SuperSAGE tags are frequently allele-specific as they are derived from the most 3' Nla*III* restriction site of the 3' UTR of each mRNA molecule (Matsumura et al., 2006). Any two tags sampled at random from the four *Boechera* SuperSAGE libraries probably originate from different genes, and are thus not expected to share sequence homology. This was supported by pairwise comparisons of all tag sequence variants obtained from *Boechera* (n = 155917), in addition to those from an in silico *Arabidopsis* SuperSAGE analysis (n = 26197), all of which demonstrated no overall deviation from expected [Figure (2a)].

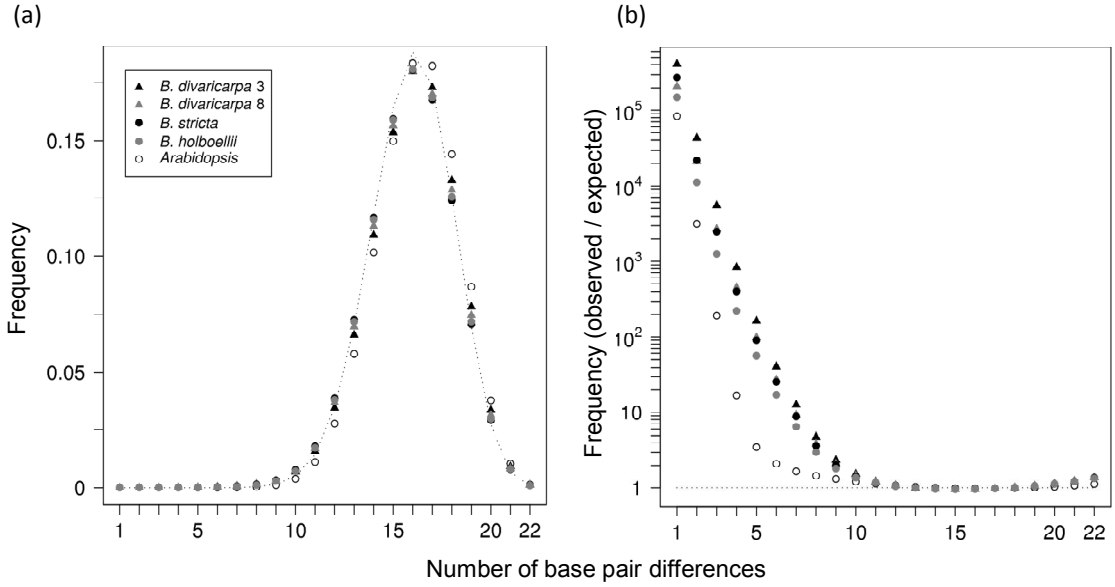


Figure 2. Frequency plots of (a) the observed and expected (dotted line) random distributions of pairwise sequence homology between SuperSAGE tags, and (b) the ratio of observed / expected numbers of SuperSAGE tag pairs that differ by 1-22 bp in all tag pairwise comparisons for four *Boechera* SuperSAGE libraries and a virtual *A. thaliana* SuperSAGE analysis.

Nonetheless, pairwise sequence comparisons of all *Boechera* and *Arabidopsis* SuperSAGE tags showed differences of 50- to >10^5-fold in the observed versus expected variation for tag pairs with 8 or fewer base pair differences, implying that the majority of these represent allelic variants of the 3' UTR regions at the same or duplicated loci [Figure (2b)].

Sex and apomeiosis specific tags, and heterochrony. Our goal was to identify transcriptomal differences between the sexual and apomictic ovules. For each SuperSAGE tag, we thus tested for significant differences in copy number between each of the four libraries using the method described by Audic and Claverie (1997), and three classes of differentially expressed tags were categorized through comparison of the four libraries (Table 4).

Table 4 Criteria based on corrected number of SuperSAGE tags per library wich were used to differentiate allele classes.

Allele class	Criteria[a]
I. Apo- or sex-specific	(a & d) = 0 & (b & c) > 0 OR (a & d) > 0 & (b & c) = 0
II. Differentially expressed[b]	(a, b, c, d) > 0; (a ≈ d & b ≈ c) & (a ≠ b & a ≠ c & d ≠ b & d ≠ c)
III. Species specific (trans regulated)	(a > 0 & d = 0) & (b ≠ a & c ≠ a) OR (a = 0 & d > 0) & (b ≠ d & c ≠ d)

[a]Letters refer to each SuperSAGE library: a, *B. holboellii* ; b, *B. divaricarpa3*; c, *B. divaricarpa8; d, B. stricta*.
[b]Tag numbers are considered not equal (≠) if P ≤ 0.05 (Audic and Claverie, 1997), otherwise tag numbers are considered almost equal (≈).

The first class comprised 1879 sex-specific and 2395 apomeiosis-specific tags that were found exclusively in both sexual or in both apomictic libraries (Table 4). Of the apomeiosis-specific tags, 83 (3.5%) were significantly differentially expressed between the two apomicts, while 47 (2%) of the sex-specific tags were significantly differentially expressed between the two sexuals libraries (P < 0.05) (Audic and Claverie, 1997). Thus, most apomeiosis- and sex-specific tags were consistently expressed between the two apomictic or sexual libraries, and, in addition, the expression of the sex-specific tags was conserved even when compared between two species (*B. stricta* and *B. holboellii*).

The presence and absence of tags (i.e. alleles) between SuperSAGE libraries could be explained by allelic differences between the accessions, variation in developmental timing between the accessions and/or tissues stages analysed here, or, alternatively heterochronic allele expression could characterize the sexual and apomeiotic ovules. In the case of allelic variation for any sex- or apomeiosis-specific tag, corresponding tags of high sequence homology (i.e. alleles) should be found in the pools of tags sequenced from other libraries. Alternatively, heterochrony would be shown if identical alleles were

found in both sexual and apomictic genomic DNA, and if they were expressed only within sexual or apomictic ovules at the developmental stage sampled here.

We tested for allelic differences by first searching for homologous SuperSAGE tags what differed by one nucleotide from the sex- and apomeiosis-specific tags, and identified an additional 899 related tags. We then calculated the transcriptional profile for each family of tags (original tag plus its homologues) by adding the SuperSAGE tag copy numbers together for each family, and, after doing so, 1809 (96%) and 1712 (71%) of the families remained as sex- and apomeiosis-specific, respectively. Thus, most of the alleles that are specifically expressed either during sexual or apomeiotic egg formation do not have a corresponding allele (as defined by a 1bp difference) expressed in the opposite reproductive mode.

In order to examine whether expression of the apomeiotic- and sex-specific tags was temporally shifted (heterochrony), we performed a sequence homology search against two normalized cDNA libraries from pooled flower stages 1 – 12 (Smyth et al., 1990) of three diploid sexual plants (two *B. holboellii* and one *B. stricta*) and three apomictic plants (two diploid and one triploid *B. divaricarpa*) using 454 (FLX) technology. Of the apomeiotic-specific tags, 180 (10%) and 177 (10%) blasted perfectly (100% homology) to the apomixis- and sex-specific 454 libraries, respectively (Figure 3). In contrast, 366 (20%) and 329 (18%) of the sex-specific tags blasted perfectly to the apomixis- and sex-specific 454 libraries, respectively (Figure 3).

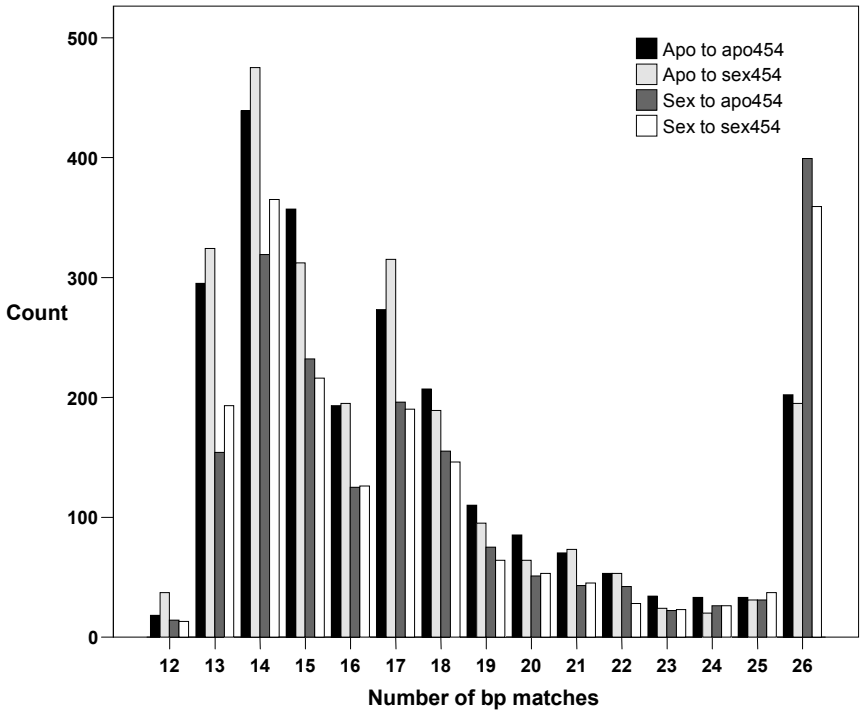


Figure 3. Number of matches between sex- and apomeiosis-specific SuperSAGE tags and two normalized cDNA libraries (sex454 and apo454), according to the number of homologous base pairs (12 - 26).

We have thus identified 177 apomeiosis-specific and 366 sex-specific tags that are found in their "opposite" cDNA libraries that contain summed transcripts expressed over many developmental stages. In other words, 543 alleles that are common to both sexual and apomictic genomes are turned on or off with respect to apomeiosis or sex at this stage of ovule development, and, when turned on, show similar expression patterns between the two sexual or two apomictic ovules.

The goal of this study was to screen the ovule transcriptomes for differentially expressed mRNAs whose expression patterns will subsequently be examined using microarrays for various *Boechera* accessions and multiple developmental stages and tissues. We nonetheless attempted to confirm our hypothesis of heterochronic expression using RT-PCR on genes corresponding to four sex- and four apomeiosis-specific SuperSAGE tags in two ovule stages (Figure 1, Table 1; Table S1). Compared to stage 1, the four genes corresponding to the sex-specific SuperSAGE tags demonstrated an overall shift to over-expression in sexual ovules in stage 2 (Table S1 and Figure S1). An overall change in expression between ovule stages 1 and 2 for the four genes corresponding to the apomeiosis-specific tags was not apparent (Table S1 and Figure S1). Although the weak correlation between our SuperSAGE and RT-PCR data sets may reflect the general problem of comparing transcript abundance across platforms (Wang, 2007), we suspect that (i) variable allele concentrations (i.e. mRNA dosage) relative to RT-PCR detection threshold levels (Grimanelli et al., 2005) in micro-dissected samples, (ii) the use of gene- rather than allele-specific primers in the RT-PCR analysis, and (iii) the compounded influence of widespread duplications in the apomictic genome (Corral et al., 2009) on allele dosage also contributed to the inconsistencies.

Differentially-expressed tags. A second class was characterized by tags that were found in all four libraries, but that were differentially expressed between sex and apomeiotic ovules, and expressed consistently within each reproductive mode (Table 4). We identified 59 tags that were expressed in all four libraries and were not different in copy number when the two sexual or two apomictic individuals were compared, but were significantly differentially expressed between the apomicts and sexual individuals in all comparisons [Figure 4(a)]. Of these, 39 and 20 tags were down- and up-regulated, respectively, and furthermore showed similar levels in both apomeiotic ovules [Figure 4(a)].

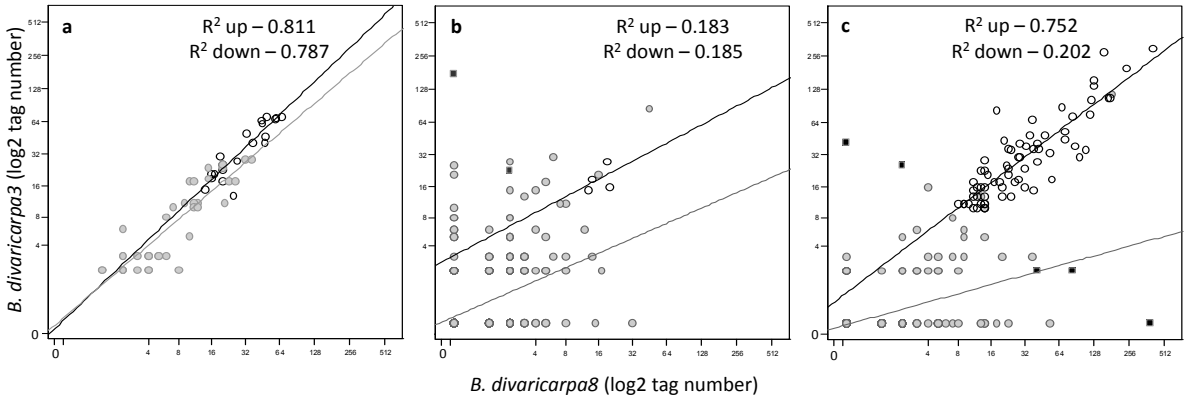


Figure 4. Scatter plots showing normalized numbers of SuperSAGE tags plotted for both apomictic ovule samples for (a) 59 differentially expressed, (b) 525 *B. stricta*-specific, and (c) 440 *B. holboellii*-specific tags
Symbols show expression relative to sexual libraries: white circle, up-regulated in both apos; grey circle, down-regulated in both apomicts; square, up-regulated in one apomict and down-regulated in one other.

Parent of origin effects. The final category of apomeiosis candidate alleles was characterized by tags which were found in one sexual library only (either *B. stricta* or *B. holboellii*) and in both apomictic libraries (Table 4). These tags support hypotheses of apomixis origins via hybridization (Carman, 1997), as *B. stricta* and a taxon related to *B. holboellii* have been shown phylogenetically to be the progenitors of the hybrid apomictic *B. divaricarpa* (Dobeš et al., 2004a; Dobeš et al., 2004b).

Regulatory changes in mRNA expression can be inferred by measuring cis- and trans- effects in expression patterns in both parents and hybrids, a comparison which is appropriate when the parental species have undergone a long period of reproductive isolation (Landry et al., 2005), as is the case with *Boechera*. With regards to mRNA tags that were found only in one of the two sexual individuals, 13581 and 7412 tags were identified as either *B. holboellii*- or *B. stricta*-specific respectively. Of these, 440 *B. holboellii*- and 525 *B. stricta*-specific tags were significantly differentially expressed in both apomictic individuals [Figure 4(b, c)], and hence probaly under trans-control (i.e. 'parent of origin' effect). Interestingly, the species-specific alleles exhibit different patterns of expression in the apomicts, with 16% (n=69) and 83% (n=366) of the *B. holboellii*-specific alleles being up- and down-regulated, respectively, while virtually all (99%, n=519) *B. stricta*-specific alleles are down-regulated [Figure 4(b, c)]. These data suggest that apomeiosis is characterized by asymmetrical up-regulation of *B. holboellii* alleles in the hybrid genome. The relatively scattered pattern of down-regulated tags in both distributions [Figure 4(b, c)] is suggestive of the number of mechanisms by which a gene can be down-regulated (Vaucheret, 2006), in addition to variability in the rate of post-transcriptional regulation (Garneau et al., 2007).

Gene copy number variation. Gene duplication was estimated by calculating allele family size in the four SuperSAGE libraries. For each SuperSAGE tag, a network of tags (i.e. allele family) from its respective library was computed such that each member of a network was related to at least one another network member by 1 bp. This analysis provides a conservative estimate of allele family size, as insertion/deletion variation is not considered, nor are allelic variants whit differences >1 bp, although there is evidence for allelic variation characterized by up to 8 bp differences in observed tag pairs [Figure 2(b)]. The number of members per allele family was calculated and plotted for each SuperSAGE library [Figure 5(a)], and demonstrates a trend of increasing frequencies of large allele families (more than nine alleles per family) in the apomictic libraries [Figure 5(a)]. Furthermore, up-regulated tags belonging to allele classes II and III (Table 4) show a trend of belonging to larger allele families compared to down-regulated tags [Figure 5(b)] Thus, it appears that many of the differentially expressed tags that are up-regulated in the apomeiotic compared with the sexual ovules may have been transcribed from duplicated loci in both apomicts. The elevated values for the apomictic libraries (Figure 5) are not consistent with a larger number of sequenced SuperSAGE tags, as the divarcarpa3 library had the least number of tags sequenced (Table 3).

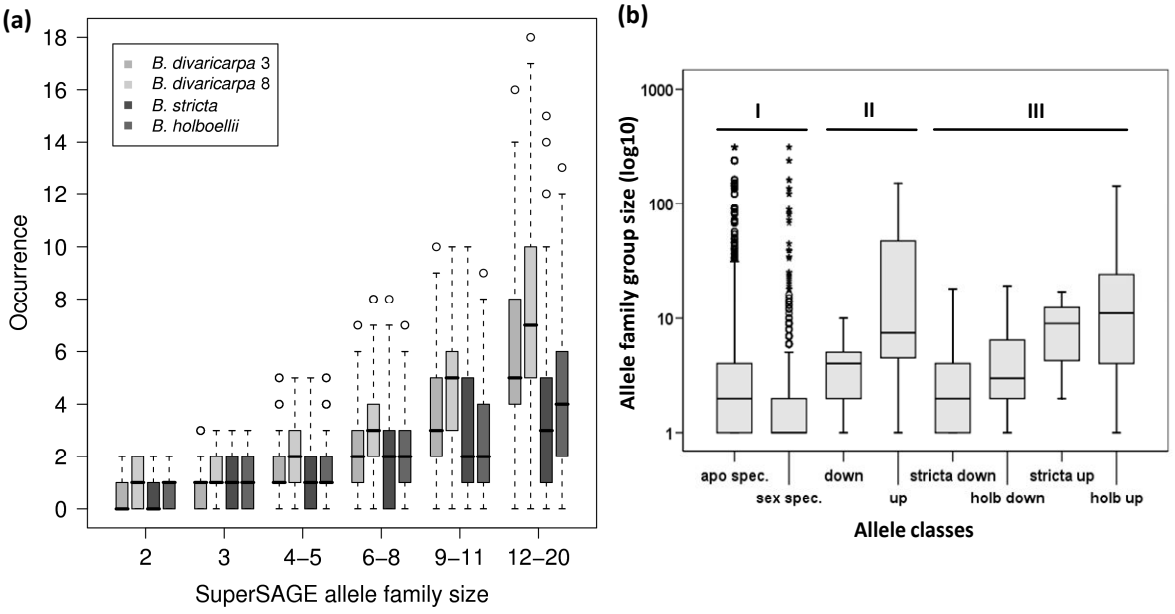


Figure 5. SuperSAGE allele family size and distribution. (a) Frequencies of SuperSAGE tag group size (i.e. allele family size, whereby all members of a family differ by 1 bp) for various size classes (i.e. number of members per group). (b) Distributions of allele family group size for three classes (see Table 4) of differentially expressed tags between sexual and apomeiotic ovules.

Gene ontology analysis. Using the Fisher's exact test (Conesa and Gotz, 2008), we analysed *Boechera* 454 sequence groups corresponding to each SuperSAGE tag class (Table 4), and found no significant enrichment for particular gene ontology (GO) classes in any comparison (data not shown). Eighteen different level 3 GO (biological process) terms

associated with reproduction were nonetheless identified from the *Boechera* cDNAs corresponding to differentially expressed *B. stricta*- and *B. holboellii*-specific SuperSAGE tags (Figure S2 and Table S2).

DISCUSSION

The data presented here were generated from live microdissected ovules in the developmental stage during which apomeiosis is hypothesized to be expressed, thus minimizing the effects of mRNA degradation through tissue fixation (Goldsworthy et al., 1999) and non-specificity of transcriptomic profiles due to analyses of mixed tissues. The effects of mRNA degradation could conceivably distort statistical comparisons of SuperSAGE tag frequencies between libraries, given that the majority of tags identified here had relatively low copy numbers (< 100) (Figure 6).

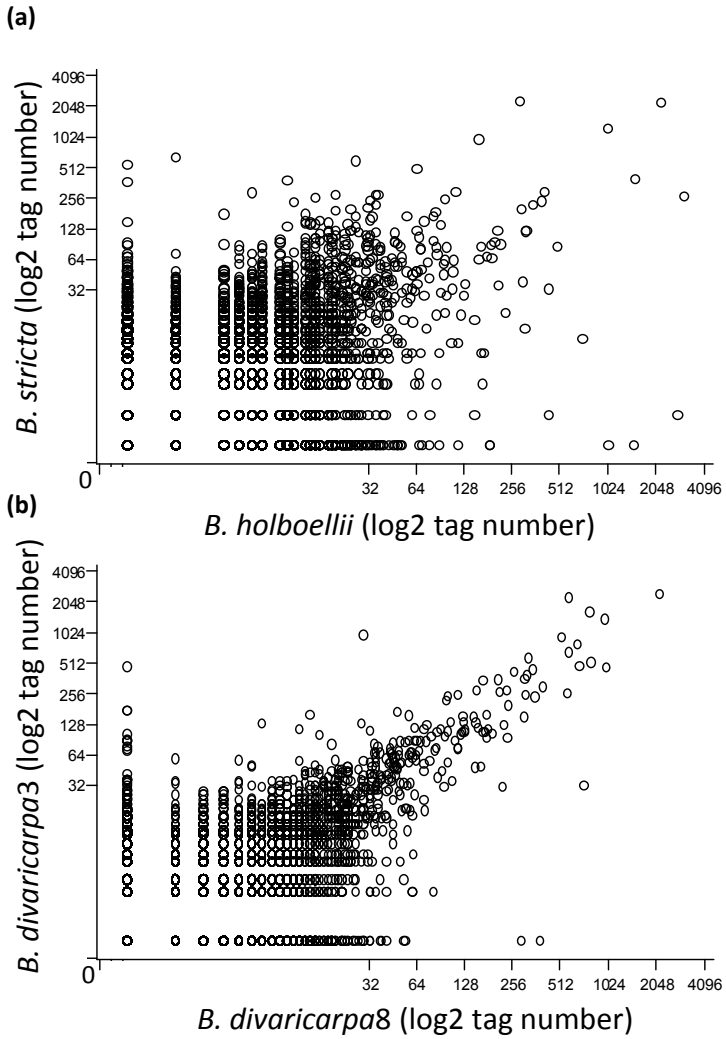


Figure 6. Comparative plots of all SuperSAGE tag numbers (normalized) shared between (a) the two sexual libraries (n = 28 271 tags) and (b) 2 apomeiotic libraries (n = 33 699 tags).

A combination of deep transcriptomic profiling of the ovules using SuperSAGE (Matsumura et al., 2006), in conjunction with sequencing of the sexual and apomictic flower transcriptomes using 454 technology, has enabled us to identify over 4000 differentially expressed mRNAs between sexual and apomeiotic ovules at this single stage of development. Finally, the use of diploid apomictic *B. divaricarpa* accessions that show high levels of apomeiosis (Table 2) has reduced the confounding effects of polyploidy.

The relatively low levels of sequence overlap (<50%; data not shown) between the SuperSAGE (3'-biased) and 454 cDNA libraries may be reflective of 5' bias in the 454 sequencing (Weber et al., 2007), although 3' bias resulting from incomplete (i.e. not full-length) cDNA synthesis has also been demonstrated (Bainbridge et al., 2006; Emrich et al., 2007). As cDNA preparation for the 454 sequencing here entailed linear amplification, 3' bias in our cDNA libraries is more likely the case. Hence an alternative explanation for the low overlap between our SuperSAGE and 454 libraries could be that ovule-specific transcripts may have been under-represented or absent from the pooled flower-specific cDNA libraries. Furthermore, as 3' UTR regions demonstrate allelic variability (Andolfatto, 2005; Eveland et al., 2008) that may involve factors that were not extensively analysed here, including insertion/deletion mutations as well as single nucleotide variation (Figure 3) at single or duplicated loci (Figure 5), our assessment of homology between the SuperSAGE and 454 datasets is probably an under-estimation.

Allelic variation.

The pairwise comparisons of all SuperSAGE tags, both in the four *Boechera* libraries and in the in silico *Arabidopsis* SuperSAGE experiment, demonstrate that allelic variation has been captured for many of the sampled mRNAs from the sexual and apomeiotic ovules (Figure 2). SuperSAGE tags are derived from the 3' UTR of each mRNA molecule (Matsumura et al., 2006), which have been shown to be allele specific in plants (Eveland et al., 2008). Although they are not translated into proteins, 3' UTRs are involved in the regulation of mRNA transcript numbers, and have been shown to be under measurable levels of selection pressure in humans (Chen and Rajewsky, 2006). In Drosophila, 3' UTRs are frequently under higher levels of selection compared to synonymous polymorphisms in codon regions (Andolfatto, 2005), a likely result of functional constraints on specific sequence motifs involved with down-regulation of gene expression through translation inhibition and rapid targeted decay of mRNAs (Shyu et al., 2008).

Thus the elevated numbers of observed tags that share high levels of sequence homology [Figure 2(b)] are reflective of multiple alleles at the same or duplicated loci. The

observed numbers of homologous (i.e. allelic) *Boechera* tag pairs were consistently higher than those from the in silico *Arabidopsis* SuperSAGE analysis [Figure 2(b)], which could reflect the different sequencing technologies used to generate the data (GS20 versus Sanger). However, it is more likely that the differences can be explained by elevated levels of homozygosity in *Arabidopsis* (Marais et al., 2004) compared to *Boechera*, whose complex history of glacial isolation, hybridization and gene flow have led to elevated heterozygosity (Roy, 1995; Dobeš et al., 2004a; Dobeš et al., 2004b).

Transcriptomal differences between sexual and apomeiotic ovules.

Remarkably, the differentially expressed profiles of many hundreds of alleles representing three classes (Table 4) are consistently similar between the apomeiotic ovules that were independently collected and prepared from two asexual lineages (Figures 4 and 6). In contrast, the global expression profiles of the two sexual *Boechera* are less similar (Figure 6), and probatly demonstrate species-specific differences at this stage of development. While expression differences between the apomictic libraries may have resulted from experimental noise, some could also reflect the dynamics of apomixis evolution from sexual ancestors; for example, whether the two *B. divaricarpa* accessions are from one or multiple asexual lineages (i.e. single or repeated origin from sexual ancestors). In the former case, the differences in the transcriptomal patterns could reflect independent mutation accumulation since the last shared asexual ancestor between them. In the latter, the independent origins of the two asexual lineages from sexual ancestors must have led convergently to highly similar global gene expression patterns at this stage of ovule development (Figure 6).

The identification of 543 genes that exhibit a developmental timing shift in expression between the sexual and apomeiotic ovules implicates sexual pathway de-regulation in apomeiosis expression (Koltunow, 1993; Grossniklaus, 2001). This is probably a conservative estimate of the number of genes that actually show heterochronic shifts, as we have sampled one developmental stage and could thus only consider presence and absence of mRNA tags (i.e. genes turned on and off). We have recently completed a second SuperSAGE analysis (over two million sequenced tags) of sexual and apomeiotic ovules sampled across four developmental stages so that we can additionally test for heterochrony in genes which exhibit regulatory changes over time.

One of the most intriguing results was the class of apomeiosis candidate alleles (class III) that displayed the ancient 'parent of origin' effect [Table 4 Fig. 4(b,c)]. *Boechera stricta* or *B. holboellii*-specific mRNAs that are differentially expressed in the *B. divaricarpa* (hybrid apomictic) nucleus could be explained by heterochronic shifts in gene expression,

'genomic shock' - like responses (McClintock, 1984) in hybrid genomes that may lead to post-transcriptional mRNA degradation, or changes to major regulatory genes that have cascade effects on transcription networks (Osborn et al., 2003). It is striking that the unfertilized apomeiotic ovule is characterized by up-regulation of *B. holboellii* alleles [Figure 4(c)], and furthermore that this pattern exists several hundred (if not thousands) of generations after the original hybridization event (Dobeš et al., 2004a). Given that maternal allele expression frequently characterizes early embryo development (Grimanelli et al., 2005; Springer and Stupar, 2007), *B. holboellii* (or a closely related taxon) was probably the maternal parent in the origin of the asexual (hybrid) lineages analysed here.

We have not performed reciprocal crosses in order to test the hypothesized 'parent of origin' effect (Springer and Stupar, 2007), nor would this be straightforward given the variable pollen and egg cell ploidy and their influence on offspring ploidy in sexual and apomictic accessions (Schranz et al., 2005; Voigt et al., 2007). Differential allelic expression is known to occur in homoeologous genes in hybrid plants (Adams et al., 2003; Adams and Wendel, 2005; Schlueter et al., 2006; Udall et al., 2006; Adams, 2007), and thus the 'parent of origin' effect could also reflect species-specific differences (i.e. between *B. holboellii* and *B. stricta*, or related ancestral lineages involved in the original hybridization event) with regard to tissue specificity or response to environmental conditions (Guo et al., 2004), rather than being the underlying factors in the switch from sexual reproduction to apomixis.

Genomic duplications in apomicts.

Duplications in the apomictic genome could provide polyploid dosages of alleles required for apomeiosis, and this would explain the rare evolutionary stability of diploid apomixis in *Boechera*. Apomictic genomes, unconstrained by selection pressure of maintain stable meiosis, are typically structurally variable compared to sexual genomes (Roche et al., 2001). In apomictic *Boechera*, this is supported by variable chromosomal synapsis in meiosis I (Böcher, 1951), as well as allele sequence divergence and genomic duplication (Kantama et al., 2007; Corral et al., 2009).

If one assumes that SuperSAGE allele families (i.e. groups of tags where any two members differ by 1 bp) comprising more than two members represent alleles transcribed from duplicated loci in diploid *Boechera*, evidence for the highest number of genomic duplications whereby all duplicated alleles are expressed is shown for both apomeiotic SuperSAGE libraries [Figure 5(a)]. Furthermore, up-regulated candidate apomeiosis alleles in classes II and III (Table 4) show a trend for larger allele family sizes

relative to down-regulated alleles [Figure 5(b)]. Taken together, it appears that many of the differentially expressed tags which are up-regulated in apomeiotic ovules have been transcribed from loci that have undergone duplication in the apomictic genome.

Evolution of apomeiosis in Boechera.

We have provided evidence that heterochrony, gene duplication, and, interestingly, ancient 'parent of origin' effects characterize the transcriptomic signature of apomeiosis in *Boechera*. Apparently allelic up-regulation is important for apomeiosis, and gene duplication is the mechanism that mediates transcript elevation in the apomictic genome. Alternatively, down-regulation of alleles expressed during sexual seed development may also be implicated in apomeiosis, although we cannot make any conclusions regarding 3' UTR motifs and their involvement in post transcriptional regulatory differences between sexual and apomeiotic ovules due to limited sequence coverage (26 bp) by the SuperSAGE data.

We have attempted to overcome the difficulties of polyploidy by using diploid apomictic accessions, but the analyses performed here still leave us with a similar conundrum when considering the effects of duplicated gene dosage versus hybridization on global gene expression patterns leading to apomeiosis. Both the 'parent of origin' effect described here and previous chromosomal analyses (Kantama et al., 2007) show that diploid apomictic *Boechera* are hybrid, and analyses of seed formation demonstrate that diploid and triploid apomictic *Boechera* generally produce unreduced gametes (Dobeš et al., 2004b; Schranz et al., 2005; Voigt et al., 2007). Hence hybridization probably preceded polyploidization during the evolution of apomixis, as the latter would have been made possible by the unreduced gametes generated by diploid apomicts. The selective pressure to maintain homologous chromosome pairing during meiosis is, by definition, relaxed in asexual organisms, and this is reflected in chromosomal rearrangements (i.e. chromosomal heteromorphy) that typically characterize asexual genomes (William Birky Jr., 1996).

On possibility is that apomeiosis was first induced through hybridization, with subsequent establishment of asexuality leading to meiotic perturbations and the accumulation of chromosomal rearrangements (including duplications). Alternatively, the potential for expressing apomeiosis could be an older characteristic of the genus (i.e. pre-Pleistocene), as there is evidence for unreduced gamete formation in other *Boechera* species (Dobeš et al., 2006). Hybridization is nonetheless a major factor which has played a significant role in the evolution of *Boechera* (Koch et al., 2003; Dobeš et al., 2004a; Dobeš et al., 2004b; Dobes et al., 2004; Schranz et al., 2005; Dobeš et al., 2006; Dobeš et

al., 2007), and hence inter-breeding between members of this genus could have led convergently to the expression of apomeiosis independently in different apomictic lineages. Analyses of apomeiosis expression in other apomictic *Boechera* are required in order to differentiate between these scenarios. Finally, the identification of differentially expressed species-specific mRNAs representing many genes involved reproduction (Table S2) is consistent with the 'hybridization-derived floral asynchrony' hypothesis of Carman (1997).

The question remains whether one or more linkage groups are involved in duplications in the apomictic genome, as apomixis has been shown to be inherited in a Mendelian pattern in some taxa (Richards, 2003). The B chromosomes of apomictic *Boechera* (Sharbel et al., 2004) might harbour duplicated regions, although their DNA sequence homology with *B. stricta* (Sharbel et al., 2004; Kantama et al., 2007) is not consistent with the 'parent of origin' effect involving *B. holboellii* alleles [Figure 4(c)]. Furthermore, B chromosomes are not found in all apomictic accessions (Sharbel et al., 2005), and hence duplication could also entail autosomal regions. Phylogenetic analyses have shown *B. holboellii* and *B. divaricarpa* to have paraphyletic Pleistocene origins (Dobeš et al., 2004a), and hence the 'parent of origin' effect may stem from taxa other than *B. holboellii*. Nevertheless, the ovule transcriptomes of apomictic *Boechera* lineages continue to show residual expression patterns reflective of hybridization after many generations of asexuality. Our analyses of diploid apomeiotic ovules have thus implicated both hybridization and gene duplication in the switch from sexual to apomictic seed formation, and have revealed unexpected levels of variation in the asexual genome.

Acknowledgements

We thank I. Schubert, J. Carman and E. Schranz for valuable comments on the manuscript. We thank the apomixis research group for enlightening discussions and technical support; and T. Mitchell-Olds, B. Roy and E. Schranz for seed material. This study was supported by funding to TFS from the IPK and Land Sachsen Anhalt, in addition to a "Pakt für Forschung und Innovation 2007" grant from the Leibniz Association.

Author Contributions: T.F.S. designed and coordinated the project, analysed data and wrote the manuscript. M.-L.V. performed seed screens on all material. M.-L.V., J.M.C.G., A.V, and J.K. collected the microdissected ovules. J.M.C.G. performed RNA extractions and RT-PCR analyses. T.T. analysed the SuperSAGE data for allelic variation and performed virtual Arabidopsis SuperSAGE analysis. H.V. performed all RNA extractions and cDNA amplifications for the SuperSAGE and 454 analyses. B.R. generated the SuperSAGE data. All authors discussed results and provided feedback on the manuscript.

LITERATUR CITED

Adams KL, Cronn R, Percifield R and Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci U S A 100: 4649-4654.

Adams KL and Wendel JF. 2005. Allele-specific, bidirectional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. Genetics, 171, 2139-2142.

Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. J Hered, 98, 136-141.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature, 437, 1149-1152.

Audic S and Claverie JM. 1997. The significance of digital gene expression profiles. Genome Res, 7, 986-995.

Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA and Jones SJ. 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics, 7, 246.

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

Besemer J and Borodovsky M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucl. Acids Res., 33, W451-454.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Det Kongelige Danske Videnskabernes Selskab, 6, 1-59.

Burge C and Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. J Mol Biol, 268, 78-94.

Camacho JPM, Sharbel TF and Beukeboom LW. 2000. B chromosome evolution. Phil. Trans. R. Soc. Lond. B, 355, 163-178.

Carman JG. 1997. Asynchromous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biological Journal of the Linnean Society, 61, 51-94.

Chen K and Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet, 38, 1452-1456.

Comai L, Madlung A, Josefsson C and Tyagi A. 2003. Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? Philos Trans R Soc Lond B Biol Sci, 358, 1149-1155.

Conesa A and Gotz S. 2008. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics, 2008, 619832.

Corral JM, Piwczynski M and Sharbel TF. 2009. Allelic sequence divergence in the apomictic *Boechera holboellii* complex. In Lost Sex, the Evolutionary Biology of

Parthenogenetic Reproduction. (Schön I, Martens K and Van Dijk PJ. eds). Berlin: Springer Verlag, p. In press.

Curtis MD and Grossniklaus U. 2007. Amphimixis and apomixis: two sides of the same coin! Vienna: A.R.G. Gantner Verlag K.G.

Dobeš C, Mitchell-Olds T and Koch MA. 2004a. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, A. xdivericarpa, and *A. holboellii* (Brassicaceae). Molecular Ecology, 13, 349-370.

Dobeš C, Mitchell-Olds T and Koch MA. 2004b. Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera x divaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers-an integrative approach. Am. J. Bot., 91, 2087-2101.

Dobeš C, Koch MA and Sharbel TF. 2006. Embryology, karyology, and modes of reproduction in the North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Ann. Missouri Bot. Gard., 93, 517-533.

Dobeš C, Sharbel TF and Koch M. 2007. Towards understanding the dynamics of hybridization and apomixis in the evolution of the genus *Boechera* (Brassicaceae). Syst Biodiversity, 5, 321-331.

Dobeš CH, Mitchell-Olds T and Koch MA. 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. x divaricarpa*, and *A. holboellii* (Brassicaceae). Mol Ecol, 13, 349-370.

Emrich SJ, Barbazuk WB, Li L and Schnable PS. 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research, 17, 69-73.

Eveland AL, McCarty DR and Koch KE. 2008. Transcript Profiling by 3'-Untranslated Region Sequencing Resolves Expression of Gene Families. Plant Physiol., 146, 32-44.

Garneau NL, Wilusz J. and Wilusz CJ. 2007. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol, 8, 113.

Goldsworthy SM, Stockton PS, Trempus CS, Foley JF and Maronpot RR. 1999. Effects of fixation on RNA extraction and amplification from laser capture microdissected tissue. Mol Carcinog, 25, 86-91.

Grimanelli D, Leblanc O, Perotti E and Grossniklaus U. 2001. Developmental genetics of gametophytic apomixis. Trends Genet, 17, 597-604.

Grimanelli D, Perotti E, Ramirez J. and Leblanc O. 2005. Timing of the maternal-to-zygotic transition during early seed development in maize. Plant Cell, 17, 1061-1072.

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

Grossniklaus U, Nogler GA and van Dijk PJ. 2001. How to avoid sex: the genetic control of gametophytic apomixis. Plant Cell, 13, 1491-1498.

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA and Smith OS. 2004. Allelic variation of gene expression in maize hybrids. Plant Cell, 16, 1707-1716.

Huse S, Huber J, Morrison H, Sogin M and Welch D. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology, 8, R143.

Kantama L, Sharbel TF, Eric Schranz M, Mitchell-Olds T, De Vries S and De Jong H. 2007. Diploid apomicts of the *Boechera holboelli* complex display large scale chromosome substitutions and aberrant chromosomes. Proceedings of the National Academy of Sciences of the United States of America, 104, 14026-14031.

Kearney M. 2005. Hybridization, glaciation and geographical parthenogenesis. Trends Ecol Evol, 20, 495-502.

Koch MA, Dobeš C and Mitchell-Olds T. 2003. Multiple Hybrid Formation in Natural Populations: Concerted Evolution of the Internal Transcribed Spacer of Nuclear Ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Mol Biol Evol, 20, 338-350.

Koltunow AM. 1993. Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell, 5, 1425-1437.

Koltunow AM and Grossniklaus U. 2003 Apomixis: a developmental perspective. Annu Rev Plant Biol, 54, 547-574.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG and Hartl DL (2005) Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. Genetics, 171, 1813-1822.

Marais G, Charlesworth B and Wright SI. 2004 Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. Genome Biol, 5, R45.

Matsumura H, Bin Nasir KH, Yoshida K, Ito A, Kahl G, Kruger DH and Terauchi R. 2006. SuperSAGE array: the direct use of 26-base-pair transcript tags in oligonucleotide arrays. Nat Methods, 3, 469-474.

Matzk F, Meister A and Schubert I. 2000 An efficient screen for reproductive pathways using mature seeds of monocots and dicots. The Plant Journal, 21, 97-108.

McClintock B. 1984. The significance of responses of the genome to challenge. Science, 226, 792-801.

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

Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V and Martienssen RA. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet, 19, 141-147.

Pichot C, El Maataoui M, Raddi S and Raddi P. 2001. Surrogate mother for endangered *Cupressus*. Nature, 412, 39.

Ravi M, Marimuthu MPA and Siddiqi I. 2008. Gamete formation without meiosis in *Arabidopsis*. Nature.

Richards AJ. 2003. Apomixis in flowering plants: an overview. Philos Trans R Soc Lond B Biol Sci, 358, 1085-1093.

Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, Varhol RJ, Fjell C, Marra M, Jones S and Siddiqui A. 2007. DiscoverySpace: an interactive data analysis application. Genome Biol, 8, R6.

Roche D, Hanna WW and Ozias-Akins P. 2001. Is supernumerary chromatin involved in gametophytic apomixis of polyploid plants? Sexual Plant Reproduction, 13, 343-349.

Roy BA and Rieseberg LH. 1989. Evidence for apomixis in *Arabis*. The Journal of Heredity, 80, 506-508.

Roy BA. 1995. The breeding systems of six species of *Arabis* (Brassicaceae). American Journal of Botany, 82, 869-877.

Savidan Y. 2007. Apomixis in higher plants. In Apomixis: Evolution, Mechanisms and Perspectives (Hörandl E, Grossniklaus U, van Dijk P and Sharbel TF., eds). Rugell: A.R.G. Ganter Verlag K.G., pp. 15-22.

Schlueter JA, Scheffler BE, Schlueter SD and Shoemaker RC. 2006 Sequence conservation of homeologous bacterial artificial chromosomes and transcription of homeologous genes in soybean (*Glycine max* L. Merr.). Genetics, 174, 1017-1028.

Schneitz K, Hulskamp M and Pruitt RE. 1995. Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. The Plant Journal, 7, 731-749.

Schranz ME, Dobeš C, Koch MA and Mitchell-Olds T. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). Am. J. Bot., 92, 1797-1810.

Sharbel TF and Mitchell-Olds T. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity, 87, 59-68.

Sharbel TF, Voigt ML, Mitchell-Olds T, Kantama L. and de Jong, H. 2004. Is the aneuploid chromosome in an apomictic *Boechera holboellii* a genuine B chromosome? Cytogenet Genome Res, 106, 173-183.

Sharbel TF, Mitchell-Olds T, Dobes C, Kantama L and de Jong, H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic and Genome Research, 109, 283-292.

Shyu AB, Wilkinson MF and van Hoof A. 2008. Messenger RNA regulation: to translate or to degrade. Embo J, 27, 471-481.

Siena L, Sartor M, Espinoza F, Quarin C and Ortiz J. 2008. Genetic and embryological evidences of apomixis at the diploid level in *Paspalum rufum* support recurrent auto-polyploidization in the species. Sexual Plant Reproduction, 21, 205-215.

Smyth DR, Bowman JL and Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. Plant Cell, 2, 755-767.

Spillane C, Curtis MD and Grossniklaus U. 2004. Apomixis technology development - virgin births in farmers' fields? Nat Biotech, 22, 687.

Springer NM and Stupar RM. 2007. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell, 19, 2391-2402.

Sturzenbaum SR and Kille P. 2001. Control genes in quantitative molecular biological techniques: the variability of invariance. Comp Biochem Physiol B Biochem Mol Biol, 130, 281-289.

Udall JA, Swanson JM, Nettleton D, Percifield RJ and Wendel JF. 2006. A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics, 173, 1823-1827.

Van Dijk P and Vijverberg K. eds. 2005. The significance of apomixis in the evolution of angiosperms: a reappraisal. Plant species-level systematics: new perspectives on pattern and process. Ruggell: A.R.G. Gantner Verlag.

Vaucheret H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. Genes Dev, 20, 759-771.

Voigt ML, Melzer M, Rutten T, Mitchell-Olds T and Sharbel TF. 2007. Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. Vienna: A.R.G. Gantner Verlag K.G.

Wang SM. 2007. Understanding SAGE data. Trends Genet, 23, 42-50.

Weber APM, Weber KL, Carr K, Wilkerson C and Ohlrogge JB. 2007 Sampling the *Arabidopsis* Transcriptome with Massively Parallel Pyrosequencing. Plant Physiol., 144, 32-42.

William Birky Jr, C. 1996. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics, 144, 427-437.

Supplementary data

Supporting information Table 1. (a) BLAST results of *Arabidopsis* sequences corresponding to 8 heterochronically expressed *Boechera* SuperSAGE tags.

Supporting information Table 2. Number of *Boechera* sequences corresponding to three SuperSAGE tag classes (see Table 4).

Supporting information Figure 1. Regression analysis of relative expression (RT-PCR) levels of sex- (grey) and apomeiosis-specific (black) genes (see supporting information Table 1).

Supporting information Figure 2. Gene ontology analysis of *Boechera* sequences corresponding to different SuperSAGE tag classes (see Table 4).

Apomixis initiation Candidates

Marie-Luise Voigt

Abstract

In Chapter III we illustrated a first understanding of apomixis expression and its complexity in the genus *Boechera.* 4 000 differentially expressed mRNAs were identified between sexual and apomictic ovules at the megaspore mother cell stage (MMC). In 2008 an additional SuperSAGE experiment was completed, looking at 4 stages of ovule development. In this chapter I would like to demonstrate how a combined analysis of both datasets yields a set of putative apomixis initiating candidate genes. Both datasets were generated without prior knowledge of genes or their function. In this chapter I first identified *Arabidopsis thaliana* genes known to be involved in reproduction. These selected *A. thaliana* genes were the basis for a sequence homology search between *A. thaliana* and *Boechera*. Identified sequences in *Boechera* were analysed with respect to their expression profiles and sequence homology between both SuperSAGE experiments. This study yielded in a list of putative apomixis candidates, which demonstrated that apomixis in *Boechera* might be influenced by chromatin remodelling, which could suppress or enhance gene transcription, and the cause for chromatin remodelling could be the heterozygous state of the apomicts. Those candidates were highly active at the time point, where the initiation of the apomixis process in *Boechera* is proposed.

KEY WORDS: apomixis, apomeiosis candidates, *Boechera holboellii*, *Arabidobsis thaliana*, SuperSAGE, 454 FLX technology, gametophyte development, transcription factor activity, chromatin remodelling

INTRODUCTION

In Chapter III we illustrated how the global expression analysis of microdissected ovules of apomictic and sexual individuals yielded an understanding of apomixis expression and its complexity in the genus *Boechera*. With the employment of a SuperSAGE (Matsumura et al., 2005) experiment we identified over 4 000 differentially expressed mRNA tags between the microdissected ovules of two sexual and two apomictic lines. Apomixis consists of three independent traits: i) the formation of unreduced gametes (*apomeiosis*), ii) the development of an embryo from an unfertitilized and unreduced egg cell (*parthenogenesis*), and iii) the formation of functional endosperm (*pseudogamy*). The approach was designed to uncover candidates involved in the first component of apomixis (apomeiosis).

Apomeiosis is the first component of apomixis, and consequently, the first component that differs from the sexual pathway. Cytological studies on *Boechera* (Böcher 1951; Naumova et al., 2001) have revealed that embryo sac formation in sexual plants is of the *Polygonum*-type. However, in apomictic plants it is of the *Taraxacum*-type. The megaspore mother cell (MMC) of apomictic plants fails to enter the reductional phase of meiosis (meiosis I) and proceeds with the second phase (meiosis II), leading to the formation of an unreduced dyad, instead of a reduced tetrad. One of the dyad cells degenerates, while the other continues, as in the case of the sexual pathway, with three mitotic steps. The mature embryo sac is morphologically-indistinguishable when compared between the sexual or apomictically-derived pathway (Naumova et al., 2001).

By dissecting ovules at the MMC stage in both sexual and apomictic plants, which corresponds to meiosis entrance (Schneitz et al., 1995), we searched for differentially expressed tag sequences (see chapter III) to identify apomixis initiation candidate genes. For further functional analyses, our goal was to identify consistently expressed genes between apomictic and sexual lines which are independent of technical and biological noise resulting from: (1) the technique (Stollberg et al., 2000; Colinge and Feger, 2001; Akmaev and Wang, 2004); (2) sample staging errors, and (3) the genetic background of the sampled lineages (genotype and environmental interaction influences phenotype and gene expression; Ungerer et al., 2003). Comparisons to *Arabidopsis thaliana* (*Boechera* is the closest wild relative; Koch et al., 2003) would facilitate the annotation of tag sequences and yield knowledge about their biological function and reduce introduced noise.

In 2008 we finished a 2nd SuperSAGE experiment (unpublished data) with the aim of analyzing gene expression though 4 stages of ovule development. To accomplish this task, one sexual and one apomictic plant were chosen and four defined ovule

developmental stages were isolated. The stages defined were based upon *Arabidopsis thaliana* ovule development (Schneitz et al., 1995; Mitchell-Olds and Schmitt, 2006) and compared with work from Böcher (1951) and Naumova (2001). As with our first SuperSAGE experiment (Sharbel et al., 2009), the mRNA tag sequence and copy number were obtained, and their expression changes were examined through time. With the discrete time points of ovule development, the gene-expression patterns could be characterized and aid in the construction of gene networks.

In this Chapter, I will demonstrate an approach whereby data from two independent high-throughput datasets are combined to yield in a set of putative apomixis initiation candidates. The advantage of the 1st SuperSAGE experiment was that the comparisons were made within and between 2 sexual and 2 apomictic lines, although the focus on one developmental stage meant that the experiment was limited by developmental single time point. By combining the data from the 2nd SuperSAGE and 1st SuperSAGE experiments, gene-expression was additionally analysed as a function of time.

The datasets were independently generated without prior knowledge of gene annotation. To examine function, we first identified *Arabidopsis thaliana* genes known to be involved with reproduction. The genera *Boechera* and *Arabidopsis* are close relatives (diverged ~10 mya, Koch 2003), and therefore, comparative genomics studies of traits between the *Arabidopsis* wild relatives can be carried out (Mitchell-Olds and Schmitt 2006; Schranz et al. 2006; Windsor et al. 2006; Mitchell-Olds et al. 2007). In the following analyses, I attempted to gain deeper insight into the transcriptomic differences between the sexual and apomictic developmental pathways, and thus, I asked the following questions: (1) how many of the genes show sequence homology between *Arabidopsis thaliana* and *Boechera*; (2) how many of the selected genes are expressed in the 1st SuperSAGE experiment; and (3) how are their expression-profiles characterized over time?

METHODOLOGY

Strategy. At first, a search of the Gene Ontology database (www.geneontology.org; Gene Ontology Consortium 2000) was performed for genes known to be involved with reproduction in *Arabidopsis thaliana* (status July 2008; Figure 1). The homology between *Arabidopsis thaliana* protein sequences of genes and two *Boechera* cDNA libraries, generated using 454 technology (454SEX & 454APO; Sharbel et al. 2009) were assessed using TBLASTN (Altschul et al. 1990) and the following criteria: E = 1e-20, minimum percent identity = 70%, alignment length >50bp, number of gaps <3, and Bit-score >120. The resulting cDNA sequences from each 454 library, were then used to map mRNA tags

from the 2 apomictic and 2 sexual SuperSAGE libraries (i.e. SuperSAGE experiment 1; Sharbel et. al 2009) with BLASTN (Altschul et al. 1990) and the following criteria: E = 1e-20, minimum percent identity of 100%, and a score of 52. Mapping SuperSAGE tag sequences resulting from this analysis were subsequently searched for in the 2nd SuperSAGE database, and only tags with 100% identity (homologs) were further studied (Figure 1).

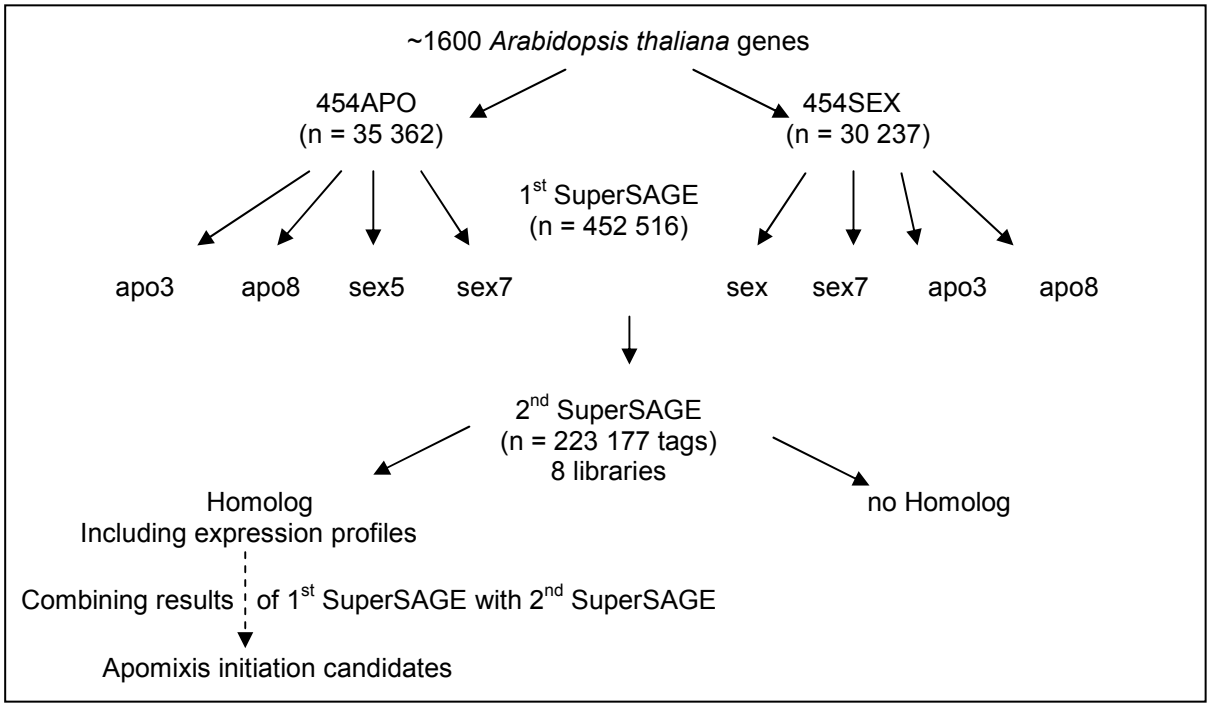


Figure 1. Overview of the analysis pipeline (n = number of sequenced reads per library).

Databases. To enhance reliable genome annotation of short sequence fragment reads (Lu et al., 2004) from serial analysis of gene expression experiments (SuperSAGE; Matsumura et al., 2005), two *Boechera* cDNA libraries were sequenced using 454 (FLX) technology (Sharbel et al., 2009). For each cDNA library, flowers from stage 1 to 12 of development (Smyth et al., 1990) were pooled from three diploid sexual plants (from now on referred to as the "454SEX" library) and three apomictic plants (from now on referred to as the "454APO" library). The generated cDNA libraries were normalized as described in Sharbel et al. (2009).

We performed two SuperSAGE experiments (Table 1; Figure 2). In the first experiment, 4 libraries were generated from two sexual and two apomictic plants from a single ovule developmental stage (Stage II; Sharbel et al., 2009). In the second SuperSAGE experiment, live ovules were microdissected from one sexual (diploid *Boechera holboellii*) and one apomictic (diploid *Boechera holboellii*) plant (unpublished data, Supplementary data A) from 4 developmental stages: 1I-1II protrusion arises, 2II-2IV megaspore mother cell (MMC) and meiosis entrance, 2V-3I tetrad to mono-nuclear embryo sac and 3IV-4IV

tetra-nuclear to mature embryo sac; (Schneitz et al., 2000; Figure 2). The SuperSAGE libraries were generated as in Sharbel et al. (2009), with the company GenXPro (Frankfurt am Main, Germany).

Table 1. Number of different SuperSAGE tag sequences per libraries and per stage of ovule development amount of tag sequences (Supplementary data A).

	Library	Stages			
		I.	II.	III.	IV.
1st SuperSAGE	Sex5		77 713[a]		
	Sex7		67 433		
	Apo3		61 692		
	Apo8		133 040		
2nd SuperSAGE					
	Sex	19 494	27 014	15 324	19 174
	Apo	41 487	36 576	25 415	38 693

[a] Number of individual tag sequences in particular library.

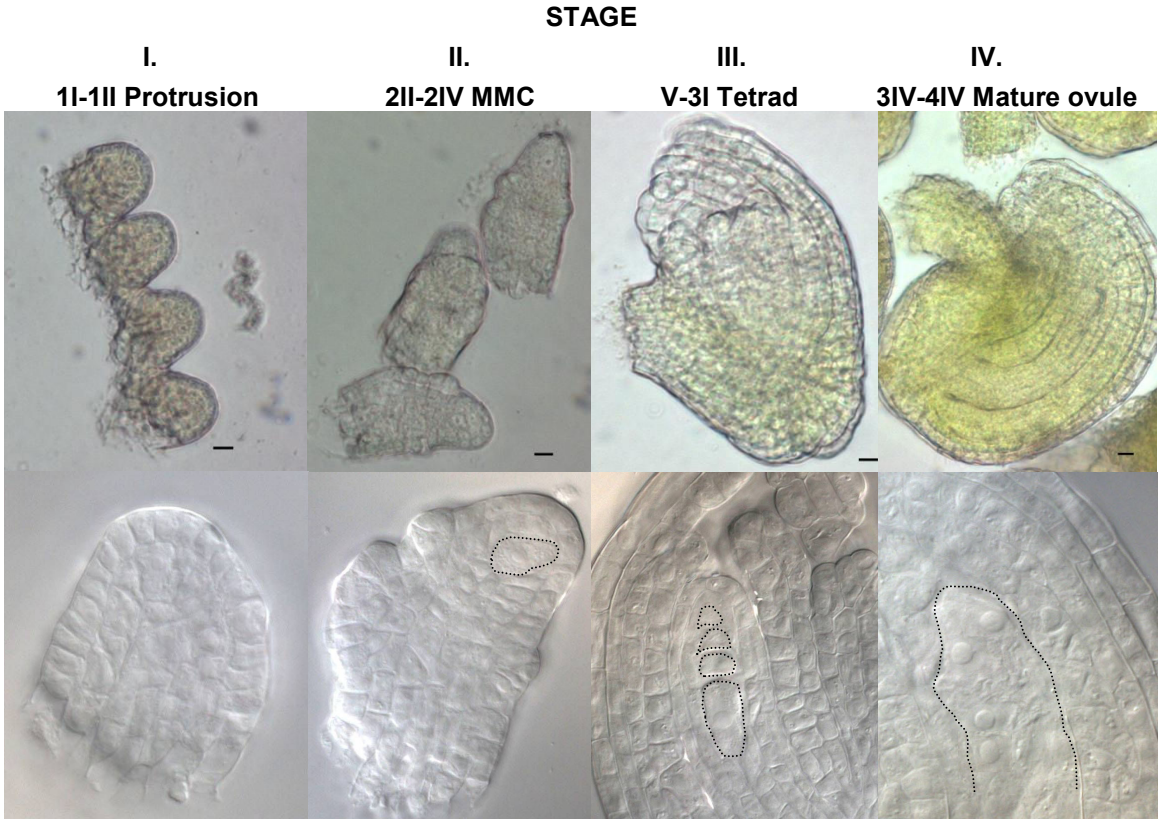


Figure 2. Pictures of isolated ovules at each developmental stage from a sexual diploid line (all bar scales = 10µm). Dotted lines illustrate surrounding of a cell (Stage II and III) or in stage IV embryo sac.

Data analysis. To calculate whether the expression profile of a gene was significantly different between two libraries, the tag copy numbers were corrected for the total number of sequenced tags per library and for multiple tests using the DiscoverySpace4.0 software (Robertson et al. 2007), which incorporates the method of Audic and Claverie (1997). A P-value < 0.01 was considered to show significant differential expression for any SuperSAGE tag between any two samples.

Candidate genes were extracted in several ways: i) in 1st the first SuperSAGE experiment any tag showing significant differences between any two (sex & apo; apo & apo) libraries were selected, ii) in the 2nd SuperSAGE experiment any tag showing significant differential expression in at least one stage between sexual and apomictic sample was selected, iii) sex- or apomixis-specific tags were selected as candidates, if they appeared in both SuperSAGE data sets, and iv) rare transcripts which appeared in both SuperSAGE data sets.

To illustrate expression profiles for selected candidates, the software VANTED - Visualization and Analysis of Networks containing Experimental Data, was employed (Junker et al., 2006). To calculate the over-representation of selected candidates (one-tailed), for small sample sets a Fisher's exact test was applied and for large data points *Chi 2 – test* (SPSS, version 11.5).

The free available software Cytoscape (version 2.6.2.; Shannon et al., 2003) with plugin BINGO (Maere et al., 2005) visualized over-representation of Gene Ontology category. Over-representation was calculated using a hypergeometric test with multiple testing correction using Benjamini & Hochberg's FDR and chosen significant level of P = 0.05 with complete annotation (from Gene Ontology) as reference.

RESULTS

1. Identification of Arabidopsis homologs in Boechera.

Gene Ontology has three organization groupings: cellular component, biological process and molecular function. I restricted the search to biological process and selected terms related to reproduction, and additionally selected one molecular function term "Transcription factor" (explanation see discussion). In total 1616 genes were selected (Supplementary data B). For a chosen Gene Ontology term (GO term) filtered for *Arabidopsis thaliana* between 12 and 772 genes were found (Table 2).

Table 2. List of Gene Ontology terms (Gene Ontology Consortium 2000; July 2008) used to identify genes associated with reproduction in *A. thaliana*.

Term No.	Gene Ontology terms	GO number	Number of genes
1	Cell cycle	GO:0007049	177
2	Floral organ development	GO:0048437	86
3	Gametophyte development	GO:0048229	101
4	Negative regulation of flower development	GO:0009910	28
5	Negative regulation of transcription	GO:0016481	44
6	Ovule development	GO:0048481	36
7	Positive regulation of flower development	GO:0009911	27
8	Positive regulation of transcription	GO:0045935	24
9	Post-transcriptional regulation of gene expression	GO:0010608	61
10	Regulation of flower development	GO:0009909	75
11	Regulation of gene expression, epigenetics	GO:0040029	94
12	Transcription factor	GO:0003700	772
13	Sequence specific DNA binding	GO:0043565	32
14	Sexual reproduction	GO:0019953	50
15	Endosperm development	GO:0009960	12

A *Boechera* flower transcript (cDNA) sequence of significant homology was ascertained for approximately 70% (1 131) of the selected genes from *A. thaliana*. In the 454APO database, 59% of genes (952 of the 1 616 of *A. thalianas* selected genes, Supplementary data C) were identified, which represented 2.9% (1 050) of the total number of cDNA sequences (total 35 362 cDNA contigs). For the 454SEX database, 52% of genes (841 of the 1 616 of *Arabidopsis* genes; Supplementary data D) were identified, which represented 2.5% (756) of the total number of cDNA sequences (total 30 237 cDNA contigs). There were significant overlaps of sequences found in both the 454SEX and 454APO libraries, with a total of 658 *Boechera* transcripts representing 40% of all identified *A. thaliana* sequences. This resulted in 292 genes uniquely identified in 454APO and 181 genes uniquely identified in 454SEX. Figure 3 details the results per GO term (Term No. 1-15, Table 1). To test whether GO terms corresponding to any of the uniquely identified genes (473) were over-represented in either the 454APO or 454SEX library, a *chi2 test* calculated an over-representation of GO term 3 (Gametophyte development) in the 454APO library (P=0.002) (Supplementary data E).

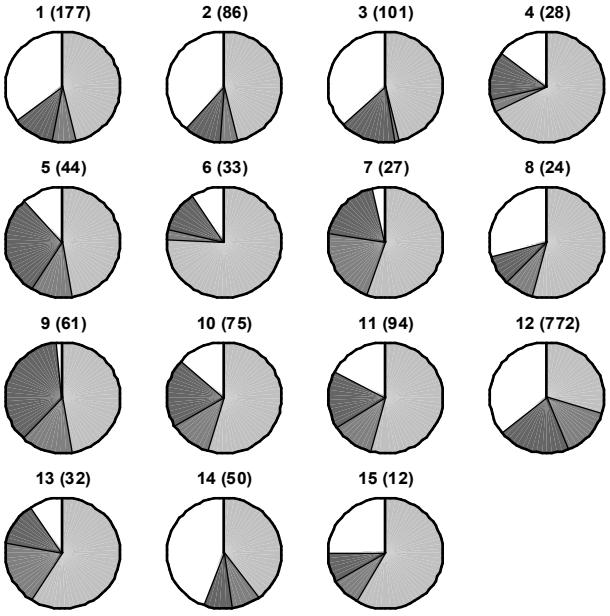


Figure 3. Each pie represents percentage of gene annotations among the cDNA flower libraries per GO term (the first number represents the GO term, while the numbers in brackets represent the total gene number; see Table 1). Light grey – *A. thaliana* genes identified in both libraries (454SEX and 454APO), grey - *A. thaliana* genes found only in the 454SEX library, black - *A. thaliana* genes found only in the 454APO library, and white - number of *A. thaliana* genes showing no significant homology to either database.

2. Boechera cDNA transcript homology with 1st SuperSAGE experiment.

The resultant 756 454SEX cDNA sequences were used as a query database and mapped against the four SuperSAGE libraries from Sharbel et al. (2009). The results for the four SuperSAGE libraries were merged and 178 cDNA transcripts (23.5% of the resultant 756 sequences) were extracted, corresponding to 250 unique tag sequences (Figure 4, blue).

The extracted 1050 454APO sequences were used as a query database and blasted against the four SuperSAGE libraries (Sharbel et al., 2009). In combining the results among the four SuperSAGE libraries for the 200 cDNA transcripts (19.04% of the resultant 1050) representing 264 unique tag sequences were identified (Figure 4, green).

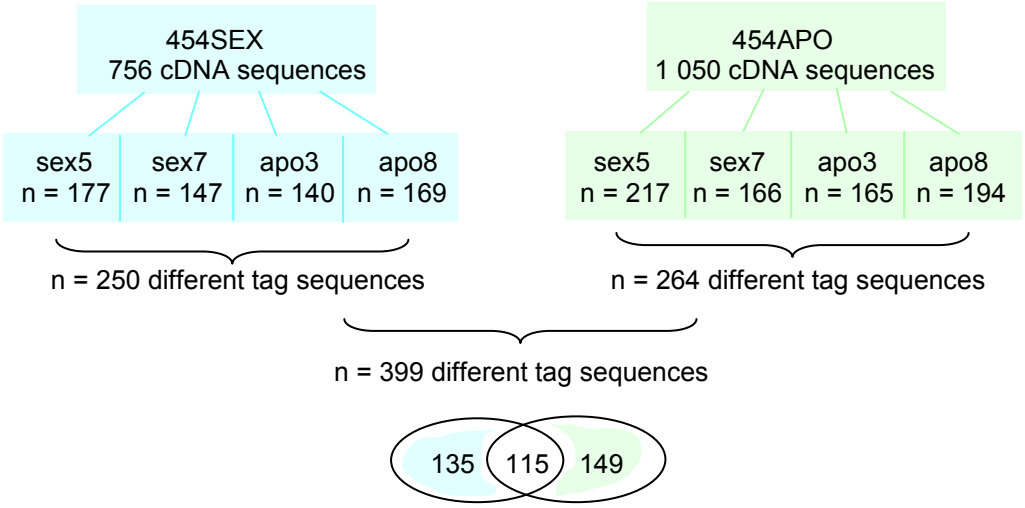


Figure 4. Outline of the mapping process followed for obtaining the 399 unique tag sequences used in further analysis. Below a Venn-diagram illustrating shared and unique identifies tag sequences between 454 libraries.

In searching for sequence homology between both the 454SEX and 454APO resultant tag sequences, 399 unique SuperSAGE tag sequences were identified (Supplementary data F). Of these, 115 tags were found in both cDNA libraries, 135 tags were exclusively found in 454SEX, and 149 tags were 454APO-specific (Figure 4, Venn-diagram).

2.1 A. thaliana candidate gene homologs identified in the 1st SuperSAGE experiment.

Examination of the expression profiles (i.e., SuperSAGE tag copy number per library) of the resultant 399 SuperSAGE tags demonstrated that 191 tags were uniquely expressed in one library. Of these, 155 tags were singletons (i.e., only one copy of the tag was found in all 4 libraries). Forty-nine (20.1 %) of the remaining 244 tags (not including singletons) were differentially expressed (Audic and Claverie, 1997) in at least one apomixis-sex comparison, and 2 additionally were exclusively differentially expressed between the two apomictic libraries. No exclusively differentially expressed tags were found between the two sexual libraries, although 24 differentially expressed tags were identified which were also differentially expressed in at least one apomixis-sex comparisons (Supplementary data G). These 51 tags, considered as our first set of candidates, contain one apomixis-specific (in both apomictic libraries, but not in any sexual one) and three sexual-specific tags (Supplementary data F). The categorization of the biological function of the 51 candidates demonstrated 2 major groups, *transcription factor* activity (Fisher's exact test over-representation $P < 0.01$) and *cell cycle* genes (over-representation $P < 0.01$) (Supplementary data E).

3. *Homologous tags in 1st and 2nd SuperSAGE datasets*

3.1 Singletons.

The 399 different SuperSAGE tag sequences identified from the 1st SuperSAGE experiment (see results section 2; Fig. 4) were used in a sequence homology search against the 2nd SuperSAGE experiment. In total, 237 (59.4 %) had perfect matches (Supplementary data H), and of the remaining 162 tags, 104 were singletons in the 1st SuperSAGE experiment. From the remaining 51 singletons identified in the 1st SuperSAGE experiment, 19 also occurred as singletons in the 2nd SuperSAGE experiment. The remaining 32 of these tags were no longer singletons in the 2nd SuperSAGE experiment, and 7 were significantly differentially expressed in at least one stage of ovule development between the sexual and apomictic samples. Of the identified 237 tags in the 2nd SuperSAGE experiment, 35 were singletons, of which 16 were not singletons in the 1st SuperSAGE experiment. Tag sequences which were in one SuperSAGE experiment identified as a singleton, but with identified higher copy number (>2) in the other SuperSAGE were classified as rare transcripts (Table 3) as one group of candidates.

Table 3. Summary of distribution of the 399 mapped tag sequences of 1st to 2nd SuperSAGE datasets.

		1st SuperSAGE		
		Singleton	Non-Singleton	total
2nd SuperSAGE	Singleton	19	16[a]	35
	Homolog	32[a]	170	202
	No homology	104	58	162
	total	155	244	399

[a] Considered as rare transcript candidates.

3.1 Apomixis and sex-specific tags.

63 tags were classified as *apomixis-specific* in the 2nd SuperSAGE experiment, appearing in one to four of the 4 apomictic libraries (i.e., stages) but never in any of the four sexual libraries. To confirm apomixis specificity, a search was made against the datasets from the 1st SuperSAGE experiment. Nineteen were found within in at least one apomictic library of the 1st SuperSAGE and not in the sexual libraries. Of the 44 tags which occurred among any sexual library, 27 of which were found exclusively in the *Boechera stricta* (sex5) library.

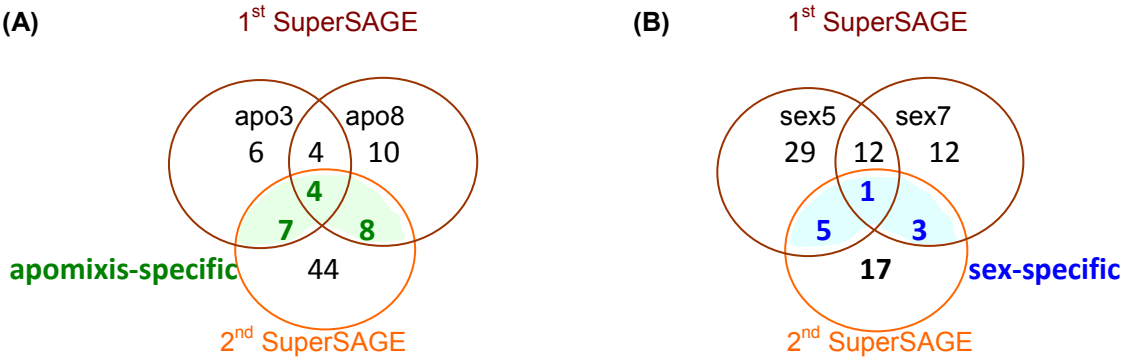


Figure 5. Venn diagram of shared tag sequences for identification of (A) apomixis-specific (green), and (B) sex-specific (blue) tags.

In the 1st SuperSAGE experiment, 8 apomictic-specific tag sequences were identified in both apomictic libraries, and one was among the 51 candidates of the 1st SuperSAGE (section 2.1). But in contrast this one apomixis-specific tag was identified in 3 (sex1; sex3; sex4) of the 4 sex libraries of the 2nd SuperSAGE experiment. Together, the expression of the 19 tags can be considered *apomixis-specific* (green numbers; Figure 5A, Supplementary data I), although only one was significantly differentially expressed from the sexual sample in three stages (stage I; III; IV) ($P < 0.01$).

Twenty six tags were classified as sex-specific in the 2nd SuperSAGE experiment, since they were only found among the four sexual libraries (i.e., stages). Seventeen of these tags were found in at least one apomictic library from the 1st SuperSAGE experiment. Furthermore, 13 sex-specific tag sequences were identified in the 1st SuperSAGE experiment, but only one was confirmed to be sex-specific among the 2nd SuperSAGE data. Together, nine tags were *sex-specific*, but none was significantly differentially expressed from the apomictic samples (Figure 5B, Supplementary data I).

3.2 2nd SuperSAGE candidates.

The 237 identified tag sequences and their expression profiles (copy number in SuperSAGE libraries) were compared between stages for the sexual and apomixis libraries. Significant differences in the expression profile ($P < 0.01$; Audic and Claverie, 1997) were found for tags at each stage: I (n = 47), II (n = 49), III (n =18) and IV (n = 23) tags (Figure 6). In total, 76 tags revealed significant differential expression between the sexual and apomictic sample in at least one developmental stage, and these were the selected candidates from the 2nd SuperSAGE experiment.

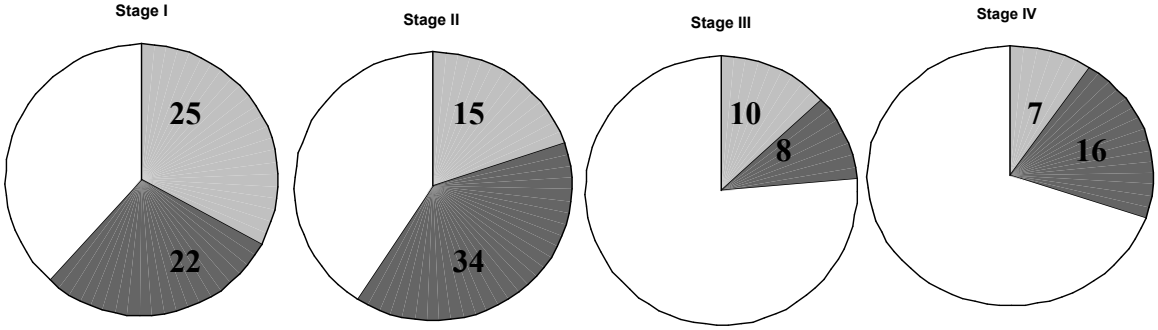


Figure 6. Pie-charts illustrating gene expression differences per ovule developmental stage of the identified 76 candidates; white similar expression between sexual and apomictic sample; grey apomictic sample is up-regulated; and black apomictic sample is down-regulated.

In stage I and stage II, 62 % (n = 47) of the candidates were differentially expressed between the apomictic and sexual sample (Figure 6). 29 % (n = 22) were down-regulated in the apomictic sample in stage I and 45 % (n = 34) in stage II. The categorization of the 76 candidates, according to biological function (GO term Table 1), demonstrates transcription factor activity as the major group (Fisher's exact test over-representation $P < 0.01$; Supplementary data E).

Among the gene expression profiles of the 76 candidates over the 4 developmental stages, similar expression patterns were recognized (Figure 7, different colors indicate different groups A to I), although no overrepresentation of any Gene Ontology term was identified within any group (Fisher's exact test $P > 0.05$). Group A contains six profiles, whereby the apomixis sample was exclusively expressed, although

only one candidate is apomixis-specific (not found in any of the 6 sexual SuperSAGE libraries). Group B was characterized by 13 profiles characterized by higher expression in the sexual sample at stage I, with similar expression between the sexual and apomictic sample in the following stages. Group C showed a significant differential expression in stage 4. Group D was characterized by differential expression between the sexual and apomictic sample in stages I and II, with similar expression levels in the following stages. Group E displayed a constantly higher expression of the apomictic, compared to the sexual sample, over the four time stages (except in one case, profile 13-2).

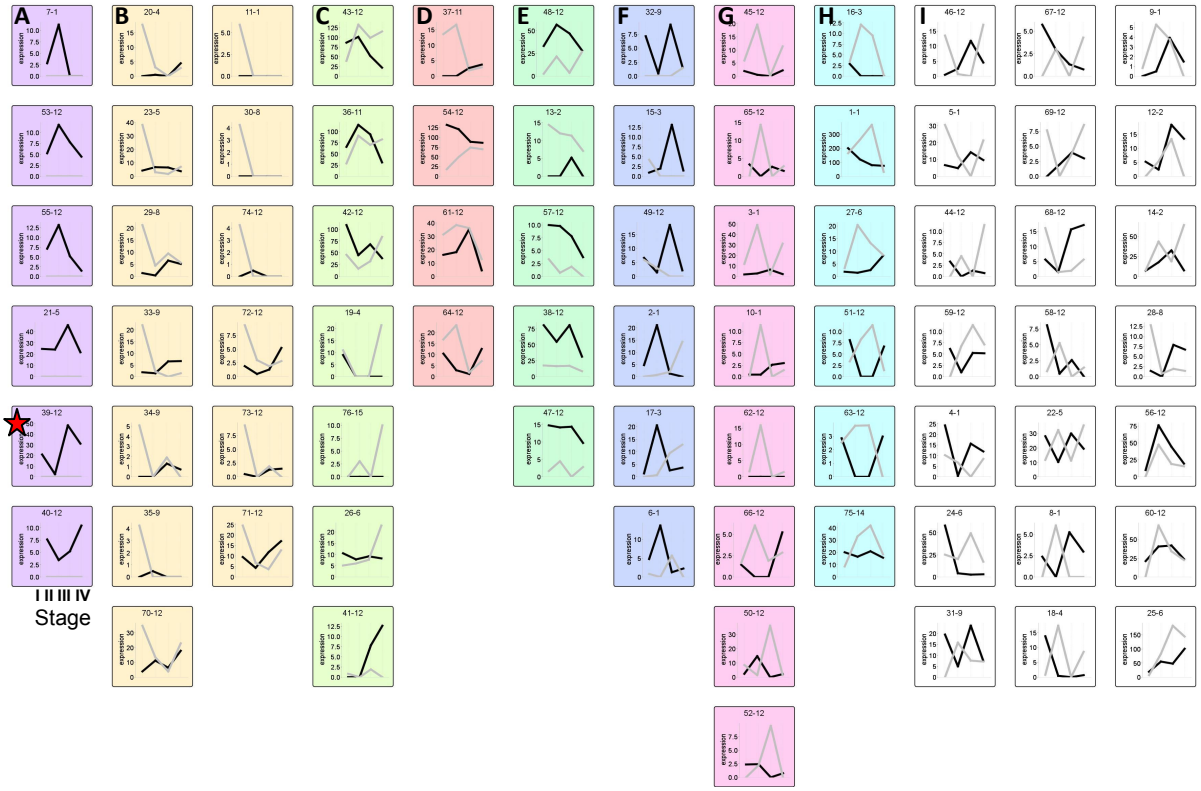


Figure 7. Gene expression profiles for 76 selected candidates of 2nd SuperSAGE are illustrated. Over four developmental time points (stages 1 to 4 on x-axis), the normalized SuperSAGE tag number of sexual (grey) and apomixis samples (black) are plotted. Colors of graphs demonstrate candidates having similar patterns as explained in text, same color correspond to same letter. The red star indicates the apomixis-specific gene (see text). The number code within each graph is the "gene number – GO term No." (corresponding to Table 1).

Group F showed elevated apomixis expression levels at stages II or III. Group G was characterized by higher expression levels in the sexual sample at stages II or III. Group H was characterized by a higher expression level of the sexual sample in stages II and III, with the apomictic sample showing a low level of expression in stages II and III. The remaining graphs revealed no specific pattern and were not assigned to any specific group.

4. *Comparison of 1st SuperSAGE candidates with 2nd SuperSAGE candidates.*

The search for tag sequence homology between selected candidates of both SuperSAGE experiments (see section 2.1 and 3.2) resulted in 30 tag sequences (Figure 8, in green) in which the GO term "Transcription factor activity" was over-represented ($P < 0.01$) (Supplementary data E; I). Those 30 tag sequences were considered as our core apomixis candidates.

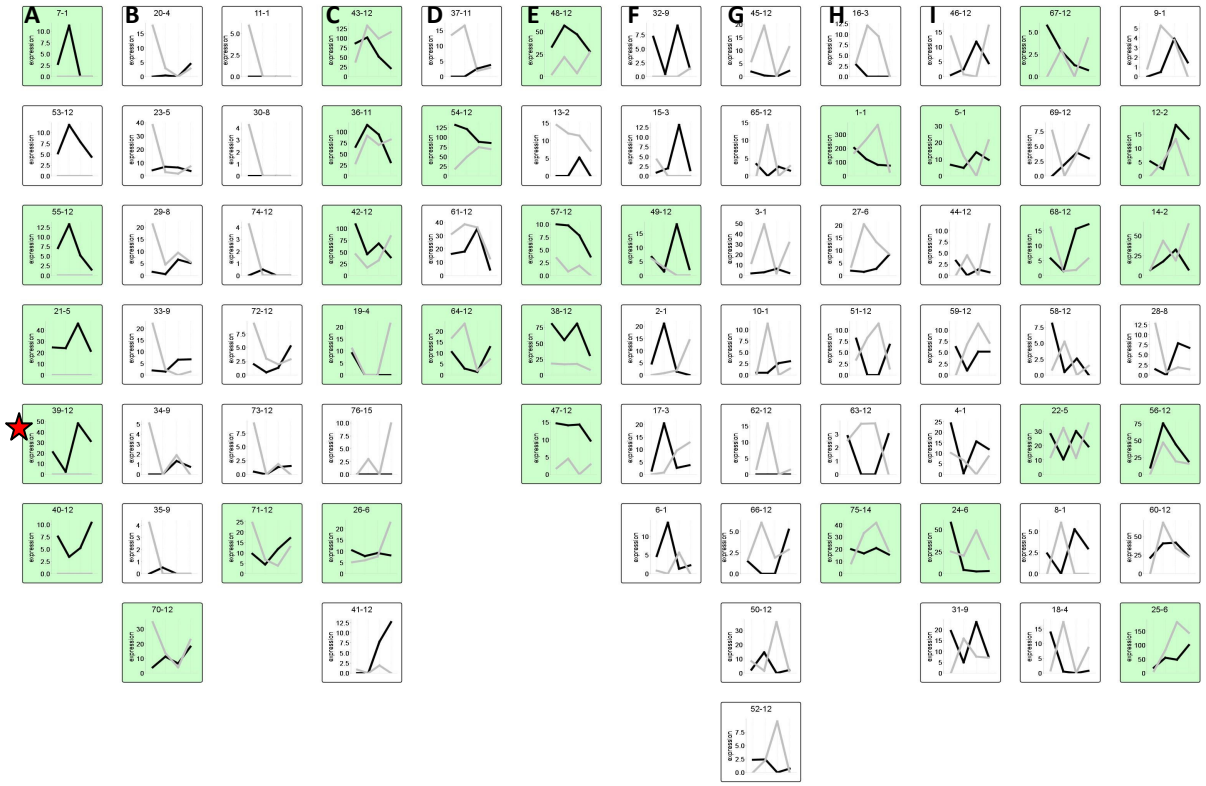


Figure 8. The order of gene expression profiles for 76 selected candidates of the 2nd SuperSAGE experiment is the same as Figure 8. Over the four developmental time points normalized expression copy number of sexual sample (grey) and apomixis sample (black) are shown. Graphs colored in green show the 30 candidates identified in both SuperSAGE experiments and the remaining in white display candidates with no sequence homolog within 51 candidates of 1st SuperSAGE experiment. The number code within each graph is the "gene number – GO term No." (corresponding to Table 1).

5. *Candidate groups.*

The minimum criteria in defining a candidate was its identification in at least one library of each SuperSAGE data set, and resulted in different groups: a) the 30 core tag sequences, which were identified in a sequence homology search between both candidates sets of the 2 SuperSAGE experiments; b) 19 apomixis-specific tags, which were not identified in any sexual library; c) 9 sexual-specific tags, which were not identified in any apomictic library; and d) 48 rare tags, which were identified in both SuperSAGE data sets, but with low tag copy number (see section 3.1, Table 3; Supplementary data I).

DISCUSSION

From Arabidopsis to Boechera.

Arabidopsis and *Boechera* belong to Brassicaceae, which is composed of approximately 340 genera and 3 400 species (Windsor et al., 2006). It has been estimated that the genus *Boechera* diverged from the genus *Arabidopsis* approximately 10 million years ago (Koch et al., 2000; Koch and Chatterjee, 2001; Koch et al., 2003). Members of this clade are the closest living relatives to *Arabidopsis* outside of the genus *Arabidopsis*. DNA sequence homology is high between *Boechera* and *Arabidopsis thaliana*, as has been shown by random shotgun sequencing of *Boechera stricta* where approximately 74% of 23 136 clone sequences demonstrated a high similarity to *Arabidopsis* (Windsor et al., 2006). Similar levels of homology were found in the work presented here, as our approach in blasting selected *Arabidopsis* genes against the two *Boechera* cDNA 454 libraries led us to the identification of 1 131 genes (i.e. about 70% overlap with a BLAST e-value < 10^{-28}).

Of the 1 131 genes, 658 (58%) were identified in both cDNA 454 libraries, while 473 (42%) were unique to the apomixis (292 of 473) or sexual (181 of 473) cDNA libraries. It was intriguing that among the 454APO library, more genes were identified than among the 454Sex library. This could reflect differences in sequence homology between libraries, differences in expression of particular genes between the 454 libraries, or gene duplication in the apomictic library (*sensu* Corral et al., 2009). The second possibility seems unlikely since both *Boechera* flower cDNA libraries were prepared in the same way, using flowers pooled from developmental stages 1 though 12 (Smyth et al. 1990) and thus the most relevant genes involved in flower development should have been caught. In terms of sequence divergence or weak sequence homology between *Boechera* and *A. thaliana* genes, two curious results stand out: i) selected *A. thaliana* genes for the Gene Ontology terms “Positive regulation of flower development” and “Post-transcriptional regulation of gene expression“ were identified to almost 100%; and ii) “Gametophyte development” gene sequences seem to be more conserved between *A. thaliana* and apomictic individuals (Figure 3, Supplementary data E).

The first curious result could be a consequence of the high conservation of post-transcriptional regulators, as has been shown in a study where 23 families of *Arabidopsis* miRNAs were tested, of which 11 were expressed in gymnosperms and 8 in ferns, thus indicating ancient origin (Axtell and Bartel, 2005). This is furthermore supported by sequence conservation of these regulators over long evolutionary distances (e.g from *Arabidospis* to moss), which diverged 400 million years ago (Arazi et al., 2005). In animals, where more data about miRNAs exists, 18-30 miRNA families seemed to be conserved in all bilaterians (Sempere et al., 2006), but sequence conservation over longer evolutionary distances (i.e. vertebrates to *Drosophila*) is rare (Chen and Rajewsky, 2007).

But even though high sequence conservation for certain biological processes exists, it remains intriguing that more genes were uniquely identified among the 454APO (apomictic) cDNA library, and that the particular GO term “gametophyte development” was characteristic of the 454APO library. Asuming that differential expression of certain genes between the sex and apomictic cDNA libraries is not the reason for this difference, weak sequence homology and/ gene duplication in the apomictic library (Corral et al., 2009) could account for it. The sexual *Boechera* used in 454library generation were by definition highly homozygous (Song et al., 2006), whereas apomictic *Boechera* are characterized by elevated heterozygosity (Chapter II). Apomictic *Boechera* are of hybrid origin (*B. stricta* X *B. holboellii*) and therefore alloploid (Roy, 1995; Dobeš et al., 2004a; Dobeš et al., 2004b; Sharbel et al., 2005; Kantama et al., 2007). Alloploidy can lead to novel changes in gene regulation, including gene silencing, including activation and /or methylation changes (Lee and Chen 2001; Wang et al. 2004; Comai, 2000; Xu et al. 2009). Such differences can arise from cis-and trans-regulatory effects, and can even lead to organ-specific gene expression whereby only one allele is expressed in a certain organ or tissue (subfunctionalization, Adams et al., 2003; 2004; Bottley at al., 2006).

There are two processes, mutation and recombination, that can change sequence information. It is predicted that apomicts (i.e. asexual genome) should accumulate mutation over time, whereby sexual organisms go though recombination which has the effect of segregating polymorphisms and purging deleterious mutations from a population. These two processes could have lead to divergent sequence information in the 454 cDNA libraries. If both mechanisms change sequence information it is still curious that we found more sequence homology between Arabidopsis and the 454APO library. The use of a triploid line within the apomictic cDNA flower library (Sharbel et al., 2009) may at least partially explain it. If a diploid apomict accumulates one mutation at a particular locus, it would still have one wild-type allele left, whereas, a triploid apomict would still have two wild-type alleles. Thus, assuming that most mutations would be deleterious, many of these would be lost in diploid apomicts, while they would be “masked” and hence still present in triploid apomicts. Even though diploid and triploids have the same rate of mutation accumulation , the chances that one wild-type allele stays unmodified is higher in triploids, and this may also extend to its expression. A gene expression experiment between diploid and triploid *Paragonimus westermani*, supports this idea, as a higher annotation success was documented with the triploid library (29.9%) compared with the diploid library (22.3%) (Kim et al., 2006). The predominance of genes involved with gametophyte development among the uniquely identified 454APO genes showing higher sequence homology to *A. thaliana* remains puzzling, as it is predicted that

gametophyte development genes would be fundamental for reproduction and hence under high conservation (i.e. selection pressure).

Boechera cDNA transcript homology with 1st SuperSAGE.

Serial analysis of gene expression (SAGE) was developed by Velculescu et al. (1995) to determine the absolute frequency of every mRNA transcript in a predefined tissue. The original SAGE tag sequence length was 14 bp, and improvements to the technique have led to increased tag lengths of 26 bp (called SuperSAGE, Matsumura et al., 2005), which increases the possibility of correct tag annotations (Lu et al., 2004). To increase reliable sequence annotation to the model species *Arabidopsis thaliana*, I blasted the selected *A. thaliana* genes against the two *Boechera* cDNA libraries generated with 454 FLX technology (Sharbel et al., 2009). Extracted *Boechera* cDNA matches were blasted against the SuperSAGE tags, and only homologs with a 100% match (26 bp) were chosen for further study (Torres et al., 2008). Our objective with the 1st SuperSAGE experiment was to examine transcriptome differences between sexual and apomictic ovules at the entrance of meiosis, and the identification of 399 sequence homologs between the cDNA flower libraries and the 1st SuperSAGE experiment was the basis for further analysis.

The reduction from the originally identified 1 131 cDNA sequences among the cDNA libraries to 399 SuperSAGE tags could be explained by i) the samples used for library construction, ii) the generation of libraries, and iii) the setting of 100% homology as criteria for successful homology between datasets. The two *Boechera* cDNA libraries were generated from pooled whole flowers, whereas the SAGE libraries were generated from very specific tissues (10 live microdissected ovules). *Boechera* cDNA sequences with no sequence homolog in the 1st SuperSAGE libraries were found to be equally distributed across the examined Gene Ontology terms, except for the GO term “Transcription factor” (GO term No. 12; Supplementary data E). It would be interesting to know whether under-representation of the identified “transcription factor” is a consequence of transcription factor sample-specificity (we used different samples for the generation of datasets, Supplementary data A), especially with respect to examining a very small time window of ovule development, or whether this is due to technical procedure or incomplete sequencing coverage of all transcripts.

An essential step for the work presented here was the mapping of SuperSAGE tag sequences to *Boechera* cDNAs. Both 454 sequencing and SuperSAGE have limitations. SuperSAGE experiments are based on *Nla III* restriction enzyme sites found within the 3’-untranslated region (UTR) of mRNAs (Pleasance et al., 2003), and thus transcripts with no *Nla III* recognition side will be missed. The bias toward the 3’UTR end can lead to difficulties in gene mapping, since not all provided sequences in public data sets (i.e., Genebank) contain 3’UTR end sequence information (Seki et al., 2002). In generating the

Boechera cDNA libraries, the advantage is that no enzyme restriction site is needed since all cDNA is randomly sheared. Both techniques nonetheless rely on sequencing success. The use of a new generation sequencing procedures for both methods (pyrosequencing platform, Branford, CT, USA), had the common advantage of avoiding cloning, although it has been shown to have a slightly higher error rate in comparison to Sanger sequencing (Moore et al., 2006, Torres et al., 2008). Furthermore, an under-representation of short (<~80bp) and long sequence fragments (>300bp – 400bp) has been shown (Torres et al., 2008), which means fragments smaller than 80bp or longer than 300bp will be missed (critical factor for *Boechera* cDNA library generation).

1st SuperSAGE versus 2nd SuperSAGE.

By comparing the 399 identified tag sequences extracted from the 1st SuperSAGE experiment with the 2nd SuperSAGE experiment, we can test reproducibility as well as artifacts introduced by the generation of the SuperSAGE libraries. Of specific interest are the detected singletons (i.e. tags of a particular sequence occurring once). The relevance of singletons is complex to estimate, and a classification into rare transcripts or artifacts would be useful. To discriminate between novel transcripts and artifacts, sequence complexity could be an indicator, for example if a sequence contains repeats or shows only one base pair difference to another (i.e. low complexity could reflect artifacts; Khattra et al., 2007). One insight into this was made by Sharbel et al. (2009), who showed that the frequency of tags having 1 to 8 bp differences were much higher relative to what was expected if there was random sequence variation, thus leading to the postulation that these were not artifacts. A second indicator for rare transcript could be its mapping to known genes (Akmaev and Wang, 2004; Khattra et al., 2007). This study is based upon annotation success of *Boechera* cDNAs to *A. thaliana* genes, and thus singletons mapping to particular *A. thaliana* genes may not be artifacts.

Candidates of 1st and 2nd SuperSAGE.

51 and 76 candidate tags were identified in the 1st and 2nd SuperSAGE experiments respectively. 90% of the identified candidates (of the 76) among the 2nd SuperSAGE showed significantly different expression between the sexual and apomixis samples in stages I and II, supporting our approach of having selected stage II in our first attempt to identify differentially-expressed genes involved in apomeiosis (Sharbel et al., 2009). The additional dataset of the 2nd SuperSAGE experiment illustrates the importance of measuring differences between samples over more than one time point, which is not sufficient for understanding underlying networks and complexity in gene regulation (Wang et al., 2008). For example, our data revealed that stage I was already characterized by major changes in gene expression between sexual and apomictic ovules (Figure 6).

Both SuperSAGE experiments were performed independently on different *Boechera* lines, an approach which we believed that would enable us to identify consistently differentially expressed candidates and circumvent sample-specific transcripts (i.e. transcriptomic noise). Strikingly, among the resultant sequences, an over–representation of the GO term "transcription factor" was found ($P < 0.01$, Supplementary data E). But how effective is it to compare and combine the candidates of both SuperSAGE experiments? An intriguing point is the high frequency of gene expression profiles with a normalized tag copy number greater than 30 (in at least at one developmental stage, Supplementary data J). Among the 76 candidates of the 2nd SuperSAGE experiment, 22 revealed this profile, and of those, 16 were identified again ($P < 0.001$; Fisher's exact test with 76 candidates as reference) among the 30 final candidates, suggesting that we have identified core apomixis initiation candidate genes. This is consistent with what has been shown for comparative experiments, for example with Gene-Chip (Affimetrix) and SAGE, where the expression of genes was mostly confirmed for those with high gene expression levels or greater changes in expression (Ishii et al., 2000; Kim et al., 2003). One is left to wonder how many low copy number/ rare transcripts are lost in such comparisons, as rare identified transcripts could also be included as candidates.

Considering the rare transcripts (Table 3) as well as sex- or apomixis-specific transcripts (Figure 5, whereby one is already among the core candidates) we have in total 105 different tag sequences, which interestingly mapped to 76 genes. Several tags seemed to map to the same gene, either to different gene regions or to multiple alleles. If several tags mapped to the same gene region the expression profile should be the same, but if they map to different alleles then their expression profiles could be different. We cannot differentiate between these possibilities based upon the data presented here.

Apomixis in Boechera.

This analysis was initiated using annotation information specific to *Arabidopsis thaliana* where, by definition, the biological process is known. It is unclear whether these annotations can be extended across species, and their involvement in biological processes in *Boechera* can only be inferred. At the beginning of the search I included the Gene Ontology *Molecular Function* term "transcription factor" in addition to all defined biological process terms involved in reproduction, as the purpose was to include as many possible gene expression regulators. In performing this approach I expected that non-relevant transcription factors would be excluded in the search, because we focused on the transcriptome of extracted ovules. In total thirteen different transcription factor families (Supplementary data L) were identified, and nine of these transcription factor families were found within the selected 30 core apomixis initiation candidates

(Supplementary data L). These 30 core apomixis candidates were selected on the basis of significant differential expression between the sexual and apomictic libraries from both SuperSAGE experiments. For four of these transcription factor families more that one member was identified.

One transcription factor stands out as it belongs to the Alfin-like family and was uniquely apomixis-specific in both SuperSAGE experiments (Figure 8, red star). Interestingly, a recent study (Lee et al., 2009) showed that this factor shows strong homology to Alfin1 identified in alfalfa (*Medicago sativa*) (Bastola et al., 1998). Alfin1 has long been considered to be a transcription factor, but it was based on the observations that it contains a zinc finger motif and exhibits DNA-binding activity (Bastola et al., 1998; Winicov and Bastola, 1999; Winicov, 2000). However, a recent study (Lee at al., 2009) showed that the zinc finger motif in Alfin1 is a canonical PHD finger in the *Arabidopsis* homolog. The homolog in *Arabidopsis AL* (alfin1-like) has been shown to be localized in the nucleus, and the PHD finger can bind to the histone H3K4me3/2. So far this group has only been identified in plants and its molecular function is unclear, although it is postulated to play a roll in chromatin regulation (Lee et al., 2009). Strikingly, we identified 4 out of 9 known *alfin-like* transcription factors.

The molecular function of all selected candidates was inferred by associating their corresponding *Arabidopsis* genes with Gene Ontology terms. The terms can be related to one another (i.e. topology) as directed acyclic graphs, which are similar to hierarchies but differ in that a more specialized term can be related to more than one less specialized term. For example in the area enclosed by a dotted red line in Figure 9, the size of the orange spots corresponds to the number of genes identified with that molecular function (e.g. the specialized term *binding* is subdivided into *protein binding* and *nucleic acid binding,* the latter of which can further divided into *RNA binding* and *DNA binding*, with DNA binding involved in *transcription factor activity,* and so on). This network shows only over-represented molecular functions of the selected candidates, as indicated by the color of the spots, while more specialized terms having no significant over/underrepresentation were excluded from this network.

Even though we selected different groups of candidates (based upon different prerequirements) most of them share common molecular functions involved in *binding*, *DNA binding* and *transcription factor activity (*Figure 9, red surrounded area, biological processes Supplementary data K). This network is influenced by the identified of a further 17 transcription factors among the rare and reproductive mode-specific candidates. One could hypothesize that these transcription factors should not be included, as their selection is based on sequence homology rather than significant differential expression between a sexual and apomictic sample, and thus they may not be relevant.

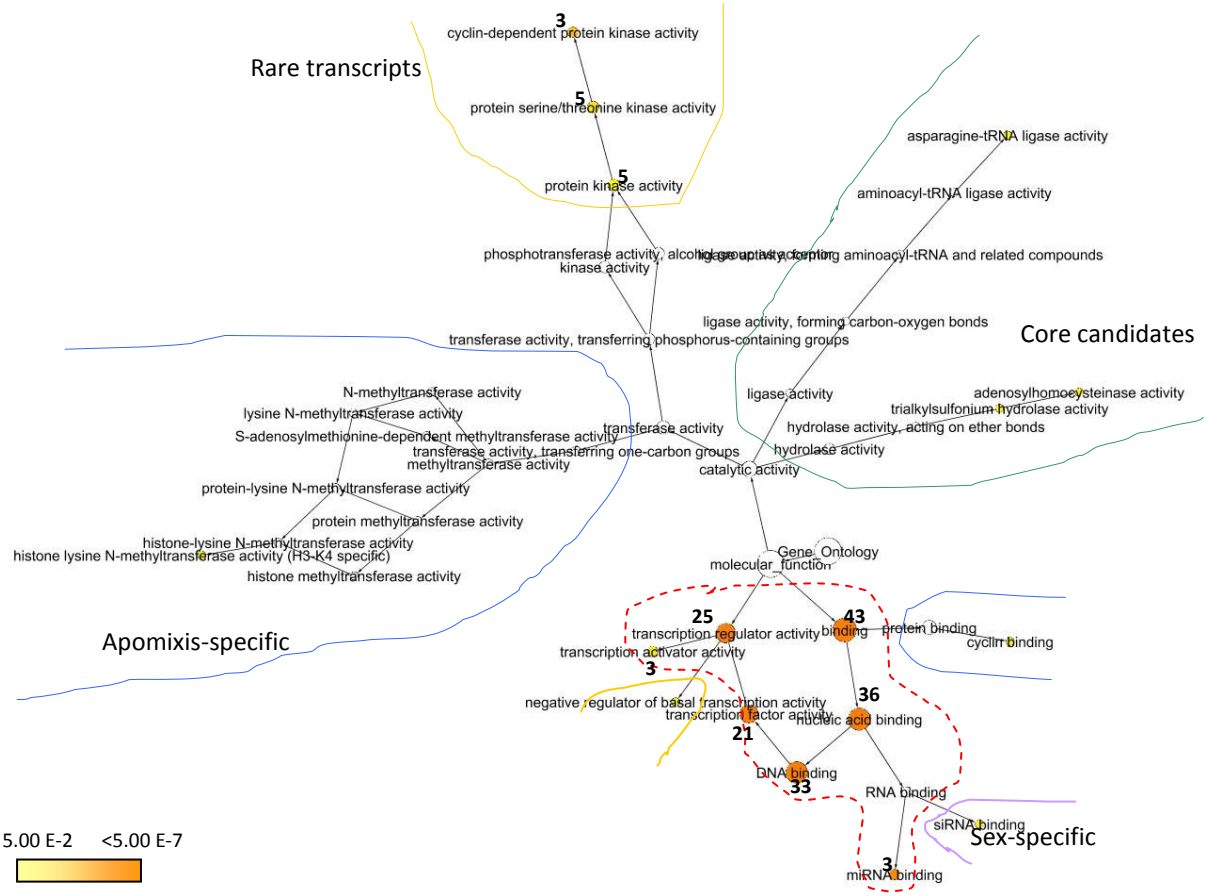


Figure 9. Molecular function network of selected candidate groups employing software Cytoscape and plugin BINGO (Shannon et al., 2003, Maere et al., 2005). BINGO calculated the over-representation (significance 5.00 E-2 to <5.00 E-7, color bar on side) of molecular function and visualized it as a network (explanation in text). The arms of the network surrounded by different colors indicate candidate group-specificity, and the arm where most candidates are clustered is surrounded by a dotted red line. The number next to each spot shows the number of candidates with that particular molecular function (no number it means 1 candidate).

Among the candidates (without the transcription factors; Figure 10), further molecular functions were identified with involvement in processes like post-transcriptional regulation (miRNA, siRNA), DNA methylation, histone modification by methylation or acetylation, cell cycle (cyclin-dependent protein kinase activity) and translation processes (asparagine-tRNA ligase activity). A *histone deacetylase,* specifically identified among core candidates, removes acetyl groups and encourages binding between histone and DNA to repress transcription. We furthermore identified a candidate involved in *histone methylation* with a link to transcriptional regulation (Wang et al., 2001, Berger, 2007). This histone lysine N-methyltransferase activity (H3-K4 specific) was exclusively identified in the apomictic sample. Chromatin remodelling can alter DNA-histone interactions and the coordination between transcription factors and specialized multi-protein complexes (review Martin and Zhang, 2005). Methylation of H3-K4 has been

shown to be enriched in actively transcribed genes (Santos-Rosa et al., 2002), which is consistent with the identification of a transcription factor binding to H3K4me3/2 (*Alfin-like* transcription factor, discussed above) structures. S-adenosyl-L-homocysteinase activity (hydrolase, *SAHH*), as identified among the core apomixis candidates, catalyzes the formation of adenosin and L-homocystein from S-adenoly-L-homocystein. It has been shown that the *Arabidopsis* homolog encodes *SAHH,* and a mutation in this gene causes genome wide DNA demethylation which leads to relaxation of transcriptional gene silencing (Rocha et al., 2005).

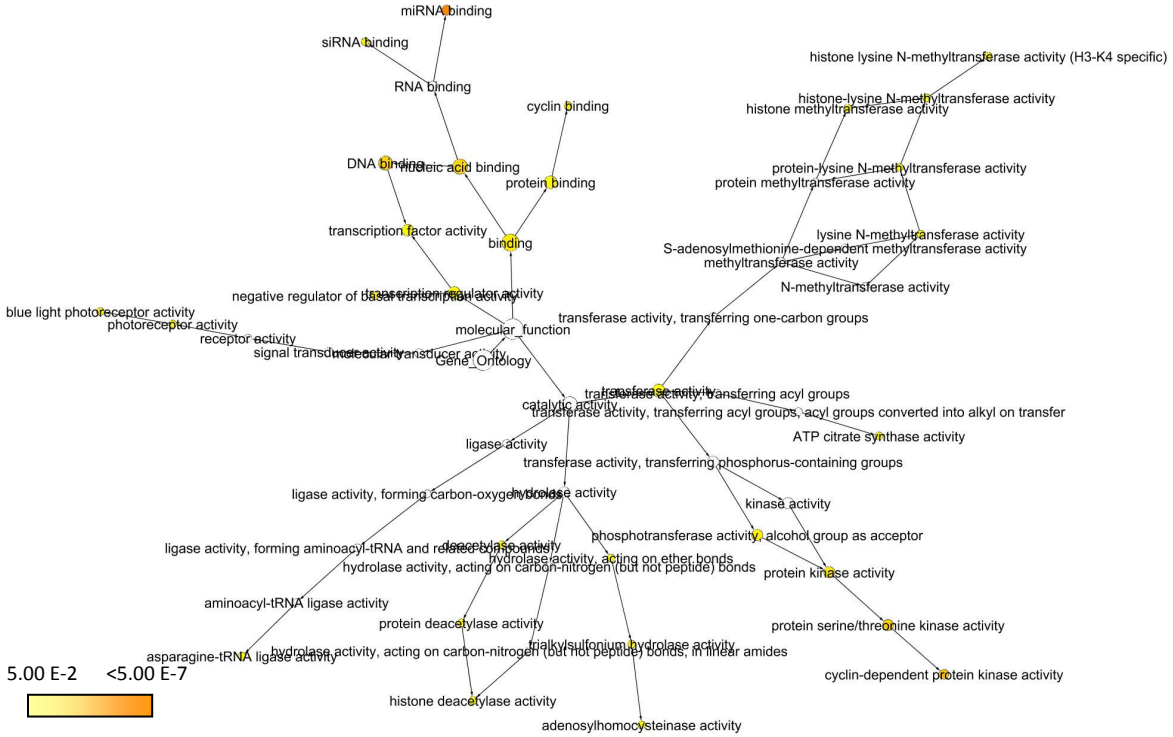


Figure 10. Molecular function network of selected candidate groups (without transcription factors) employing software Cytoscape and plugin BINGO (Shannon et al., 2003, Maere et al., 2005). BINGO calculated the over-representation (significance 5.00 E-2 to <5.00 E-7, color bar on side) of molecular function and visualized it as a network (explanation in text).

As discussed in chapter II, one possibility is that an “apomixis factor” is present in sexual parents, and misexpression of this factor somehow leads to apomeiosis after hybridization. In the same light, its “normal” expression or posttranscriptional regulation could depend on the homozygous state of genes, and on successful somatic pairing of homologous chromosomes (most sexual are homozygous, Song et al., 2006). Misexpression could ensue if pairing is disturbed, which has been shown to affect gene expression (Grant-Downton and Dicknson, 2004). It has been proposed that pairing-regulated genes in plants have a sequence context, as has been found in mammals (Allen et al., 2003), that promotes direct interactions. Heterozygosity can inhibit homolog allelic interactions involved with gene expression, leading to suppression of expression of one of

the alleles (Kusaba et al., 2002). Interestingly we found a similar effect in the 1st SuperSAGE experiment, whereby *B. stricta* specific-alleles were significantly lower expressed (Sharbel et al., 2009) in the apomictic samples. It could be that these alleles were in the homozygous state in *B. stricta* (Song et al., 2006), but in the apomictic heterozygous hybrid (Böcher, 1951; Rollins, 1983; Sharbel and Mitchell-Olds, 2001; Koch et al., 2003; Dobeš et al., 2004a,b; Kantama et al., 2007) inhibited homologous chromosome pairing may have suppressed or enhanced the expression of specific alleles.

Additionally, it has been shown that unpaired DNA could be targeted by methylation (Shiu et al., 2001; Dalmay et al., 2000; Bean et al., 2004). Studies suggests that DNA and histone methylation are very important in regulating expression and epigenetic changes (Kinoshita et al. 2004; Jullien et al. 2006; Mull et al. 2006), for example in transcriptional and post-transcriptional gene silencing (Morel et al. 2000; Jones et al. 2001). DNA is wrapped around an octamer of four core histone proteins, the nucleosome, and with increasing compactness the DNA is less available as a template for transcription. Histone proteins can be modified by methylation and acetylation (Martin and Zhang, 2005). Several of the selected candidates here exhibited such molecular functions (Figure 10), which supports the idea that the initiation of apomixis could be a consequence of changes and modification of gene expression and regulation.

This study has demonstrated that apomixis in *Boechera* might be influenced by chromatin remodelling, which could suppress or enhance gene transcription, and that the cause for chromatin remodelling could be the heterozygous (i.e. hybrid) state of the apomictic genome. Recent studies have shown that chromatin remodelling is a dynamic process and that the activating or repressing functions of particular modifications are variable in different genomic contexts (reviewed in Berger, 2007). The identified 30 core apomixis candidate genes from this study are highly expressed during the ovule stages hypothesized to be characterized by apomeiosis (Sharbel et al., 2009). In stage I of ovule development, before entrance into meiosis, most of these candidates were significantly more highly expressed in the apomictic sample than in the sexual sample, and in proceeding to stage III (after meiosis, tetrad formation) most of these candidates become equally expressed between the sexual and the apomictic sample (Figure 6). It is assumed that genes having the same pattern of expression share regulatory factors (Eisen et al., 1998; Wang et al. 2008), and both *trans* acting elements and disturbed chromosome pairing could have combined effects (Grant-Downton and Dicknson, 2004).

For future projects the selected candidates should be confirmed in additional *Boechera* lines. Microarray experiments, including analyses of species-specific allele expression could give further insight into the association between hybridization and apomixis evolution.

ACKNOWLEDGEMENTS

I am grateful to T.F. Sharbel and J.M. Coral for critical reading of the manuscript and helpful comments to improve the chapter. I would like to thank T. Thiel for assistance in computer performance and to the many people who were involved in the generation of the two SuperSAGE experiments and flower cDNA library construction.

LITERATUR CITED

Adams KL, Cronn R, Percifield R and Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proceedings of the National Academy of Sciences of the United States of America 100: 4649-4654.

Akmaev VR and Wang CJ. 2004. Correction of sequence-based artifacts in serial analysis of gene expression. Bioinformatics (Oxford, England) 20: 1254-1263.

Allen E. Horvath S, Tong F, Kraft P, Spiteri E, Riggs AD, and Marahrens Y. 2003. High concentrations of long interspersed nuclear element sequences distinguish monoallelically expressed genes. Proceedings of the National Academy of Sciences of the United States of America 100: 9940–9945.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

Arazi T, Talmor-Neiman M, Stav R, Riese M, Huijser P and Baulcombe DC. 2005. Cloning and characterization of micro-RNAs from moss. The Plant Journal 43: 837-848.

Axtell MJ and Bartel DP. 2005. Antiquity of microRNAs and their targets in land plants. The Plant Cell 17: 1658-1673.

Audic S and Claverie JM. 1997. The significance of digital gene expression profiles. Genome Research 7: 986-995.

Bastola DR, Pethe VV and Winicov I. 1998. Alfin1, a novel zinc-finger protein in alfalfa roots that binds to promoter elements in the salt-inducible MsPRP2 gene. Plant Molecular Biology 38(6): 1123-35.

Bean CJ. Schaner CE, and Kelly WG. 2004 Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. Nature Genetics. 36: 100–105.

Berger SL. 2007. The complex language of chromatin regulation during transcription. Nature 447: 407-412.

Bienz M. 2006. The PHD finger, a nuclear protein-interaction domain. Trends in Biochemical Science 31(1): 35-40.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Det Kongelige Danske Videnskabernes Selskab 6: 1-59.

Bottley A, Xia GM and Koebner RM. 2006. Homoeologous gene silencing in hexaploid wheat. The Plant Journal 47: 897-906.

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

Chen K and Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews 8: 93-103.

Colinge J and Feger G. 2001. Detecting the impact of sequencing errors on SAGE data. Bioinformatics (Oxford, England) 17: 840-842.

Comai, L. 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Molecular Biology 43387-399.

Corral JM. Piwczynski M and Sharbel TF. 2009. Allelic sequence divergence in the apomictic *Boechera holboellii* complex. In Lost Sex, the Evolutionary Biology of Parthenogenetic Reproduction., I. Schön, K. Martens, and P.J. Van Dijk, eds (Berlin: Springer Verlag), pp. In press.

Dalmay T, Hamilton A, Rudd S, Angell S, and Baulcombe DC. 2000. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101: 543–553.

Dobeš C, Mitchell-Olds T and Koch MA. 2004a. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* x*divericarpa*, and *A. holboellii* (Brassicaceae). Molecular Ecology. 13: 349-370.

Dobeš C, Mitchell-Olds T and Koch MA. 2004b. Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera xdivaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers-an integrative approach. American Journal of Botany 91: 2087-2101.

Doerks T, Copley R and Bork P. 2001. DDT -- a novel domain in different transcription and chromosome remodeling factors. Trends in Biochemical Science 26(3): 145-6

Eisen MB, Spellman PT, Brown PO and Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America 95: 14863-14868.

Eulgem T, Rushton PJ, Robatzek S and Somssich IE. 2000. The WRKY superfamily of plant transcription factors. Trends in Plant Science. 5(5): 199-206

Grimanelli D, Garcia M, Kaszas E, Perotti E and Leblanc O. 2003. Heterochronic Expression of Sexual Reproductive Programs During Apomictic Development in Tripsacum. Genetics 165: 1521-1531.

Ishii M, Hashimoto S, Tsutsumi S, Wada Y, Matsushima K, Kodama T and Aburatani H. 2000. Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. Genomics 68: 136-143.

Jones L, Ratcliff F and Baulcombe DC.. 2001. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Current Biology 11:747-757.

Jullien PE, Kinoshita T, Ohad N and Berger F. 2006. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. The Plant Cell 18: 1360-1372.

Junker BH, Klukas C and Schreiber F. 2006. VANTED: a system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7: 109.

Kantama L, Sharbel TF, Schranz ME, Mitchell-Olds T, De Vries S and De Jong H. 2007. Diploid apomicts of the *Boechera holboelli* complex display large scale chromosome substitutions and aberrant chromosomes. Proceedings of the National Academy of Sciences of the United States of America 104: 14026-14031.

Kappen C. 2000. The homeodomain: an ancient evolutionary motif in animals and plants. Journal of Computational Chemistry 24(1): 95-103

Khattra J, Delaney AD, Zhao Y, Siddiqui A, Asano J, McDonald H, Pandoh P, Dhalla N, Prabhu AL, Ma K, Lee S, Ally A, Tam A, Sa D, Rogers S, Charest D, Stott J, Zuyderduyn S, Varhol R, Eaves C, Jones S, Holt R, Hirst M, Hoodless PA and Marra MA. 2007. Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines. Genome Research 17: 108-116.

Kim HL. 2003. Comparison of oligonucleotide-microarray and serial analysis of gene expression (SAGE) in transcript profiling analysis of megakaryocytes derived from CD34+ cells. Experimental & Molecular Medicine 35: 460-466.

Kim TS, de Guzman JV, Kong HH and Chung DI. 2006. Comparison of gene representation between diploid and triploid Paragonimus westermani by expressed sequence tag analyses. The Journal of Parasitology 92: 803-816.

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL and Kakutani T. 2004. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303: 521-523.

Koch E and Chatterjee SS. 2001. Hyperforin stimulates intracellular calcium mobilisation and enhances extracellular acidification in DDT1-MF2 smooth muscle cells. Pharmacopsychiatry 34. Suppl 1: S70-73.

Koch MA, Haubold B and Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, Arabis, and related genera (Brassicaceae). Molecular Biology and Evolution 17: 1483-1498.

Koch MA, Dobeš C and Mitchell-Olds T. 2003. Multiple Hybrid Formation in Natural Populations: Concerted Evolution of the Internal Transcribed Spacer of Nuclear Ribosomal DNA (ITS) in North American *Arabis divaricarpa (*Brassicaceae). Molecular Biology and Evolution 20: 338-350.

Koltunow AM and Grossniklaus U. 2003. Apomixis: a developmental perspective. Annual Review of Plant Biology 54: 547-574.

Kusaba M, Tung CW, Nasrallah ME, and Nasrallah JB. 2002. Monoallelic expression and dominance interactions in anthers of self-incompatible Arabidopsis lyrata. Plant Physiology 128: 17–20.

Landschulz WH, Johnson PF and McKnight SL. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240(4860): 1759-64

Lee HS and Chen ZJ. 2001. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. Proceedings of the National Academy of Sciences of the United States of America 98: 6753-6758.

Lee JY, Colinas J, Wang JY, Mace D, Ohler U and Benfey PN. 2006. Transcriptional and posttranscriptional regulation of transcription factor expression in *Arabidopsis* roots. Proceedings of the National Academy of Science of the United States of America 103: 6055-6060.

Lee WY, Lee D, Chung WI, and Kwon CS. 2009. Arabidopsis ING and alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. The Plant Journal In press.

Li XY, Mantovani R, Hooft R, van Huijsduijnen R, Andre I, Benoist C and Mathis D. 1992. Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Research 20(5): 1087-91.

Lu J, Lal A, Merriman B, Nelson S and Riggins G. 2004. A comparison of gene expression profiles produced by SAGE, long SAGE, and oligonucleotide chips. Genomics 84: 631-636.

Maere S, Heymans K and Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics 21: 3448-3449.

Man MZ, Wang X and Wang Y. 2000. POWER_SAGE: comparing statistical tests for SAGE experiments. Bioinformatics (Oxford, England) 16: 953-959.

Martin C and Paz-Ares J. 1997. MYB transcription factors in plants. Trends in Genetics 13(2): 67-73.

Martin C and Zhang Y. 2005. The diverse functions of histone lysine methylation. Nature Reviews Molecular Cell Biology 6: 838-849.

Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Kruger DH and Terauchi R. 2005. SuperSAGE. Cellular microbiology 7: 11-18.

Mitchell-Olds T and Schmitt J. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature 441: 947-952.

Mitchell-Olds T, Willis JH and Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nature Reviews 8: 845-856.

Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM and Soltis DE. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biology 6:17.

Morel JB and Vaucheret H. 2000. Post-transcriptional gene silencing mutants. Plant Molecular Biology 43: 275-284.

Mull L, Ebbs ML, and Bender J. 2006. A histone methylation-dependent DNA methylation pathway is uniquely impaired by deficiency in Arabidopsis S-adenosylhomocysteine hydrolase. Genetics 174: 1161-1171.

Naumova TN, van der Laak J, Osadtchiy J, Matzk F, Kravtchenko A, Bergervoet J, Ramulu KS and K Boutilier. 2001. Reproductive development in apomictic populations of Arabis holboellii (Brassicaceae). Sexual Plant Reproduction 14: 195-200.

Pleasance ED, Marra MA, and Jones SJ. 2003. Assessment of SAGE in transcript identification. Genome Research 13: 1203-1215.

Riechmann JL and Meyerowitz EM. 1997. MADS domain proteins in plant development. Journal of Biological Chemistry 378(10): 1079-101.

Riechmann JL and Meyerowitz EM. 1998. The AP2/EREBP family of plant transcription factors. Journal of Biological Chemistry 379(6):633-46.

Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, Varhol RJ, Fjell C, Marra M, Jones S and Siddiqui A. 2007. DiscoverySpace: an interactive data analysis application. Genome Biology 8:R6.

Rocha PS, Sheikh M., Melchiorre R, Fagard M., Boutet S, Loach R, Moffatt B, Wagner C, Vaucheret H and Furner I.. 2005. The Arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. The Plant Cell 17: 404-417.

Roy BA. 1995. The breeding systems of six species of *Arabis (*Brassicaceae). American Journal of Botany 82: 869-877.

Santi L, Wang Y, Stile MR, Berendzen K, Wanke D, Roig C, Pozzi C, Müller K, Müller J, Rohde W and Salamini F. 2003. The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. The Plant Journal 34(6): 813-26.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J and Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. Nature 419: 407-411.

Schneitz K, Hülskamp M and Pruitt RE. 1995. Wild-type ovule development in *Arabidopsis* thaliana: a light microscope study of cleared whole-mount tissue. The Plant Journal 7: 731-749.

Schön I and Martens K. 1998. DNA repair in an ancient asexual: a new solution for an old problem? Journal of Natural History 32: 943-948.

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

Schranz ME, Kantama L, de Jong H and Mitchell-Olds T. 2006. Asexual reproduction in a close relative of *Arabidopsis*: a genetic investigation of apomixis in *Boechera* (Brassicaceae). New Phytologist 171: 425-438.

Sempere LF, Cole CN, McPeek MA, and Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. Journal of experimental Zoology. Part B 306: 575-588.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498-2504.

Sharbel TF, Mitchell-Olds T, Dobes C, Kantama L and de Jong H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic and Genome Research 109: 283-292.

Sharbel TF, Voigt ML, Corral JM, 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. In press.

Shiu PK, Raju NB, Zickler D, and Metzenberg RL. 2001. Meiotic silencing of unpaired DNA. Cell 107: 905–916.

Smyth DR, Bowman JL and Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. The Plant Cell 2: 755-767.

Song BH, Clauss MJ, Pepper A, and Mitchell-Olds T. 2006. Geographic patterns of microsatellite variation in Boechera stricta, a close relative of Arabidopsis. Moleculare Ecology 15: 357-369.

Stollberg J, Urschitz J, Urban Z and Boyd CD. 2000. A quantitative evaluation of SAGE. Genome Research 10: 1241-1248.

Suzuki M, Kao CY, McCarty DR. 1997. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. The Plant Cell 9(5): 799-807.

The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25: 25-29.

Tiwari SB, Wang XJ, Hagen G and Guilfoyle TJ. 2001. AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. The Plant Cell 13(12): 2809-22.

Torres TT, Metta M, Ottenwalder B and Schlotterer C. 2008. Gene expression profiling by massively parallel sequencing. Genome Research 18: 172-177.

Ungerer MC, Linder CR and Rieseberg LH. 2003. Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. Genetics 163: 277-286.

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW. 1995. Serial analysis of gene expression. Science 270: 484-487).

Wang H, Cao R, Xia L, Erdjument-Bromage H, Borchers C, Tempst P and Zhang Y. 2001. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Molecular Cell 8: 1207-1217.

Wang H, Zheng H, and Azuaje F. 2008. Clustering-based approaches to SAGE data mining. BioData mining 1:5.

Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L and Chen ZJ. 2004. Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics 167: 1961-1973.

Windsor AJ, Schranz ME, Formanova N, Gebauer-Jung S, Bishop JG, 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: 1169-1182.

Winicov I. 2000. Alfin1 transcription factor overexpression enhances plant root growth under normal and saline conditions and improves salt tolerance in alfalfa. Planta 210: 416-422.

Winicov II and Bastola DR. 1999. Transgenic overexpression of the transcription factor alfin1 enhances expression of the endogenous MsPRP2 gene in alfalfa and improves salinity tolerance of the plants. Plant Physiology 120: 473-480.

Wray GA. 2003. Transcriptional regulation and the evolution of development. The International Journal of Developmental Biology 47: 675-684.

Xu Y, Zhong L, Wu X, Fang X and Wang J. 2009. Rapid alterations of gene expression and cytosine methylation in newly synthesized Brassica napus allopolyploids. Planta 229: 471-483.

Supplementary data:

A) Additionally information about lines included in this study.

B) Protein sequences of selected *A. thaliana* genes (Table 2), (fasta.).

C) Annotated 454APO cDNA to A. thaliana sequences.

D) Annotated 454SEX cDNA to A. thaliana sequences.

E) Summary of identified genes in SuperSAGE data and calculated p-values.

F) F01- Mapped TAG sequences of 1st SuperSAGE to identified cDNA sequences; F02 – 244 TAG sequences without sinflotons; F03 – 51 candidates.

G) Bar-chart illustrating frequency of significant differential expressed TAG sequences among comparison of SuperSAGE libraries.

H) H01 - Summary of sequence homology search between identified TAG sequences of 1st SuperSAGE with TAG sequences of 2nd SuperSAGE experiment;
H02 – 76 candidates.

I) Summary of selected candidates; I01 – I04.

J) Comparison between 76 candidates of 2nd SuperSAGE and 30 core candidates with relevance for high copy number expression.

K) Network of candidates and their involvement in Gene Ontology category "Biological process".

L) Transcription factor families identified.

ACKNOWLEDGEMENTS

The completion of this project would not have been possible without the assistance, courtesy, and encouragement of many people.

I am grateful to my first supervisor, Professor Marcus Koch (HIP, University of Heidelberg) – Thank you for giving me the opportunity to defend my work at Heidelberg and for your commitment to the *Boechera* phylogenetics.

Dr. Timothy F. Sharbel – I am deeply indebted to you for your work as my supervisor, Tim. Not only did you open your lab to me for the completion of my work, but also, during difficult times, you let me believe that everything is possible. Thank you for your enthusiasm, your motivation, your advice and your support – all of which went beyond the scope of this thesis.

The technical support given by Heidi, Leane and Vera "Gold wert" – What would I have done without you? I will certainly miss the breakfast breaks and the laughs we shared. Special thanks to Heidi for your sense of humor and continuous ambition to keep the lab running, to Vera for taking care of my plants, and to Leane, without whom the amount of seed measurements would have not been possible. Thank you!!!

Deep thanks to all members of the lab for the exchange of research ideas and new exciting discoveries: Jose, thank you so much, you always had time for me and to answer my questions. I will miss your detailed knowledge about so many different topics; to list them all would expand the limitation of my thesis;-)
Marta – My most beautiful friend, you are amazing. It was so much fun to share an office with you. Marta, Thomas and Maria, all the best for the three of you!!!! I will miss you!
Marcin- What happened to you? Why did you leave me alone? The lab was not the same anymore (I only say: "You killing me with your smile!").
Wale, Martin - Don't ever stop asking questions. Lorenzo, Guilo, Moha, Inma – I haven't heard anyone speaking faster than you, you enriched the lab so greatly – thank you guys!
Marco and Samuel – Enjoy the stay and take the advantage of all the IPK has to offer!

Thanks to many groups of the IPK, which were involved in the presented work in providing technique, samples and sharing expertise. I would like to express my gratitude to Prof. Ingo Schubert and Jörg Fuchs, Jochen Kumlehn and Alok Varshney, Andreas Houben and the service provided by the bioinformatics department and the graphic department.

Astrid, Andrea, Eszter and Marta - Thanks for sharing happiness and challenges during motherhood.

Meinen Eltern möchte ich auf diesem Wege meinen tiefsten Dank aussprechen, die kontinuierliche Unterstützung und "du schaffst das schon" haben besser gewirkt als ihr jemals ermessen könnt – DANKE!!!

Und nun zu meinen beiden Liebsten Marco and Alma, ihr seid mein Leben und ich freue mich auf unsere nächsten Herausforderungen!

E. Contents supplementary material and raw data on the DVD included in thesis.

Supplementary data Chapter I

includes all supplementary data mentioned in the text in chapter I.

Supplementary data Chapter II

includes all supplementary data mentioned in the text in chapter II.

Supplementary data Chapter III

includes all supplementary data mentioned in the text in chapter III.

Supplementary data Chapter IV

includes all supplementary data mentioned in the text in chapter IV.

Thesis_Marie-Luise Voigt_2009.pdf

Eidesstattliche Erklärung

„Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Stellen, die anderen Werken wörtlich oder sinngemäß entnommen sind, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat."

Gatersleben, 23.04.2009

Marie-Luise Voigt