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	Martin Mau
born in:	Pasewalk, Germany
Oral-examination:	04.07.2013

Identification, molecular analysis and phylogeographic distribution of the chimeric *UPGRADE-2* gene, a candidate for the initiation of unreduced pollen formation in apomictic *Boechera* (Brassicaceae)

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

Gatersleben, 2013

I Table of contents

II List of figures

III List of tables

IV Abbreviations

AK	ancestral karyotype	lncRNA	long non-coding RNA
ASGR	apospory-specific genomic region	natsiRNA	natural antisense siRNA
ATP	adenosine triphosphate	min	minutes
AUC	area under the receiver operating characteristic curve	miRNA, MIR	microRNA
BAC	bacterial artificial chromosome	mm	millimeter
BIO(X)	bioclimate variable	MMC	microspore mother cell
BLASTN	Basic Local Alignment Search Tool against nucleotide sequences	mRNA	micro RNA
BLASTX	Basic Local Alignment Search Tool against protein sequences	Msp	microspore
bp	base pair	ncRNA/ npcRNA	non-protein coding RNA
BSA	bovine serum albumin	nt	nucleotide
Bsp*HRD*	*Boechera* species 3-hydroxy-3-methylglutaryl-coenzyme A reductase degradation	ORF	open reading frame
Bsp*UPG*	*Boechera* species unreduced pollengrain development	PBS	phosphate buffered saline
Bsp*UPG-1*	sequence of original locus of *Boechera* species unreduced pollengrain development	PCA	principal components analysis
Bsp*UPG-2*	=Bsp*UPG/UPGRADE*, active transcript and duplicated locus of *Boechera* species unreduced pollengrain development	PD	pairwise distance
BstLG	*Boechera stricta* linkage group	PgC	polycomb group complex
cis-nat-siRNA	natural *cis*-antisense siRNA	piRNA	P-element induced wimpy testis (Piwi)-interacting RNA
CDK	cyclin-dependent kinase	PMC	pollen mother cell
cm	centimeter	RACE	rapid amplification of cDNA ends
CNV	copy number variation	RISC	RNA-induced silencing complex
CpG	phosphodiester bond between the cytosine and the guanine	ROC	receiver operating characteristic curve
DAPI	4',6-diamidino-2-phenylindole	rRNA	ribosomal RNA
DEPC	diethyl phosphorocyanidate	SAM	shoot apical meristem
DSB	double-strand break	SDR	second division restitution
EBN	endosperm balance number	SDS	sodium dodecyl sulfate
EDTA	ethylenediaminetetraacetic acid	SI	self incompatibility
EFTU	Elongation factor Tu/EF-1A	SINE	short interspersed nuclear

	protein		elements
ER	endoplasmic reticulum	siRNA	small interfering RNAs
FCSS	flow cytometric seed screen	Sp	sporocyte
FDR	first division restitution	sRNA	small RNA
GAPDH	glycerinaldehyd-3-phosphat-dehydrogenase	SSC	saline-sodium citrate
GO	gene ontology	T4	DNA-Ligase from T4 phage
GSP	gene-specific primer	TAIR	The Arabidopsis Information Resource
HDGS	homology dependent gene silencing	ta-siRNA	*trans*-acting siRNA
HFA	hybridization-derived floral asynchrony	TBE	TRIS-Borat-EDTA
hr	hour	TBLASTX	basic local alignment search tool compares *six-frame* translation of nucleotide sequence against *six-frame* translation of nucleotide database
IR	inverted repeat	TE	transposable element
JA	jasmonic acid	TFBS	transcription factor binding site
kb	kilobase	*trnL-F*	region of transfer RNA genes of chloroplast genome between the 3`-end of 5`*trnL* (UAA) exon and the 5`-end of the *trnF* (GAA) exon
km	kilometer	TSP	target specific primer
LCB	locally colinear block	Tween	polyoxyethylen(X)-sorbitan-monolaurat
LINE	long interspersed nuclear element	UTR	untranslated region
LB	lysogeny broth	WGS	whole genome sequencing
LTR	long terminal repeat	X-gal	5-bromo-4-chloro-indolyl-*β*-D-galactopyranoside

V Publications and Presentations (during PhD study)

Parts of the PhD study hereto were published elsewhere or were presented on conferences.

Publications

1. **Martin Mau**, José M. Corral, John Lovell, Christiane Kiefer, Marcus Koch, Aliyu Olawale, John McKay and Timothy F. Sharbel. 2013. *APOLLO* and *UPGRADE*: Phylogeographic association study of two novel genetic factors for female and male apomeiosis in the genus *Boechera*. (in preparation).

2. **Martin Mau**, José M. Corral, Heiko Vogel, Michael Melzer, Jörg Fuchs, Markus Kuhlmann, Nico de Storme, Danny Geelen, and Timothy F. Sharbel. 2013. The novel chimeric transcript *UPGRADE* is associated with unreduced pollen formation in apomictic *Boechera.* The Plant Journal (under review).

3. Nico De Storme, Linda Zamariola, **Martin Mau**, Timothy F. Sharbel, and Danny Geelen. 2013. Volumetric pollen analysis as a rapid assessment of somatic and gametophytic ploidy and the mode of reproduction in flowering plants. Plant Reproduction DOI 10.1007/s00497-012-0209-0.

4. Marco Pellino, Timothy F. Sharbel, **Martin Mau**, Samuel Amiteye and José M. Corral. 2011. Selection of reference genes for quantitative realtime PCR expression studies of microdissected reproductive tissues in apomictic and sexual *Boechera*. BMC Research Notes 4 (303).

5. Giulio Galla, Sara Zenoni, Gianpiero Marconi, Giada Marino, Alessandro Botton, Francesco Pinosa, Sandra Citterio, Benedetto Ruperti, Klaus Palme, Emidio Albertini, Mario Pezzotti, **Martin Mau**, Timothy F. Sharbel, Nico De Storme, Danny Geelen and Gianni Barcaccia. 2011. Sporophytic and gametophytic functions of the cell cycle-associated *Mob1* gene in *Arabidopsis thaliana* L. Gene. 484 (1-2): 1-12.

Oral and poster presentations

1. **Martin Mau** and Timothy F. Sharbel. Unreduced pollen formation in *Boechera* - a transcriptomic study. 9th IMPRS Symposium. February 15-16, 2010. Altes Schloss, Dornburg (Poster presentation).

2. **Martin Mau** and Timothy F. Sharbel. Unreduced pollen formation – a transcriptomic study. Plant Science Student Conference. June 15-18, 2010. IPK, Gatersleben (Oral presentation).

3. **Martin Mau** and Timothy F. Sharbel. Unreduced pollen formation in *Boechera* - phenotyping and global transcriptome analysis. 2nd International PhD School on Plant Development. Oktober 07-09, 2010. Conference center Benediktushöhe, Retzbach (Oral presentation).

4. **Martin Mau** and Timothy F. Sharbel. Unreduced pollen formation – a histological and transcriptomic study in the genus *Boechera.* 11th IMPRS Symposium. February 07-08, 2011. Altes Schloss, Dornburg (Oral presentation).

1.1 Summary

In sexual plants meiosis gives rise to recombined and reduced gametes. In diplosporous accessions of the North American genus *Boechera* meiotic processes are circumvented (*i.e.* apomeiosis) which lead to a supressed recombination and to production of clonal unreduced male and female gametes. Unreduced male gametes in diplosporous *Boechera* are required to produce balanced endosperm. The objective of this study was to identify and characterize candidate genetic factors for unreduced pollen formation, and to analyze their genus-wide dynamics in order to contrast the hypotheses whether apomeiosis expression was induced through interspecific hybridization or if it could be an older characteristic of the genus (*i.e.* pre-Pleistocene). Apomictic *Boechera* demonstrated high variability for reduced and unreduced meiocyte production. Early flower developmental staging and flow-cytometric analyses together led to the selection of a single pollen mother cell stage at the onset of meiosis for microarray-based comparative gene expression analyses between diploid sexual and diploid apomictic genotypes, and led to the identification of a single highly-upregulated factor (Bsp*UPG-2*) which is highly conserved among apomictic *Boechera* but has no homologue in sexual *Boechera* or in other taxa. Bsp*UPG-2* exhibits four intron-exon structure variants which suggest alternative splicing, and lack of a prominent open reading frame and no overall sequence homology to known genes suggests that Bsp*UPG-2* belongs to the novel class of long regulatory non-coding mRNA-like RNAs. Bsp*UPG-2* has apparently arisen through a three-step process initiated by ancestral gene duplication of the original non-genic Bsp*UPG-1* locus, followed by sequential insertions of segmentally duplicated gene fragments which led to its chimeric structure and neofunctionalization in apomictic *Boechera*. Its genesis reflects the hybridization history which characterizes the genus *Boechera*. Computational analysis demonstrated that portions of Bsp*UPG-2* form secondary structures which were classified as potential primary microRNAs (pri-miRNAs) according to their minimal folding free energy index (MFEI≥0.70) and their A+U content (≥56.70%). One such structure is highly similar to the third exon of a GTP-binding elongation factor Tu/1-A family homolog from *Arabidopsis* (AT4G02930, *E*-value=7.00E-24) which could be a potential regulatory target. Two other sequence fragments at the 5`-end of Bsp*UPG-2* are also homologous to known protein-coding genes (AT5G19960 and AT1G18260), hence supporting the hypothesized regulatory function of Bsp*UPG-2*. The apomixis-specificity of Bsp*UPG-2* enabled a genus-wide analysis which demonstrated its ubiquity in all *Boechera* lineages, including the ancient AB haplotype and two single individuals each of one closely-related genus. These results attest to Bsp*UPG-2's* hypothesized importance unreduced pollen formation and hence for balanced endosperm formation, and point to a single origin of this factor which coincides with the root of the genus *Boechera* dating to the middle of the Pleistocene.

1.2 Zusammenfassung

In sexuell reproduzierenden Pflanzen führen Meiosen zu Gameten mit rekombinierten und reduzierten Chromosomensätzen. Bei diplosporen Akzessionen der nordamerikanischen Gattung *Boechera* bleibt die Reduktionsteilung und die damit einhergehende Rekombination mütterlicher und väterlicher Chromosomensätze aus (d.h. Apomeiose), und führt zu klonalen und unreduzierten weiblichen und männlichen Gameten. Diplospore *Boechera* Akzessionen benötigen unreduzierte männliche Gameten zur Balancierung der elterlichen Chromosomensätze im Endospermgewebe. Ziel dieser Dissertation war es, Kandidatengene für die Formierung unreduzierter männlicher Gameten zu identifizieren, charakterisieren und mittels geeigneter Kandidategene die Hypothese zu testen, ob Apomeiose in *Boechera* durch interspezifische Hybridisierungen induziert wurde oder ein älteres Charakteristikum der Gattung *Boechera* ist (d.h. pre-Pleistozän). Die Formierung von reduzierten und unreduzierten Meiocyten war hoch-variabel in apomiktische *Boechera* Akzessionen. Die Analyse von Stadien der frühen Blütenentwicklung, zusammen mit durchflusszytometrischer Analyse von Pollenkernen und Samen führte zur Identifikation eines spezifischen Pollenmutterzellstadium am Eintritt zum Meioseprozess, welches zur Mikroarray-gestützen vergleichenden Transkriptomanalyse von diploiden sexuellen und diploiden apomiktischen Akzessionen genutzt wurde. Es konnte ein einziges hochgradig exprimiertes Gen (Bsp*UPG-2*) identifiziert, welches mit hoch konservierter Sequenz spezifisch in apomiktischen Akzessionen vorkommt und dort prevalent in Antheren exprimiert wird. Bsp*UPG-2* zeigte weder Homologien zu Genen in sexuellen *Boechera* Akzessionen noch zu Genen anderen Taxa. Insgesamt vier Transkriptvarianten wurden für das Bsp*UPG-2* Gen identifiziert, was auf differenzielle Spleißvorgänge bei der Transkription schließen lässt. Obwohl Bsp*UPG-2* auch Merkmale protein-kodierender Gene (z.B. Polyadenylierung) besitzt, deutet ein fehlender Leserahmen bei einer Transkriptlänge von ca. 2648 Nukleotiden auf eine Klassifizerung als neuartige regulatorische lange nicht-kodierende mRNA-typische RNA hin. Bsp*UPG-2* entwickelte sich offensichtlich aus einem dreistufigen Prozeß, der durch eine Genduplikation des originalen Lokus Bsp*UPG-1* initiiert wurde. Anschließend kam es zur schrittweisen Insertion duplizierter Genfragmente und zur Ausprägung einer Intron-Exon-Struktur. Dieser Vorgang führte zu einer Chimärstruktur von Bsp*UPG-2* und zu einer potentiellen Neofunktionalisierung des Gens in apomiktischen *Boechera* Akzessionen. Mittels bioinformatischer Methoden wurde gezeigt, dass bestimmte Sequenzbereiche des Bsp*UPG-2* Gens Sekundärstrukturen formen können deren Eigenschaften primären microRNAs (pri-miRNAs) gleichen (MFEI≥0.70, A+U content ≥56.70%). Eine dieser Strukturen ist homolog zum dritten Exon des GTP-bindenden Elongationfaktor Tu/1-A Homolog aus *Arabidopsis* (AT4G02930, *E*=7.00E-24) welches vermutlich ein regulatorisches Zielgen ist. Die Homologie zweier weiterer Fragmente des Bsp*UPG-2* Gens zu bekannten protein-kodierenden Genes (AT5G19960 und AT1G18260), unterstreicht nochmals die

vermutete regulatorische Funktion des Bsp*UPG-2* Gens. Das spezifische Vorkommen von Bsp*UPG-2* in apomiktischen *Boechera* Akzessionen ermöglichte die marker-gestützte Klassifizierung von apomiktischen und sexuellen *Boechera* Akkzessionen und zeigte, dass Bsp*UPG-2* in allen *Boechera* Chloroplasten-Haplotypenlinien vorkommt und sowohl im evolutionär ältesten Chloroplasten-Haplotyp (AB) als auch in zwei Individuen nahverwandter Gattungen detektiert wurde. Diese Ergebnisse unterstützen die vermutete potentielle Bedeutung von Bsp*UPG-2* für die Formierung unreduzierter männlicher Gameten und damit auch für die Ausbildung eines funktionierenden Endosperms in apomiktischen *Boechera* Akzessionen und deuten darüber hinaus auf einen singulären Ursprung des Kandidatengens Bsp*UPG-2* hin, welcher mit dem Ursprung der Gattung *Boechera* koinzidiert.

2. Introduction and Literature Review

2.1 Development and function of the male germline in flowering plants

2.1.1 Developmental steps

The "quartet model", a revised version of the classical ABC model, explains the coordinated developmental transition from the vegetative to the reproductive phase in flowering plants (Haughn and Sommerville 1988; Theissen and Saedler 2001). The determined differentiation pattern of flower organs occurs after the plant hormone-mediated switch of the shoot apical meristem (SAM) and its lateral organs from vegetative to generative growth, and involves a few homeotic genes (Steeves and Sussex 1989; Coen and Meyerowitz 1991; Yanofsky 1995; Laux et al., 1996; Blázquez et al., 1998; Fletcher et al., 1999; Brand et al., 2000; Lenhard et al., 2002). Thereby sepals are derived from the A function (*APETALA* genes *AP1* and *AP2*), petals from the A and B function, the gynoecium from the C function and stamens from the B and C functions. All interact with the E function, containing the *SEPALLATA1, 2, 3* and *4* (*SEP*) genes, which are flower-specific MADS-box transcription factors as described for the B (*DEFICIENS* (*DEF*) and *GLOBOSA* (*GLO*) in *Antirrhinum* and *AP3* and *PISTILLATA* (*PI*) in *Arabidopsis*) and C functions (*AGAMOUS (AG)* and *PLENA* genes). In Brassicaceae, flower buds are generally simply structured with a calyx of four separated sepals, a corolla of four petals, and four long centered and two short lateral stamens surrounding the gynoecium (Smyth et al., 1990). Flower buds typically emerge at the flanking side of apical meristems in a whorl pattern, whereby the age of a bud is related to its distance from the whorl center (Smyth et al., 1990).

The developmental reference points of male reproductive organs (stamens) were described in detail for *Arabidopsis thaliana* (Smyth et al., 1990), *Brassica napus* (Scott et al., 1991) and tobacco (Koltunow 1990; Goldberg 1993). Thereby, stamen development is initiated by the stamen primordia which are located in the third whorl within the bud (Coen and Meyerowitz 1991). Stamen development is characterized by the elongation of its filament after differentiation in an epidermal cell layer and vascular bundle, but is ultimately dominated by differentiation of the multiple specialized cell types in the anthers which are linked to the filament *via* the anther connective. In *Arabidopsis* the *SPOROCYTELESS* (*SPL*)/*NOZZLE* (*NZZ*) gene, which is partly under the control of the *AG* gene, is one of the few examined genes promoting early stamen differentiation (Schiefthaler et al., 1999; Yang et al., 1999). Anther development can be divided into several stages which can be grouped into two phases. Phase one includes

early stages, where anthers undergo major histodifferentiation events such as cell specification, tissue differentiation, meiosis and microspore formation, whereas in phase two anthers massively elongate, pollen grains differentiate while other tissues degenerate (*e.g.* tapetum), and anther development finally ends with dehiscence and pollen release (Koltunow 1990; Smyth et al., 1990). All stamen tissues derive from three initial "germ" layers: L1 giving rise to the epidermis, L2 developing into the pollen mother cells (PMCs), the endothecium and the outer tapetum, and L3, which differentiates into the connective, the inner tapetum and the vascular bundle (Goldberg 1993). The fully differentiated anther is structured in the vascular bundle surrounding anther connective which joins four microsporangia consisting each of four somatic layers: epidermis, endothecium, middle layer(s) and tapetum. Two microsporangia are clustered in a lobe and separated by two additional nonreproductive tissues: the circular cell cluster and the stomium. Each of the four microsporangia forms a locule that houses the gametophyte initials (PMCs or meiocytes). Differentiation of the tapetum and the meiocytes is thereby putatively controlled by antagonistic expression levels of the *EXCESS MALE SPOROCYTES1* (*EMS1*) gene and the *TAPETUMDETERMINANT1* (*TPD1*) gene in the precursor cells (Zhao et al., 2002; Yang et al., 2003; Ma 2005).

The essential step from the sporophyte to the gametophyte, which hosts the male germline, occurs during gametogenesis, a two-part process. It starts with microsporogenesis where the enlarged, interconnected (*e.g. via* plasmodesmata) and dedifferentiated diploid PMCs undergo one round of DNA replication followed by one meiotic reductional and one mitosis-like equational division to form four haploid (reduced) microspores. The onset of meiosis is accompanied with callose deposition along the plasma membranes which is guided by microtubule arrays within the cytoplasm (Cresti et al., 1992). During the second part, microgametogenesis, thickening of the callose wall at the end of the second meiotic division and final endo-(1,3)-*ß*-glucanase-mediated callose dissolution separates the four haploid microspores from each other. Each undergoes a vacuolation- and nuclear migration phase, whereby unequational mitotic division leads to a binucleate pollen grain (Scott et al., 2004). Thereby one reduced gamete (generative cell or sperm cell) is separated from the vegetative cell. In some angiosperms, including *Arabidopsis*, the generative cell additionally divides into two isomorphic sperm cells before germination of the pollen tube, but in about 70% of the angiosperms the second division appears after germination (pleisomorphic bicellular pollen), which has been shown to be a phylogenetically

rudimentary form of pollen formation (Brewbaker 1967; McCormick 1993). During microgametogenesis pollen grains form a multilayered pollen wall supporting its function as carrier of the male gametes. The pollen wall consists of a pectocellulosic intine layer which is surrounded by a tapetum-derived sporopollenin-based exine, containing two layers, the inner nexine and the outer sexine. The tectum, as the outermost sexine layer, is highly variable between species and reflects the mode of dispersal. The tectum is fragmented by precisely positioned germinal apertures and arrays of columellae forming interlocking lacunae which function as sites of germination and play a part in harmomegathy (Erdtman 1947, 1952; Cresti et al., 1992; Scott et al., 2004).

Anther development ends with dehiscence, a process of tissue degeneration (*e.g.* of the septum, the tapetum and the circular cell cluster) that forms a bilocular anther in the first step and disruption of the stomium in a second step. In *Arabidopsis* the *DEFECTIVE IN ANTHER DEHISCENCE1* (*DAD1*) gene, which initiates the biosynthesis of jasmonic acid (JA), and the transcription factor *MYB26* are involved in dehiscence. JA may regulate programmed cell death (Zhao and Ma 2000) and is involved in water transport and up-take in anthers, synchronizing pollen maturation and anther dehiscence (Ishiguro et al., 2001). In interacton with the JA function as regulator of water fluxes, *MYB26* activates the phenylpropanoid pathway in endothecial cells to provide lignin residues for wall thickening, allowing the endothecium cells to become turgid, thereby disrupting the neighboring stomium and leading to pollen release (Scott et al., 2004).

2.1.2 The role of male gametes

The role of pollen grains is to deliver the generative (sperm) cells to the embryo sac to undergo cell fusion with the egg and central cells. This task is divided into three phases, (1) transfer of the pollen from the anther to the stigma of the pistil by biotic and abiotic vectors (pollination), (2) pollen tube initiation and guided growth (germination), and (3) fusion of the male and female gametes (fertilization). Competition between single pollen during pollination is thereby one major driver for the evolutionary success of flowering plants.

Pollen-pistil recognition on the stigma surface of the gynoecium involves at least three essential pollen coat-specific genes (*CER1, CER3* and *CER6*) which were identified in the *eceriferum* mutant of *Arabidopsis* and which represent various classes

of long-chain lipids whose combinations are necessary for pollen recognition (Hülskamp et al., 1995). Additional lipophilic molecules in the exine wall are assumed to play a role in species-specific adhesion between pollen and stigma cells (Zinkl et al., 1999). Moreover, strong pollen competition underlies these steps to inhibit incompatible pollen, and increase progeny fitness (Armbruster and Gobeille Rogers 2004), although this is strongly modulated by the progeny growth environment (Kalla and Ashman 2002). The various levels of pollen competition include pollen position on the stigma (Thomson 1989), the number of pollen grains (Mulcahy and Mulcahy 1975), the distance of pollen growth (McKenna and Mulcahy 1993), pollen mentor effects (Michurin 1950), heteromorphic self-incompatibility (HSI, Darwin (1877); East (1940)) or various other types of self-incompatibility (SI) barriers (*e.g.* cryptic- (CSI), sporophytic- (SSI), gametophytic- (GSI) or gametophytic-sporophytic-SI (GSSI)). In many Brassicaceae SSI prevents self-fertilization after deposition of the pollen grain on the stigma surface (Schierup et al., 1998). Similar to the tightly linked *S-RNase* gene and the *S*-locus F-box (*SLF*) gene as determinants of SI at the *S* (sterility) locus in *Petunia* (Indriolo and Goring 2010), the *S* locus of outcrossing *A. lyrata* contains various *S* haplotypes of three tightly linked genes, the stigma-specific *S* locus receptor kinase (*SRK*) gene, the pollen-specific *S* locus cysteine-rich protein (*SCR*) gene and the papilar cell-specific *S* locus glycoprotein (*SLG*). The *S* haplotype variants of these genes with similar specificity, lead to inhibition of pollen germination on the stigma surface (Schopfer et al., 1999; Kusaba et al., 2001). In contrast, this process of competitive interaction is limited in the self-compatible *A. thaliana*, which is caused by a lack of polymorphism at the *SCR* locus (Bechsgaard et al., 2006).

After circumvention of the SI barrier the vegetative cell, which is a storage cell, initiates pollen tube formation (germination) for delivery of the nonmobile male germ unit (MGU), composed of the two sperm cells. Thereby, the pollen tube emerges after rehydration and aperture opening of the pollen grain and penetrates into the transmitting tissues of the stylar canal. A complex network of genes is involved in pollen tube guidance through the stylar canal *via* the septum and the funiculi to the ovules. In tomato, pectin degrading enzymes (Wing et al., 1989) allow pollen tube growth, and the plasma membrane localized leucine rich repeat (LRR) receptor-like kinases, LePRK1 and LePRK2 (Muschietti et al., 1998), play a role in signal transduction important for growth coordination and direction. In *Arabidopsis* the *TIP1* and *KIP* genes play a role in initiation and maintenance of growth in pollen tube tip-growing cells (Schiefelbein et

al., 1993; Ryan et al., 1998; Procissi et al., 2003). Mutant screening in *Arabidopsis* additionally identified genes influencing γ-aminobutyric acid (GABA) concentration in female tissues (POP2) or the Ca^{2+} gradient in pollen tubes (D-serine activated glutamate receptors (GLRs)), all of which are necessary to avoid pollen tube guidance defects.

The involvement of intracellular messengers like GABA and Ca^{2+} shows that, similar to the female gametophyte, the male gametophyte must control a complex molecular machinery of signal perception, signal transduction and cellular response to control chemical signals from sporophytic tissues and the female gametophyte (Geitmann and Palanivelu 2007). Signal perception, for example is mediated from a large group of pollen-specific LRR receptor kinases (Muschietti et al., 1998; Kim et al., 2002), *via* channel-guided mechanical-sensitive calcium uptake (Feijó et al., 1995; Malhó et al., 1995), inward K^+ channels (Fan et al., 2001) or other putative voltage gated ion channels (Wang et al., 1989). Cytosolic calcium plays a major role in signal transduction (Malhó et al., 1994; 1995) *via* activation of calcium-dependent protein kinases (CDPK; Estruch et al., (1994)) and induction of the synthesis of secondary-messengers like calmodulin (Vogel 1994) or inositol 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG) *via* phophatidylinositol 4,5-bisphosphate (PIP_2) hydrolysis (Dowd et al., 2006). Another class of signal transducers are Rho-related GTPases, which play a critical role for pollen tube tip growth (Lin and Yang 1997; Li et al., 1998). Geitmann and Palanivelu (2007) reviewed the manyfold cellular signal responses of the male gametophyte, and allude to changes in growth rate, growth direction and adhesion ability of the pollen tube involving the spatial control of the cytoskeleton *via* actin-binding proteins (*e.g.* Profilin; Staiger et al. (1997)), or *via* redirecting of cell wall softeners, such as methyl-esterifed pectin.

In angiosperms the pollen tube enters the ovule through the micropylar pole into one of the two synergids. The *MYB98* gene (Higashiyama et al., 2001), which encodes a transcription factor modulating the secretion of chemoatractants from the filiform apparatus, and ZmEA1 (Márton et al., 2005), a small secreted protein in maize, are two of only a few known genetic factors involved in chemotaxis guiding of the pollen tube to the synergids. Finally, the pollen tube arrests at the synergids, which is mediated by the FERONIA/SIRENE (FER/SRN) signaling pathway (Huck et al., 2003; Escobar-Restrepo et al., 2007) and multiple other genes, like the *LORELEI* gene (Capron et al., 2008). Hereafter, ion channels, *e.g.* KZM1 in maize (Amien et al., 2010) and *ACA9* in *Arabidopsis* (Schiott et al., 2004), are strongly involved in the rupture of the pollen tube

tip. The pollen tube rupture initiates the released of both sperm cells into the eight-nucleate embryo sac which contains the egg cell, the binucleate central cell and three antipodal cells at the chalazal pole (Reiser and Fischer 1993).

Similarly to pollen tube guidance through the style, cell-cell communication plays a mandatory role during double fertilization (Borges et al., 2008), with a complex signalling network between the pollen tube, the synergids, sperm cells, the egg cell and the central cell, preventing polyspermy and misguidance of the male gametes inside the female gametophyte. Double fertilization refers to the process where one haploid generative nucleus migrates towards and fuses with the haploid egg cell to form the diploid (2C) zygote, whereas the second generative nucleus fertilizes the binucleate central cell to form the triploid (3C) endosperm.

Double fertilization starts with the entrance of the pollen tube into the synergids, which is controlled by secreted chemical attractants (see above; Higashiyama et al. (2001)). Besides chemical signaling from synergids, signaling from the egg cell plays a central role in selective fertilization of the egg cell with a single sperm cell, as was shown using the polyspermic *tetraspore* (*tes*) mutant of *Arabidopsis* (Scott 2008) and observed in an *Arabidopsis* mutant having only one sperm cell based on a CDC2A (*i.e.* gene *cell division cycle 2 homolog A* for the catalytic subunit p34 of a conserved protein kinase in *Arabidopsis* which is a key regulator of the cell cycle) deficiency, which showed exclusive fertilization of the egg cell (Nowak et al., 2006). Nowak et al. (2006) provided three alternatives to explain this preferential fertilization: advantageous positioning of the egg cell *versus* the central cell, predetermined fertilization of the egg cell or active signaling from the egg cell. The strong influence of egg cell signaling on the whole embryo sac during double fertilization was demonstrated using the *LACHESIS* mutant from *Arabidopsis* in which failure of homotypic nuclei fusion in the central cell blocks endosperm formation (Völz et al., 2012). Despite the strong influence of egg cell signaling, the acquisition of specific functions by the central cell was documented through the molecular characterization of the *Arabidopsis* mutant *glauce*, in which the expression of a BAHD acyl-transferase enables the active promotion of the central cell's own fertilization by the sperm cell through a signaling mechanism (Leshem et al., 2012).

Besides the activity of the female gametophyte in sperm cell guidance, the male gametes shed their immobility, which they gained during pollen tube guidance, and initiate actomyosin-mediated active migration towards the female gamete (Márton and

Dresselhaus 2008). Once the sperm cell reaches the egg cell, the fusion of both gametes is initiated. Little is known about the molecular background of this step, but the sperm cell-specific and membrane-localized HAP2/GCS1 of *Arabidopsis* was identified as one key regulator for the membrane merger step between male and female gametes in higher plants, *Chlamydomonas*, and *Plasmodium* (Liu et al., 2008). Cell cycle arrest prior to fertilization of the central cells seems to be controlled by a Polycomb group complex (PgC) which mediates epigenetic gene silencing, whereas the egg cell arrest might be mediated by the tumor suppressor retinoblastoma (RBR; reviewed in Guitton & Berger (2005)). No preferential egg cell fusion with either sperm cell could be detected (Ingouff et al., 2009; Hamamura et al., 2011), but fast blocking of the egg cell (polyspermy block) is proposed as a model to explain the directed fertilization of the endosperm with the remaining sperm cell (Hamamura et al., 2011). The endosperm, which develops in parallel with the embryo, has mainly a storage and nutritive function, but could be a source of signals important for embryogenesis by regulating trophic interactions between mother and embryo (Berger 1999).

After fusion of the parental genetic material during karyogamy, re-initiation of cell processes within the zygote takes place and the paternal genome is clearly involved in these processes in some plant species (Meyer and Scholten 2007) rather than silenced as proposed earlier (Vielle-Calzada et al., 2000). Tracking of the paternal histone H3.3 during double fertilization in *Arabidopsis* for example, suggests parental genomic imprinting based on the observed divergent development of H3.3 in the zygote and the endosperm (Ingouff et al., 2007). The expression of imprinted paternal alleles after fertilization is sufficient to provide biological functions during early plant embryogenesis (Weijers et al., 2001), and their activity hence underlines the importance of the male gamete.

2.1.3 The genetic control of male meiosis and pollen fate

Despite a relatively simplified morphological structure of the haploid male gametophytes of most gymnosperms and angiosperms, the various “tasks” of pollen grains described in the previous section demonstrate complex molecular programs in pollen which differentiate them from somatic cells.

Comparative transcriptome approaches were used to detect these programs and have mainly focused on the gametophytic control of postmeiotic developmental steps. It is speculated that the number of anther-specific genes involved in gametogenesis reaches

up to 37% of the mRNA population at any particular phase (Kamalay and Goldberg 1980). Besides a 60-90% overlap in genes expressed in mature pollen and somatic tissue, 20 000 to 24 000 unique mRNAs were estimated for mature pollen grains in maize (Willing et al., 1988). More recent transcriptome analyses in *Arabidopsis*, excluding the male gametophytic-sporophytic overlap, estimate approximately 5.6% (789 genes; Twell et al., (2006)) to 33% (7 177 genes; Borges et al. (2008)) of the total number of expressed genes in pollen to be pollen-specific. In line with this, comparative transcriptome analysis in *Arabidopsis* revealed higher estimates for pollen-specific genes compared to any other plant tissue, where approximately 3% of genes show tissue-specific expression (Ma 2005). Including the complete pollen developmental series in *Arabidopsis*, the expression of approximately 14 000 genes was reported, with decreasing gene activity as pollen mature (Honys and Twell 2004).

Comparative transcriptome profiling of male gametes in *Arabidopsis* has shown (1) enrichment of genes responsible for DNA repair, cell cycle and chromosome organization, (2) underrepresentation of the RNA-processing machinery (Pina et al., 2005; Borges et al., 2008), and (3) a developmental stage-specific switch of active transcript sets progressing from the uninuclear microspore towards the mature pollen grain (Mascarenhas 1989; Honys and Twell 2004). Separate sets of active transcripts in pollen grains and sporophytic tissues likely reflects natural selection for nonoverlapping sets of genes to avoid the proliferation of deleterious changes *e.g.* from uncontrolled pollen competition (Haldane 1932; Mulcahy and Mulcahy 1987).

Mutagenesis analyses have provided the most insights into the molecular genetic mechanisms of early gamete formation (*e.g.* meiosis, Harrison et al. (2010)). Such studies have focused on the genetic control of germline cell fate and differentiation patterns on the one hand, and cell cycle processes and control of meiosis on the other.

Meiosis is the central process defining the alternation between the diploid and haploid phases of plants, and generates genetic diversity and plasticity in progeny through chromosomal recombination. Male and female meiosis are conserved processes which are often not synchronized, in contrast to meiosis between the PMCs and between the megaspore mother cells (Armstrong and Jones 2001). Cytological events differentiating meiosis from ameiotic and somatic mitosis are (1) homolog pairing, (2) recombination, (3) the suppression of sister-chromatid separation during the first division, and (4) the absence of chromosome replication at the start of the second division (Wilkins and Holliday 2009).

Important components of meiosis have been studied *via* loss of function mutants, many of them from *Arabidopsis*, maize and lily (*e.g.* Baker et al. (1976); Ross et al. (1997); Siddiqi et al. (2000) and Mercier et al.(2001)), and have revealed effects on homeologous chromosome recombination (Lu et al., 2008), cell cycle defects (Yang et al., 2003), and defects on meiotic cell fate (Sundaresan et al., 1995; Yang et al., 1999), all of which lead to reduced fertility or death of the meiocyte (Fig. 1). Many genetic regulators for meiosis are shared between female and male sporophytes (Liu and Qu 2008).

Most reported male meiotic mutants have defects during meiosis I (Fig. 1; see Supplemental Table 1). Baker et al. (1976) divided them into two classes: mutants causing asynapsis and mutants affecting chromosome integrity. Asynaptic mutants produce partial or complete failure of homologous chromosome synapsis during prophase I, with lower frequencies of chiasmata and elevated frequencies of univalents. Structural abnormalities, like multiple spindles, misdividing univalents, formation of micronuclei and/or restitution nuclei, lead mostly to reduced fertility or sterility of the meiocytes.

Genetic factors underlying these phenotypes were examined primarily in *Arabidopsis* and have revealed mostly structural genes, as for example those necessary for sister chromatid cohesion such as *SYN1/DIF1*, which encodes a meiosis-specific cohesion subunit (Parisi et al., 1999) and *SWITCH1-2/DYAD* (*SWI1-2/DY*), which has some similarity to the structural maintenance of chromosomes (SMC) family proteins in humans (Mercier et al., 2001). Analysis of the *asynaptic1* (*asy1*) mutant of *Arabidopsis* suggests that *ASY1*, which is a homolog of the yeast *HOP1* gene (Ross et al., 1997; Armstrong et al., 2002), is involved in the formation of the synaptonemal complex (SC), as its absence leads to failure of homologous chromosome synapsis during prophase I. Homolog pairing is also affected by mutating the *AHP2-1* gene, which is a meu13^{+} homolog (Schommer et al., 2003). Additionally, male meiosis-specific histone-like proteins (*e.g.* meiotin-1) have been observed to play a role in chromatin packing during meiosis (Sasaki et al., 1990; Riggs and Hasenkampf 1991). The key feature of meiosis is recombination between homologous chromosomes. The yeast homolog of *SPO11* in *Arabidopsis* is one of the central genes coordinating meiotic recombination as it functions as a catalytic subunit of a topoisomerase causing double-strand breaks (DSB) in the genome. Additional proteins, like the MRE11-RAD50 (MR) proteins, are involved in the SPO11 pathway, being capable of tethering the ends of DNA molecules

or possessing various DNA nucleases, and are hence required for meiotic recombination, DSB and processing mis-folded DNA structures (Connelly and Leach 2002). *RAD51* is a counterpart of *RAD50*, as it is important for DSB repair (Shinohara et al., 1992; Doutriaux et al., 1998). Mutation of *Atspo11* causes defects in homolog pairing, recombination, and bivalent formation. Other genes encode for functionally redundant kinesins, like the AtNACK1/HINKEL and STUD/TES genes, which are

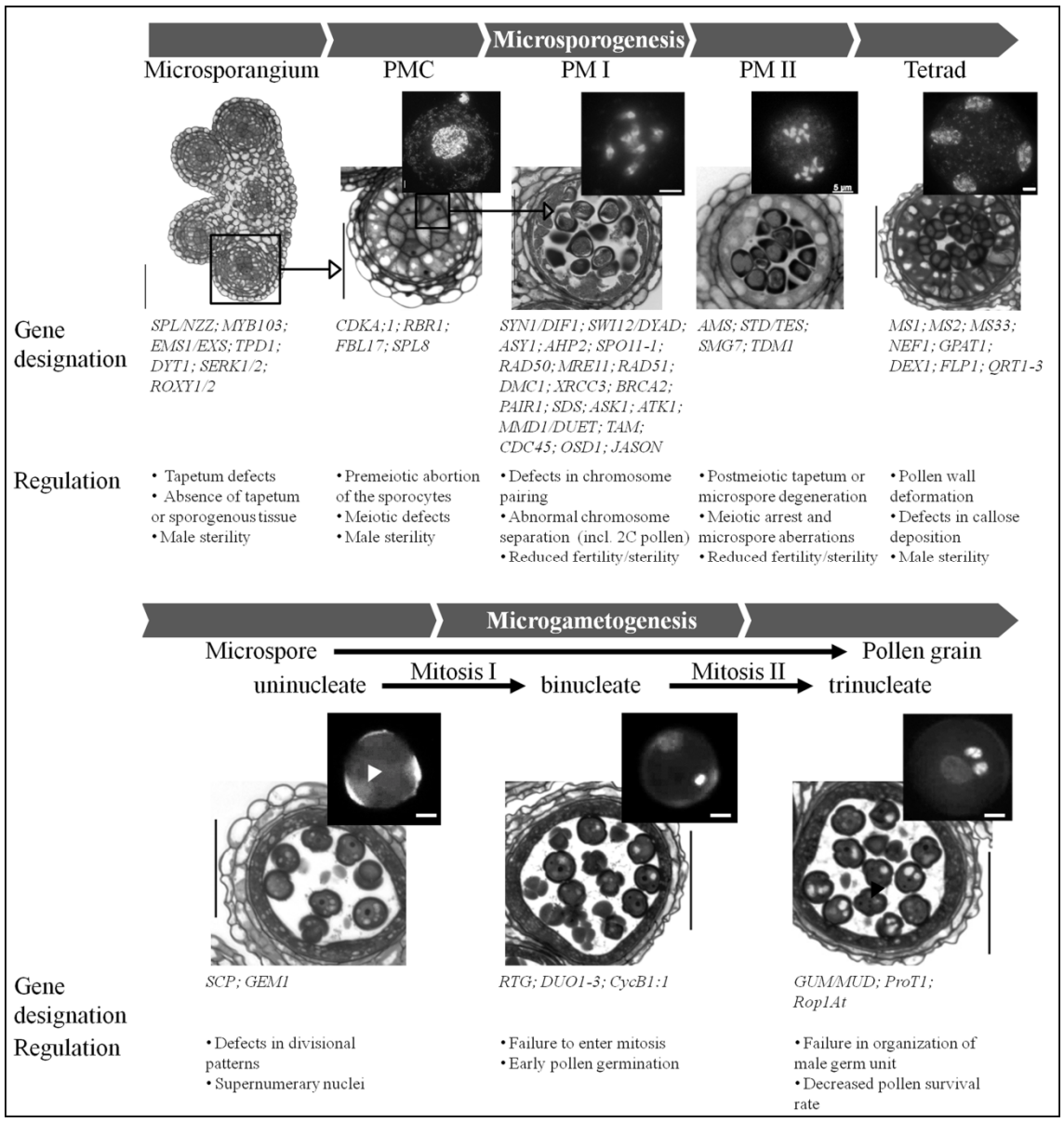

Figure 1. Cytological events of male germline development in *Boechera*.
Transversal sections of antherheads and DAPI-stained meiocyte spreads were prepared from sexual *Boechera* genotypes. Key regulators and their regulatory effects in malfunction mutants were assigned from *Arabidopsis* (see reviews Ma (2005) and Berger and Twell (2011)). Supplemental Table 1 contains the extended reference list of genes and their specifications according to the gene designations. PMC = pollen mother cell; PMI and II = pollen meiosis I and II. Bars of transversal sections = 50 µm; Bars of meiotic spreads = 5 µm.

believed to serve as a microtubule-associated motor during cytokinesis of both male and female meiosis inhibiting potential mislocalization of the nuclei (Hulskamp et al., 1997; Spielman et al., 1997; Tanaka et al., 2004).

A rare phenomenon is the involvement of a male and female meiosis-specific cyclin in homologous recombination during prophase I, as reported for the *SOLO DANCERS* (*SDS*) gene (Azumi et al., 2002). *SDS*, for which a phosphorylation activity of proteins involved in sister chromatid cohesion is proposed, is required for homolog synapsis and bivalent formation, similar to the function of the *Arabidopsis* genes *AtSPO11* and *AtDMC1* (Couteau et al., 1999; Grelon et al., 2001).

Usually, cyclins play a major role in cell cycle progression during mitosis and meiosis. The meiotic cell cycle is vital for the production of reduced pollen and any defect heavily disrupts proper gamete formation. Meiosis starts at the transition from G_2 to M-phase, which requires high levels of cyclin-dependent kinases (CDK) followed by their massive degradation once the chromosomes are in correct order at metaphase I (Brownfield and Köhler 2010). The main kinase in *Arabidopsis* is the A group CDKA;1 among 50 others, all of which can be divided in the A, B and D groups of plant cyclins (Renaudin et al., 1996). Mutagenesis of an A-type cyclin (CYCA1;2) in the *tam1* mutant in *Arabidopsis* illustrates their importance for cell cycle progression in male meiosis (Magnard et al., 2001; Wang et al., 2004). Mutation of *TAM1* causes asynchronous PMC meiosis, allowing only a subset of the meiocytes to enter meiosis II. The other subset produces, based on the missing second division, unreduced gametes with sister chromatid formation. A similar *Arabidopsis* mutant, *osd1*, lacks promoting activity for the transition from meiosis I into meiosis II. Although the molecular function of OSD1 is currently unknown, a CDK activity similar to CYCA1;2/TAM is suggested (d'Erfurth et al., 2010). CDK domain containing genes include *MALE STERILITY5* (*MS5*), which contains a fragment with similarity to a CDK subunit from *Xenopus laevis* (Glover et al., 1998). Other genes not representing cyclins, such as the *MALE MEIOCYTE DEATH1* (*MMD1/DUET*) gene (Reddy et al., 2003; Yang et al., 2003), are additionally hypothesized to regulate meiotic cell cycle progression during late male meiosis I, but mutation of both lead ultimately to pollen cell death.

Besides the array of defective meiotic mutants, others have been observed for the production of elevated frequencies of unreduced (diploid) gametes (see *osd1*). Various mechanisms leading to unreduced gamete formation (2C gametes) have been reported in the Rosaceae, Solanaceae and Brassicaceae (Veilleux 1985), the Poaceae (Harlan and

deWet 1975; Gallo 2007) and Ericaceae (Vorsa and Rowland 1997), with premeiotic, meiotic, and postmeiotic abnormalities being the most frequent correlates (Kaul and Murthy 1985). In general, the two chromosome sets in unreduced gametes can be either non-sister chromatids, due to a first division restitution (FDR), or sister chromatids, due to second division restitution (SDR). Depending on the recombination frequency non-sister chromatids are heterozygous from the centromere to the first chiasma, while sister chromatids are homozygous between the centromere and the first chiasma (Bretagnolle and Thompson 1995). Peloquin et al. (1999) estimated that the percentage heterozygosity transmitted by gametes is approximately 80% with FDR and less than 40% with SDR. Assuming a complete lack of recombination for unreduced gametes, they are desired for breeding purposes like the generation of new hybrids because (1) they offer complete homozygosity (through SDR) or complete heterozygosity (through FDR), and (2) they provide the ability to fix hybrid vigor (Brownfield and Köhler 2010).

Recently, a couple of genes were detected whose mutation cause an array of different male meiotic products in *Arabidopsis*, including high frequencies of unreduced gametes. One example for an elevated production of balanced, unreduced pollen is the mutation of *AtPS1*, which initiates nuclear restitution in male meiosis II through disruption of the spindle orientation, producing a mix of dyads and triads (d'Erfurth et al., 2008). The highly conserved AtPS1 contains domains putatively involved in protein–protein interactions, but its specific function during male meiosis is unknown. Different to the *Atps1* and *osd1* mutants, where the formation of unreduced pollen occurs by parallel spindles during meiosis II or *via* omission of the second meiotic division (SDR pollen), the loss of the male meiosis-specific gene *JASON* causes the formation of parallel arranged and fused spindles in male meiosis II, resulting in the production of unreduced FDR pollen (Erilova et al., 2009; de Storme and Geelen 2011).

Pollen mitosis I, which follows cytokinesis is a highly asymmetric cell division which critically depends on microspore polarization (Park et al., 1998). Determinants of microspore polarity were detected in the *gemini pollen* mutants of *Arabidopsis* (Park et al., 1998). Phenotypic behaviour of these mutants suggests that microspore polarity could be established during meiotic phragmoplast development, which involves specialized microtubuli. These microtubuli are putatively under control of the *GEM1* gene, a MAP215 family of microtubule-associated proteins that plays critical roles in the assembly and function of the meiotic/mitotic spindles (Park et al., 1998; Twell et al.,

1998). Another gene whose function is postulated to be necessary in setting up cell polarity is *SIDECAR POLLEN*, which might encode a repressor of cell division until cell polarity is achieved (Chen and McCormick 1996).

In their review Berger & Twell (2011) dissected the network of genetic factors involved in male germline cell fate and cell differentiation during microgametogenesis. Their model explains putative molecular pathways initiated after asymmetric division of the microspore, characterized by the exit of the vegetative cell from the cell cycle and formation of two sperm cells through symmetrical mitotic division. They illustrate that cell cycle progression from the G1 to the S phase in the generative cell is initiated by the F-BOX-LIKE 17 (FBL17), which functions as adaptor protein for the SKP-CULLIN-F-BOX (SCF) containing complex, known to act in cell-cycle control in mammals (Morgan 2006). The transient expression of the *FBL17* gene mediates the proteasomal degradation of the CDK inhibitors KRP6 and KRP7, leading to a loss of the CDK inhibition. Activated CDK's phosphorylate the retinoblastoma protein (RBR), which suppresses the E2F/DP pathway. Phosphorylated RBR inactivates supressors of E2F/DP-type transcription factors, which subsequently activate S phase genes to start the cycle for the second pollen mitosis. As opposed to the generative cells, *FBL17* is not expressed in the vegetative cell, leading to continuous repression of the E2F/DP pathway and exit from the cell cycle to G0 phase. Subsequently, germ cell differentiation genes are activated by the expression of the MYB family transcription factors *DUO1* and *DUO3*, whereby *DUO1* promotes the entry into pollen mitosis II (G2/M transition) *via* the activation of a cyclin-dependent pathway (CYCLINB1;1) and *DUO3 via* the activation of a cyclin-independent pathway. Chromatin remodeling, *e.g.* of the histone H3 variant HTR10, which is under control of *DUO1*, is supposed to play a role during germ cell specification *via* changes in CG methylation, and thus RNA dependent methylation pathways are supposed to be involved in the histone dynamic besides their role in regulation of transposable element activity (Berger and Twell 2011).

Nonetheless, beside the described activities of protein-encoding genes, the role of RNA-mediated gene regulation during male gamete formation is largely unknown.

2.1.4 The role of RNA-mediated gene regulation in the male germline

Only ~1.2% of the mammalian genome (The ENCODE Project Consortium *et al.*, 2007), ~2% of the mouse genome (FANTOM-Consortium 2005) and less than 50% of

the *Arabidopsis* genome (Yamada et al., 2003) are capable of coding proteins. The vast numbers of transcripts with low or no protein-coding potential were long referred to as "transcriptional noise", "junk" or "selfish" DNA (Ohno 1972; Crick 1979). Among them, long non-coding RNAs (lncRNAs) constitute a significant fraction of non-protein coding transcripts in humans, although some of them function as precursor molecules that are later processed into small RNAs (sRNAs) (Esteller 2011). As the name predicts and in contrast to sRNAs (<200 nt), long non-coding RNAs can range in size from ~200 to several thousand nucleotides and are most deeply studied in mammals (especially in human disease, see review Esteller (2011)), *Drosophila* and yeast (Mercer et al., 2009). LncRNAs can bear many signatures of mRNAs including 5`capping, splicing and polyadenylation, but have few or no ORFs (Au et al., 2011). It was shown that many of the characterized lncRNAs are, in contrast to sRNAs, not strongly conserved between mouse and human (Wang et al., 2004), but nevertheless they do exhibit a wide array of functions both in *cis* and *trans* and are a key player of the regulation of gene expression (Pang et al., 2006; Derrien et al., 2012). Explanations for this conflict were pointing to different evolutionary constraints on lncRNAs compaired to sRNAs. One reason could be that lncRNAs interact with a limited number of targets, hence increasing co-variable interactions through evolutionary time (Pang et al., 2006). In addition, the fact that some of the lncRNAs act as carriers of smaller functional products (see above) implies that only portions of the overall transcript might be under selection (Pang et al., 2006).

In plants knowledge about active lncRNAs compared to sRNAs is poor. Nevertheless, it is clear that lncRNAs and multiple sRNA pathways (*e.g. via* microRNAs (miRNAs), PIWI-RNAs (piRNAs) and small silencing RNAs (siRNAs)) are active in male gametophytes (Berger and Twell 2011; Dickinson and Grant-Downton 2011).

Similar to animals, sRNA systems in the plant germline are generally used for the control of genome stability, gene expression and defense (Bonnet et al., 2006). Small RNAs are 21 – 27 nt sized non-coding RNAs (ncRNAs) with a preference for transcription factors as targets, but other mechanisms, including response to stress or environmental changes and modulation of DNA methylation and/or repressive heterochromatic histone modifications, are reported. Three major sRNA synthesis pathways in plants are known (Bonnet et al., 2006).

The biogenesis of miRNAs involves transcription from their own locus (from 50-350 nt long precursor miRNAs) by RNA-polymerase II (POL II). The resulting hairpin-

like secondary structure is further processed by the RNase III catalytic enzyme DICER (*e.g.* Dicer-like1, DCL1) in several steps to produce miRNA:miRNA* duplexes, requiring the HYPONASTIC LEAVES1 (HYL1) protein as nuclear localization signal. Then the duplexes are methylated by RNA methylase HUA ENHANCER 1 (HEN1) and exported to the cytoplasm. Here the duplex is unwound and the miRNA is associated with the RNA-binding ARGONAUTE1 protein (AGO1), the core component of the RNA-induced silencing complex (RISC). The miRNA-RISC binds specifically to a target messenger RNA mediating its cleavage or translational repression.

The synthesis of siRNAs slightly varies from miRNA biogenesis. Small interfering RNAs are synthesized from various origins like viruses, transposons or transgenes by converting longer dsRNAs into 21 nt long siRNAs by DICER enzymes (DCL4), which are subsequently loaded into RISC and associated with AGO proteins (*e.g.* AGO4). This complex binds to the same messenger RNA from which it originates. Hence, the complex intitiates its silencing or enhances siRNA production by inducing the transformation of single-stranded RNA (ssRNA) into dsRNA *via* RNA-dependent RNA polymerases (RDRs) RDR1 and RDR6.

Some miRNAs first cleave a target mRNA expressed from trans-acting siRNA (TAS) loci into two fragments. After cleavage, both the 5`- or 3`-terminus can be converted into dsRNA by RDR6, and then DCL4 catalyzes the process into 21 nt ta-siRNAs that guide degradation of a target mRNA that is different from the ta-siRNA transcript from which they originated.

Although a general lack of sRNA expression in pollen has been detected (Borges et al., 2008), microarray analyses have revealed pollen-specific expression of key genes representing the three major gene families involved in miRNA and siRNA biogenesis pathways: *AGO*, *DCL*, and *RDR* (Grant-Downton et al., 2009a). Except for *DCL1*, all *DCL* genes loose their expression after pollen mitosis II, whereas many of the *AGO* and *RDR* family genes are constitutively expressed throughout microgametogenesis. Some sRNA pathway genes, such as *AGO5* and *AGO9*, or genes required to maintain the DNA methylation status like *METHYLTRANSFERASE1* (*MET1*) and *DEFICIENT IN DNA METHYLATION1* (*DDM1*), are even highly enriched in sperm cells (Borges et al., 2011). In this context the enriched expression of *AGO9* is interesting, since analysis of female gametogenesis reveals a significant role of *AGO9* in restriction of pre-meiotic differentiation of single sub-epidermal cells into single ovules, implying a similar function on male gametogenesis (Olmedo-Monfil et al., 2010). Independent microarray

analyses have revealed ~58 known miRNAs specifically or enriched expressed in pollen of *Arabidopsis* (Grant-Downton *et al.*, 2009a; Borges et al., 2011). One example for miRNA targeting of transcripts during pollen formation is the miR160 mediated degradation of *AUXIN RESPONSE FACTOR (ARF)* transcripts regulating auxin perception during pollen maturation (Grant-Downton et al., 2009a). Another prominent example is miR172-targeting of the transcription factor *Target of Eat2* (*TOE2*) in mature pollen, which is essential in earlier stages *e.g.* for repression of ectopic expression of floral development pathway genes (Grant-Downton et al., 2009a). In contrast, other miRNAs functions, such as the predicted function of the pollen-specific miR159 in the cleavage of a regulator (*DUO1*) of cyclin-CDK activity, as well as the role of the pollen-specific miR156a-f in the degradation of the floral meristem identity *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene family of transcription factors, remain unexplored (Borges et al., 2011). Furthermore it is known, that some miRNAs do not target coding mRNAs, but instead recognize TAS loci (see above). In pollen, miR173 and miR390 are known *Arabidopsis* miRNAs targeting *TAS* genes (*e.g. AtTAS3*; Montgomery et al. (2008)), and recently more pollen-specific ta-siRNAs were identified from the four *TAS* genes: *TAS1a*, *TAS1b*, *TAS1c* and *TAS2* (Grant-Downton et al., 2009b).

The interaction with sRNAs demonstrates one of the multifaceted functions of lncRNAs in plant reproductive tissues, which, as is known from humans, includes transcriptional and post-transcriptional regulation, chromatin modifications, cargo functions for subcellular protein trafficking, organelle biogenesis and as endogeneous RNAs (ceRNA) competing with miRNAs (Mercer et al., 2009; Zhu and Wang 2012). LncRNAs are transcribed from introns or intergenic regions; they can overlap with exons and/or introns of protein-coding regions in sense or antisense direction and can act both in *cis* and/or *trans* to modulate the expression of their target loci (Kim and Sung 2012).

Various lncRNA types exist in plants, with lncRNAs which act as precursors for sRNAs being the first to be discovered in plant reproductive tissues (Hirsch et al., 2006). Hirsch et al. (2006) discovered both inflorescence-specific lncRNAs (npcRNAs 58 and 155) and a miRNA-containing lncRNA that might act as ceRNA (npcRNA 78/miR162a, also called pri-miRNAs). The combination of functional lncRNAs with RNA-mediated gene silencing pathways in reproductive tissues was furthermore demonstrated for two subunits of the plant-specific multi-subunit nuclear enzyme RNA

polymerase IVb/PolV in *Arabidopsis*. RNA polymerase IVb/PolV is not essential for the viability in *Arabidopsis*, but participates in multiple small RNA-mediated gene silencing pathways, such as the siRNA-directed DNA methylation pathway (RdDM) (Pikaard et al., 2008). Two large subunits, NRPD1 and NRPE1 (Nuclear RNA Polymerase E1/D1) were identified as lncRNAs which serve as scaffolds for a siRNA-biogenesis pathway and subsequently mediate the silencing of overlapping and adjacent genes like short (SINEs; *e.g. AtSN1*) and long interspersed nuclear elements (LINEs; *e.g.* At5g27845; Wierzbicki et al. (2008)).

Other lncRNA types acting in reproductive tissues regulate gene expression by a range of mechanisms, including mediating chromatin-modifying activity, as has been discovered for the floral repressor *FLOWERING LOCUS C* (*FLC*) in *Arabidopsis*. Perennial plants need vernalization prior to flowering and the MADS-box transcription factor FLC directly represses the floral promoter *FLOWERING LOCUS T (FT)* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) in leaves and apices preventing premature flowering (Helliwell et al., 2006). Two cold-induced FLC long non-coding *antisense* transcripts, *COOLAIR* and *COLDAIR*, have a role in the epigenetic silencing of *FLC* (Zaratiegui et al., 2007). Early *FLC* silencing is mediated by *COOLAIR*, whereas *COLDAIR* causes epigenetic repression of *FLC* by physically associating with components of the Polycomb Repressive Complex 2 (PRC2), which is a key regulator of epigenetic states during the development of endosperm, catalyzing histone H3 lysine 27 trimethylation (H3K27me3), a repressive chromatin mark (Schwartz and Pirrotta 2008).

A third lncRNA-type, natural *antisense* transcript (NAT), represents a significant portion of the transcriptome and has partial complementary to protein-coding transcripts (reviewed in Kim & Sung (2012) and Au et al. (2011)). Both, *cis*- and *trans*-NATs are known. A portion of NATs induce translation of transcription factors (TF). Subsequently the cytoplasmatic localization of NAT-TF protein complexes is directed by 5S ribosomal RNAs. Such NATs are generally considered to fine tune the expression of genes. Other NATs can form double-stranded RNAs (dsRNA), which subsequently serve as templates to generate endogenous siRNA (*cis*-nat-siRNA) to induce transcriptional silencing. The detection of *Sho*-specific *sense* and *antisense* transcripts in *Arabidopsis* is one prominent example of the potential role of a flower-specific overexpressed *cis*-NAT serving as template for the synthesis of *cis*-nat-siRNAs (Zubko and Meyer 2007). The *Sho* gene encodes an enzyme that is responsible for the synthesis

of plant cytokinins, which are important for shoot meristem growth, flower differentiation and male and female gamete development. The *Sho* Nat regulates *Sho* transcripts *via* induction of a gene silencing mechanism through a RNA interference pathway involving the DICER-mediated formation of nat-siRNAs. Additionally, various NATs from self-incompatibility genes of *S* loci in maize and in *Brassica oleracea* have been observed in reproductive tissues with particular interest in the developing male sexual tissues (Cock et al., 1997; Ansaldi et al., 2000). In male gametophytes the *S* NATs and *S* genes show complementary expression patterns, with *antisense* RNAs decreasing in abundance in maturing anthers, whereas *sense S*-like transcripts accumulate to the highest level in mature pollen. The reciprocal relationship of *S* NATs and *S* genes in mature pollen supports the possibility of a regulatory role for the NATs in control of male reproductive development.

So far, only a few lncRNAs have been reported to be essential for male gamete development. In maize, a putative lncRNA, *Zm401*, is expressed specifically in pollen and regulates the expression of genes critical for pollen development, including *ZmMADS2*, *MZm3-3* and *ZmC5* (Au et al., 2011). Overexpression of *Zm401* leads to abnormal tassels and degenerate anthers. The mechanism of differential gene regulation is still undetermined, but interaction of transcript fractions with different members of transcriptional protein complexes is assumed. Research on lncRNAs in humans suggested the induction of lncRNA activity to be generally related to the organisms' response to environmental changes (Au et al., 2011). One example for a pollen-specific lncRNA whose activity is interrelated with an abiotic factor is the long-day–specific male-fertility–associated RNA (LDMAR) in *Arabidopsis*, which regulates photoperiod-sensitive male sterility (PSMS) and is thus required for male fertility (Ding et al., 2012). The *MALE FERTILITY11* (*BcMF11)* gene is another lncRNA described for *Brassica campestris*, and is expressed during pollen formation (Song et al., 2007). Its function is unknown, but mutagenizing *BcMF11* causes serious morphological defects on pollen grains leading to reduced male fertility (Song et al., 2012).

A (pre) -meiotically acting lncRNA in plants is yet undiscovered. Furthermore, the few examples illustrate, that the discovery and functional characterization of lncRNAs in plants is still in its infancy. This might be caused by the fact, that lncRNA are difficult targets for mutational screens, because RNA genes are immune to frameshift or nonsense mutations, and are often small and multicopy (Eddy 2001). Thus, comparative

analyses, including unannotated sequences might give a better insight into the RNA world, reducing the bias towards discovery of protein-coding genes.

2.1.5 Reproduction types

Male gamete formation is tightly bound to the type of reproduction of its host species. In flowering plants reproduction can be divided into two types: sexual and asexual reproduction. Sexual reproduction or amphimixis (greek, "amphi" for "on both sides" and "mixis" for "mixing"), including vivipary, occurs in the overwhelming majority of plants and involves meiotic reduction of the somatic chromosome number and its restoration during the fertilization. Sexual reproduction is the most successful type of reproduction, although it often holds the well-cited "2-fold cost" (*i.e.* not true for isogamous organisms), which is best interpreted as cost of producing males which do not produce own offspring and the cost of meiosis segregating co-adapted alleles (Charlesworth and Charlesworth 1978; Bell 1982). Advantagous features of sexual reproduction are genetic recombination and syngamy, both which produce genetic variation upon which natural selection can act to allow rapid adaptation to environmental variability (Fryxell 1957). An alternative explanation for its success refers to recombination as a DNA repair process (Bernstein 1977) which minimizes the long-term accumulation of disadvantageous mutations (Kondrashov 1985).

Asexual reproduction in plants, including pseudovivipary, involves the inheritance of genetic material from a single parent and occurs in two types: vegetative propagation and apomixis. In vegetatively propagating plants all organs can be used to produce new ramets by mitotic division. The vegetative propagation of angiosperms *via* horizontal runners (*e.g. Fragaria*), stems (*e.g. Rubus*), leaves (also called bryophyllum, *e.g. Kalanchoe*) and roots (*e.g. Ranunculus ficaria*) is prevalent in grasses (*Poa*) and aquatic vascular plants (*e.g. Elodea*). Only a minority, approximately 0.1%, of the angiosperm species reproduce asexually *via* seeds, also called apomixis or agamospermy, which is by definition a female trait (Nogler 1984).

Some apomictic species do not suffer the costs of producing males due to purely clonal reproduction *via* the female germline. Nonetheless, approx. 90% of apomictic species are pseudogamous and produce pollen (see *2.2.2*, Mogie (1992)), potentially having some of the disadvantages of sexuality (*e.g.* requirement for a mate in case of outbreeders) without the counterbalancing advantages (*e.g.* recombination). Accumulating mutational load is one of the proposed disadvantages of apomicts and is

characterized by the irreversible random loss of the lowest mutational class (*i.e.* fewest mutation) of genotypes in apomictic populations (Muller's ratchet, Muller (1964)). Thus, counterbalancing this process requires tolerance mechanisms, such as polyploidization or a potential reduction of the rate of mutation accumulation, similar to the approx. ten-fold reduction of the mutation rate in land plant organelles relative to nuclear mutation rates (Lynch 2010) or the mutation rate variances between highly and low expressed genes (Martincorena et al., 2012), which support long-standing theories predicting that selection could modulate the rate of mutation itself (Kimura 1967; Levins 1967). In apomicts none of these mechanisms were reported so far. In bdelloid rotiferes however, an increased DNA repair efficiency and massive gene conversion (*i.e.* repair of original gene copies using a second gene copy) have been observed (Mark Welch and Meselson 2000; Gladyshev and Meselson 2008; Mark Welch et al., 2008). On the contrary, both apomixis and self-fertilization are major advantages for colonization (Baker's law; Baker (1955)), and are characteristic of many invasive species (Rambuda and Johnson 2004), for rare or endangered species and for edges of species' geographical ranges (Reproductive Assurance Hypothesis; Stebbins (1950), Jain (1976)). Indeed, similar to parthenogenetic animals, a screen of asexual taxa in higher plants revealed tendencies towards larger ranges, higher latitudes and elevations compared to their sexual relatives, and towards colonization of previously glaciated refugia ('geographical parthenogenesis', Vandel (1928) and Bierzychudek (1985)). Therefore, common habitats for apomicts include arctic and alpine sites and disturbed environments (Hörandl and Paun 2007).

2.2 Fundamentals of apomixis research

2.2.1 Types of apomixis

Hans Karl Albert Winkler, a german botanist (1847-1945), shaped not only the term "Genome", but used as one of the first the term "Apomixis" (greek, "apo" for "from, away from" and "mixis" for "mixing") for the production of seeds without fertilization by pollen and for vegetative propagation (Winkler 1908). At present the term apomixis is used solely for asexual reproduction *via* seeds, which is by definition a female trait (*e.g. Cupressus dupreziana* is an androgeneous exception; Pichot et al. (2001)) where all progeny are genetically identical copies (clones) of the mother plant (Nogler 1984). General characteristic features of apomicts include (1) facultative expression of the trait in the majority of apomictic species, (2) polyploidy (with some exceptions *e.g.*

Boechera), and (3) hybridization (between sexuals or between sexuals and apomicts). Apomixis refers to three major developmental alterations of the sexual pathway: (1) apomeiosis (unreduced gamete formation), (2) parthenogenetic development of the egg cell with autonomous embryo formation (*i.e.* without fertilization), and (3) the formation of a fully functional endosperm. Apomixis is observed in about 40 angiosperm families, whilst 75% of them are found in Asteraceae, Rosaceae, Poaceae (Hanna and Bashaw 1987; Asker and Jerling 1992; Carman 1997). Three types of apomictic seed formation exist: sporophytic and gametophytic apomixis, the latter of which is subdivided into diplospory and apospory (Nogler 1984). All three types are "short-circuited" versions of the sexual pathway (see Fig. 2; see Koltunow (1993)). Sporophytic apomixis, also called adventitious embryony, usually occurs late in mature ovules were the embryo

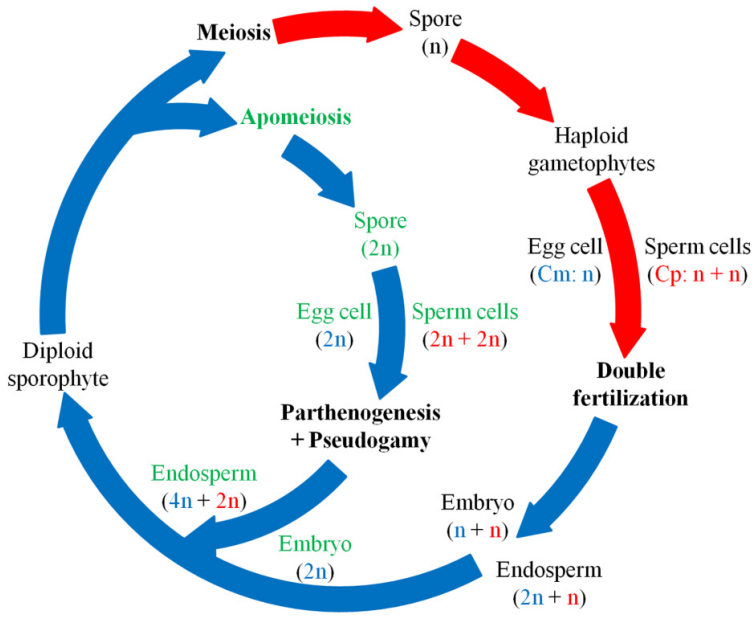

Figure 2. The diplo-haplontic plant life cycle. The switch from the diploid sporophyte stage (blue) towards the haploid gametophyte stage (red) which is typical for sexual species is bypassed ("short-circuited") in the germline of gametophytic apomicts *via* parthenogenetic development of the egg cell and a pseudogamous production of endosperm. Modified from Berger and Twell (2011).

is derived not from the megaspore mother cell, but from one of the somatic cells of the nucellus, whereas the pseudogamous endosperm is derived by fertilization of the central cells of the embryo sac by a pollen nucleus (Koltunow 1993). Adventitious embryony is mainly found in diploid tropical genera like *Tulipa* and *Lilium* (Liliaceae; Marasek et al. 2004), *Nigritella* (Orchidaceae; Teppner (1996), Asker & Jerling (1992)) and *Opuntia* (Cactaceae; Asker & Jerling, (1992)). In apospory the alternation of the sexual pathway occurs earlier, resulting in an unreduced embryo sac initiated from somatic cells of the nucellus (aposporous initials) and subsequent parthenogenetic development of the

embryo. Similar to adventitious embryony, multiple embryo sacs (polyembryony) can occur within one ovule depending on the species (Koltunow 1993). Pollination of the central cell is generally required, whereas autonomous endosperm formation is rare in aposporous species with exception of *Hieracium* (Bashaw and Hanna 1990). Apospory is present in many species of Rosaceae (*e.g. Sorbus* and *Rubus*; Asker & Jerling (1992)) and *Poaceae*, (*e.g. Panicum*; Bashaw & Hanna (1990)), but less frequent in Asteraceae (*e.g. Hieracium*, *Achillea*; Terziiski *et al.* (1995)). In diplosporous apomicts, the sporogenous tissue gives rise to an archesporial initial cell which differentiates into a megaspore mother cell (MMC) with the somatic chromosome number. The MMC develops either directly *via* three mitotic divisions (Antennaria type) or indirectly *via* abortion of meiosis I (apomeiosis; Taraxacum, Ixeris type) and develops into an unreduced embryo sac (Nogler 1984). Additional types of diplospory include premeiotic chromosome doubling (Allium type; Nogler (1984)). Diplospory occurs in a wide range of species, is most common in the Asteraceae but best studied in *Tripsacum* (Grimanelli et al., 1998) and *Taraxacum* (Mogie 1992). Similar to aposporous apomicts, autonomous endosperm formation is rare in diplosporous apomicts, and hence the central cell requires fertilization via pollen (pseudogamy). Unless otherwise specified we subsequently refer to pseudogamous diplospory when using the term “apomixis”.

2.2.2 Modifications for balancing the endosperm

The observation of high seed abortion rates in intra- and interspecific crosses Johnston et al. (1980) was initially hypothezised to arise *via* parent-of-origin effects that accompanied deviations from the 2 maternal to 1 paternal genome ratio in the endosperm (Quarin 1999). The 2 to 1 ratio reflects the parental conflict theory of Haig and Westoby (1991), which predicts selection for maternal and paternal alleles (“parental imprinting”), which are involved in the acquisition of resources, depending upon their opposing effects on endosperm growth. Considering the 2 to 1 endosperm balance number (EBN, *i.e.* an effective ploidy), and that an apomictic plant produces meiotically-unreduced ovules, viable seed formation depends strongly on the ploidy of the pollen.

Hence, pseudogamous apomicts display a wide range of adaptations to cope with problems during fertilization, such as the switch of selective fertilization from egg cell to central cell, or relaxation of genomic balance during endosperm formation (Nogler 1984; Savidan 2000; Spielman and Scott 2008). For example, a mechanisms to prevent

egg fertilization could include pre-fertilization induction of embryo development or thickening of the egg cell wall, as has been shown in grasses (*e.g. Dichanthium annulatum*; Spielman & Scott (2008)).

Disruption of the endosperm balance number occurs in apomicts that produce unreduced egg cells (*i.e.* apomeiosis) and reduced pollen, leading to an unbalanced endosperm ratio *via* fertilization of the unreduced double-nucleated central cell. In pseudogamous species like *Boechera*, the Polygonum type of embryo sac is characterized mostly by stable endosperm formation with a 4m:2p ratio by fertilization of the two fused unreduced polar nuclei by one unreduced generative sperm cell, although some deviation from the 2:1 balanced endosperm ratio in *Boechera* (Voigt et al., 2007; Aliyu et al., 2010) and other species (*e.g. Paspalum*, Quarin (1999)) is known.

In other species the deleterious effects of deviations from the stable EBN are circumvented by early sperm cell degeneration (*e.g. Zephyranthus* (*Atomosco*) *texana*; Pace (1913)) or fertilization of a uninucleate, respectively a fused secondary uninucleate central cell with one haploid pollen, as observed in mature four-nucleate aposporous embryo sacs in *Pennisetum* (Ozias-Akins et al., 2003) and in some apomictic Pyrinae (*e.g. Sorbus*, Jankun and Kovanda (1987)). Additional strategies circumventing endosperm imbalance include polyspermy (*e.g. Ranunculus auricomus*, Rutishauser (1954)), autonomous (*i.e.* without fertilization) endosperm formation (*e.g. Hieracium* species, Nogler (1984); Tucker et al. (2003)) or relaxation of the genetic requirements for a balanced endosperm (Nogler 1984).

2.2.3 The role of pollen in pseudogamous diplospory

Various mechanisms exist to stabilize the EBN and avoid uniparental imprinting in apomicts, most of which involve unreduced pollen (see *2.2.2*). Alternatively to their effects on endosperm balance number, unreduced pollen formation in pseudogamous species may be selected for by nature of their ability to mask deleterious mutations (Nogler 1984). The mechanism of selective guidance and fusion of one sperm cell with the central cell is unknown, whereas several mechanisms are proposed for prevention of egg cell fertilization (see *2.2.2*). Analysis of the pollen-pistil interaction of the non-pseudogamous apomict *Commiphora wightii* (Mukul myrrh tree) revealed disruption of pollen tube growth half-way through the pistil when pollinated with its own pollen, suggesting a disruptive mechanism induced by the female gametophyte (Gupta et al., 1998). On the contrary, in pseudogamous apomicts fertilization with both self and

foreign pollen is tolerated, exhibiting a strong relaxation of the fertilization-dependent barriers known from sexual systems (loss of self-incompatibility) and which suggests the participation of pollen in female fitness (Noirot et al., 1997).

The participation of pollen in female fitness can be furthermore understood as contribution to the maintenance of asexual lineages existing in sympatry with sexual individuals (Carrillo et al., 2002), while the female function (apomixis *sensu stricto* female trait) is most effective during the establishment of apomixis (Mogie et al., 2007). Considering both sexual and asexual individuals as simultaneous hermaphrodites in sympatry (*e.g. Taraxacum* and *Boechera*), pollen contribution to the establishment and persistence of apomictic individuals within sexual populations could be three-fold. Firstly, pollen from apomicts, especially from highly selfing taxa such as *Boechera* (Roy 1995), could provide newly recombined asexual genotypes through crosses with outcrossing sexuals, thereby reducing their own cost of pollen production (Charlesworth and Charlesworth 1978; Mogie 1988). Secondly, the production of unreduced male gametes in asexual lineages contributes to the classical cost of sex, if progenies of crosses with sexuals are viable, because it provides the benefits associated with sex for the asexual population (Charlesworth 1980; Lynch 1984; Mogie et al., 2007). Finally, if asexual individuals produce poor quality pollen (*i.e.* from mutation accumulation or polyploidization) which fertilizes sexual mothers, the cost for sexuality and the reproductive success of sexual competitors decrease. Hence, pollen production in asexuals in conjunction with the ability of asexual lineages to sire progeny in crosses with sexual lineages can influence the conditions under which long-term coexistence of sexuals and asexuals can be maintained (Britton and Mogie 2001).

2.3 The genetic control of apomixis

2.3.1 Theories

Apomixis is hypothesized to be under genetic control, but its characterization at the molecular level is still in its infancy (Nogler 1984; Savidan 2000; Grossniklaus 2001). While some early geneticists proposed dominant inheritance of apomixis based on crossing in *Hieracium* (Ostenfeld 1910), others explained variation in the expression of apomixis by a sensitive genetic balance of a few recessive genes with dosage effects (Müntzing 1940). Referring to the three-fold alternation of the sexual pathway (apomeiosis, parthenogenesis, and altered selective fertilization for functional endosperm development) it has furthermore been hypothesized that recessive genes

might be responsible for each apomixis component (Powers 1945; Spillane et al., 2001). Nogler (1984) criticized theories involving recessive genes, as the requirement of homozygosity and a superior assertiveness of those putative recessive genes was not mirrored by the apomictic cases examined at that time. Simple Mendelian factors could not explain the reappearance of parthenogenesis from sexual parents (*e.g.* observed in *Ranunculus auricomus*, Nogler (1984)), thus leading to a model which describes apomixis as a quantitative trait under polygenic control (Asker and Jerling 1992; Matzk et al., 2005). On the contrary, asexuality as a dominant trait induced by a single master gene regulating the different apomixis elements is an attractive idea, as random mutations (*i.e.* Darwinian evolution) leading to the independent "coordinated" evolution of several genes regulating each apomixis component is unlikely (Mogie 1992). Co-occurrence of sexual and apomictic processes in apomicts exhibiting adventitious embryony or apospory (*e.g.* in *Brachiaria*) suggests independently operating cell- and region-specific gene expression programs (heterotopy), retaining the order and timing of events derived from the sexual pathway (Koltunow 1993). Temporal deregultation or rearrangement (heterochrony) of those subprograms that normally constitute the sexual pathway could additionally explain the onsets of apospory, diplospory and adventitious embryony (Grimanelli et al., 2001; Grossniklaus et al., 2001; Spillane et al., 2001). The assumption of a dominant mutant allele for apomixis and recessive wild type alleles, as proposed in initial models from Mogie (1992), for the regulation of diplospory in *Taraxacum* and from Nogler (1984), for apospory in *R. auricomus*, could explain flexibility in terms of temporal and spatial expression of the various types of apomixis. On the contrary, this hypothesis is challenged by segregation analyses in diplosporous species like *Taraxacum, Erigon* and *Boechera*, (van Dijk et al., 1999; Noyes and Rieseberg 2000; Schranz et al., 2005), which suggest complexes of at least 2-3 coadapted cosegregating genes controlling the three traits of apomixis. On the other hand in some aposporous species like *R. auricomus* and *Panicum maximum* apomixis could indeed be regulated by a single master regulatory gene influencing the other apomixis elements in a cascade-like fashion. Considering that apomixis has independently arisen in various taxa, it is likely that no single model applies to all (Grossniklaus et al., 2001).

Coadapted apomixis factors could be tightly linked in a complex which acts as non-recombining unit (Jefferson 1993), as is evidenced by the presence of apospory markers in hemizygous chromosomal regions in some Poaceae (*Brachiaria decumbens*,

Tripsacum dactyloides and *Paspalum simplex;* reviewed in Grossniklaus et al. (2001)), as well as in a number of dicots (Grossniklaus et al., 2001).

Besides the view of apomixis as derived version of the sexual pathway, the hybridization-derived floral asynchrony (HFA) theory proposes that the asynchronous expression of duplicated gene sets which are initially necessary for female gamete development may cause apomixis (Carman 1997). This hypothesis is opposed to Harlan's observations in *Bothriochloa* and *Dichantium* which led him to conclude that "…apomixis and sexual reproduction are not alternative modes of reproduction, either genetically or operationally, but are simultaneous and independent phenomena…. The genes controlling normal sexual reproduction are not allelic to those controlling apomixis in the conventional sense" (Harlan et al. (1964), p. 46). Nonetheless, the HFA theory was recently supported by heterochronic expression profiles in four ovule developmental stages between sexual and apomictic *Boechera* (Sharbel et al., 2010). The overrepresentation of transcription factors among the expression profiles of apomicts points to a network of epigenetic and post-transcriptional regulation during germline specification (Twell 2010) and apomixis expression (Sharbel et al., 2010; Grimanelli 2011). Gene expression changes or the occurrence of epialleles could be the result of hybridization between distinct genotypes (Carman 2001). Such epialleles would not only be stable inherited (genomic imprinting), but could additionally be induced rapidly on a genome-wide level, providing the raw material needed to accumulate the genetic factors necessary to control the different apomixis elements (Koltunow and Grossniklaus 2003).

2.3.2 Genetic factors causing apomixis elements

The genetic control of apomixis components has been mostly observed in female tissues (apomixis is *sensu stricto* a female trait). Examining natural aposporous apomicts with mapping strategies, numerous molecular markers linked to apospory have been detected in *Paspalum notatum* (Martínez et al., 2001), *Panicum maximum* (Ebina et al., 2005)*, R. auricomus* (dominant allele A^-; Nogler (1984)), *Hypericum perforatum* (HAPPY locus; Schallau et al. (2010)) and *Pennisetum squamulatum* (Ozias-Akins et al., 1998), all of which support a single locus model for this trait (ASGR, apospory-specific genomic region; Ozias-Akins et al.(1998)). In contrast, a segregation study in aposporous *Poa pratensis* suggested that apomixis could be inherited by five different loci, demonstrating that apomixis in natural systems might be more complex than

previously reported (Matzk et al., 2005). Here it was postulated that asexual seed formation is induced by dominant alleles of two genes, the *Apospory initiator* (*Ait*) gene and *Parthenogensis initiator* (*Pit*) which are solely expressed in apomicts, two preventer counterparts, the *Apospory preventer* (*Apv*) and *Parthenogenesis preventer* (*Ppv*), in addition to a homozygous recessive *Megaspore development* (*Mdv*) in apomicts.

Marker analysis in diplosporous apomictic dandelions (*Taraxacum officinale*) suggested that a single sex-specific dominant locus, *DIPLOSPOROUS* (*DIP*), causes unreduced MMC formation (van Dijk and Bakx-Schotman 2004), while in diplosporous *Erigeron annuus* two independent loci were proposed to control apomixis (Noyes and Rieseberg 2000).

Mutant analyses have led to the identification of numerous factors which can mimic apomictic components (Spillane et al. (2001) and Grossniklaus et al. (2001)), and include genes causal for the spontaneous induction of embryo formation (*SOMATIC EMBRYOGENESIS RECEPTOR KINASE* (SERK), carrot; *LEAFY COTYLEDON* (*LEC1*), *Arabidopsis*), cell fate decisions (*e.g. multiple archesporial cells 1* (*mac1*), *multiple sporocytes1* (*msp1*), maize), induction of apomeiosis (*e.g. switch1* (*swi1*); *Arabidopsis*), induction of parthenogenesis (*e.g. haploidy initiator* (*hap*); barley) and fertilization independent endosperm formation (*e.g.* through the PcG proteins *medea* (*mea*), *fertilization-independent endosperm* (*fie*) and *fertilization independent seed* (*fis*), *Arabidopsis*). So far no stable apomictic mutant was recovered from sexual species which implies that this trait may require gains in functions that are solely present in recurrent natural apomicts (Vielle-Calzada et al., 1996).

Discovery of gain of function genes involves comparative expression profiling studies between natural apomicts and sexuals. Evolving technology (*e.g.* laser microdissection method (LAM)) has facilitated the comparative expression analyses of gametes in early developmental stages. Differential expression analyses between natural apomictic and sexual genotypes were conducted mostly in monocots, and have revealed a plethora of candidate apomixis factors (Chen et al., 1999; Pessino et al., 2001; Rodrigues et al., 2003; Albertini et al., 2004; Albertini et al., 2005; Singh et al., 2007; Polegri et al., 2010). One promising example involves the *SERK* family members in *P. pratensis* and *Hieracium*, whereby mutations in sexual systems were independently identified in apomicts (Tucker et al., 2003; Albertini et al., 2005). Pp*SERK1* and Pp*SERK2* were differentially expressed between sexual and apomictic individuals, and according to its function in *Arabidopsis* it is hypothesized to induce embryo sac

development and redirect gene products to different compartments by altered signaling (Albertini et al., 2005). A second candidate is *APOSTART*, which is downregulated in apomicts compared to sexuals during both male and female meiosis (Albertini et al., 2005). Similar to *MALE MEIOCYTE DEATH1* (*MMD1*) in *Arabidopsis* it may be a candidate for cell death in male meiocytes in *Poa pratensis*.

In *Hieracium* the floral organ-identity gene *DEFICIENS* (*DEF*) was isolated from an apomictic line and showed predominant expression in early stamen development and petal primordia, but was not detected in ovules of either sexuals or apomicts (Guerin et al., 2000). In other taxa DEF is expressed in ovules and functions in heterodimers together with the B-class MADS-box transcription factor protein GLOBOSA (see *2.1.1*), which together maintain continuity of the sporophytic developmental program of the ovule. Thus, the absence of *HPDEF* in sexuals could activate cellular differentiation programs of the megaspore mother cell, in contrast to apomicts were down regulation might initiate reprogramming of sporophytic cells towards a gametophytic fate (Guerin et al., 2000). Recently, time series differential expression analyses of ovule development during meiosis in *Boechera* revealed heterochronic gene expression profiles between sexual and apomictic genotypes, with the peak of differentially expressed sequence tags (*e.g.* transcription factors) at premeiotic stages (Sharbel et al., 2010).

2.3.3 Apomictic crops

Except for citrus, none of the major crops reproduce apomictically, and furthermore no natural apomicts are closely related to crop species (except *Tripsacum* with maize; Bicknell & Koltunow (2004) and *Boechera* with *Brassica* species; Beilstein et al. (2006)). The induction of apomixis in the major crop plants could hypothetically allow the fixation of any desired trait (*e.g.* hybrid vigor or plant disease resistance), provide faster and less cost intensive plant breeding, and enable seed propagation of vegetatively-propagated crops to lower rates of pathogen transmission (Dresselhaus *et al.*, 2001). Moreover, small farm businesses and farmers of the developing world could benefit by reproducing their own cultivars from locally adapted varieties (niche breeding).

Progress on introducing fully operational apomixis into sexual systems has been limited. Ploidy barriers (gene dosage) and to a lower extent mentor effects (*e.g.* genomic imprinting) often hinder the replacement of sexuality by apomixis, as crosses in *R.*

auricomus revealed, using diploid sexual mother plants and tetraploid or hexaploid apomictic pollen donors (Hörandl and Temsch 2009). Limited crossability between sexuals and apomicts was shown in *Boechera*, were the seed production of F_1 polyploids derived from nonreduced pollen and reduced eggs was low, suggesting a failure in the transmission of apomixis (Schranz et al., 2005).

Attempts to introgress apomixis components from apomictic wild relatives into crops using back crossing (BC) strategies were conducted in wheat (with *Elymus rectisetus*; Liu et al. (1994)), pearl millet (with *Pennisetum purpureum*; Savidan (2001)) and maize (with *T. dactyloides*; Savidan (2001)). Major limitations seemed to be high or complete male sterility in F_1 hybrids, BC_3 lines with low agronomic value, and the restriction of the introgression strategy to close related species.

The mutagenesis approach is based on the assumption that apomixis could be induced in sexual plants by perturbation of gene activity (Grossniklaus 2001). In line with this, many mutants with phenotypes corresponding to elements of apomixis have been described (see *2.1.3* and *2.3.2*). Recently, an attempt was made to introduce clonal reproduction into maize (Marimuthu et al., 2011). Crosses of a triple mutant that controls meiosis (*spo11 rec8 osd1* resp. *cyca;1/tam*; *MiMe* mutant) as female or male parent with another mutant controlling for chromosome segregation (*CENH3*; *GEM* mutant) as female or male counterpart led to some level (<50%) of diploid progeny lacking the maternal or paternal contribution, depending on the parental mutant line.

Nonetheless, these results show that modification of sexual species to propagate clonal seeds via apomixis mimicking is hindered by low penetrance of the trait.

2.4 The genus *Boechera* (Brassicaceae)

2.4.1 A model plant system for studying apomixis

The perennial *Boechera* (Löve & Löve (1975); rockcress; formerly *Arabis* (x=8); Al-Shebaz & Windham (1993+); Fig. 3) is a genus of the Brassicaceae almost exclusively found in North America. In addition, one species was found in Siberia and one in Greenland. The genome size of *Boechera* is heterogeneous (Aliyu et al., unpublished results) and about 1.8 times the size of its close relative *Arabidopsis* (~ 125 Mb). Their divergence is estimated to have taken place about 10-14 million years ago (mya; Koch et al. (2001)). Comparisons of a sexual *Boechera stricta* to *Arabidopsis* revealed a relatively high genome conservation of approx. 71.4% (Windsor et al., 2006).

Karyological analysis of the chromosome composition revealed 1.2-5.8 µm sized chromosomes with median to subterminal centromere positions (Kantama et al., 2007). Additionally, two types of independently-evolving extra chromosomes (*Het* and *Del*) were observed in diploid and triploid apomicts, with differences in morphology, size, and levels of heterochromatin (Sharbel et al., 2004; Kantama et al., 2007).

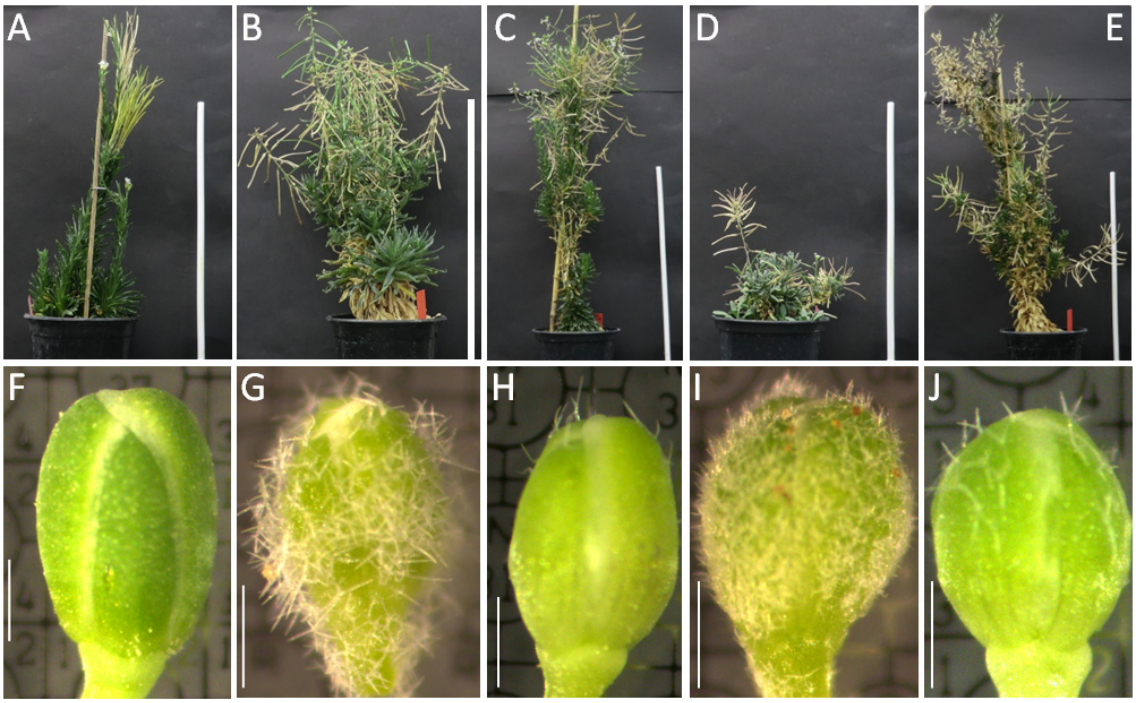

Figure 3. Example photos of *Boechera* genotypes used for transcriptome analysis. Flower buds with different shapes and trichome classes and patterns (F-J) correspond to *Boechera* genotypes above (A-E). A, F – *B. stricta* (ES 865); B, G – *B. retrofracta* (300.9); C, H – *B. stricta* X *retrofracta* (120.6); D, I – *B. lignifera* (ES 753); E, J – *B. divaricarpa* (ES 524.2). A-E, bar = 25 cm; F-J, bar = 500 µm.

The approximately 110 species comprising the genus *Boechera* are characterized by an enormous genomic plasticity and high ecological diversity which reflects a hybrid background (Schranz et al., 2006; Kantama et al., 2007; Schranz et al., 2007), polyphyletic speciation (Koch et al., 1999; 2003; Kiefer et al., 2009) and quantitative variation for penetrance of apomixis (Aliyu et al., 2010). *Boechera* can reproduce either sexually or apomictically, and allopolyploid (mostly 2n=3x=21), aneuploid and diploid apomixis occurs (Böcher 1951; Windham et al., 2004; Windham and Al-Shehbaz 2007). As triploids represent the majority of polyploids and only a few tetraploids are known, triploidy is probably generated through crosses between diploids (fusion of unreduced and reduced gametes), although triploid origin via crosses between diploids and tetraploids cannot be excluded (Schranz et al., 2005). Apomicts are highly correlated

with interspecific hybridization (Böcher 1951, 1954; Roy 1995; Dobeš et al., 2004a; Schranz et al., 2005). A postglacial (Wisconsin glaciation 18,000 years ago) range extension and recolonization from the center of species diversity in the southwestern United States is hypothesized (Koch et al., 2003; Kiefer et al., 2009). The origin of *Boechera* is placed into the Pleistocene (Koch et al., 2003; Dobeš et al., 2004a), but a diverse "ancient" gene pool in the southern Great Basin is hypothesized, members of which contain older ancestral genetic variation (0.7-2 mya) and which are characterized by unreduced gamete formation (Koch et al., 2003; Dobeš et al., 2004a; Dobeš et al., 2006). The relative contributions of hybridization and polyploidy to apomixis remain a topic of debate (Winge 1917; Buxton and Darlington 1932; Gustafsson 1946, 1947a, 1947b; Harlan and deWet 1975; Grimanelli et al., 2001; Beck et al., 2011), although diploid apomixis in *Boechera* suggests the former as the causal link (Sharbel et al., 2009; 2010; Beck et al., 2011).

Diploid apomixis is a rare condition due to greatly reduced fitness in early generation hybrids (reviewed in Soltis and Soltis (2009)), and is well-described only in a few species *e.g.* in *Boechera*, *Helianthus*, *Iris* (Soltis and Soltis 2009) and *Paspalum rufum* (Siena et al., 2008). Diploid apomicts facilitate comparisons between sex and apomixis without the added complexity of different ploidy levels. In *Boechera* the facts that a majority of apomicts are diploid and that most newly formed triploids are not apomictic imply that apomixis in *Boechera* may not be under simple genetic control by a single factor and/or requires certain gene dosage ratios (Schranz et al., 2005).

Apomictic *Boechera* are characterized by a *Taraxacum*-type of pseudogamous diplospory producing unreduced male and female gametes (Rollins 1941; Böcher 1951; Rollins 1993; Naumova 2001; Dobeš et al., 2006). Unreduced pollen formation is demonstrated by the fact that diploid and triploid apomicts produce seeds almost exclusively with hexaploid ($6C = [4C_{maternal}] + [2C_{paternal}]$) and nonaploid ($9C = [6C_{m}] + [3C_{p}]$) endosperm respectively (Voigt et al., 2007; Aliyu et al., 2010; Voigt-Zielinski et al., 2012). Unreduced pollen of apomictic *Boechera* is characterized by great morphological variablity and higher percentages of infertility, causing higher levels of seed sterility (Böcher 1951, 1954; Koch et al., 2003; Voigt et al., 2007). Observations of male gametogenesis in apomicts are limited (Böcher 1951; Naumova 2001; Sharbel et al., 2005), and suggest either complete asynapsis or irregular pairing of chromosomes during meiosis I, followed by the formation of a restitution nucleus resulting in gametes with unreduced chromosome numbers. Additional observations demonstrated

accompanied reduced and unreduced pollen formation within populations (Böcher 1951; Sharbel et al., 2005). Although tolerance of endosperm imbalance has been described for some *Boechera* (Voigt et al., 2007), castration experiments (Böcher 1951) and extensive flow cytometric analyses of seeds (Aliyu et al., 2010) strongly support selection pressure for the maintenance of unreduced pollen development to fulfill endosperm balance requirements.

Taken together, these data are suggestive of male gametophytes as excellent model system for studies of apomeiosis as initial element of apomixis in *Boechera*.

2.4.2 Genome modulation in the light of interspecific hybridization

Chromosome number reduction from ancestral karyotypes with x=8 chromosomes (*A. lyrata* or *Capsella*), to x=5 (*e.g.* in *A. thaliana*), x=6 (*e.g.* in *Arabis glabra*) or x=7 (*e.g.* in *Boechera*) characterizes genome evolution among the Crucifereae (Koch et al., 1999; Schranz et al., 2007). The transition from the ancestral karyotype (AK) to sexual *Boechera* genomes was accompanied by fragment exchanges *via* reciprocal translocations of their centromeres in four of eight ancestral chromosomes (forming *B. stricta* linkage group 1 (BstLG1) and BstLG2 from AK1 and AK2, and BstLG3 from AK3 and AK8), in addition to chromosome fusion (AK8 and AK5 forming BstLG5; Schranz et al. (2007)).

Besides the long term genome evolution that formed sexual lineages, virtually all *Boechera* apomicts were generated by homoploid (*i.e* diploid in this case) hybridization and allopolyploidization (Roy 1995; Koch et al., 2003; Dobeš et al., 2004a; Schranz et al., 2005). Thereby, allodiploids could have resulted from two 2n gametes rather than *via* the diploid-tetraploid-dihaploid cycle considering the rare occurrence of tetraploids in this genus (Schranz et al., 2005), while allotriploids resulted from one reduced gamete and one unreduced (2n) gamete. Karyotypes of the hybrid progeny (*i.e.* apomictic) exhibit diversity for parental contributions, putatively caused by large-scale homeologous chromosome substitutions (Kantama et al., 2007). These substitutions could be hallmarks of the "genomic shock" that follows hybridization, caused by incompatibilities between the different genomes, by the excess of one parental chromosome set (Kantama et al., 2007) and/or by differences in repetitive elements between the two parental genomes (McClintock 1984). Both transposable element (TE) classes, through RNA intermediates (class I) and through DNA forms (class II), play a major role in the genome dynamics of interspecific hybrids, in which both

activation/mobilization (Lisch 2009) and depression (Kentner et al., 2003; Hazzouri et al., 2008) have been reported.

Enhanced TE activity can have severe effects on the hybrid genome through gene duplication, insertional mutagenesis or unequal homologous recombination (reviewed in Kazazian Jr (2004)). In natural hybrids of sunflower, genome enlargements of greater than 50 % were detected due to massive amplification of a unique retrotransposon class (Baack et al., 2005). Another result of TE activation in hybrids is the production of chimeric gene transcripts. For example the well characterized Wis 2-1A retrotransposon in wheat hybrids produces chimeric transcripts which can be involved in homology dependent gene silencing (HDGS) by producing antisense sequences relative to a coding gene (Kashkush et al., 2003).

On the other hand, elevated TE levels in hybrids may have induced gene silencing mechanisms such as methylation and heterochromatin formation, as adaptive response to the selfish behaviour of TEs (McDonald 1998). Such TE-induced silencing was observed for the Mutator element (*MuDR*) derivative *Mu killer* (*Muk*) gene, whose expression results in a long hairpin RNA molecule that triggers silencing of other *MuDR* elements (Lisch 2009).

In hybrids, especially in allopolyploids, where simultaneous duplication of many genes occurs, gene silencing and DNA sequence elimination are mechanisms to remove redundant DNA sequences or maladaptive loci from their genomes. These mechanisms could have shaped for example the aberrant extra chromosomes in aneuploid apomictic *Boechera* (Kantama et al., 2007). One hypothesis is that incompatibilities between parental genomes could have led to synapsis failure, followed by epigenetic changes leading to stepwise chromosome degeneration, heterochromatinization and ultimately the *Het* chromosome (Kantama et al., 2007). The observed imbalances in parental contributions to different hybrid lineages could also lead to nucleolar dominance, whereby one parental set of ribosomal RNA (rRNA) genes is silenced, and which was observed in crosses of *A. thaliana* with *Cardaminopsis arenosa* (Chen et al., 1998). Furthermore, gene silencing in allopolyploids could also lead to genetic diploidization, in which duplicate genes are either silenced or expressed at reduced levels (Feldman et al., 1997), while the presence of homeologous chromosomes could force exclusive bivalent pairing of homologous chromosomes leading to (partial) chromosome "diploidization". In both cases the consequences are massive gene loss and genome rearrangements (Arabidopsis Genome Initiative 2000). Diploidization mechanisms are

proposed for maize (*e.g.* caused by *Ph1* locus in *Zea maize*; Sears (1977), Gaut and Doebley (1997)), *Sorghum bicolor* (Chen et al., 1997), and also for *Arabidopsis* (Lagercrantz 1998). Thereby, gene inactivation could be enhanced by epigenetic shaping of gene expression through epiallele formation (Jacobsen and Meyerowitz 1997).

Besides replicative translocation, hybridization is a driver of gene duplication (Lynch et al., 2001). If duplicated genes in hybrids are not silenced by degenerative mutations (nonfunctionalization), they could gain new (neofunctionalization) or additional functions (subfunctionalization; Walsh (2003)). Together, recurrent gene duplication and alternative silencing represent mechanisms for generating microchromosomal rearrangements (Lynch et al., 2001; Walsh 2003).

Chromosomal rearrangements in diploid hybrids could occur simply through homeologous recombination, although this mechanism could be genetically repressed in allopolyploids (*e.g.* by *Ph1* in wheat, see above). Homeologous recombination is generally considered as deleterious (Comai 2000) by elevating the rate of chromosomal segment loss and duplication *e.g.* in *Brassica napus* causing reciprocal translocations (Udall et al., 2005; Nicolas et al., 2007). Such rearrangements could also cause chromosome shrinkage, as proposed for *Del* chromosome formation in *Boechera* aneuploids, which is considered to be a translocation or recombinant chromosome between parental chromosomes (Kantama et al., 2007).

Apomixis might be the concomitant consequence and cause for some of the restructurations in hybrid genomes (*sensu* Carman (2001)), whereby the induction of apomixis could enhance some of the mechanisms initially caused by hybridization, for example by hindering homologous chromosome pairing in diploids (Comai 2000). Thus, the absence of recombination in apomicts may enhance again homologous chromosome heteromorphy, whereas in sexual systems the structural divergence between homologues is held low by random drift (Birky, 1996). Additionally, apomicts might not be able to suppress TE activity compared to their sexual counterparts, as has been shown in *Hieracium* apomicts, which typically contained higher frequencies of transposon insertions and gene rearrangements (Bicknell and Koltunow 2004). The lack of TE control together with a disrupted repair mechanism (*i.e.* origin of recombination as a DNA repair process; Bernstein (1977)) could, in addition to hybridization, enhance extensive gene duplication and genome rearrangements in apomictic *Boechera*.

2.5 Conclusion and aims of the dissertation

The maintenance of a balanced endosperm seems to constitute a strong selection pressure for functional unreduced pollen in apomictic *Boechera* (see *2.2.2*). The *de novo* engineering of apomixis in crops requires a fundamental insight of the molecular nature of apomixis and of unreduced pollen formation in natural apomicts. The implementation of functional unreduced pollen formation from natural apomicts in crops would hypothetically solve the "endosperm problem" by balancing maternal and paternal genome ratios *via* central cell fertilization (Birchler 1993).

However, underlying genetic factors controlling unreduced pollen formation in the natural apomictic *Boechera* are unknown. Thus far, the male gametes in *Boechera* were mainly observed for reproductive mode related morphological variation, like pollen grain size (Böcher 1951; Koch et al., 2003; Voigt et al., 2007), pollen fecundity and viability (Voigt et al., 2007), the occurrence of seed sterility (Böcher 1951) and the quantitative variation for apomixis penetrance (Aliyu et al., 2010). The observation of large scale homeologous chromosome substitutions, variable reproduction and the hybrid nature of apomictic *Boechera* (Kantama et al., 2007; Schranz et al., 2007) lead to the hypothesis, whether the potential of expressing apomeiosis was first induced through interspecific hybridization or it could be an older characteristic of the genus (*i.e.* pre-Pleistocene; Sharbel et al. (2009)).

In our study we apply comparative microscopic and molecular methods on a wide range of *Boechera* genotypes to test this hypothesis and to identify both the mechanism of unreduced pollen formation and candidate factors responsible for this trait *via*:

1. Characterization of early male gametophyte development in sexual and apomictic *Boechera* for the identification of male gametophyte stages suitable for comparative gene expression analyses.
2. Microarray-based identification and isolation of candidate genes for the initiation of unreduced pollen formation.
3. Characterization of the molecular structure and function of these candidate genes to reveal apomixis-specific polymorphisms.
4. Elucidation of the phylogeographic distribution of detected apomixis-specific polymorphisms.

3. Material and Methods

3.1 Materials

3.1.1 Plants

In the frame of this study seed material (IPK Gatersleben, AG Apomixis) from 22 genotypes was used (Plant/genotype IDs refer either to Sharbel et al. (2005) or Schranz et al. (2005)). Numbers at the first position indicate original seed sets from the provider, whereas numbers behind the dot indicate seed sets from daughter plants. The geographic information of *Boechera* genotypes used for the identification of candidate genes in the course of the comparative gene expression analysis are summarized in Supplemental Table 2, and illustrated in Figure 4. Details of *Boechera* genotypes used

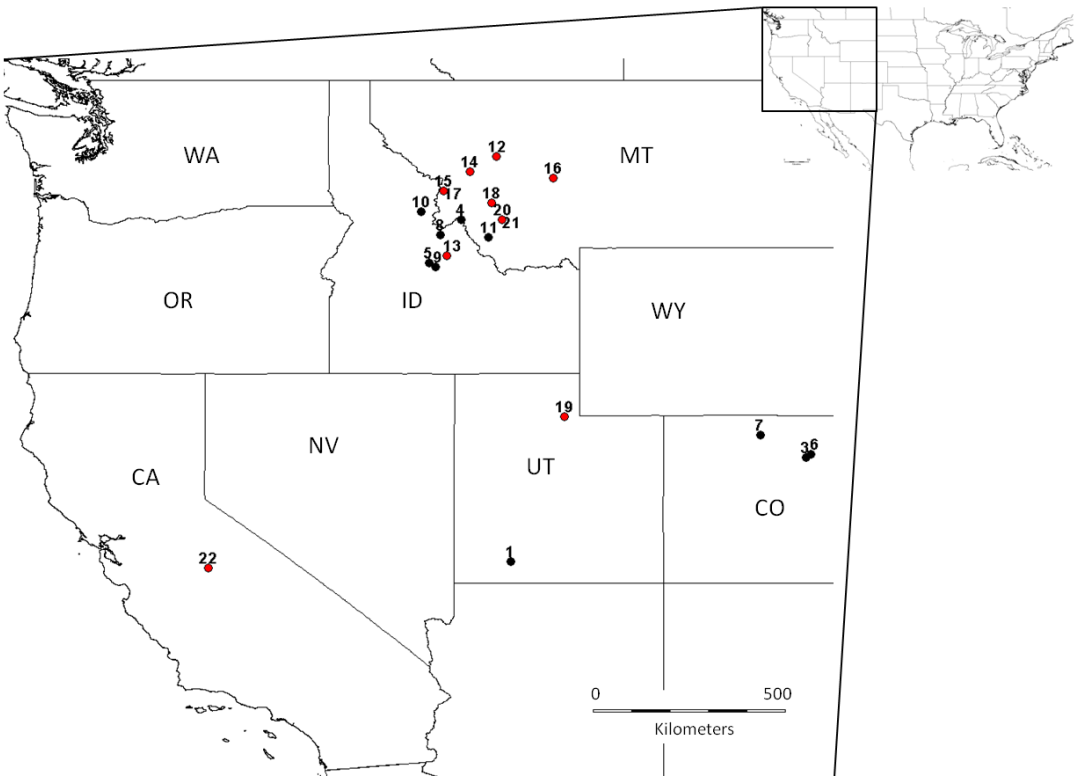

Figure 4. Distribution map of *Boechera* genotypes used for transcriptome analysis.
Diploid sexual (black) and apomictic (red) genotypes from western North America were used in this study for discovery of genes involved in unreduced male gamete formation. The numbers refer to genotype names and localities given in Supplemental Table 2. Map designed with Diva GIS (v7.5, http://www.diva-gis.org/).

for the phylogeographic analyses were published by Christiane Kiefer (PhD thesis, University of Heidelberg, 2008) and can be downloaded from accession database of the The Heidelberg *Boechera* Group (http://ephedra.hip.uni-heidelberg.de/boechera/content/accession_db/). *Boechera* were classified using the

latest nomenclature (*sensu* Al-Shebaz and Windham (1993+)), but as the taxonomic relationships are undergoing revision, here all references were made to genotypes/lines.

3.1.2 Chemicals and Enzymes

Chemicals having generally the "Extra pure", "Molecular biology reagents" or "For Analytical Purpose" grade were ordered from Agilent Technologies (Böblingen, D), Applichem (Darmstadt, D), Fluka (Buchs, CH), Life Technologies (Carlsbad, Ca., USA), Hartmann Analytic (Braunschweig, D), Macherey-Nagel (Düren, D), Merk (Darmstadt, D), Qiagen (Hilden, D), Roche Molecular Biochemicals (Mannheim, D), Roth (Karlsruhe, D), Seegene (Eschborn, D), Serva (Heidelberg, D), Sigma-Aldrich (Taufkirchen, D), Takara Bio Europe (St. Germain-en-Laye, F) and TedPella (Redding, Ca., USA). Enzymes used for cytochemical and histological preparations of flower material and molecular analyses were ordered from Bioline (Luckenwalde, D), Fermentas Life Sciences (St. Leon-Rot, D), Invitrogen (Karlsruhe, D), New England BioLabs (Frankfurt a. M., D), Qiagen (Hilden, D) and Sigma-Aldrich (Taufkirchen, D). Solutions, buffers and media were prepared with distilled water of "aqua tridest."-quality (Milli-Q Water System, Millipore, Bedford, MA, USA) according to Sambrook et al. (1989). If necessary, solutions and media were autoclaved (20 min, 120°C, $2x10^5$ Pa) or sterile filtrated (Rotilabo® Spitzenfilter, pore diameter 0.22 µm, Roth, Karlsruhe, D).

3.1.3 Oligonucleotides

Oligonucleotides were designed with Primer3 (v0.4.0; http://frodo.wi.mit.edu/primer3/) and were synthesized by Eurofins MWG Operon (Ebersberg, Germany).

Chromosome walking

Sharb1199059

TSP1_1L	GCTTACACATTGGGTTGCTT
TSP1_1R	GGGTTAAAGGTACAACTCACCA
TSP1_2L	AAGAAAACGCTTCGGACAGG
TSP1_2R	CAACTCACCAAATCCCTATTGC
TSP1_3L	GCAATAGGGATTTGGTGAGTTG
TSP1_3R	CCTGTCCGAAGCGTTTTCTT

Sharb0931225

TSP2_1L	CGATATGACCTGCACAACC
TSP2_1R	CTAAATTTGCACACCACCTG
TSP2_2L	GGCGAATATTTGCAGGTGGT
TSP2_2R	CCTGCAAATATTCGCCCAGA
TSP2_3L	GCAGGTGGTGTGCAAATTTAGT
TSP2_3R	GGGTTGTGCAGGTCATATCG

Sharb0425060

TSP5_1L	AATTTACAGACCCTGCGATCT
TSP5_1R	CTTTCTCCTCGATTTCTGATTG
TSP5_2L	CCATAAGCACAACCAATCGAAA
TSP5_2R	TTTCGATTGGTTGTGCTTATGG
TSP5_3L	CAGAAATCGAGGAGAAAGAGACA
TSP5_3R	TCGCAGGGTCTGTAAATTGG

Sharb0501554

TSP3_1L	ACATCATCCACAACCCAAAA
TSP3_1R	CGCGTACCTTAGGCTAAATTC
TSP3_2L	TCTTTGGACTTCAGTGGATCG
TSP3_2R	TGTAAGTGCACAGCAAACACAA
TSP3_3L	TGCCGGGGACCAATGTAAT
TSP3_3R	GGTTCGATCCACTGAAGTCCA

Sharb0690829

TSP4_1L	AAGTCGATCGAACACCACAT
TSP4_1R	CGGAAGTAAACATGAACGATG
TSP4_2L	CGATCGAACACCACATGAGAA
TSP4_2R	CATGAACGATGGCGAAGAAG
TSP4_3L	TCATGTCTTCTTCGCCATCG
TSP4_3R	TGTTCGATCGACTTCCTCCTC

Bsp*HRD3*

TSP9_1L	TTCTCAGGCTTGCTTGTTGA
TSP9_1R	CCAAAGCAAGCCAAAACATT
TSP9_2L	GCCTTCAAATGCAGGCAAGAG
TSP9_2R	TCCAAGGTTAAAATGCCCACT
TSP9_3L	GGTGGGCGTGGCAGAGATTA
TSP9_3R	AAATCAACAAGCAAGCCTGAGA

Identification of Bsp*UPG* polymorphisms

Bsp*UPG* in apomictic genomes

CON234X5L	TCCGACCTAAATCCTACCAAACTGA
CON234X5R	TGCTCAATTTTGAACATCTTATTTGC
CON234X2L**	CTGGAATTGGGTACTTGTATGTCAA
TSP3_3R**	GGTTCGATCCACTGAAGTCCA
4RBAC_L**	ATGAACGATGGCGAAGAAGA
4RBAC_R**	TGGATTGCTGTTAAGACCATGT
PC1pol1_L**	CTTTTCCGTTGACTTTCCGACAAAT
PC1pol1_L**	TCGATCAATCTCATTCGGGATCTAT

Bsp*UPG* in sexual *B. stricta* genomes

CON234X11L	CAAAAATAAAAGATTTGATGTAGATTGC
CON234X5R	TGCTCAATTTTGAACATCTTATTTGC
CON234X10_L**	GCTGCCCTGACTCTCTCTCC
CON234B2_R**	TTCACAAATCTAGATGAAGAACCCT
CON234X8_L * *	GAATGTCGCAATCTTCCAAAAC
TSP3_1R**	CGCGTACCTTAGGCTAAATTC

Bsp*UPG* in genomes of other sexual genotypes

FLTsexX2L	GAAGAAAGAGCTACGGCGGTGAT
CON234X5R	TGCTCAATTTTGAACATCTTATTTGC
FLTsexX5L**	CCCTAGACGTTCAAGCCTCGTTA
FLTsexX5R**	AATTGCATTGGTCCTGGCAAC
FLTsexX6L**	TGAGGCTGCTTTTCATTACGTGT
TSP4_1R**	CGGAAGTAAACATGAACGATG

RACE

(Rapid amplification of cDNA ends)

Bsp*UPG-2*

GSP3	TCTTCGCCATCGTTCATGTTTACTTCCG
GSP4	TCATCATGTCTTCTTCGCCATCGTTCA

Bsp*HRD3*

GSP11	TAATGCCCACTGGGGTCGTCATTGT
CON234X14L	ACTGGAATTGGGTACTTGTATGTCA

QRT-PCR analysis

Probe no.	Primer name	Primer sequence 5` → 3`	Efficiency ± SD
Sharb0931225	CON234B4L CON234B4R	TTGCTTTGGTTGAATGCAATAC AATTACTAAATTTGCACACCACCTG	0.91 ± 0.02
Sharb0501554	CON234B3L CON234B3R	TGTGTTTGCTGTGCACTTACAG TCTCAAGAGAACCTGAGACACAAA	0.76 ± 0.01
Sharb0690829	CON234B2L CON234B2R	TCTTCTTCGCCATCGTTCAT TTCACAAATCTAGATGAAGAACCCT	0.90 ± 0.01
Sharb0425060	CON5B9L CON5B9R	TGGATGAGAAATACAAACTTGG AGGAACACGCCCTCAAATTC	0.95 ± 0.01
Sharb0350102	Single0350102L Single0350102R	TGATGCACCAAGGTTGCCATA CAGAAAACCGAGTGCGAATGC	0.96 ± 0.01
Sharb1627083	Single1627083L Single1627083R	ACTCGGCCAACTTGCTCGTC TCATGGGTCGACTCGGTGAG	0.95 ± 0.01
ACTIN2	RT_Act2_T7L RT_Act2_T7R	GTTCCACCACTGAGCACAATGTTACC AGTCTTGTTCCAGCCCTCTTTTGTG	0.91 ± 0.01
EFα1	RT_EFα1_M13L RT_EFα1_M13R	CCAAGGGTGAAAGCAAGGAGAGC CACTGGTGGTTTTGAGGCTGGTATCT	0.95 ± 0.01
GAPDH	GAPDH_For GAPDH_Rev	CAAGGTCATCCATGACAACTTTG GTCCACCACCCTGTTGCTGTAG	-

Sequencing of vector inserts

pCR™4-TOPO® vector

T3	ATTAACCCTCACTAAAGGGA
T7	TAATACGACTCACTATAGGG
M13for	GTAAAACGACGGCCAG
M13rev	CAGGAAACAGCTATGAC

pJET™2.1/blunt cloning vector

pJET™1.2For	CGACTCACTATAGGGAGAGCGGC
pJET™1.2Rev	AAGAACATCGATTTTCCATGGCAG

Boechera BAC library screen

Sharb1199059

TSP1_1L	GCTTACACATTGGGTTGCTT
1LBAC_R	TGCTTCTTCCGTTTCCACTT

Sharb0931225

TSP2_1L	CGATATGACCTGCACAACC
2LBAC_R	TGGGCATGTATTTGTGTGCT

Sharb0501554

TSP3_1L	ACATCATCCACAACCCAAAA
3LBAC_R	CTGGGCATGTATTTGTGTGC

Sharb0690829

4RBAC_L	ATGAACGATGGCGAAGAAGA
4RBAC_R	TGGATTGCTGTTAAGACCATGT

Sharb0425060

TSP5_1R	AATTTACAGACCCTGCGATCT
5RBAC_R	GACACATTAACAAGACAAGGCTCT

Small RNA Northern Blot

Indel9plus	TTTTCACCTTAGCTTAATTTGAATATA
Indel9minus	TATATTCAAATTAAGCTAAGGTGAAAA
Indel9longplus	AACCTTATATTTTCACCTTAGCTTAATT TGAATATAATAAACTCGTT
Indel9longminus	AACGAGTTTATTATATTCAAATTAAGCT AAGGTGAAAATATAAGGTT
ACT2plus	AAGTCTTGTTCCAGCCCTCGTTTGTGG
ACT2minus	CCACAAACGAGGGCTGGAACAAGACTT

Biogeographic distribution of Bsp*UPG-2*

PC1pol1_L	CTTTTCCGTTGACTTTCCGACAAAT
PC1pol1_R	TCGATCAATCTCATTCGGGATCTAT
RT_Act2_T7_L	GTTCCACCACTGAGCACAATGTTACC
RT_Act2_T7_R	AGTCTTGTTCCAGCCCTCTTTTGTG

*Primer sequences listed in 5`- to 3`-orientation

**Bsp*UPG-2*-specific internal sequencing primers

3.1.4 Polymerase chain reaction (PCR), Bacteria and Plasmids

Standard PCR reactions were performed in 0.2 ml eppendorff tubes or 96-well plates with BioTaq™ *Taq* DNA polymerase (Bioline) according to the manufacturer`s protocol. Sanger sequencing of PCR products generated with proofreading polymerase Phusion® High-Fidelity (Thermo Fisher Scientific) was conducted by cloning the PCR products into the vector of the CloneJET™ PCR Cloning Kit (Fermentas). All other products with A-overhangs synthesized by the BioTaq™ *Taq* DNA polymerase were cloned into the pCR®4-TOPO® vector of the TOPO TA Cloning® Kit for sequencing (Invitrogen). Plasmids were transformed into chemical competent cells of the *Escherichia coli* TOP10 strain (Invitrogen) typically grown on X-gal-LB plates containing 100 µg/ml ampicillin with shaking (200 rpm) at 37°C. Positive clones were selected based on a blue/white screen and a colony PCR-based screening according to the manufacturer's protocol of the CloneJET™ PCR Cloning Kit (Fermentas) or TOPO TA Cloning® Kit for sequencing (Invitrogen).

3.1.5 Molecular weight markers

DNA sizing of DNA fragments in Agarose gels was conducted using HyperLadder™ I for products with >1 kb size, HyperLadder™ IV for products <1kb size and HyperLadder™ V for products <500 bp size (Bioline).

3.2 Methods

3.2.1 Plant cultivation conditions

Several sets of the same diploids, ten apomictic and twelve sexual *Boechera* genotypes, were used for the entire experimental pipeline (Supplemental Table 2; note recent taxonomic information, (Koch 2010)). Ten seeds per genotype were cultured on moist filter paper in sealed Petri dishes and vernalized at 4°C in the dark for two weeks until germination. Seedlings were transplanted to plastic pots (11x11x13 cm) containing autoclaved substrate, and transferred to a phytotron and grown without insecticide and herbicide treatment under long-day conditions (16h light and 8h dark, 21°C) and constant relative humidity at 70%.

3.2.2 Ploidy measurements

Relative nuclear DNA content (referred to as ploidy) in leaf, seed (Matzk et al., 2000) and pollen (de Storme and Geelen 2011) was quantified using leaf tissue from a

diploid, sexual *B. stricta* (ES 558.2; Supplemental Table 2) as an external control. Leaf and seed material (20 seeds) were chopped with a razor blade in a drop of Galbraith's buffer containing 4 µg/ml DAPI (4′,6-Diamidin-2-phenylindol; Galbraith et al. (1983). Approximately 20 mature open flower buds at dehiscence stage were harvested for pollen grain extraction by shaking gently for 15 min in 500 µl distilled water, followed by centrifugation at 13200 rpm for 6 min. After decanting the supernatant the pollen pellets were resuspended in 100 µl Galbraith's buffer containing 4 µg/ml DAPI and ground with two 6 mm steel balls in 2 ml eppendorf tubes using a Retsch mixer mill MM 400 for 20 sec at 30 Hz. Extracted nuclei from leaf, seed and pollen tissues were suspended in 1 ml of Galbraith's buffer. Seed and leaf nuclei were filtered through a 30 µm nylon mesh (Partec, Münster, Germany) and pollen nuclei were filtered through a 10 µm nylon net filter (Millipore, Billerica, MA,USA).

Tissue-specific ploidy measurements were performed on a FACSAria II (BD Biosciences, Franklin Lakes, NJ, USA) equipped with a 375 nm near UV laser. Data were measured using the FACSDiva Software (v6.1, BD Biosciences) and analyzed using WinMDI v2.9 software (The Scripps Research Institute, http://facs.scripps.edu/software.html); "C" referes to DNA content of a haploid anaphase cell, and "x" the basic chromosome number.

3.2.3 Isolation and cytochemical preparation of Boechera *flower material*

Pollen developmental stages were defined (Fig. 6, Table 1; Regan and Moffatt (1990)) from each of twelve flower bud size stages (S1-S12) which differed in 100 µm length ranges (*sensu* Smyth et al. (1990) and Sanders (1999)). Whole flowers of stage S3 and antherheads of flower developmental stages S8 – S12 were dissected under a Zeiss Discovery V20 (Carl Zeiss, Jena, Germany) stereomicroscope using sterile glass needles (self-made using a Narishige PC-10 puller, Narishige Group, Kasuya, Japan) bent to an angle of appr. 100-120° and tip size of 50-100 µm. Antherheads and flower buds were collected and pooled per size stage into 2 ml eppendorf tubes containing 2% Formaldehyde (FA) + 2% Glutaraldehyde (GA) in 0,5 M phosphate buffer (pH 7.2). The plant material was fixed overnight at 4°C, then washed twice with 0.5 M phosphate buffer and distilled water, and postfixed in 1% osmium tetroxide. The fixed material was dehydrated in a graded ethanol series (30%, 40%, 50%, 60%, 75%, 90%, 2×100%), transferred to 100% propylenoxide, embedded in Spurr's epoxy resin (Spurr (1969); TedPella, Redding, Ca) and sectioned in 3 µm slices using a Leica RM 2255 microtome

(Leica, Wetzlar, Germany). Transverse antherhead sections were stained with 1:1 methylene blue/azure blue II at 60°C for 80 sec, and photographed on a Zeiss Axioplan 2 Imaging (Carl Zeiss) microscope under differential interference contrast optics and 20-fold magnification.

Transverse sections of antherheads in LR White resin, appr. 3 µm thick, were cut and dehydrated for immunofluorescence staining of callose as described above. Slides with LR White resin embedded antherhead sections were pre-incubated (14 µl drop/well) in 3% (w/v) bovine serum albumin (BSA) (Fraction V, A 2153, Sigma) in PBS at pH 7.4 for 20 min, then incubated for 1h in primary antibody. Anti-callose (1,3)-*ß*-glycan primary antibody (Biosupplies) was used at a 1:100 dilution and the secondary antibody (Alexa Fluor 488 anti-mouse, Invitrogen) was diluted 1:200 in PBS, 1% BSA, 0.5% Tween. Slides were rinsed five times for 5 min with PBS Tween, and then incubated for 1h in the dark with secondary antibody. Slides were then rinsed twice with PBS Tween, two times with PBS and three times with water. Samples were analysed on a confocal microscope Zeiss LSM 510 META equipped with a Zeiss AxioCam HRc camera.

3.2.4 Cytological observation of meiocytes in fixated Boechera *anthers*

Cytological observations of meiosis and PMCs were made from sexual and apomictic *Boechera* flower bud stages S9 to S10. Meiotic chromosome spreads of *Boechera* anthers were prepared according to Ross et al. (1996) with minor changes. Prepared spreads were stained with 40 µl of DAPI (4 µg/ml), mounted with a 24 X 50 mm cover slip and sealed with Fixogum (Marabu, Tamm, Germany). The slides were incubated for 30 min in the dark at 4°C, and cytological analysis was performed under a Zeiss Axioplan 2 imaging microscope (Carl Zeiss). Photographs were taken with a high resolution camera (AxioCam HRc Rev. 2) under 100-fold magnification using a 49 DAPI BP reflector block. Meiocytes at the tetrad stage were examined by separate squashes of all six antherheads per flower bud, from one sexual and six apomictic genotypes, to determine the number of monads, dyads, triads and tetrads in each antherhead. The anthers were squashed according to Peterson et al. (2010), stained (Alexander's stain, Alexander (1969)) and examined under a Zeiss Axioplan 2 imaging microscope (Carl Zeiss). All available meiocytes per anther were counted, and statistical analyses of meiocyte behaviour and anther size correlations of gametophyte stages were evaluated with SPSS v11.5 (LEAD Technologies, Charlotte, NC, USA).

3.2.5 Live-microdissection of antherheads and isolation of total RNA

Microdissections and total RNA extractions were prepared using tools and dissection area which were cleaned with ethanol, treated with RNAseZap® (Ambion, Carlsbad, CA, USA) and washed with DEPC treated distilled water (DEPC, Diethyl phosphorocyanidate). Approximately 30 antherheads of each genotype, corresponding to the PMC stage, were live microdissected from fresh whole flower buds under a Zeiss Discovery V20 stereomicroscope (Carl Zeiss) using sterile glass needles (see above), and directly collected in 500 µl RNAlater (Qiagen, Hilden, Germany), which was substituted by the lysis buffer of the extraction kit (RNeasy® Micro kit, Qiagen) prior to RNA extraction. 18 µl of RNAse-free DNAse I (Qiagen) digested total RNA extracts were eluted through RNeasy MinElute Spin® columns (Qiagen) and quantified on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). RNA quality of all samples was assessed with the RNA 6000 Nano LabChip Kit II (Agilent Technologies, Santa Clara, CA, USA) on the Agilent 2100 Bioanalyzer. Purified total RNA was stored at -80°C.

3.2.6 Flower-specific Boechera *454 cDNA libraries*

The *Boechera* 454 FLX cDNA libraries (Sharbel et al., 2009; 2010) used in this study were sequenced from pooled flower stages 1–12 (Smyth et al., 1990) of three diploid sexual plants (genotypes ES 910.2.2, 105.6.1 and ES 616.2) and three apomictic plants (genotypes 67.5, 300.6.1 and 218.2.2).

3.2.7 Microarray design

The 454 cDNA sequences were assembled using the CLC Genomics workbench using standard assembly parameters for long-read high-throughput sequences, after trimming of all reads using internal sequence quality scores. In total, 36 289 contig sequences and 154 468 non-assembled singleton sequences were obtained. This data was then sent to ImaGenes GmbH (Berlin, Germany) for microarray development using their Pre-selection strategy (PSS) service. Thereby, 14 different oligonucleotides (each 60 bp in length) per contig and 8 oligonucleotides per singleton, including the "anti-sense" sequence of each oligo, were bioinformatically designed and spotted onto two 1 million-spot test arrays. These test-arrays were probed using (1) a "complex cRNA mixture" (obtained by pooling tissues and harvesting all RNA from them), and (2) genomic DNA extracted from leaf tissue pooled from a sexual and an apomictic

individual. Based upon the separate hybridization results from the cRNA and genomic DNA samples, and after all quality tests, a final 2x105k spot array was designed containing in total 103 747 oligos including 37 974 non-assembled sequences (singletons) and 65 773 contig sequences (clusters). According to a BLASTX search against the TAIR9 cDNA database, the 103 747 oligonucleotide targets represent 31051 annotated genes (6.23% singletons, 93.77% clusters), and approx. 72 696 (49.57% singletons, 50.43% clusters) non annotated *Boechera* genomic regions which are known to have transcriptional activity (Sharbel et al., 2009). This array hypothetically contains multiple oligonucleotides (*i.e.* technical replicates) of every gene expressed during *Boechera* flower development and was uploaded on the eArray platform (Agilent Technologies, https://earray.chem.agilent.com/earray/) for *in situ* synthesis of the microarray.

3.2.8 Microarray probe preparation, hybridization, data analysis and validation

To ensure optimal cRNA yield after labeling, 200ng total RNA with absorbance readings of A260/A280 >1.8 and A260/A230 >1.8 (recommended by Agilent) were used for the labeling procedure. Approximately 1.5 µg of the generated Cy3-labeled cRNA per sample (One-color Quick Amp Labeling kit, Agilent Technologies) was hybridized for 18hrs at 65°C to the custom *Boechera* whole flower-105k-Agilent microarrays, which were scanned at 5µm double pass resolution with an Agilent G2565BA Microarray Scanner. The One-color RNA Spike-in kit (Agilent Technologies) was used to assure optimal microarray processing. Microarray hybridization quality was assessed with Feature Extraction 10.1 software (Agilent Technologies), whereas quantile normalization with baseline to median transformation and gene expression analysis was performed with GeneSpring GX 10 software (Agilent Technologies). Differentially expressed microarray probes were considered validated with $p \leq 0.05$ as assessed by an unpaired *t*-test with a mean difference ≥2-fold. *P*-values were corrected for the family-wise error rate (FWER) as control for false positives using the Bonferroni method. Principal components analysis (PCA) is a method of data reduction (Manly 1994). If the data are highly correlated, a plot of the microarray probe signal intensity values against the first few principal components will account for a large portion of their total variance. Such a plot would effectively summarize the structure contained in the full data set. We applied PCA to our microarray datasets

formed by different fold change levels using GeneSpring GX 10 software (Agilent Technologies).

Array probes which, (1) were highly expressed (FC≥10) or (2) demonstrated significant differential genexpression (FC≥2, $p \leq 0.05$) between sexual and apomictic genotypes were traced back to their original *Boechera* cDNA sequences (based upon assemblies of 454 reads; Sharbel et al. (2009)) and these full-length sequences were used in a blast search to find *Arabidopsis* cDNA homologues (TAIR9) using the following parameters: blastall –p blastn -m8 -e1e-3 -W7 -r1 –q -1 -i (Altschul et al., 1997).

Corresponding TAIR9 hits of dataset (1) with highest *E*-value and Bit score were then used for a Gene Ontology (GO) analysis using AgriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) for GO term enrichment, using the annotated genes (N=31051) present on the 2x105k *Boechera* flower-specific array as background comparison set, with the following analysis parameters: Hypergeometric test, Yekutieli (FDR under dependency) adjustment for multiple tests, significance level 0.05, minimum number of mapping entries = 5, and complete plant GO gene ontology.

3.2.9 qRT-PCR validation of genes demonstrating differential gene expression

SYBR® Green quantitative PCR (qRT-PCR) (SYBR® Green PCR Master Mix, Applied Biosystems, Carlsbad, CA, USA) was used as an independent validation of the microarray data. The RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Waltham, MA, USA) and oligo$(dT)_{18}$ primer were used to conduct a cDNA synthesis from 300 ng total RNA from microarray samples. CDNA quantification was performed with Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) on a ND-3300 fluorospectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Primers were designed on the candidate microarray probe homologous cDNA reads ET5PU7E01BTR6J, ET5PU7E01DDH28 and ET5PU7E01DERC2 (Supplemental Table 3) using Primer3 v0.4.0 (T_m ~ 60°C, 40% < CG content < 80% and PCR product size < 200 bp). 1 ng of cDNA in a 5 µl SYBR® Green PCR Master Mix containing microarray probe specific primer combinations was initially denatured for 2 min at 50 °C and 10 min at 95°C and amplified by 40 cycles of PCR, with each cycle consisting of 15 sec at 95 °C, 1 min at 60 °C. The melting curve for each PCR product was achieved by a temperature ramp (95 °C, 60°C, 95 °C each 15 sec). QRT-PCR was performed on an ABI-PRISM 7900HT FAST Real-Time PCR System (Applied

Biosystems) using the SDS software v2.4 (Applied Biosystems). Seven biological and four technical replicates were run for each probe and tissue in a 384-well plate together with two endogenous control genes tested on *Boechera* anther material (*ACTIN2 (ACT2)* and *EFα1*, (Pellino et al., 2011)), negative template and reverse transcriptase controls. PCR efficiencies and normalized C_t values of each set of four technical replicates were processed with the Real-time PCR Miner v2.0 software (Zhao and Fernald 2005). Relative quantification and normalized cycle threshold (C_t) values of the amplified targets were calculated separately with reference to the expression levels of each of the two housekeeping genes employing the ΔΔCt method (Pfaffl 2001) using a calibrator sample (ES 910.2 and ES 913.3 (Sharb0350102); Supplemental Table 4). The corresponding mean relative expression ratio for each genotype was calculated with SPSS (v11.5; LEAD Technologies) and significant differences between samples were evaluated using a one-way ANOVA ($\alpha = 0.05$) with a Tuckey-HSD *post hoc* test for differences between multiple pairs of means.

3.2.10 DNA preparation, chromosome walking, and sequence analysis in Boechera

DNA was extracted from young leaves using the Cleanplant DNA kit with minor protocol modifications (CleanNA, www.cleanna.com). Chromosome walking on microarray candidate probes and Bsp*HRD3* (homolog to At1g18260) in sexual and apomictic *Boechera* genomes was conducted using the DNA Walking SpeedUp Premix Kit I (Seegene Inc.) according to the manufacturer's protocol. Chromosome walking products and PCR products were cloned (*see 3.1.4*), subsequently amplified using the TempliPhi™ DNA Sequencing Template Amplification Kit (Reagin 2003) and Sanger sequenced on the ABI 3730 XL sequencing system at the GenomeCentre, IPK Gatersleben in forward and reverse reaction. Sequence analysis and assembly was carried out with Lasergene (v8.0.2; DNAStar).

3.2.11 BAC probe preparation and screening of Boechera *BAC library*

Candidate 60-mer microarray probes were mapped against *Boechera* 454 FLX cDNA libraries (Sharbel et al., 2009) using CLC Genomics Workbench v4.5.1 (CLC Bio, Aarhus, Denmark, standard parameters). Specific primers were designed for chromosome walking on flanking regions to the mapped microarray probes (Primer3 v0.4.0; see *3.1.3*). The sub-cloned (see *3.1.4*), PCR amplified and finally gel purified DNA walking products (NucleoSpin® Extract II kit, Macherey-Nagel, Düren,

Germany) hybridized against a gridded macro array of a bacterial artificial chromosome (BAC) library (48 x 384 spotted wells onto a 22 x 22 cm filter membrane, binary vector pCLD04541, Bancroft (1997)). Individually radio-labelled and pooled probes were hybridized as a group using the overgo hybridization method (Ross et al., 1999). Vector inserts of 500ng pure BAC clone DNA extracts were restriction digested to completion with *Hin*dIII, *Bgl*II and *Bam*HI (Amersham Pharmacia Biotech) at 37°C (*Hin*dIII, *Bgl*II) and 30°C (*Bam*HI) for 8h and examined on a 1% agarose gel. Based on their partial overlapping restriction patterns DNA was isolated from four chosen BAC clones (A4O22, E7K5, C8B11 and F8G11) using Nucleobond Xtra Midi Kits (Macherey-Nagel), BAC DNA was randomly sheared (Hydroshear, Digilab, Marlborough, MA, USA) and size-fractionated by agarose gel electrophoresis in ~1kb and ~4-5kb size classes. These fragments were end-repaired, blunt-end ligated into pUC19 (Life Technologies), transformed into *E. coli* ELECTROMAX DH5α-E electro-competent cells (Invitrogen) and sequenced on an ABI 3730 XL automatic DNA sequencer (PE Applied Biosystems). Vector clipping, quality trimming and sequence assembly using stringent conditions (*e.g.* 95% sequence identity cutoff, 25 bp overlap) was done with the Lasergene 8 (DNAStar) and Staden (http://staden.sourceforge.net/). Assembly of the complete BAC clone sequences was performed using Seqman and the Gap4 algorithm implemented in Staden. Remaining gaps in the contiguous BAC sequences were manually inspected and closed with primer walking or PCR products crossing the gaps from adjacent contigs. The resulting sequences were assembled using Lasergene 8 (DNAStar) set to an overlap minimum of 20 bp with 95% identity. All BAC assemblies were annotated, based on a combined BLASTN and BLASTX search against the non-redundant GenBank nucleotide and protein databases, respectively (http://www.ncbi.nlm.nih.gov/genbank/).

3.2.12 Rapid Amplification of cDNA Ends (RACE)

The SMARTer RACE method (Clontech, Palo Alto, CA, USA) was employed to obtain 5`-end and poly(A)-site information from Bsp*UPG-2* and Bsp*HRD3*, an *Arabidopsis* homologous of *HRD3* in *Boechera*. Primers (Primer3 v0.4.0) for *BspUPG-2* RACE were derived from cDNAs homologous to the microarray probe read ET5PU7E01BTR6J (5`-end primer GSP3: 5`-TCTTCGCCATCGTTCATGTTTACTTCCG-3`; 3`-end primer GSP4: 5`-TCATCATGTCTTCTTCGCCATCGTTCA-3`), and those for Bsp*HRD3* RACE were

derived from the LCB2 of BspUPG-2 (5`-end primer GSP11: 5`-TAATGCCCACTGGGGTCGTCATTGT-3`; 3`-end primer CON234X14_L: 5`-ACTGGAATTGGGTACTTGTATGTCA-3`). PCR reactions were performed with the Advantage 2 PCR Kit (Clontech) and PCR fragments were cloned (pCR4-TOPO TA; Life Technologies) and Sanger sequenced (see *3.1.4* and *3.2.10*).

3.2.13 BAC annotation and analysis of sequence polymorphisms

The allelic constitution of Bsp*UPG-2* in nine each sexual and apomictic genotypes was determined sequencing proof-reading polymerase amplified (Phusion high-fidelity polymerase, Thermo Scientific), gel purified (NucleoSpin® Extract II kit, Macherey-Nagel) and multiply-cloned PCR fragments (CloneJet™ PCR cloning kit, Fermentas) using specific primers (5`-end primer in apomictic genotypes CON234X5L: 5`-TCCGACCTAAATCCTACCAAACTGA-3`; in sexual *B. stricta* CON234X11L 5`-CAAAAATAAAAGATTTGATGTAGATTGC-3` and in other sexual genotypes FLTsexX2L: 5`-GAAGAAAGAGCTACGGCGTGAT-3`; 3`-end primer CON234X5R: 5`-TGCTCAATTTTGAACATCTTATTTGC-3`) according to the manufacturer's protocol (Phusion high-fidelity polymerase protocol, Thermo Scientific). The amplifications were run under the following conditions: 0.5 min initial denaturation at 98°C; 32 cycles of amplification with 10 s at 98°C, 30 s at 58°C, and 1.5 min at 72°C; and 10 min of final elongation at 72°C. PCR success was checked with electrophoresis in a 1% agarose gel in TBE-buffer and staining with ethidium bromide. Lasergene 8 software (DNAStar) was used for assembly and similarity analysis of Phred 20 quality-trimmed sequences. Open reading frames (ORF) in all six frames of Bsp*UPG-2* with minimum 100 nt length were identified with Geneious (v5.3.6; Biomatters, Auckland, NZ). Geneious was used for pairwise sequence alignment using CLUSTALW (IUB cost matrix, gap open cost = 15, gap extend cost = 6,66, free end gaps) and maximum-likelihood method comparison (Tamura and Nei 1993) were conducted in MEGA5 (Tamura et al., 2011) using standard parameters.

Pairwise CLUSTALW comparisons in the presence of rearrangements were performed with Mauve (v2.3.1; progressiveMauve, default parameters, Darling et al. (2004)) to detect collinear sequence blocks (LCBs; conserved sequence segments, which are internally free from genome rearrangements) in Bsp*UPG-1* and Bsp*UPG-2* between different genotypes.

BAC sequence assemblies were annotated for transposable elements by screening the green plant section (Viridiplantae) of the Repbase repetitive element database (Jurka (1998), http://www.girinst.org/server/Maps/AT/index.html) using CENSOR (Kohany 2006) and Repeatmasker (Smit et al., 1996-2004). Programs, einverted and EMBOSS (Rice et al., 2000) were employed to identify inverted repeats with ≥80% matches and the Pipmaker software for simple repeats and CpG islands (Schwartz et al., 2003). LTR analyses of Assembly 1 and Assembly 2 were performed with LTR FINDER software (Xu and Wang 2007).

*3.2.14 Computational analysis of RNA folding probabilities of Bsp*UPG-2

Structural RNAs are usally characterized by an unusual thermodynamic stability and a conserved secondary structure. The minimum folding energy (MFE) as a measure of thermodynamic stability for a sequence (*i.e.* negative values indicate that a sequence is more stable) was calculated using the RNAfold and RNAz version 1.0 software for Windows of the Vienna RNA package (Hofacker et al. (1994), http://www.tbi.univie.ac.at/ivo/RNA). All MFEs were expressed as negative kcal/mol. To better classifiy npcRNAs from candidate genes (*e.g.* into miRNA, tRNA, random mRNA or rRNA), the adjusted MFE (AMFE), the minimal folding free energy index (MFEI) and the A+U content were calculated and only the optimal folding structure was used, which must not be necessarily the biological correct structure for RNAs (Zhang et al., 2006). The optimal structure was conducted by the lowest ensemble diversity which is a measure to indicate how much time the secondary strucure stays in the actual "target" shape (Hofacker et al., 1994).

The presence of statistically significant secondary structures of Bsp*UPG-2* was monitored using Z-score values as described by Crespi et al. (1994) and Bonnet et al. (2004). Bsp*UPG-2* and its complementary strand were scanned every 10 nt using sliding windows of sizes (= step size, impact of different step sizes was examined in Kavanaugh and Dietrich (2009) showing strong decrease in detection sensitivity with step sizes ≥25) ranging between 50 and 300 nt (by increments of 10 nt = window delta). For each window, usually shuffled sequences (in mono- or dinucleotides) are generated to estimate the mean and standard deviation (SD) of the MFE for all possible sequences (Workman and Krogh 1999; Washietl et al., 2005). The Z-score is the number of SDs between the actual MFE of the sequence and the mean value of the energies of folding of the shuffled sequences, whereby negative values indicate that a

sequence is more stable than expected by chance (Crespi et al., 1994). RNAz does not actually sample random sequences but approximates Z-scores using Support Vector Machine (SVM) regression (Washietl et al., 2005). In order to augment the energy model for Bsp*UPG-2* secondary structures for covariance information a consensus MFE (E_A) for Bsp*UPG-2* secondary structures was conducted from a CLUSTALW alignment of Bsp*UPG-2* homologs from six different *Boechera* genotypes (ES 776.1, 300.9, 28.6, ES 753, ES 514 and ES 524.2). A comparison of E_A with the individual MFEs (E) results in the structural conservation index (SCI=E_A/E). Based on Z-score and SCI, RNAz calculates a combined score, the so-called "RNA class probability" which also is referred to as "*p*-value". If $p>0.5$, RNAz classifies an alignment as "RNA", meaning that RNAz has detected an unusually stable and/or unusually conserved RNA structure. A Z-score threshold of ≤-3.5 was used according to the RNAz software manual (http://www.tbi.univie.ac.at/~wash/RNAz/manual.pdf). The "npcRNA [number]"-names were chosen to follow the naming convention established by previous investigators (Hirsch et al., 2006).

3.2.15 Copy Number Variation (CNV) and whole genome sequencing (WGS) read analysis

Duplicated sites on Bsp*UPG-2* were tested by mapping a set of sequence tags from an array-based comparative genome hybridization (aCGH) experiment in sexual and apomictic *Boechera* (Aliyu et al., unpublished results) using a BLASTN search on sliding windows of Bsp*UPG-2* (blastall -p blastn -m -e-10 -W7 -r1 -q-1; 200 bp, window length, 20 bp, step size). Local BLASTN search (Genomics Workbench v4.5.1, CLC Bio; parameters: costs = match 1, mismatch 3, existence 5, extension 2; *E*-value = 10; word size 11, filter complexity = yes) between Bsp*UPG-2* and the complete genomic Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/, 454 GS FLX) of a sexual *B. stricta* (SRP007750) was conducted to test the presence of Bsp*UPG-2* fragments in sexual *Boechera* genomes.

3.2.16 Small RNA Northern blot

Total RNA and small RNA were extracted from 0.150 – 0.250g flowers at different stages (pooled together from each four sexual and apomictic genotypes separately) using a mirVana™ miRNA Isolation kit (Ambion) according to the manufacturer's protocol. Small RNA preparations were separated on a 15% polyacrylamide-gel containing 7 M urea and transferred by a semi-dry electroblotting system onto Zeta-

probe® GT genomic tested blotting membranes (Bio-Rad, Hercules, CA) using a miRNA size marker (NewEngland Biolabs, Ipswich, MA, USA). RNA was cross-linked for 2 h at 80°C under vacuum. DNA-oligonucleotide probes were end-labeled with [γ-32P] ATP by T4 polynucleotide kinase (Fermentas). The hybridization was performed in Church buffer at 42°C (1% BSA fraction V, 1 mM EDTA, 0.5 M $NaPO_4$ pH 7.2; 7% SDS), and membrane washing in 2x SSC / 2% SDS at room temperature.

*3.2.17 Phylogeographic distribution analysis of Bsp*UPG-2

Total DNA of 1685 accessions was extracted according to protocol of the Cleanplant DNA kit (CleanNA, www.cleanna.com). PCR amplification of a 645bp spanning Bsp*UPG-2* gene fragment using primers mapping each on one of the LCBs (*see 3.2.13*) was performed in a volume of 10 µL, using 10 µM of each primer, a total of 2.0 mM MgCl 2 and 0.5 U of BioTaq polymerase (Bioline, Luckenwalde, Germany). The housekeeping gene *ACTIN2* was used as external template control. The amplifications were run under the following conditions: 5 min initial denaturation at 95°C; 32 cycles of amplification with 30 s at 95°C, 30 s at 60°C, and 1 min at 72°C; and 10 min of final elongation at 72°C. PCR success was checked with electrophoresis in a 2% agarose gel in TBE-buffer and staining with ethidium bromide. Sample coordinates of most of the 1576 successfully screened *Boechera* accessions for map reconstructions (DIVA GIS v7.5, http://www.diva-gis.org/) were taken from Kiefer et al. (2009) and Schranz et al. (2005). Latitude and longitude data were unavailable for some accessions, and thus only a subset (N=1502) of the complete data could be analysed. Pairwise distances from each pair of accessions were calculated with GENALEX (v6.5; Peakall and Smouse (2012)) and used for distribution analysis with SPSS v11.5 (LEAD Technologies, Charlotte, NC, USA). (Supra)haplotype designations based on *trnL-F* sequence data of all accessions were taken from Kiefer et al. (2009). Network reconstruction was conducted on the TCS 1.21 software (Clement et al., 2000). The *trnL-F* dataset was split in three subalignments according to parsimony analysis in Kiefer et al. (2009). Then the pseudogene region was excluded for the analysis and the connection limit was set to 95%. After network reconstruction the subnetworks were rejoined according to the parsimony analysis. Haplotype node sectors indicate the partition of this haplotype between accessions carrying Bsp*UPG-2* and accessions lacking the candidate gene. Ecological habitat modelling and jackknife analysis for variable contribution for apomictic and sexual *Boechera* genotypes was performed

using the 2.5 arc-minute (ca. 5 km^2) climate and elevation grids (WorldClim version 1.4 provides monthly temperature and precipitation data which were are compiled and interpolated for all land regions across the world during a period of 1950–2000, http://www.worldclim.org/bioclim, Hijmans et al. (2005)) for the maximum entropy calculation in Maxent (standard parameters with 15 replicates, random seed, 5000 iterations; Phillips et al. (2006)). The Maxent model generates a threshold-independent, continuous output for climatic suitability range (0–1). The model performance was then evaluated using the receiver operating characteristic (ROC) analysis (Zweig and Campbell 1993) with the area under ROC curve (AUC) index (Fielding and Bell 1997). Discrepancies in *Boechera* classification (Kiefer et al., 2009) were inferred in excluding all *Boechera* accessions from lineages IV and V. All statistical analyses were carried out with SPSS v11.5 (LEAD Technologies, Charlotte, NC, USA).

4. Results

4.1 Staging of early floral development in *Boechera* genotypes

4.1.1 Morphometric analysis of flower organs

Gene expression profiles in reproductive tissues showed spatial and temporal variability (Mascarenhas 1989; Honys and Twell 2004). To optimize for comparative gene expression analyses between reduced (sex) and unreduced (apomeiosis) pollen at a single developmental stage, the morphological development of antherheads was staged depending on the flower bud length. The identification of an allometric relationship between the flower bud size and the gametophyte stage of anthers would set up a fast and non-destructive method to screen for appropiate anther developmental stages for RNA isolation and subsequent transcriptome comparisons.

Analyses of early flower development from about 780 flower buds in six diploid *Boechera* genotypes showed a linear relationship between bud length (*i.e.* from the receptaculum to the tip of the outermost sepal) and flower organ size (*e.g.* stage S4 to S12 antherheads: R^2_{apo}=0.87, F=1273.50, p<0.001; R^2_{sex}=0.95, F=4237.23, p<0.001; Fig. 5). Only minor variations among genotypes, but no differences between sex and apomixis were detected (general linear model for regression line slopes (b) of stage S4 to S12 antherheads: H_0: $b_{apo}=b_{sex}$; R^2=0.95, $F(44, 8) = 0.68$, p=0.804; Fig. 5). Bud stages S1 to S3, which most likely can be assigned to *Arabidopsis* flower developmental stages 1 to 5 (*sensu* Smyth et al. (1990)) were excluded from flower organ analysis because the majority of flower organs are at primordial stages or are prior to emergence.

In a second step we examined the correlation of antherhead length with male gametogenesis stages according to their flower bud correlates. Different from other tissues, such as the tapetum, microsporogenous tissues undergo several systematic histodifferentiation steps, and was thus reported as appropiate “marker” for anther development (Scott et al., 1991). We decided to focus on four major histodifferentiation steps of male gametophyte development which are characterized by their pre-meiotic, meiotic and post-meiotic appearance (Fig. 6, Table 1). Despite some marker-specific variations in the correlation quality, flower bud length predicts the gametophytic stage of anthers with relative high accuracy (N=455; Sp: 100%, PMC: 100%, Me: 52.27% and Msp: 58.24%). Thereby, the sporocyte (Sp) stage is prevalently correlated with flower buds of 0.2 to 0.3 mm length (antherhead length: <100µm), flower buds between0.3 and 0.9 mm length (~100 to ~430µm) exhibit a high correlation with pollen mother cells (PMCs), and microspores (Msp) are highly enriched in flower buds greater

than 1.1 mm (>610µm). Meiotically-active gametophytes (Me) are predominant in flower buds with 0.9 to 1.1 mm length (Sex: ~550 to 610 and 760µm; Apo: ~430 to 610µm).

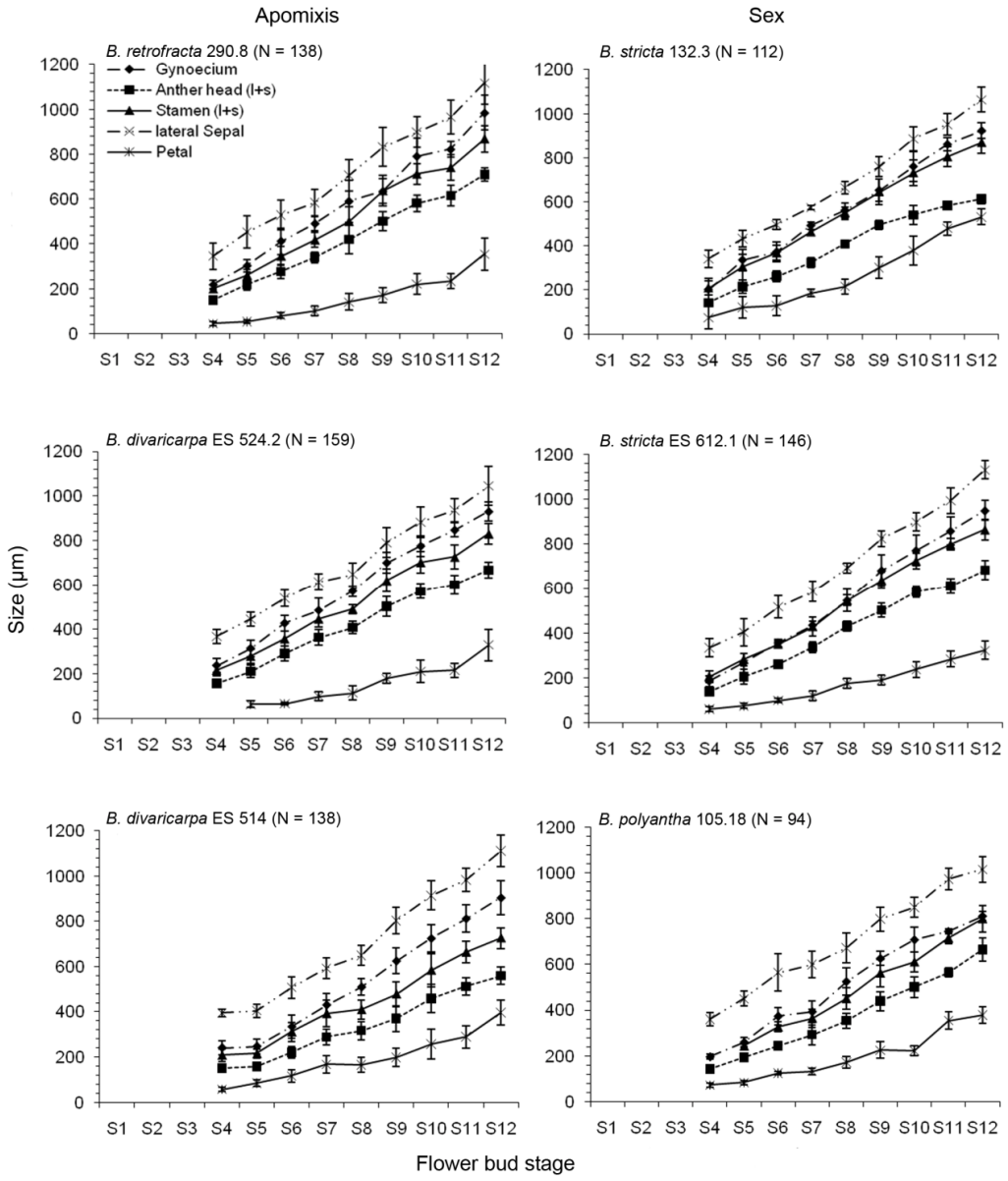

Figure 5. Correlation of flower organs and flower bud length for sexual and apomictic *Boechera* genotypes.

Error bars show the standard deviation (SD) of single flower organ sizes for each flower bud stage (see also Fig. 6 and Table 1). Total number of examined flower buds per genotypes given in parentheses.

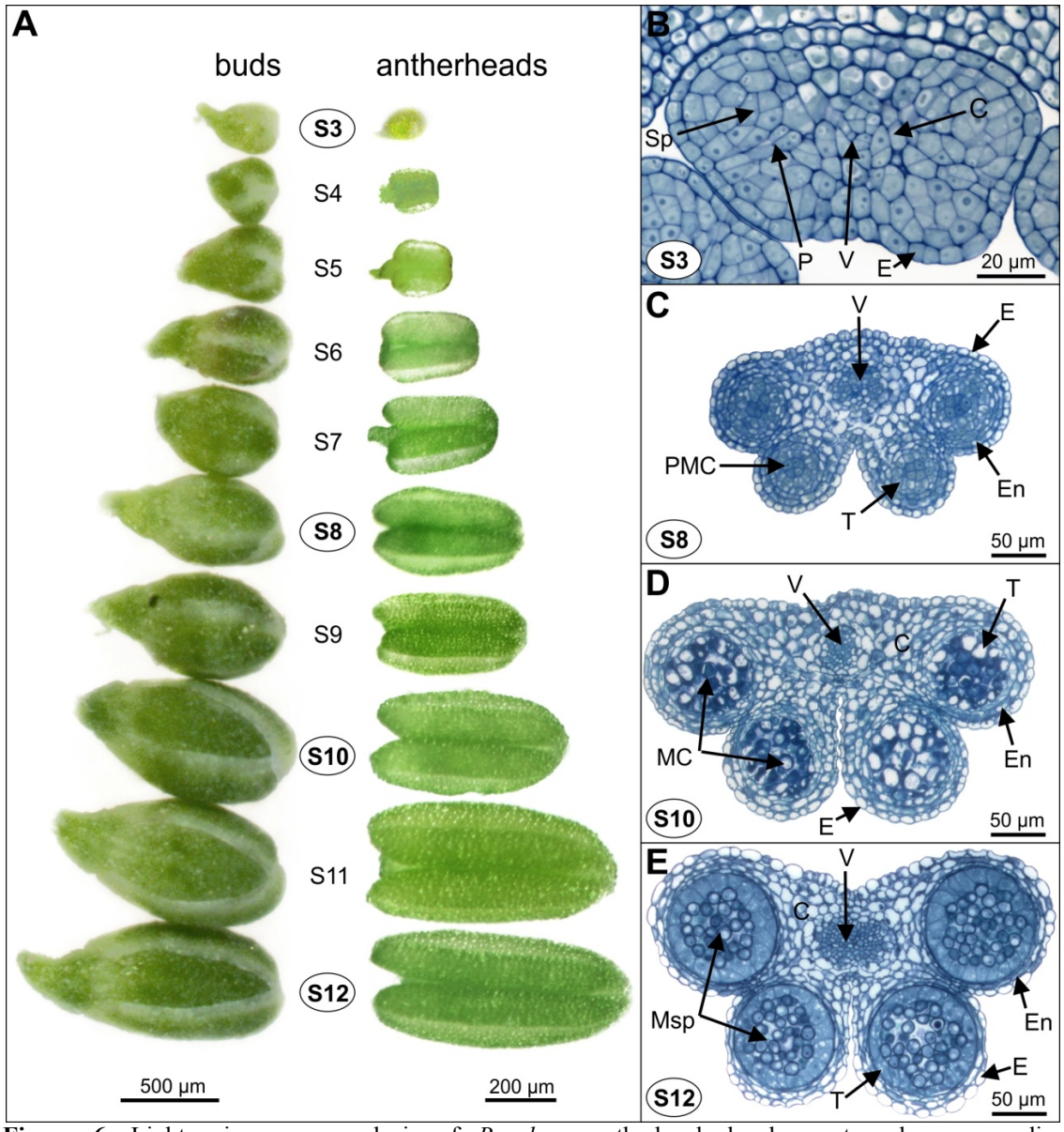

Figure 6. Light microscope analysis of *Boechera* antherhead development and corresponding gametophyte stages during pollen formation.

(A) Flower buds and appropriate antherheads of different developmental stages S3 to S12 in a sexual genotype. (B-E) Light microscopy images of semi-thin sections of resin-embedded antherheads after histological staining displaying the development of the sporocyte into a mature microspore, used for cytological validation of major gametophyte stages (B – Sp; C - PMC; D - Me; E – Msp, Table 1). E, endodermis; P, parietal cells; Sp, sporocyte; V, vascular region; C, connective; T, tapetum; PMC, pollen mother cell; En, endothecium; MC, meiotic cell; Me, meiosis; Msp, microspores.

Table 1. Developmental markers of *Boechera* pollen formation corresponding to gametophyte and flower stages.

Flower stage	Size range (mm)	*Arabidopsis* flower stage[a]	Pollen stage[b]	Developmental marker ID	Developmental marker and histological events[c]
S1	0.0 – 0.1	n/a	n/a	n/a	n/a
S2	0.1 – 0.2	n/a	n/a	n/a	n/a
S3	0.2 – 0.3	7-8	n/a	Sp	Sporocytes, premeiotic, anther in differentiation stage, occurrence of 1° and 2° parietal cells, vascular region initiated.
S4	0.3 – 0.4	9	3		Pattern of anther defined, all four locules present.
S5	0.4 – 0.5				Rapid lengthening of all flower organs, especially antherheads and filament.
S6	0.5 – 0.6				
S7	0.6 – 0.7				
S8	0.7 – 0.8			PMC	Pollen mother cell (PMC) enlarged and clear separated from tapetum, prior to meiosis, no callose deposition.
S9	0.8 – 0.9		4		PMC enters pollen meiosis, crushed middle layer, vacuolated tapetum, initial callose deposition, binucleate tapetum.
S10	0.9 – 1.0		4 – 6	Me	Meiosis, tetrads of microspores, accumulated callose deposition.
S11	1.0 – 1.1	10	5 – 6	Msp	Microspores in interphase, early-vacuolated, minorities of tetrads or bi-nucleate microspores.
S12	1.1– 1.2	11	6 – 8		Microspores vacuolated and undergo first mitotic division, callose degradation, bi-nucleate microspores.

[a] *Arabidopsis* flower developmental stages taken from Smyth et al. (1990).
[b] *Arabidopsis* pollen developmental stage after Regan and Moffatt (1990).
[c] Observed histological events characterized according to Sanders et al. (1999).
n/a, not applicable.

4.1.2 Heterochronic development of male gametophytes in sexual and apomictic Boechera

Divergent apomictic taxa (*e.g. Panicum*, *Tripsacum*) share an accelerated development of their apomeiotic ovules relative to their sexual counterparts (Savidan 2007), an observation which is mirrored by temporal shifts (heterochrony) between both reproductive modes on the transcriptome level (Sharbel et al., 2010). Such a desynchronization of the flower development in apomicts is speculated to result from the temporal alteration of the sexual pathway without disrupting it (Grimanelli 2003). Under this viewpoint we tested, whether microgametogenesis and anther elongation growth are synchronized in the same way between sexual and apomictic *Boechera* flower buds. Therefore, the data set was analysed per genotype and the comparison was focussed on anthers of flower bud stage S8 to S12. Statistical analysis exhibited, that in apomictic genotypes antherhead length predicts all gametophytic stages of anther development, whereas in all sexual genotypes meiotic and microspore stages strongly overlapped in antherhead size class, which together point to heterochronic development between sexual and apomictic antherheads (one-way ANOVA with Tukey-HSD *post hoc* test, between $*p<10^{-2}$ and $***p<10^{-9}$, Fig. 7A). Similar to the observations in ovule

tissues, the onset of meiosis in apomictic antherheads appears to be accelerated compared to sexual anthers (Fig. 7B and C).

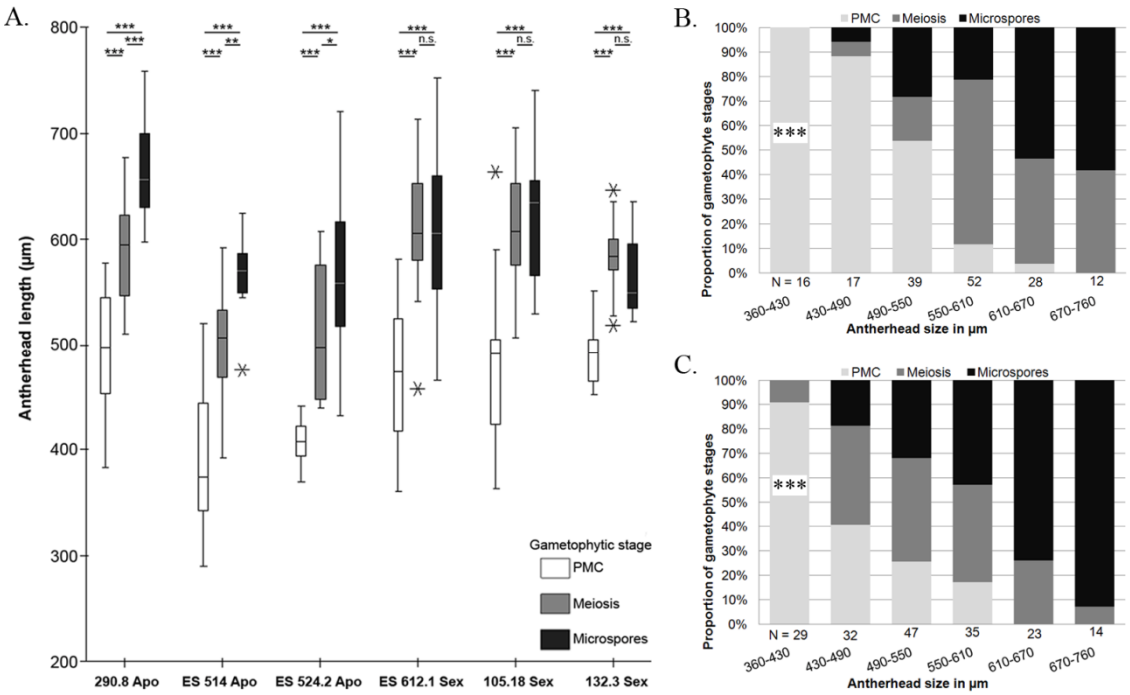

Figure 7. Correlation of antherhead length with different gametophyte stages among sexual and apomictic *Boechera*.

(A) Horizontal bars above boxplots demonstrate significant comparisons between gametophytic stages within each genotype (*p<0.05; **p<0.01; ***p<0.001, n.s. = not significant), as conducted for a 95% binomial proportion confidence interval with a one-way ANOVA including Tukey-HSD *post hoc* test (Supplemental Table 5). Despite observations of antherhead-related gametophyte stages for flower bud stages S3 to S12, the correlation analysis focuses only on S8 to S12 (refer to Figure 1 and 2). Graphs display frequencies of microspores, meiotic and pollen mother cells (PMC) relative to antherhead length bins in (B) sexual and (C) apomictic *Boechera*. The threshold for statistical significance of a stage-specific PMC enrichment according to a two-tailed Fisher's exact test was ***p<0.001 (white boxes).

4.1.3 PMC stage suitable for comparative gene expression analyses

Several criteria were proposed for selection of apomixis candidate genes, such as the specific expression of the candidate at the onset of meiosis and the ability of producing unreduced gametes by circumvention of meiosis checkpoints, which otherwise would override the abnormal behaviour (Grimanelli et al., 2001). In addition, comparative gene expression profiling of sexual and apomictic ovule transcriptomes demonstrated a spike in gene expression change at the megaspore mother cell stage (Sharbel et al.,

2010). Thus, the observation of altered male meiocyte fate at the diplotene stage of prophase I in facultative and obligate apomicts (see *4.2.1*, Fig. 8), pointing to deregulation of genes at premeiotic stages, together with the high correlation between antherhead length and gametophyte stage, led to the selection of antherheads with a length of 400±30 µm (S8) for collection of total RNA. Anthers at bud stage S8 are significantly enriched for PMCs being close to meiosis in both sexual and apomictic genotypes (one-tailed Fisher's exact test, ***p<0.001, Fig. 7B and C, Supplemental Table 5).

4.2 Cytological examination of the *Boechera* male meiocyte development

4.2.1 Chromosome behavior during meiosis in sexual and apomictic genotypes

For the characterization of apomeiosis initiation candidate genes and their products by comparative microarray-based gene expression analysis it is necessary, in addition to a well defined gametophyte stage (see. 4.1), to study their phenotypic variation. Therefore a comprehensive comparative atlas of meiotic and apomeiotic chromosome behaviour for correlation with putative apomeiosis initiation candidates on meiosis I and/ or meiosis II is an essential prerequisite. Hence, the nature of the chromosome behaviour during microsporogenesis was assessed by microscopic analyses of meiotic spreads from antherheads. Meiotic chromosome behaviour during pollen formation differed between apomictic and sexual genotypes (Fig. 8). The sexual *B. stricta* genotype ES 612.1 exhibited the expected homologous chromosome pairing with juncture of non-sister chromatids leading to seven bivalents (Fig. 8A). Subsequent two nuclear divisions lead to a tetrad with four haploid nuclei (Figs. 3B-G). In contrast, the aneuploid obligate apomictic *B. polyantha* (ES 776.2, 2n=2x=15) and the euploid *B. lignifera* (ES 753, 2n=2x=14) exhibited complete or partial chromosomal asynapsis resulting in a high level of univalents at metaphase I, which do not (or only partially) segregate during meiosis I (Figs. 8I, J and P, Q). Subsequently the frequently fused metaphase II plates (Figs. 8K and R) develop into dyads with balanced and unreduced chromosome numbers (Figs. 8L, M and S, T). Low levels of chromosomal synapsis leading to tetrad formation were also observed in all examined diploid obligate apomicts (Figs. 8N and U). Compared to obligate apomicts the facultative apomictic *B. divaricarpa* ES 514 exhibited higher levels of nuclei with bivalents (Figs. 3W and Y) which proceeded through both meiotic cell divisions (Figs. 8A', C' and E'),

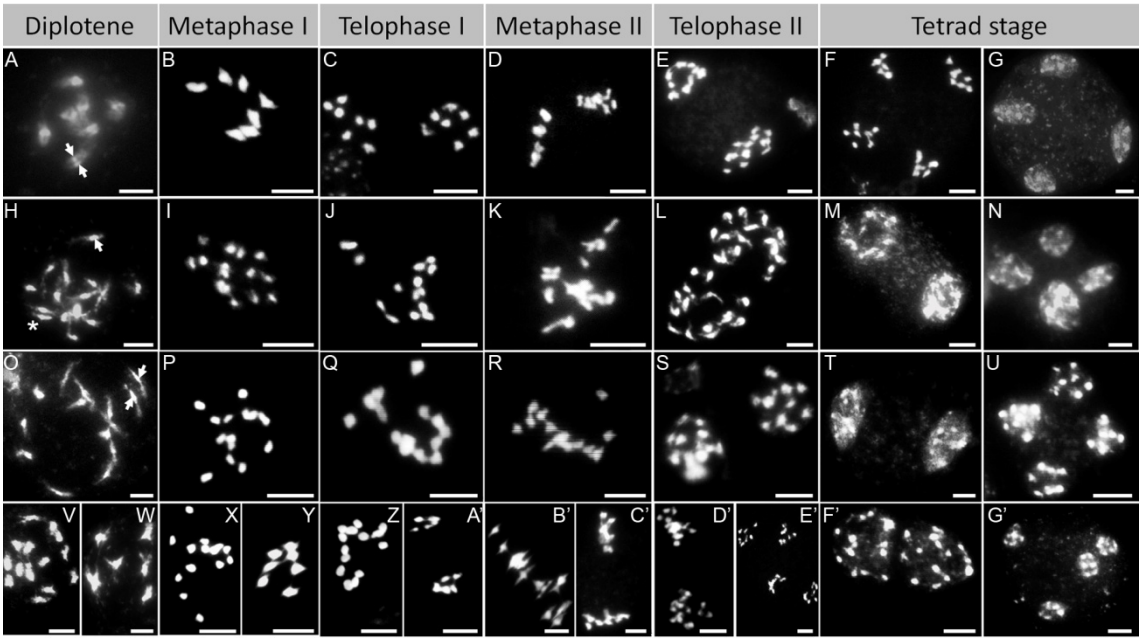

Figure 8. Meiotic chromosome behaviour of male meiocytes.
Chromosome spreads of a diploid obligate sexual (A - G), an aneuploid (H – N) and a euploid (O - U) obligate apomictic and a high facultative apomictic *Boechera* genotype (V – G') are displayed. Fig. H shows two homologues which have not synapsed and arrowheads point to single and adjacent pericentromeric heterochromatin (asterisk). Scale bars, 5 µm.

generating tetrads with four haploid cells (Fig. 8G'). Nonetheless, nuclei with univalents were frequently observed (Figs. 8V, X and Z) whose sisterchromatids were equally separate at metaphase II (Fig. 8B') to develop balanced dyads (Figs. 8D' and F'). Meiotic chromosome counts of obligate and high facultative apomicts strongly support univalent-induced first division restitution without crossover. Univalents remain together during meiosis I and disjoin in meiosis II, where sister chromatids are equationally separated to opposite poles leading to two balanced sets of diploid chromosomes and the formation of balanced dyads.

4.2.2 Male meiocyte constitution at tetrad stage

Previous studies demonstrated low level apomeiotic gamete formation in sexual members of *Boechera* (*e.g. B. stricta*) and a lack of intermediate-frequency apomicts between low-facultative (*sensu* Aliyu et al. (2010); 1-3% apomeiosis) and high-facultative or obligate apomicts (>87-100% apomeiosis) led to the hypothesis, that high apomeiosis levels may have been induced by global gene regulatory changes associated with hybridization (Kantama et al., 2007; Aliyu et al., 2010). Thereby, indirect extrapolations from FCSS data revealed that diploid apomicts produce prevalently

unreduced diploid pollen with low frequencies (<1%) of pollen ranging from reduced haploid (1C) to unreduced hexaploid (6C) genome content.

Here we used a simple anther squashing technique to examine the potential of high-facultative and obligate apomicts to produce reduced and unreduced gametes. The meiotic atlas gave first insights into pollen meiosis demonstrating the production of tetrads carrying reduced chromosome sets in all tested apomicts (Fig. 8G-G'). In order to quantify meiotic product frequencies, about 28 000 meiocytes from single antherheads of six apomictic and a sexual reference genotype were counted (Figs. 9 and 10). The sexual reference genotype produced the expected high frequency of tetrads (88%), whereas the tetrad frequency varied among apomicts (0 to 87%) between individuals of one genotype and even between different flower buds from the same individual (Fig. 9, Supplemental Table 6). Low levels of triads and tetrads were observed for apomictic genotypes ES 805.2 (0.55% and 0.35%) and ES 776.2 (0%), both of which showed high levels of monad and dyad formation. While part of the triad population could be pseudotriads, having lost one nucleus through the preparation procedure, this is inconsistent with the variability of triad number detected between genotypes (Fig. 9). Surprisingly high levels of triads and tetrads in both obligate (300.9 - 24.98% and 57.98%; ES 753 - 32.78% and 62.16%) and high-facultative apomicts (ES 514 - 11.22% and 87.00%; ES 524.2 - 15.67% and 64.53%) were found (Fig. 9). Distortions in chromosome segregation resulting in the formation of extra nuclei are low (in total 0.19% meiocytes with micronuclei and 0.02% polyads) and were mainly detected in the apomict ES 753 (0.88% meiocytes with micronuclei, Fig. 10). Interestingly, FCSS of all obligate apomictic *Boechera* suggests successful fertilization with only unreduced pollen (Fig. 10).

The low abundance of segregation distortions during meiosis in apomictic male gametes points to a coordinated genetic program underlying the process. Considering the highly variable pattern of meiocyte frequencies between individuals and between flowers of single specific apomictic genotypes (Fig. 9), only plants producing exclusively unreduced (2C) pollen were chosen for comparative gene expression analyses. Plant selection was therefore based upon measurements of nuclear DNA content of pollen and seed nuclei for all individuals (examples in Fig. 10).

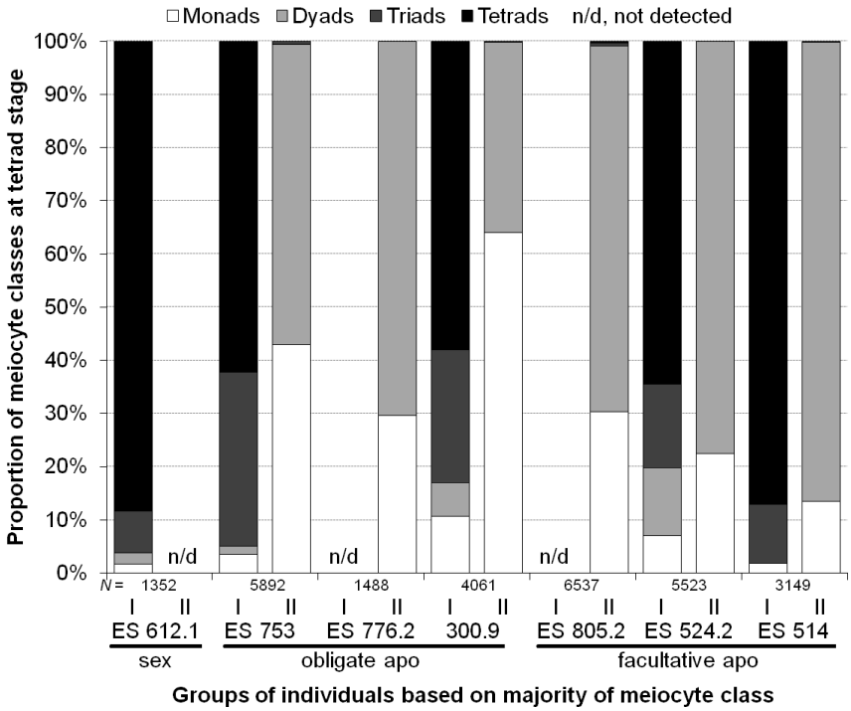

Figure 9. Meiocyte constitution at the tetrad stage in diploid sexual and apomictic *Boechera*. Roman numbers denote (I) groups of individuals primarily (>50%) producing reduced gametes *versus* (II) groups of individuals primarily (>50%) producing unreduced gametes.

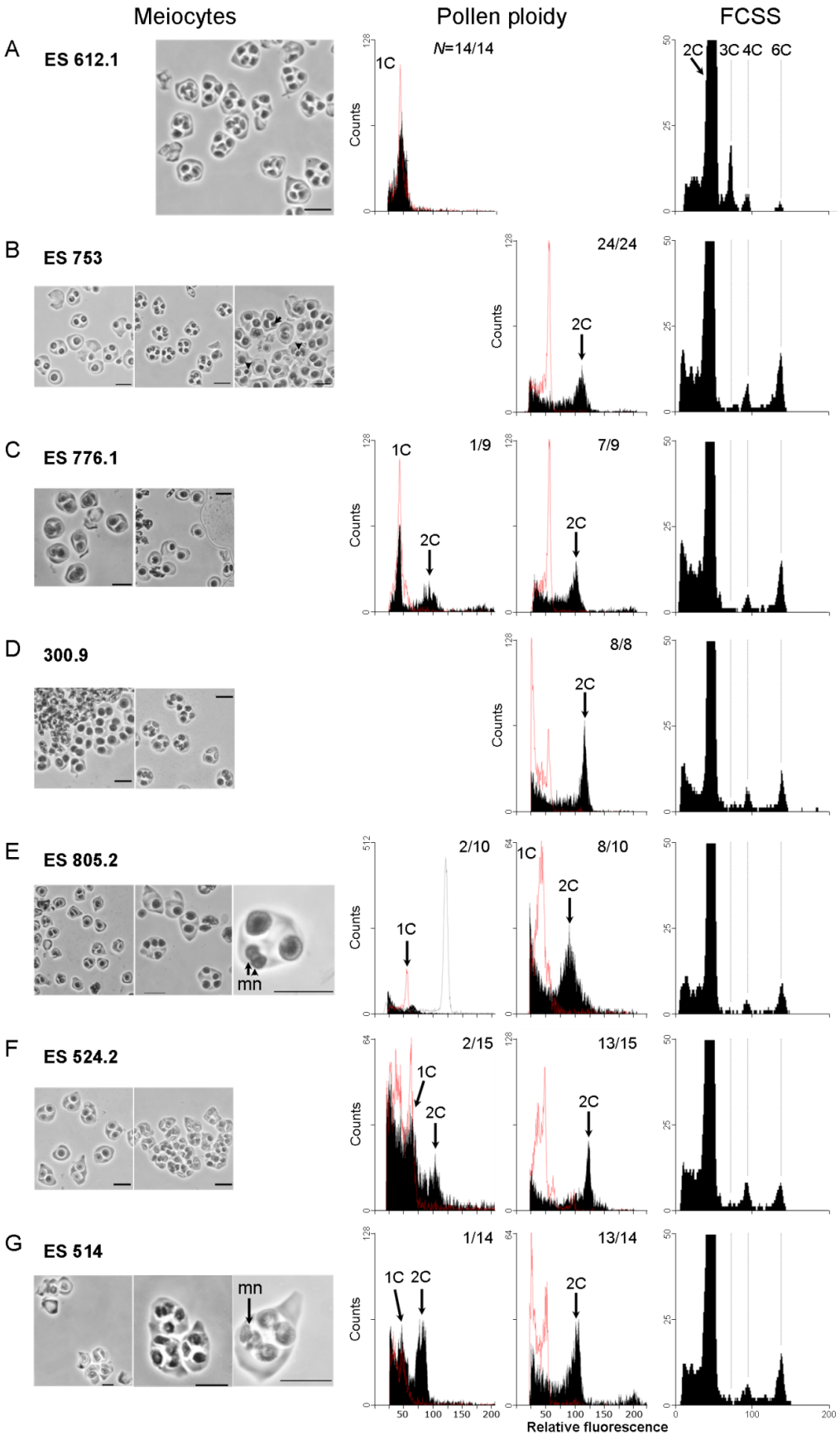

Figure 10. Representative meiocytes at tetrad stage and flow cytometric ploidy confirmation. Ploidy of pollen and seed material of diploid sexual (A) and apomictic *Boechera* genotypes (B-G), respectively was confirmed. (A) Double fertilization of the reduced embryo (1C) and

the binucleate central cell (1C+1C) by a reduced pollen (1C) leads to formation of a 2C=[$1C_{maternal}$ +$1C_{paternal}$] embryo and a 3C=[$1C_m$ + $1C_m$] + [$1C_p$] endosperm in sexual genotypes (*e.g.* ES 612.1, *B. stricta*). (B-G): In apomictic *Boechera* the 2C pollen fertilizes only the unreduced central cell, whereas the unreduced egg cell develops towards the embryo *via* parthenogenesis, giving a 2C=[$2C_m$] embryo and 6C=[$2C_m$ + $2C_m$] + [$2C_p$] endosperm. Although results from the pollen ploidy screen and the meiocyte squashes were not always congruent, both results together show that in all of the tested facultative and obligate apomictic genotypes, some individuals produce predominantly reduced or both reduced and unreduced pollen. One 1C pollen external control from the diploid sexual *Boechera* genotype ES 558.2 was used for all pollen ploidy measurements (red profile) and a leaf external control from the same diploid sexual genotypes was used for the flow cytometric seed screen (FCSS). X-axis: linear fluorescence, y-axis: events count. The 4C peak represents the G_2 of the embryo. *N*=individuals per genotype; mn/black arrows=micronuclei.

4.2.3 Callose distribution during pollen meiosis

In sexual species the deposition of callose, a β-(1,3)-linked glucan, can be used as a marker for the onset of meiosis. Callose accumulates during meiosis reaching its peak during the tetrad stage and dissolutes in subsequent stages leading to the separation of the microspores. In contrast, a lack of callose synthesis during megagametogenesis was reported for a number of apomicts (Koltunow 1993; Barcaccia et al., 1996; Tucker et al., 2001). Observation of callose distribution during microsporogenesis were previously performed in diploid and triploid apomictic *Boechera* and showed contrasting results for both ploidy levels (*i.e.* lack of callose cross walls for dyads of the diploid *B. divaricarpa* genotype, Taskin et al. (2009)).

In order to reveal the nature of callose distribution during the complete male gametogenesis in diploid apomicts we used anti-(1,3)-β-glucan primary antibodies to track the callose distribution profile from pre-meiosis to binucleate microspore stages between one flow cytometrically confirmed diploid sexual *B. stricta* and one diploid apomictic *B. divaricarpa* genotype.

In general, callose distribution across the different stages is similar between the sexual and the apomictic genotype (Fig. 11). Callose deposition before bud stage S9 (anther length <0.5mm) was not detected in anthers of both reproductive modes (Fig. 11A, E, I and M), but was first recognized in sexual PMCs of antherhead lenghts between 0.55 – 0.6 mm (bud stage S9) compared to ~0.50mm in *B. divaricarpa* (Fig. 11B, F, J and N). Similar to the formation of tetrads in the sexual genotype, dyads and tetrads of the facultative apomict were separated by a thick callose layer, including the

formation of callose cross walls in all tested antherheads (Fig. 11C, G, K and O). Besides proceeding callose degradation, callose deposition seems to persist through the first stages of microgametogenesis and was detected in uni- and bi-nucleate microspores, which in parellel exhibited mature pollen walls (Fig. 11D, H, L and P, white arrowheads).

Interestingly, at this stage the callose is encapsulated by the pollen wall layers (Figs. 11S and T), which is different to observations in other closely related species *e.g.* *Brassica napus* L. (Cresti et al., 1992) or *Arabidopsis* (Suzuki et al., 2008), and to reported models *e.g.* based on *Lilium* (Scott et al., 2004), where the developing pollen wall is enveloped by the callosic wall which dissolves from the outside while pollen wall development proceeds inside of the callosic wall towards the cytoplasm. As no typical callose cover of the developing pollen wall was detected during pollen grain maturation, but instead an irregular and relatively thick layer covering the entire inner side of the pollen wall including the aperture sites was observed, it is hypothesized that the secondary callosic layer is initiated within the intine (ca_2, Figs. 11S and T), the so called 'outer intine', as has previously been observed gymnosperm pollen (*e.g.* *Pinus sylvestris*, Rowley et al. (2000)). Alternatively, the inner callose layer could be a remnant of the originally thick layer which separates the microspores at tetrad stage, and which subsequently was more quickly degraded from the outer side compared to the inner side. In summary, different from the altered callose deposition in MMCs in a number of apomicts, callose deposition and dissolution during microsporogenesis of the apomictic *Boechera* genotype is unaffected and comparable to the sexual genotype. Hence, callose deposition seems not to be an interrelated trait with male apomeiosis. One reason could be that the role of callose during pollen development is not solely reduced during microsporogenesis (*i.e.* separation of the developing microspores), but in addition plays a role later in pollen tube formation (Müller-Stoll and Lerch 1957). Whether the inverse layering of callose into the inner side of the pollen wall represents a *Boechera*-specific chronology of pollen wall development, or represents the intine layer in which callose is integrated beside cellose and hemicelluloses, remains to be examined.

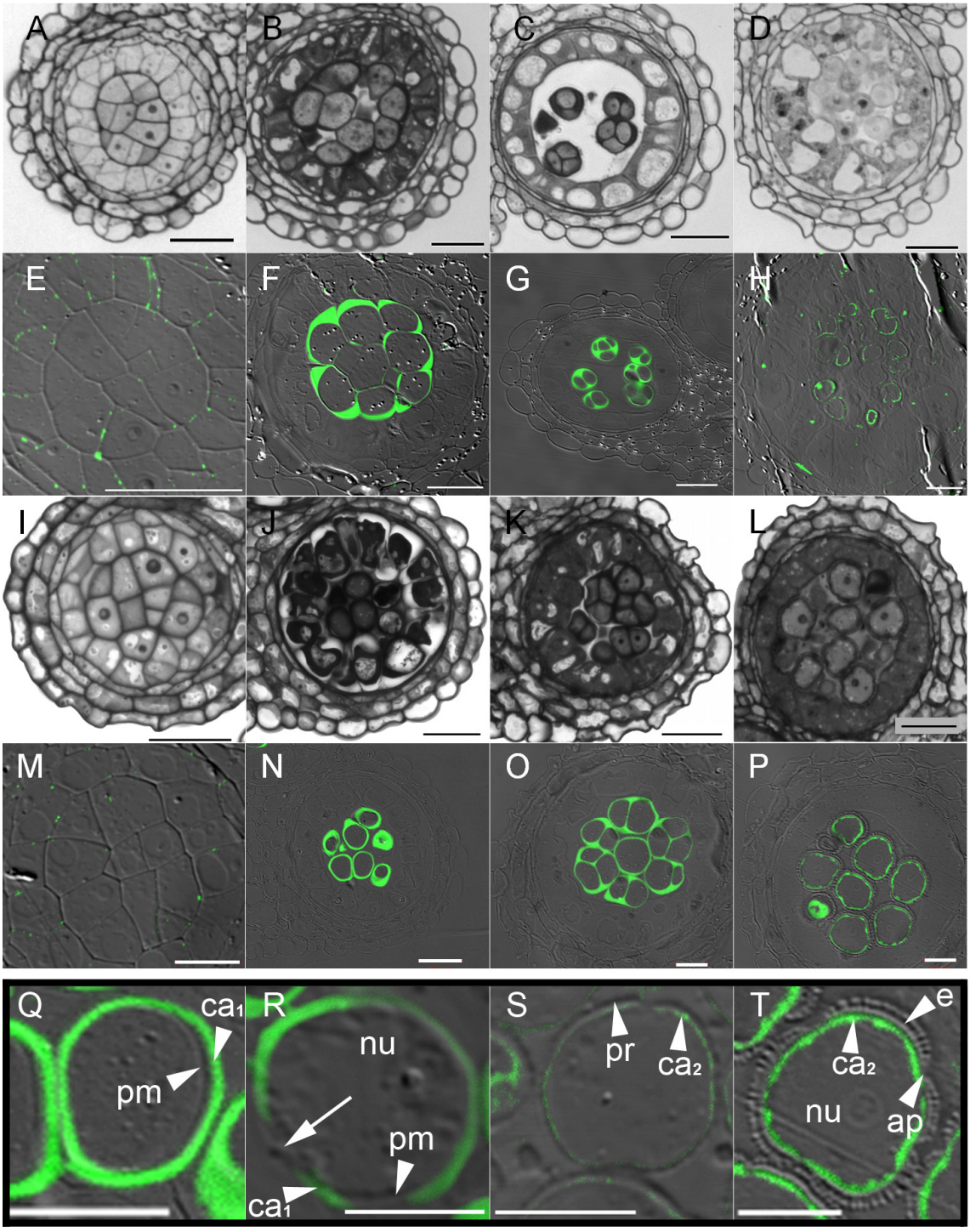

Figure 11. Immune staining of callose wall formation during microgametogenesis with anti-(1,3)-*β*-glucan antibody.

Panel displays three of the four major histodifferentiation stages of male gametes (PMC: A, E, I and M; Me: B, C, F, G, J, K, N and O; Msp: D, H, L and P) from sexual *B. stricta* (ES 612.1; A-D) and from apomictic *B. divaricarpa* (ES 524.2; I-L) and associated callose distributions (ES 612.1 – E-H; ES 524.2 – M-P). Black boxed graphs show dynamics of callose distribution which covers the entire young microspore after separation from the tetrad (Q). With proceeding maturation of the micronuclei the outer callose degraded (R) and an inner callose layer was formed, probably in the intine layer (S, T). ap- aperture, e – exine, ca_1 – external

callose, ca_2 – internal callose, nu – nucleus, pm – plasma membrane, pr – primexine. White arrow marks degrading external callose. (A-P) Bar scale, 20 µm and (Q-T) bar scale, 10 µm.

4.3 Microarray-assisted screen for apomeiosis-inducing candidate genes

4.3.1 Pre-selection strategy for apomeiosis candidate genes

The switch from sexual to apomictic seed production is associated with deregulation of the sexual developmental pathway in female and male gametes (Koltunow 1993; Grossniklaus 2001) and is characterized by wide-ranging chromosomal (Kantama et al., 2007) and global gene expression changes (Sharbel et al., 2010), which are hypothesized to result from hybridization and/or polyploidy (Grossniklaus 2001). The goal of this experiment was to perform a comparative transcriptomic analysis of sexual and apomeiotic antherheads enriched for PMCs at a developmental stage prior to meiosis (Figs. 2 and 3I'; Supplemental Table 5) to identify the factor(s) responsible for apomeiotic pollen formation in *Boechera*. A large number of diploid biological replicates were used in order to statistically correct for genetic- and ploidy-mediated background noise in subsequent analyses of differential gene expression. Based upon haplotype designation, geographic distribution and pollen ploidy, 7 different diploid sexual genotypes, and 7 different diploid obligate and facultative apomictic genotypes were chosen for a microarray-assisted gene expression analysis of microdissected live antherheads (Supplemental Table 2). Following the Agilent One color microarray protocol (version 5.7; Palo Alto, CA, USA) we revealed high quality microarray data (Supplemental Figure 1) which subsequently were normalized using the quantile normalization scenario with median to baseline transformation. The 'processed signal' values were plotted genotype-wise as a relative fold change on a log2 scale (Fig. 12). In a first attempt we examined the mean and distribution of normalized log2 intensity values between the genotypes without considering significance levels of single array probes between the reproductive modes. Therefore, gene expression ratios above a fold change (FC) cut-off value ≥2 were filtered without examining any *p*-value cut-off. Taking a FC≥2 significantly decreases the total number of probes from 103747 to 9496. Although the means of normalized log2 intensity values between the genotypes differ slightly (one-way ANOVA: $F(13, 132930)=178.533$, $p<0.001$; Fig. 12B), no overall differences of the distribution of the normalized log2 intensity values were detected across all genotypes (Tuckey-HSD *post hoc* test: -0.023±2.69 log2 intensity values, $p=0.803$). Increasing the fold change to ≥10, decreases the number significantly to 522 array probes. Interestingly, the mean and distribution of these highly expressed

normalized intensity values are statistically different between three groups: sexual *B. stricta*, all remaining sexuals and apomictic genotypes (one-way ANOVA, $F(13,7294)=157.545$, $p<0.001$; Fig. 12C and F). A PCA of the 522 highly differential expressed array probes exhibited a closer relation of apomictic with other sexual genotypes compared to sexual *B. stricta* genotypes (Fig. 2F). A Tuckey-HSD *post hoc*

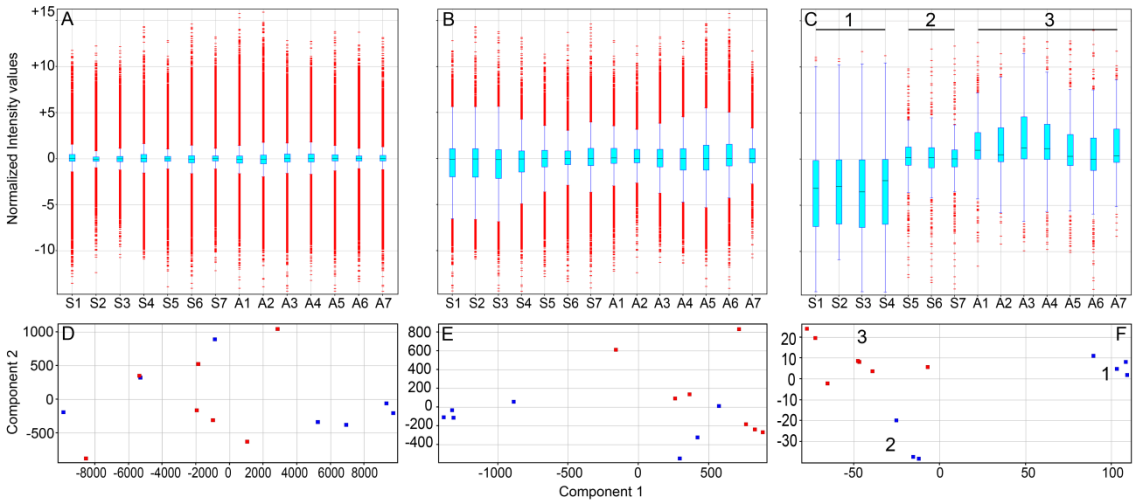

Figure 12. Microarray sample intensity values and 2D-PCA projection.
(A-C) Distributions of log2 ratios of normalized data for 7 sexual (S1 to S7) and 7 apomictic (A1 to A7) *Boechera* genotypes revealed a similar distribution across all genotypes considering all values above background (A) with fold change cut-off value (FC)≥2 (B). Entities with log2 intensity values beyond 1.5 times the inter-quartile are show in red. For expressed normalized intensity log2 values with FC≥2 significantly different means, but no statistical significant differences between the three groups were detected (one-way ANOVA, $F(13, 132930)=178.533$, $p<0.001$; Tuckey-HSD *post hoc* test: -0.023±2.69 log2 intensity values, $p=0.803$). Highly expressed normalized intensity values are statistically significant different expressed between sexual *B. stricta*, other sexual genotypes and apomictic genotypes (FC≥10, one-way ANOVA, $F(13,7294)=157.545$, $p<0.001$; C). A Tuckey-HSD *post hoc* test revealed, that mean intensity values in sexual genotypes are lower or down regulated (sexual *B.stricta*: -3.418 to -2.896 log2 intensity values, $p<0.001$; other sexual genotypes: -0.243 to -0.131 log2 intensity values, $p=0.027$) compared to apomictic genotypes (0.806 to 2.033 log2 intensity values, $p<0.001$; except for genotype A6: 0.328±4.14 log2 intensity values, $p=0.793$). (C-F) 2D-PCA projection demonstrated clustering progress of the biological replicates into the three subgroups. In 2D-PCA for FC≥10 log2 intensity values, the cumulative variance of 80.97% was reached for the first two principle components.

test revealed that mean log2 intensity values in sexual genotypes are lower or down regulated (sexual *B.stricta*: -3.418 to -2.896 log2 intensity values, $p<0.001$; other sexual genotypes: -0.243 to -0.131 log2 intensity values, $p=0.027$) compared to apomictic

genotypes (0.806 to 2.033 log2 intensity values, $p<0.001$; except for genotype A6: 0.328±4.14 log2 intensity values, $p=0.793$).

The 522 array probes corresponded to 311 cDNAs for which gene annotations in *Arabidopsis* were available, and for which a gene ontology (GO) analysis was performed. In total, 250 (80.39%) probes were GO annotated and could be assigned to the second level molecular function GO classes binding (*N*=67, 6.80%), catalytic activity (*N*=86, 34.40%) and transcription regulator activity (*N*=13, 7.60%; Fig. 13A; Supplemental Table 7). Cellular (*N*=78, 31.20%) and metabolic processes (*N*=33, 32.80%) are the most prominent biological processes (Fig. 13B). Overall, a GO analysis for overrepresentation of particular classes of genes between the annotated highly expressed probes in apomictic genotypes and all annotated probes on the 2x105k-*Boechera* flower-specific microarray (*N*=31051) shows that only genes involved in the catalytic activity, especially with mitogen-activated protein (MAP) kinase activity (GO:0004707, FDR-corrected *p*-value=0.0076), are overrepresented on a low level among the '522 highly expressed probes' dataset (Fig. 13C, Supplemental Table 7).

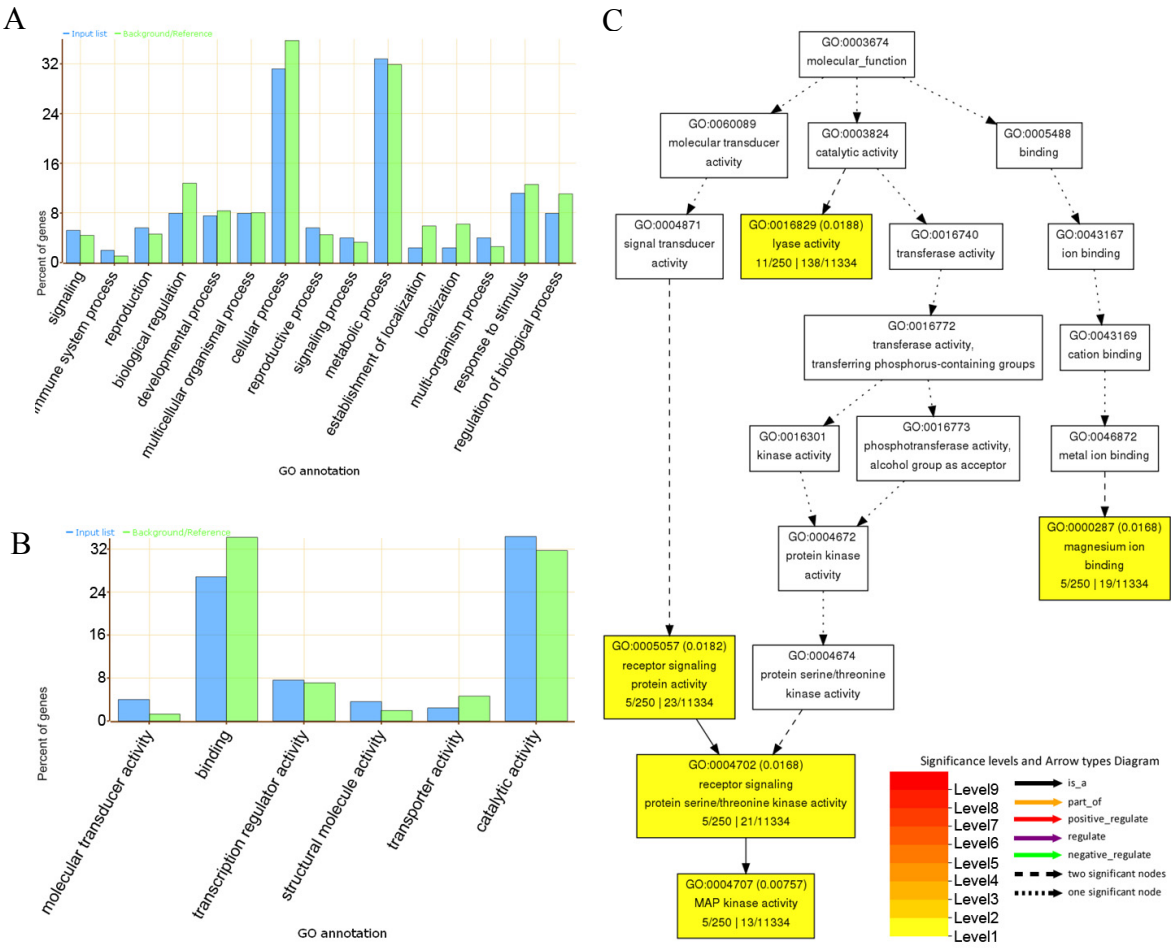

Figure 13. Gene ontology (GO) analysis of 522 higly expressed probes with cut-off value FC≥10.

Bar charts for the GO analysis show key molecular functions (A) and biological processes (B) of all annotated sequences characterized by FC≥10 in apomictic and sexual Boechera genotypes (blue) versus the total annotated genes on the microarray (green). (C) A directed acyclic graph (DAG) for the GO analysis showing key molecular functions of all enriched probes of the 522 higly expressed probes with cutt-off value FC≥10 in apomictic Boechera genotypes.

Using a foldchange threshold ≥2 and a Bonferroni corrected *p*-value ≤0.05 no single differentially expressed probe was detected, but when allowing different single outlier genotypes for each round of analysis, 13 (0.013%) microarray probes were identified to be consistently differentially-regulated between sexuals and apomicts (*i.e.* by exclusion of four different genotypes significantly differentially-regulated probes were detected, Table 2, group A-D). Except for one probe (Sharb1627083, -3.92 AFC), all are highly upregulated in apomicts (+40.00 to +849.30 AFC; Table 2, Figure 14). Homologous cDNAs corresponding to the 13 candidate array probes were obtained from a flower-specific cDNA library (Supplemental Table 3; Sharbel et al. (2009)). A BLASTN search (Altschul et al., 1997) of these cDNAs against the whole GenBank nucleotide collection (http://www.ncbi.nlm.nih.gov/genbank/) revealed hits for only five of the 13 candidate probes (Table 2). The annotation includes the *A. thaliana* TTR1/AtWRKY16 (AT5G45050) and RRS1/AtWRKY52 proteins of the WRKY protein family (AT5G45260), which have a general binding preference for genes containing W box promoter elements including the WRKY genes themselves as well as a large variety of defense-related genes (*e.g.* pathogenesis-related (*PR*) genes); a *PHENOLIC GLUCOSIDE MALONYLTRANSFERASE 2* (*PMAT2*) gene which encodes for a HXXXD-type acyl-transferase-like protein (AT3G29670) of the BAHD acyltransferase family, which is required for the synthesis of acylated anthocyanins; and an uncharacterized protein (AT1G31130). Only two of the annotated differentially-expressed genes between sexual and apomictic genotypes showed pollen-specific functions; the *A. thaliana S-locus* lectin protein kinase (AT1G11410) which is involved in pollen recognition, and a *SRK* gene (S-12 and S-15 type; AB180901.1) encoding a *Brassica oleraceae S-locus* receptor kinase involved in pollen self-recognition specificity (Table 2).

Table 2. Microarray probes demonstrating significant absolute fold change regulation in apomictic *versus* sexual genotypes.

ID[a]	Microarray probe no.[b]	Corrected *p*-value	AFC	Reg[c]	Outlier acc. no.	AGI[d]	Predicted function	*E*-value[e]	Species	Strand
A	Sharb0350102	0.0057	229.10	up	132.3	AT5G45050	WRKY transcription factor 16 (TTR1)	2.00E-93	*A. thaliana*	−
						AT5G45260	WRKY transcription factor 52 (RRS1)	5.00E-69	*A. thaliana*	+
B	Sharb0834363	2.52E-07	219.28	up	29.1	n/a	n/a	n/a	n/a	n/a
	Sharb0791806	1.77E-09	292.06	up		n/a	n/a	n/a	n/a	n/a
	Sharb1627083	1.80E-07	-3.92	down		AT3G29670	HXXXD-type acyl-transferase-like protein	4.00E-44	*A. thaliana*	+
	Sharb0789018	8.39E-08	303.42	up		n/a	n/a	n/a	n/a	n/a
	Sharb0789016	1.74E-08	354.04	up		n/a	n/a	n/a	n/a	n/a
C	Sharb0700754	5.10E-05	36.79	up	ES 805.2	AT1G31130	Uncharacterized protein	3.00E-05		−
	Sharb0505763	0.015	40.00	up		-	Chromosome 5, complete sequence	3.00E-20	*A. thaliana*	n/a
D	Sharb1199059	7.08E-06	158.53	up	ES 753	AT1G11410	*S*-locus lectin protein kinase	2.00E-30	*A. thaliana*	+
	Sharb0931225	7.12E-05	849.30	up		n/a	n/a	n/a	n/a	n/a
	Sharb0501554	1.00E-05	188.29	up		AB180901.1	*S*-12 SRK gene for S-locus receptor kinase	6.00E-05	*Brassica oleracea*	
	Sharb0690829	3.12E-04	76.50	up		n/a	n/a	n/a	n/a	n/a
	Sharb0425060	4.90E-08	48.73	up		n/a	n/a	n/a	n/a	n/a

[a] ID's correspond to single plots in Fig. 14.

[b] Microarray probe numbers correspond to cDNA database IDs in Supplemental Table 3.

[c] Direction of regulation is apo *versus* sex.

[d] Arabidopsis Genome Initiative (AGI) gene identification number from TAIR10.

[e] *E*-values corresponds to the best hit of the microarray probe corresponding *Boechera* cDNAs in a BLASTN analysis to GenBank nucleotide collection. n/a, not applicable. AFC, absolut fold change.

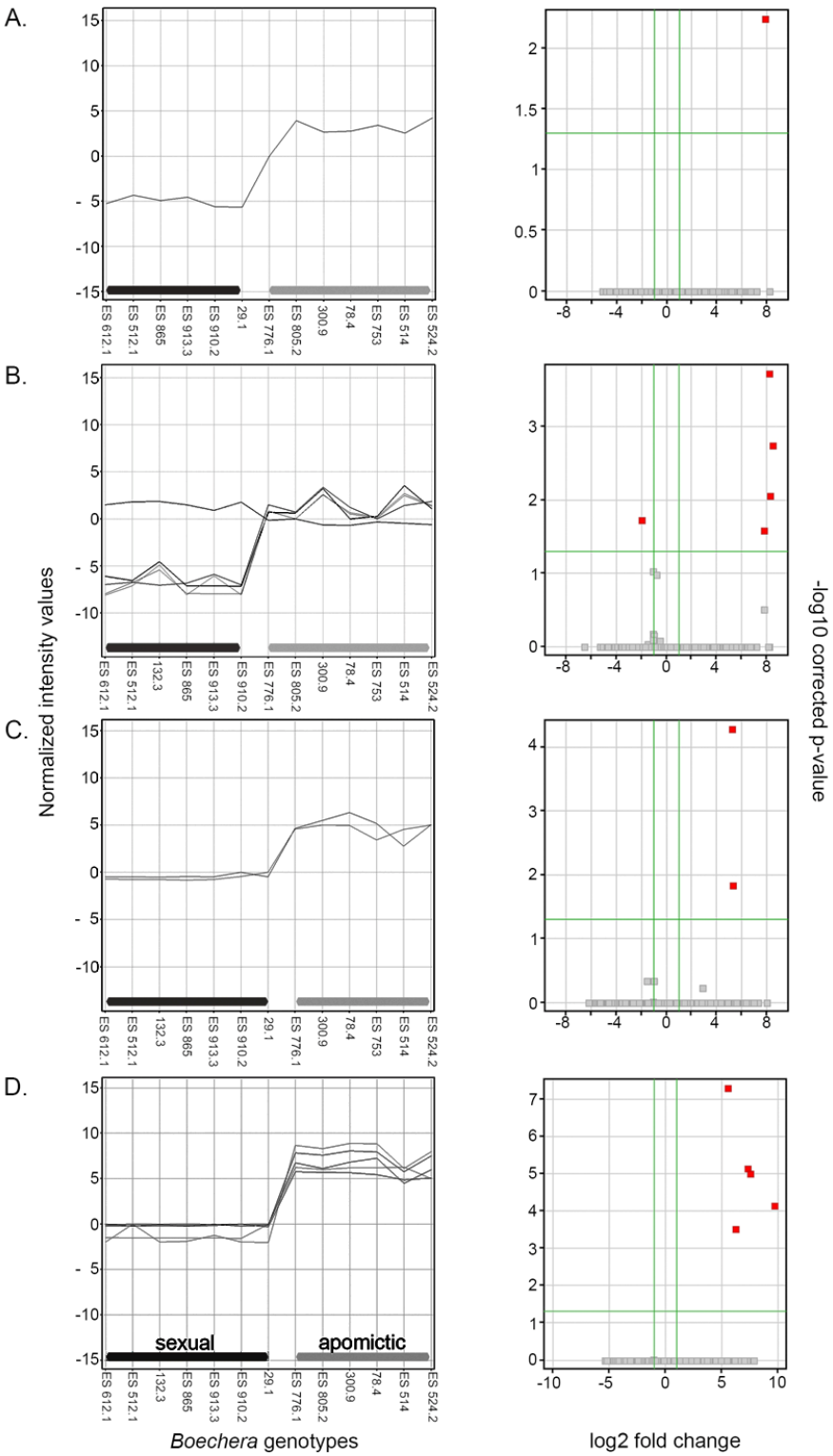

Figure 14. Constantly differentially regulated microarray probes in apomictic compared to sexual *Boechera* genotypes.

Line charts (left side) and Volcano plots (right side) display relative expression ratios of candidate microarray probes when allowing different single outlier genotypes (see Table 2). The Y-axis of the line chart displays the averaged Log2 normalized fluorescence intensity, while the X-axis show single *Boechera* genotypes and their specific mode of reproduction. The volcano plot displays the negative log (base 10) of *p*-values from unpaired *t*-tests on the

Y-axis, while the X-axis show the log (base 2) of the fold differences of the microarray probes between the sexual and the apomictic genotypes. The identified differentially expressed probes are marked in red. Green lines mark the treshholds for the fold change (≥2) and the significance level ($p \leq 0.05$).

4.3.2 Validation of differentially expressed candidate microarray probes

QRT-PCR was employed to validate the five annotated candidate genes found in the microarray analysis of 14 *Boechera* genotypes, except for the probe Sharb0700754 which blasted to an uncharacterized protein. Furthermore three unannotated microarray probes of the group D (Sharb0931225, Sharb0690829 and Sharb0425060; Table 2, Figure 14) were included due to the pollen-specifity of the two other genes in the same group (Sharb1199059 and Sharb0501554) to give the total of seven candidate genes which were tested by qRT-PCR. In general, relative mRNA expression values obtained by qRT-PCR showed the same trend as those obtained by the microarray (Figure 15, Supplemental Table 8). However, the sequence Sharb1199059 (lectin homolog), which is furthermore characterized by a negative BAC screen result (Table 3), was not considered for further analysis due to technical reasons. In addition, the *PMAT2* gene (Sharb1627083) is down regulated in apomictic genotypes according to the microarray experiment (-3.92, p=1.80E-07) but not uniformly differentially expressed across all genotypes of both reproductive modes according to qRT-PCR (one-way ANOVA, $F(1,11)=2.331$, $p=0.155$).

QRT-PCR confirmed the microarray expression profiles of the other five microarray probes (Figure 15; Supplemental Table 8). With exception of the sexual outlier sample (132.3), the microarray probe Sharb0350102 is highly upregulated in apomictic antherheads (one-way ANOVA, $F(1,11)=23.903$, p=4.70E-04). Similarily, the remaining four probes of group D are highly upregulated in apomictic antherheads (one-way ANOVA, Sharb0931225: $F(1,11)=8.886$, $p=0.012$; Sharb0501554: $F(1,11)=15.883$, $p=0.002$; Sharb0690829: $F(1,11)=13.287$, $p=0.004$; Sharb0425060: $F(1,11)=8.909$, $p=0.012$).

Quantitative RT-PCR techniques were used not only to validate the microarray data, but also to investigate the expression level of the selected genes in different tissues at PMC stage. In contrast to sexuals, in which the relative mRNA levels were close to the detection limit, ubiquitous expression of all probe sequences was detected in apomictic somatic (whole flower without anthers and leaf) and reproductive tissues (antherhead). Thereby, for probes Sharb0350102, Sharb1627083, Sharb0931225, Sharb0501554 and

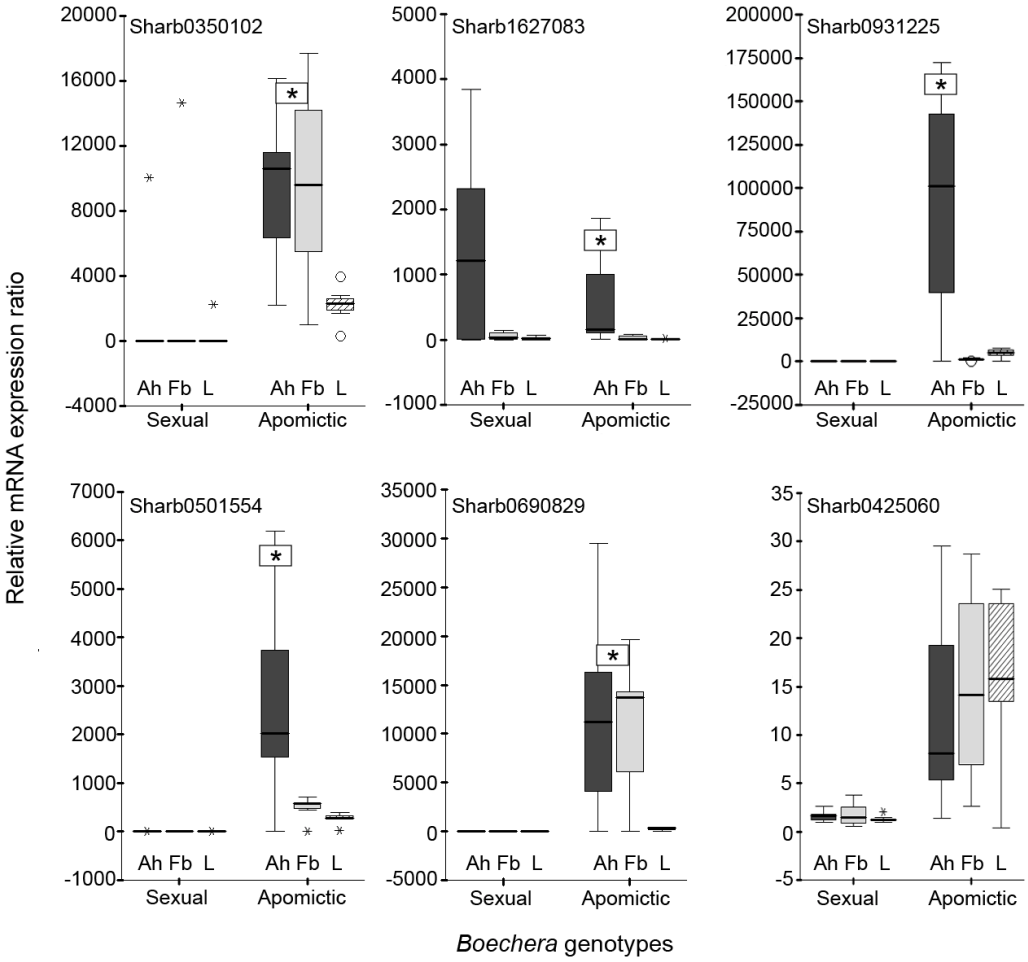

Figure 15. Validation of differentially-expressed microarray probes by qRT-PCR.
Expression of candidate microarray probes was tested in antherhead (Ah), flower bud without antherhead (Fb) and leaf tissue (L) of seven sexual *versus* seven apomictic *Boechera*. Error bars display standard errors for tissue-specific relative mRNA expression levels of the microarray probes in sexual and apomictic genotypes. Free asterisks and circles mark genotype-specific outlier values. Significant differentially-expressed probes between tissues are marked with boxed asterisks (*p<0.05).

Sharb0690829, expression was substantially higher in antherheads compared to levels in leaf tissue (Tukey-HSD *post hoc* test, p=0.021, p=0.045, p=0.022, p=0.008 and p=0.027, respectively; Fig. 15; Supplemental Table 8). Microarray probe Sharb0425060 exhibited no significant differences between any of the tested tissues (one-way ANOVA, $F(2,18)=0.254$, p=0.779).

QRT-PCR on validated microarray probes Sharb0350102, Sharb0931225, Sharb0501554 and Sharb0690829 in antherhead tissue was extended to three additional apomictic and four additional sexual genotypes which were not previously used for the microarray experiment (Fig. 16, Supplemental Table 9). Interestingly, relative

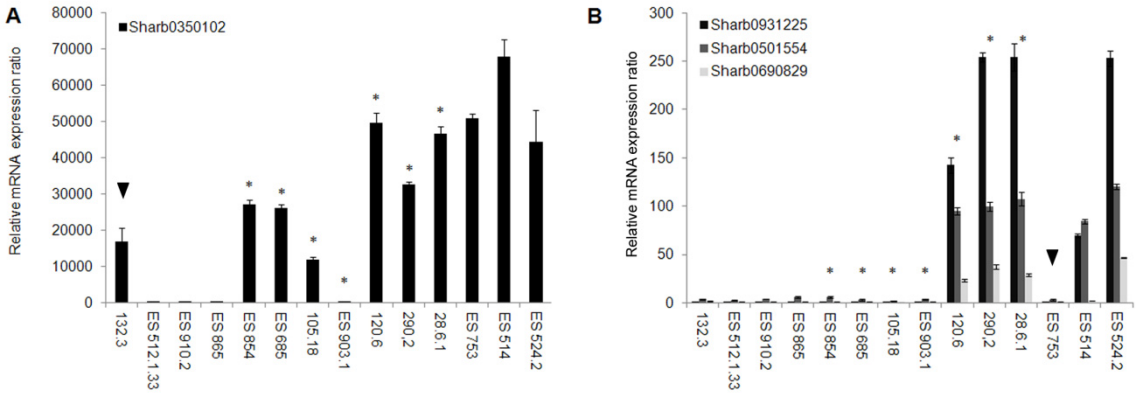

Figure 16. Validation of differentially-expressed microarray probes by qRT-PCR in antherhead tissue in seven microarray and seven additional *Boechera* genotypes.
Standard error bars are based on four technical replicates per genotype. Free asterisks mark additionally tested genotypes, while all other genotypes were used for microarray and serve as internal reference. Black arrows mark outlier genotypes identified by microarray-based gene expression analysis.

expression levels of microarray probe Sharb0350102 in sexual and apomictic genotypes showed substantial variability despite generally lower expression in sexual compared to apomictic genotypes (one-way ANOVA, $F(1,12)=10.621$, $p=0.007$, Fig. 16). Three of four additional sexual genotypes show similar upregulation as demonstrated for the outlier genotype 132.3 while the other sexual genotypes are completely downregulated. In addition one apomict (ES 903.1) was completely downregulated compared all other apomicts. In contrast, microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 exhibit constant relative differential expression between all sexual and all apomictic genotypes, including the additional non-microarray genotypes. The constantly lower average expression levels of these three probes in this second-level analysis compared to the initial analysis could be explained by the extended period of total RNA storage between the two analyses.

All told, preferential upregulation of microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 in apomictic antherhead tissue was confirmed for 10 apomictic compared to 11 sexual genotypes with exception of the apomictic outlier sample ES 753. Microarray probe Sharb0350102 was not considered for further analyses due to ambiguities of its expression pattern between genotypes of each reproductive mode.

4.4 *UPGRADE* is a primary candidate for unreduced pollen formation

4.4.1 Molecular structure of UPGRADE

In search of an apomeiotic regulator for unreduced pollen formation in *Boechera* the qRT-PCR-validated differentially expressed microarray probes Sharb0931225, Sharb0501554, Sharb0690829 and Sharb0425060 were selected for further analyses, considering their constant upregulation in apomictic reproductive tissues of virtually all tested genotypes (Figs. 15 and 16; Supplemental Tables 8 and 9). Chromosome walking along the 3`- and 5`-ends of the five candidate probes extended the range of sequence length to 843bp and 1337bp, and identified priming sites for gene specific primers used for the rapid amplification of cDNA ends (RACE). Chromosome walking products of microarray probe Sharb0931225 and Sharb0501554 overlapped with one another (Supplemental Table 3).

A RACE pre-test using primers specific for Sharb0501554 on a single apomictic genotype exhibited that microarray probe Sharb0690829 belongs in addition to microarray probes Sharb0931225 and Sharb0501554 to the same transcript (data not shown). Hence RACE was performed using gene specific primers from microarray probe Sharb0690829 (GSP3 for 5`-end RACE and GSP4 for 3`-end RACE) to obtain full-length cDNA including the 3`- and 5`-cDNA ends from ten genotypes. Although total RNA of antherheads (data not shown) and whole flower tissue from two sexuals and eight apomicts was used, 5`- and 3`-RACE generated for each tissue DNA fragments of similar size exclusively in all apomicts, except for the apomictic outlier ES 753 (Fig. 17).

All fragments from five apomicts were cloned and sequenced, and led to the identification of a single polyadenylated full-length transcript, Bsp*UPG* (*Boechera* species *UPGRADE; unreduced pollen grain development*, EMBL No. HF930769) with appr. 2648 nt length (*i.e.* for *B. divaricarpa* ES 524), excluding the poly(A) sequence. Interestingly, the microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 were found in the most 3`-exon of the Bsp*UPG* transcript (Fig. 18, Supplemental Table 10). Mapping of different Bsp*UPG* full-length cDNA variants onto genomic DNA by isolation of 3`-RACE and the various 5`-RACE fragments in different *Boechera* genotypes showed that the gene has an overall length of 3156 nt (*i.e.* for *B. divaricarpa* ES 524) and contains two putative alternative splicing sites, 61 bp (Intron1) and 303 bp size (Intron2), in addition to a 144 bp intron common to all apomicts (Intron3, Fig. 17, white arrows, Fig. 18).

The best, albeit short, sequence matches from GenBank are located on *Arabidopsis* chromosomes At1, At4 and At5, and encode respectively a GTP-binding elongation factor Tu protein (EFTU/EF-1A; AT4G02930), a RNA recognition motif-containing

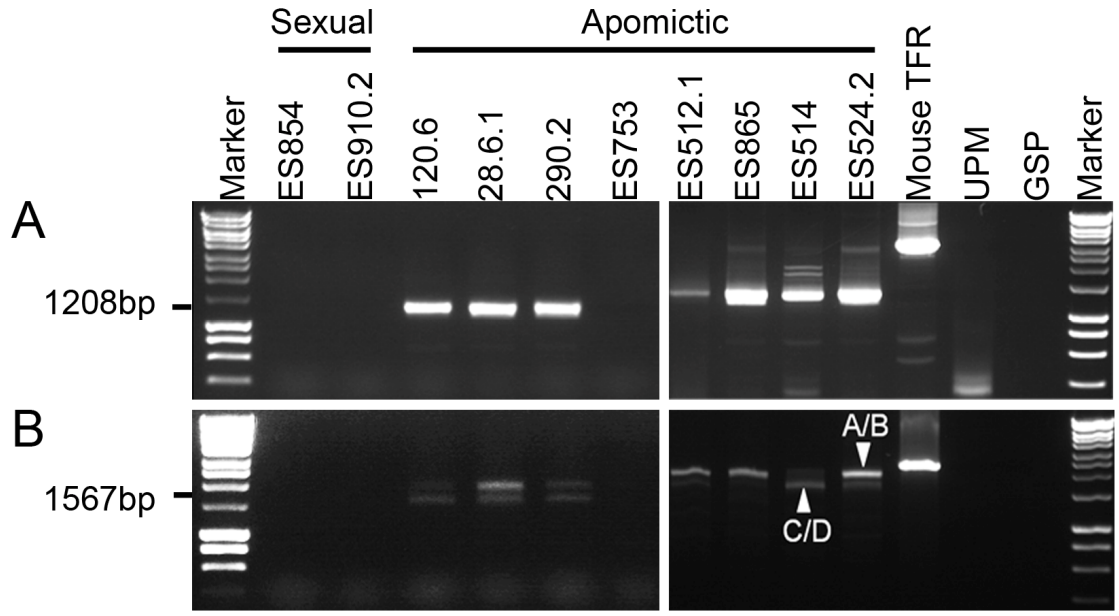

Figure 17. Rapid amplification of cDNA 3`- and 5`-ends of the candidate gene *UPGRADE*. 3`-end (A) and 5`-end (B) RACE cDNA fragments were obtained from sexual and apomictic *Boechera* whole flower cDNA. Transferrin receptor (TFR) primers on cDNA from mouse heart total RNA with optimal cycling parameters were used as positive control. A/B and C/D marks indicate the different putative splicing forms of the candidate transcript.

protein (RNAR; AT5G19960), and the HRD3 protein (HMG-coA Reductase Degradation) which is homologous to components of the yeast HRD1 complex (AT1G18260; Fig. 18A, Table 4, Supplemental Table 11). From these candidates, *AtHRD3*, which plays a central role in endoplasmic reticulum (ER)-associated protein degradation, exhibits the highest similarity to Bsp*UPG,* with fragments of 145 nt (83% identity, p<5.00E-31), 118 nt (87% identity, p<6.00E-30) and 47 nt length (89% identity, p<0.0002).

Bsp*UPG* is characterized by U2-dependent classified introns (5`-GT/3`-AG splicing site; Simpson and Filipowicz (1996.)) which are located towards the 5`-end of the transcript. Although no information about a putative promoter is available, the PlantPAN database (Chang et al., (2008); http://plantpan.mbc.nctu.edu.tw/) was screened for putative regulatory features, such as tandem repeats and transcription factor binding site (TFBS), on a 1.1 kb fragment upstream of Bsp*UPG-2* (position +47809 - 48921 nt on Assembly 2, see chapter *4.4.2*). A single tandem repeat (-726 to -985 nt, 82

nt consensus size, 3.2 copies) and multiple TFBSs were identified (Supplemental Table 12). Furthermore, a single CpG island was detected at the 5`-end of the transcript locus (+140 to 312 nt, C+G >60%; Fig. 18C). The detection of CpG/CpNpG islands, regions of genomic sequences that are rich in the CpG/CpNpG patterns, is important, because such regions are predominantly nonmethylated and tend to be associated with genes which are frequently switched on (Deaton and Bird 2011). In addition, the transcript

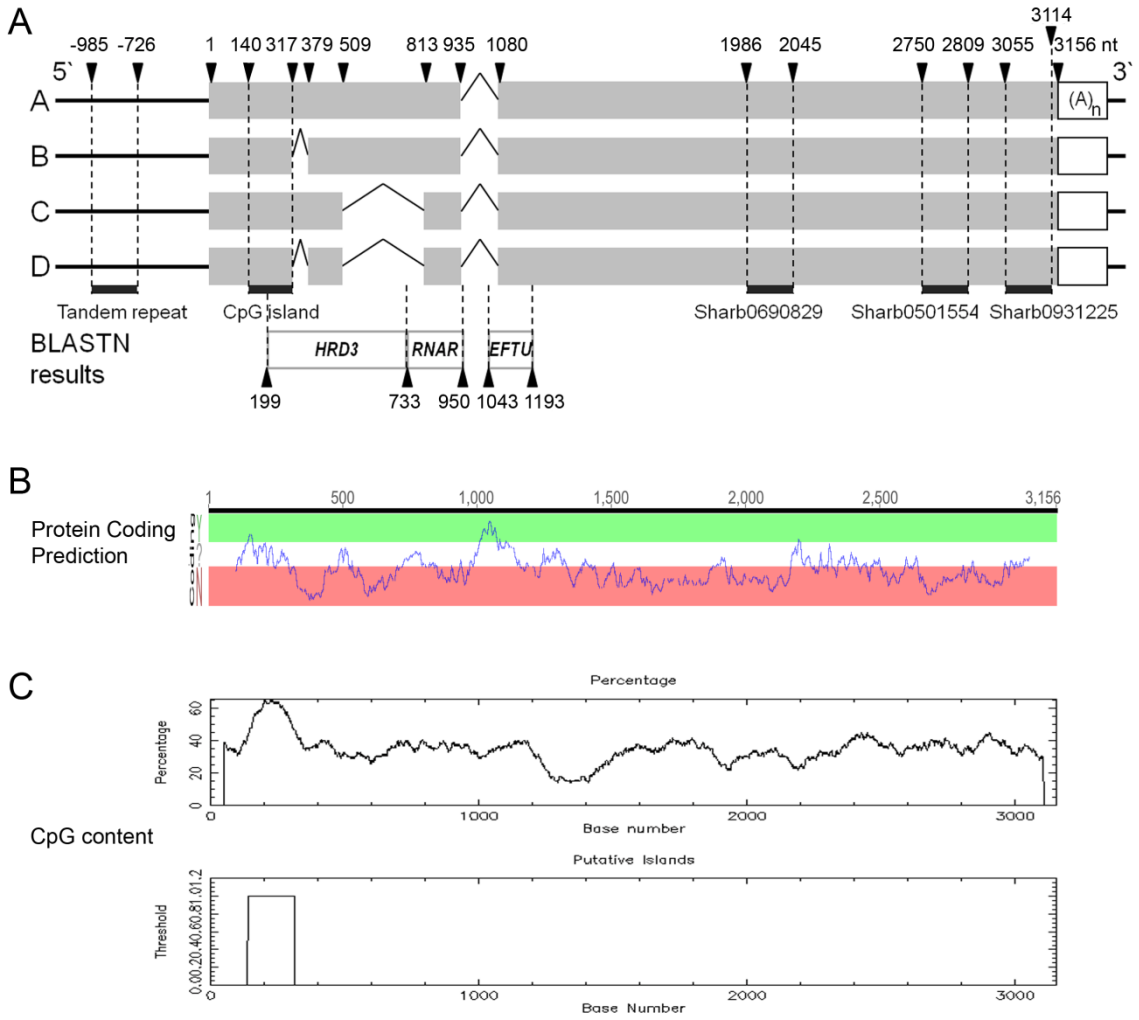

Figure 18. Schematic representation of putative splicing forms and coding potential of Bsp*UPG*.

(A) Exons are represented by grey boxes and numbers indicate intron-exon boundary positions. Promotor motives like tandem repeats were detected with a PlantPAN database search (Chang et al., 2008). (B) Coding potential graph of Bsp*UPG* was retrieved from Geneious Pro 5.3.6 (Biomatters) indicating low coding (red) and high coding potential (green). The coding potential score (cps) was calculated with the Coding Potential Calculator software (Kong et al., 2007). (C) CpG/CpNpG islands are potential promotor motifs which are important sites for DNA methylation and thus crucial for the gene regulation.

carries a poly(A)-tail with some variation in length and initiation position. A putative near upstream element (NUE, AATAAA) of a polyadenylation signal, which is the most common signal in plants (Hunt 1994), was identified 31 nt upstream of the polyadenylation site (Fig. 18, Supplemental Table 10).

Coding Potential Calculation of Bsp*UPG* (for ES 524; Kong et al. (2007)), excluding the constitutive intron 3, revealed a single short open reading frame (ORF) for each of both strands. The sense strand gave an ORF with 196 nt (66 amino acids, pos. 895 to 1090 nt, coding potential score (cps)=-0.0298) and the reverse complement strand revealed a 202 nt sized ORF (68 aa, pos. +1958 to 2159 nt, cps=-1.180). The lack of an ORF length typical for a protein encoding mRNA suggested a long non-coding RNA (lncRNA) product of Bsp*UPG* (Fig. 18B, ORF typically >100 aa; Kondo (2007)).

Apart from other lncRNA types in plants, some lncRNA act as precursors for small non-protein-coding RNAs (ncRNAs or npcRNAs) such as so far unidentified miRNAs or endogenous ta-siRNAs and may regulate translation in *cis* or *trans* (Hirsch et al., 2006; Pikaard et al., 2008). A search for miRNA binding sites of known *Arabidopsis* (www.mirbase.org, default parameter, 1 mismatch) and *Boechera* miRNAs (Amiteye et al., 2011) revealed no mature miRNAs matching the candidate transcript. In default of known miRNAs matching to Bsp*UPG* and due to the obvious lack of a long ORF on Bsp*UPG*, we used a bioinformatical approach to predict other thermodynamic stabile npcRNA structural elements that could function as an RNA molecule from the genomic background sequence Bsp*UPG*.

The minimum free energy (MFE) is a measure of the thermodynamic stability of a structure, but also depends on the length and the base composition of a folded sequence and is, therefore, difficult to interpret in absolute terms (Washietl et al., 2005). To test for an exceptional stability of a candidate npcRNA structure from Bsp*UPG* the thermodynamic stability of potential npcRNA candidates were evaluated using a Z-score, which represents the number of standard deviations that the MFE of a native sequence deviates from the mean MFE of a large number of random sequences of the same length and base composition (see *3.2.14* in Materials and Methods). Negative Z-scores indicate that a sequence is more stable than expected by chance. As npcRNAs vary in length and structure, no single sequence size is expected to be optimal for npcRNA identification. Therefore, variable window sizes between 50 and 300nt were selected and these windows were used to scan the genomic sequence of Bsp*UPG*, starting at the beginning of the sequence and moving each time by the step size (our

analysis used a step size of 10nt). Any windows producing a significant Z-score during the scanning process were considered candidate regions for a structural npcRNA.

As shown in Figure 19, the Bsp*UPG* forward and reverse strand contains eight candidate regions (*i.e.* overlapping windows are combined to determine the total length of a candidate region, Supplemental Tables 13 and 14) that display significant Z-score values (potential secondary structures of windows with most negative Z-score for each candidate region are proposed in Supplemental Figure 2), suggesting that parts of Bsp*UPG* are able to form stable secondary structures. In order to classify these newly generated secondary structures a widely accepted system (Seffens and Digby 1999; Bonnet et al., 2004; Zhang et al., 2006) was used for differentiating sRNA classes which comprises among others, (1) the minimal folding free energy, (2) the adjusted minimal folding free energy (AMFE), (3) the minimal folding free energy index (MFEI), and (4) the A+U content. Interestingely, all detected secondary structures met most of the criteria for miRNA precursors (Supplemental Table 15). Comparable with pri-miRNAs in *Boechera* (Amiteye et al., 2011) and other plant species (Zhang et al., 2006), npcRNA 1 to 8 demonstrate a lower minimal free energy (ranging from -18.27 to -73,72 kcal/mol), have an elevated A+U level (ranging from 56.70 to 70.00%), and have, with exception for npcRNA 5 (0.70), a MFEI greater than 1.07, which is much higher than in tRNAs (0.64), rRNAs (0.59), or random mRNAs (0.65). The candidate regions span 25% and 27% of the total length of each strand of Bsp*UPG* respectively. Only the 70nt-sized npcRNA 5 (position +1090nt and +1160nt on the reverse strand of Bsp*UPG*) overlapped with a genomic region of Bsp*UPG* that is homologous to a known protein-coding gene (*i.e.* EFTU/EF-1A (AT4G02930) maps to genomic region of Bsp*UPG* at position +1043nt and +1193nt). NpcRNA 5 covers approx. 75% of the third exon of EFTU/EF-1A in *Arabidopsis* (pos. +536nt to +606nt on EFTU/EF-1A full length genomic DNA, Supplemental Figure 3). NpcRNAs 1/6, 2/7 and 4/8 cover similar regions on both strands of the genomic sequence. NpcRNA 1, 6 and 7 have significant Z-scores for all tested window sizes.

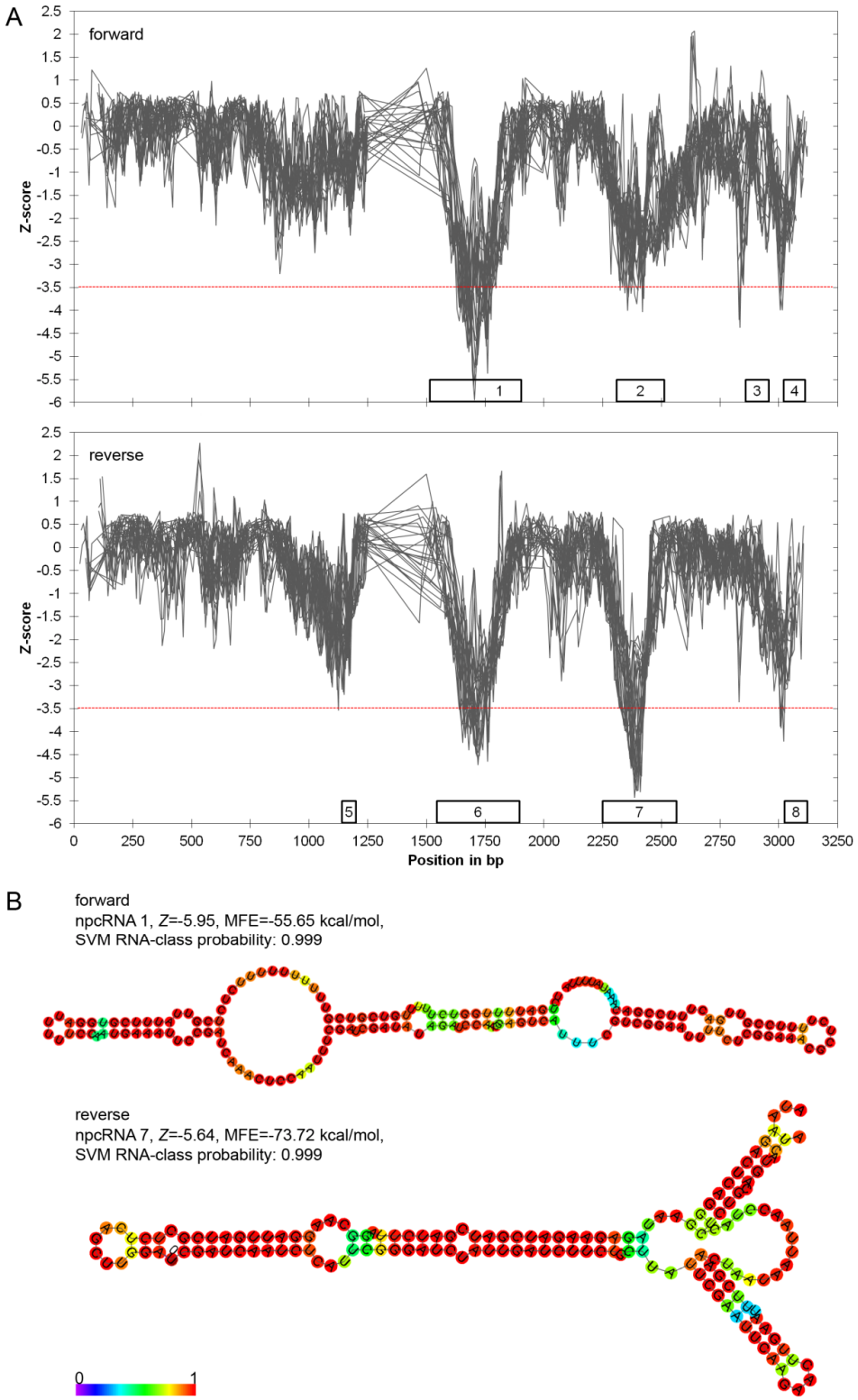

Figure 19. Z-score *versus* position on Bsp*UPG*.

(A) The Z-score for the sliding window with length between 50nt and 300nt (step size=10) is plotted *vs.* position. Any window producing a significant Z-score during the scanning process was considered candidate region for a structural npcRNA. In case that multiple, overlapping windows of several lengths produced significant Z-scores, the region encompassed by all the overlapping windows defined the candidate region (represented as black boxes). Numbers in black boxes are putative npcRNAs having most negative Z-score (Supplemental Tables 13, 14 and 15, Supplemental Figure 2). (B) Proposed

minimum free energy structures (MFE) of putative npcRNAs with most negative Z-scores detected for the forward and reverse strand of Bsp*UPG*. Negative MFEs and Z-scores indicate that a sequence is more stable than expected by chance. The structures above are colored by base-pairing probabilities from zero (violet, see legend) to hundred percent (red). For unpaired regions the color denotes the probability of being unpaired.

4.4.2 Detection of a duplicated variant of UPGRADE

Although the entire genomes of *B. stricta* and *B. divaricarpa* are being sequenced (DOE Joint Genome Institute; www.jgi.doe.gov) and comparative genomic analysis revealed partial genome information (Windsor et al., 2006), while partial genome information (Windsor et al., 2006) and a genetic linkage map (Schranz et al., 2007) are available, thus far no complete physical map of *Boechera* exists. Hence the novel Bsp*UPG* was localized by screening a *B. divaricarpa* bacterial artificial chromosome (BAC) library. Although the qRT-PCR validation was not successful due to technical reasons, probe Sharb1199059 was included due to its pollen-specific molecular function, in addition to the microarray probes mapping onto Bsp*UPG* and probe Sharb0425060, for screening of the *Boechera* BAC library. Chromosome walking products of the five candidate microarray probes hybridized to 80 *Boechera* BAC clones (Supplemental Figure 4). Colony-PCR using target-specific primers from the chromosome walking approach (see method section *3.1.3 Boechera* BAC library screen) led to the identification of twelve *Boechera* BAC clones which hybridized with one or several probes simultaneously (Table 3). None of the clones was positive for probe Sharb1199059, whereas three single hits for probe Sharb0425060 and nine triple hits for probes Sharb0931225, Sharb0501554 and Sharb0690829 were detected (Table 3). Restriction digests of the twelve BACs suggested partial overlap of their DNA inserts (Supplemental Figure 5). Sanger sequencing was thus performed on BAC clones A4O22, E7K5, C8B11 and F8G11, each being positive for three of five microarray probes (Sharb931225, Sharb501554 and Sharb690829). The C8B11 BAC sequence contig (Assembly 1, Supplemental Table 16), which could not be aligned together with the sequence contigs of the other three BAC clones, reached 57 458 bp length with 33.5% average GC content increasing to 40.1% in genic regions. BAC sequence contigs from clones A4O22, F8G11 and E7K5 overlapped and were assembled into the 58 769 bp-sized Assembly 2 (Supplemental Table 17) with 32.9% average GC content increasing to 42.1% in genic regions. Annotation of Assembly 1 identified one transposon-related gene, five protein-encoding genes and two fragments of protein-encoding genes, all of

Table 3. Overview of BAC clones carrying one to three candidate microarray probes. DNA contig lengths are given for BAC clones which were selected for Sanger sequencing.

No.	BAC ID	Hybridization signal intensity[a]	Microarray probe[b]					BAC contig length (bp)
			1	2	3	4	5	
1	D1L12	very weak	-	-	-	-	+	-
2	E3C19	very strong	-	+	+	+	-	-
3	A4O22	weak	-	+	+	+	-	55 616[d]
4	B4G11	very strong	-	+	+	+	-	-
5	B5B16	weak	-	-	-	-	+	-
6	E5O18	weak	-	-	-	-	+	-
7	E6A11	weak	-	+	+	+	-	-
8	E7K5	very strong	-	+	+	+	-	54 682 + 3 998
9	A8D8	strong	-	+	+	+	-	-
10	C8B11	weak	-	+	+	+	-	57 458 + 21 311[c]
11	D8G22	strong	-	+	+	+	-	-
12	F8G11	weak	-	+	+	+	-	55 616[d]

[a] Subjective estimation based upon observation of high-density hybridization membrane after 19 hrs hybridization period.

[b] 1 - Sharb1199059; 2 - Sharb0931225; 3 - Sharb0501554; 4 - Sharb0690829; 5 - Sharb0425060 (see Table 2). Present (plus) and not present (minus) microarray probes were detected *via* colony PCR on *Boechera* BAC clones.

[c] 21 kb-sized fragment shows no similarity with any other BAC clone sequences.

[d] 100% identical.

which (except for TER4) were homologous to genes located on *Arabidopsis* chromosome At1 (Figs. 20A and B, Table 4). In contrast, Assembly 2 contained a higher level of transposon-related genes (3) and gene fragments (9) whose homologs were found on all *Arabidopsis* chromosomes except for At3 (Figs. 20A and C, Table 4). Comparison of both assemblies showed that both fragments share several highly homologous sequences; *e.g.* for the two protein-encoding genes *MtN21* and *RRP4*, for a fragment of the protein-encoding gene *MBOAT*, and for the 3`-end of Bsp*UPG*. Allowing for rearrangements, Assembly 1 and Assembly 2 aligned along orthologous regions covering 27.15 kb of Assembly 1 and 46.20 kb of Assembly 2 (Fig. 20A). Synteny gaps in Assembly 2 contained transposon-related genes (*e.g. TER1* and *TER2*) and insertions composed of short protein-encoding gene fragments flanked by inverted repeat (IR) sequences (*e.g. IR 7*; Figs. 20B and C, Supplemental Table 18), suggesting

that parts of Assembly 2 arose *via* partial duplications of Assembly 1. Hence, considering the partial presence of Bsp*UPG* at both loci, we subsequently labelled the original locus in Assembly 1 Bsp*UPG-1*, and the duplicated variant on Assembly 2 Bsp*UPG-2* (Fig. 20).

Table 4. Gene annotation of *Boechera* BAC clone Assembly 1 and Assembly 2.

Symbol	Description	Best Hit[b]	Accession ID	Position[d]	Expect	Strand	Algorithm	Assembly
RRP4	Exosome complex component RRP4	Ath	AT1G03360	3163 - 4068	8.00E-136	+	BLASTN	1
MBOAT	Membrane bound O-acyl transferase-like protein	Ath	AT1G57600	6339 - 6460	9.00E-40	+	BLASTN	1
MtN21	Nodulin MtN21 /EamA-like transporter protein	Ath	AT1G43650	18280 - 22394	1.00E-95	-	BLASTN	1
Bsp*UPG-1*	Unreduced pollen grain development (original locus)	Boe	n/a	23071 - 25374	0.00	+	MAUVE	1
TER4	non-LTR retroelement reverse transcriptase-like protein	Ath (BAB08714)	AT5G35540[c]	30148 - 31596	6.00E-133	+	BLASTX	1
TLP5	Tubby-like F-box protein 5	Ath	AT1G43640	42429 - 45148	0.00	-	BLASTN	1
UP1	Uncharacterized protein	Ath	AT1G43630	46255 - 47862	0.00	+	BLASTN	1
TIR	TIR-NBS class of disease resistance protein	Ath	AT1G66090	50446 - 50976	2.00E-67	+	BLASTN	1
UGT	Sterol 3beta-glucosyltransferase	Ath	AT1G43620	53271 - 57458	3.00E-121	-	BLASTN	1
TER5[a]	Putative LTR retroelement polyprotein	Ath (BAB10790.1)	AT5G34980	12427 - 15165	0.00	-	BLASTX	1 (fragment)
NPC1	Niemann-Pick C1 protein	Ath	AT1G42470	3634 - 3790	3.00E-34	+	BLASTN	2
TER3[a]	Putative LTR retroelement polyprotein	Ath (AAG10812.1)	AT1G34967	4915 - 6534	0.00	+	BLASTX	2
RRP4	Exosome complex component RRP4	Ath	AT1G03360	4335 - 12075	1.00E-133	+	BLASTN	2
TER1[a]	Hypothetical transposable element-related protein	Vvi (AM456232.2)	AT4G03810	13249 - 17298	0.00	-	BLASTX	2
MBOAT	Membrane bound O-acyl transferase-like protein	Ath	AT1G57600	20588 - 20707	7.00E-35	+	BLASTN	2
MtN21	Nodulin MtN21 /EamA-like transporter protein	Ath	AT1G43650	31590 - 35114	0.00	-	BLASTN	2
TPR	Tetratricopeptide repeat domain-containing protein	Ath	AT5G02590	39831 - 39928	7.00E-16	+	BLASTN	2
DY2A	Dynamin-2A	Ath	AT1G10290	39930 - 40093	8.00E-28	-	BLASTN	2
GRV2	DNAJ heat shock N-terminal domain-containing protein	Ath	AT2G26890	40102 - 40468	9.00E-78	+	BLASTN	2
TER2	Putative TNP2-like transposon protein	Ath (AAD20646.1)	At2G13000	42155 - 44066	0.00	-	BLASTX	2
HRD3	HRD3 like protein	Ath	AT1G18260	49324 - 49651	1.00E-31	-	BLASTN	2
RNAR	RNA recognition motif-containing protein	Ath	AT5G19960	49652 - 49868	9.00E-34	+	BLASTN	2
EFTU	GTP binding Elongation factor Tu/EF-1A family protein	Ath	AT4G02930	49961 - 50111	9.00E-40	-	BLASTN	2
Bsp*UPG-2*	Unreduced pollen grain development (duplicated locus)	Boe	n/a	48921 - 52073	0.00	+	MAUVE	2

[a] LTR retroelement function confirmed with LTR FINDER software (Xu and Wang 2007).

[b] Plant species: Ath, *Arabidopsis thaliana*; Boe, *Boechera species*; Vvi, *Vitis vinifera*.

[c] *Arabidopsis* locus identifier refers to neighbouring gene MOK9.17; for TER4 = MOK9.16 no locus identifier is available.

[d] Position of annotated gene on designated assembly.

n/a, not applicable.

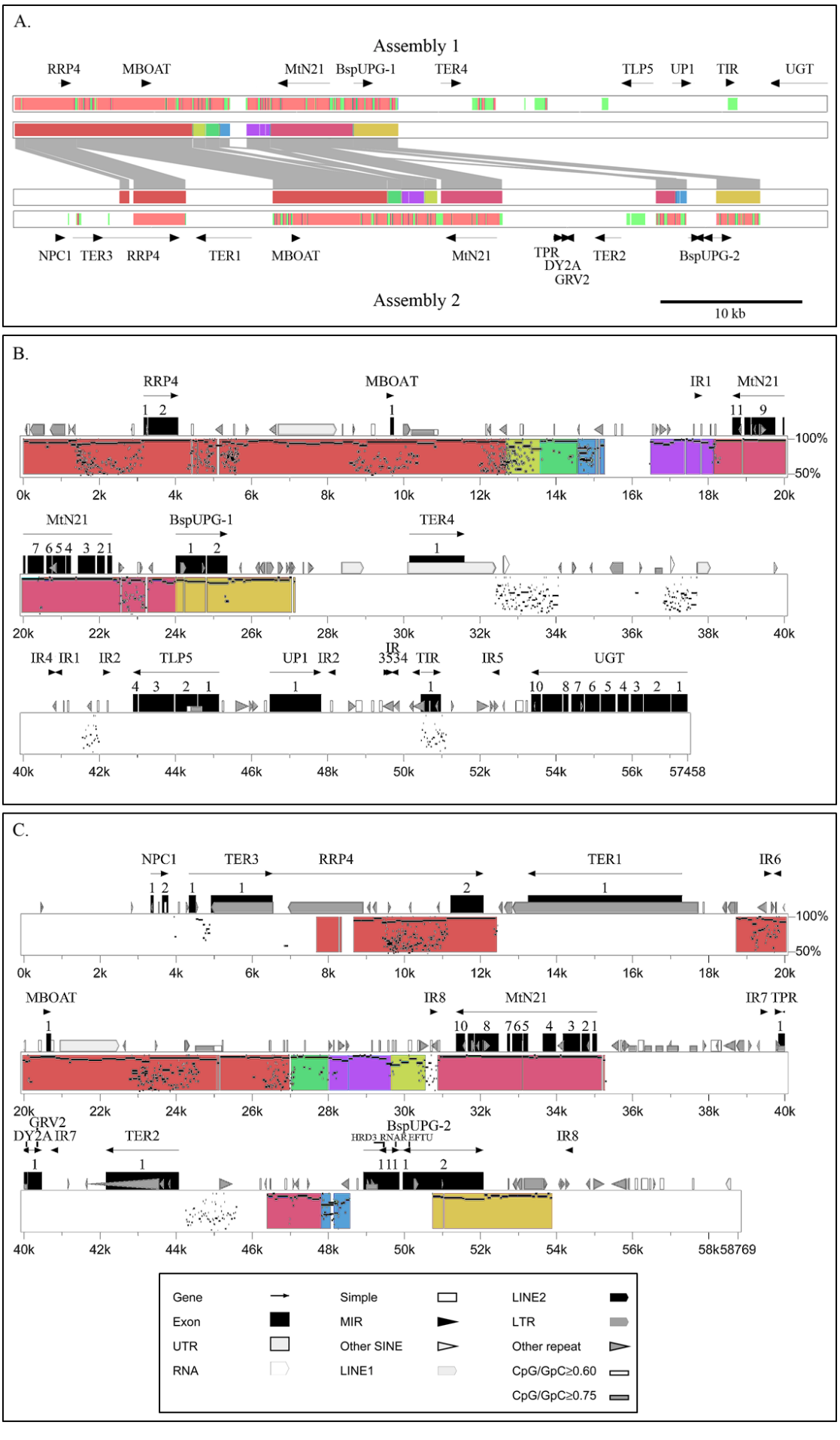
A.
Assembly 1
RRP4
MBOAT
MtN21
BspUPG-1
TER4
TLP5
UP1
TIR
UGT
NPC1
TER3
RRP4
TER1
MBOAT
MtN21
TPR
DY2A
GRV2
TER2
BspUPG-2
Assembly 2
10 kb
B.
RRP4
MBOAT
IR1
MtN21
100%
50%
MtN21
BspUPG-1
TER4
IR4
IR1
IR2
TLP5
UP1
IR2
IR
3534
TIR
IR5
UGT
C.
NPC1
TER3
RRP4
TER1
IR6
MBOAT
IR8
MtN21
IR7
TPR
GRV2
DY2A
IR7
TER2
BspUPG-2
IR8
Gene
Exon
UTR
RNA
Simple
MIR
Other SINE
LINE1
LINE2
LTR
Other repeat
CpG/GpC≥0.60
CpG/GpC≥0.75

Figure 20. Percent identity plots showing the original *UPGRADE* locus and its duplicated variant.

(A) Percent identity plot (pip) showing the BAC clone Assembly 1 region in the *B. stricta* linkage group 1 (BstLG1) block C_1, which is associated with the original locus of Bsp*UPG* aligned with BAC clone Assembly 2 containing the duplicated locus Bsp*UPG-2*. Green bars represent all regions within an alignment with at least 50% nucleotide identity and red bars are those regions that align at a high level of similarity (at least 100 bp without a gap and with at least 70% nucleotide identity). (B) and (C) illustrate the alignments with their associated gene, repetitive element and promotor motif annotations on a pip. The aligning segments are drawn according to their percent identity, which is shown on the vertical axis from 50% to 100%. Arrows indicate the motif orientation on the assemblies. Each aligning segment on both assemblies is displayed as a series of horizontal lines whose positions correspond to the second assembly. The colored sequence stretches in (B) and (C) specify rearranged sequence blocks between both assemblies and correspond to colors given in (A). Genes were annotated according to BLASTN and BLASTX search in GenBank. Annotation of transposable elements were retrieved from the Repbase databank (Jurka 1998), simple repeats from RepeatMasker (Smit et al., 1996-2004), inverted repeats (IR) from einvert software and EMBOSS (Rice et al., 2000) and pip with positions of CpG islands were provided with the Pipmaker software (Schwartz et al., 2003).

Bsp*UPG-1* is flanked by four genes centered in the centromeric region on the lower arm of *Arabidopsis* chromosome At1 (Figs. 20 and 21, Table 4; ~ 0.57 megabases distance to T3P8-sp6, Hosouchi et al., (2002)). Interestingly, two genetic markers (Bst006701 and BSTES0032) from the synthetic F_2 linkage map of sexual *B. stricta* (Schranz et al., 2007), homologous to *Arabidopsis* locus identifier At1g51310 and At1g43245, flank Bsp*UPG-1*. Both markers span the interval of genomic block C_1, which localizes Bsp*UPG-1* on the BstLG1 linkage group of *B. stricta* for which very low levels of recombination were detected (Schranz et al., 2007). The putative location of the duplicated Bsp*UPG-2* is unknown.

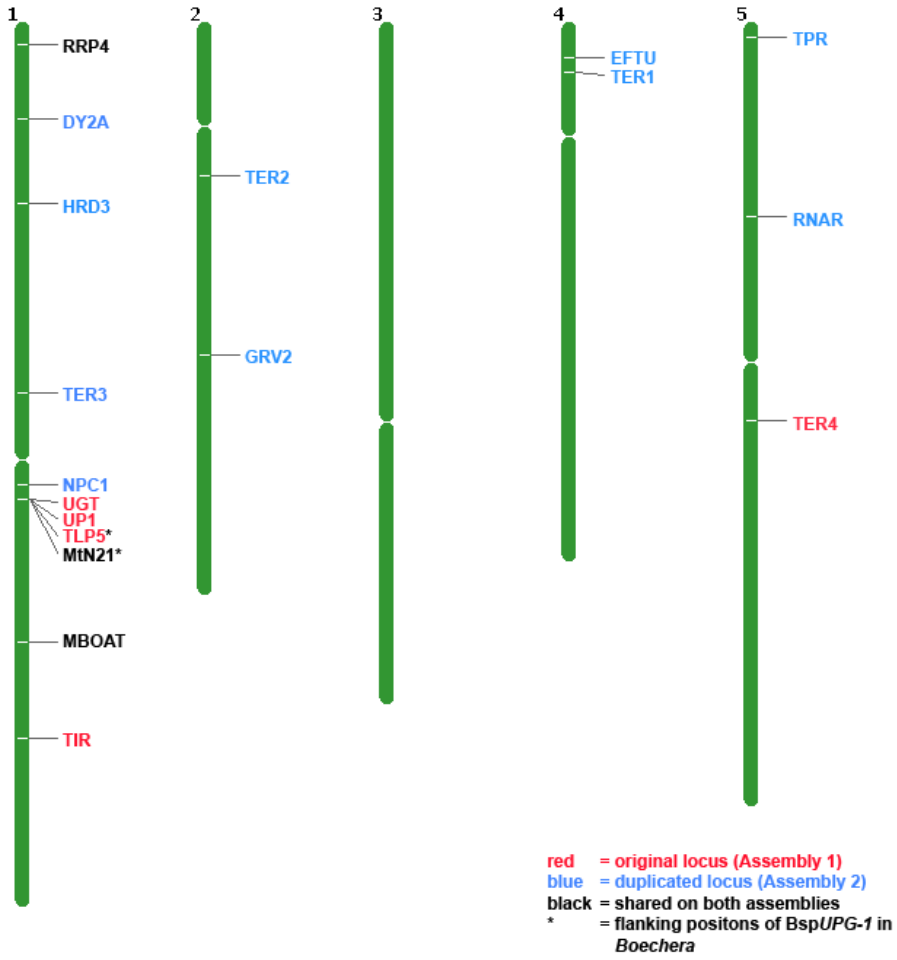

Figure 21. Distribution of homologous loci surrounding Bsp*UPG-1* and Bsp*UPG-2* along the *Arabidopsis* genome.

A distribution map of homologous loci corresponding to (red) the original locus (Assembly 1), (blue) the duplicated locus (Assembly 2) and (black) to both loci in *Boechera*, along *Arabidopsis* chromosomes were generated by the chromosome map tool of TAIR (The Arabidopsis Information Resource). Asterisks denote flanking loci of Bsp*UPG-1*. *Arabidopsis* loci AT1G43245 and AT1G51310 refer to boundaries of genomic block C_1 on BstLG1 in the synthetic F_2 genetic map of *B. stricta* (Schranz et al., 2007).

4.4.3 Distribution of transposable elements on the original and the duplicated locus

Transposable elements (TEs) are mobile, repetitive DNA sequences that constitute a large component of eukaryotic nuclear genomes, ranging from ~21% (*A. thaliana* genome; Ahmed et al. (2011)) to ~88% (wheat; Choulet et al. (2010)) of the total genome. Since TEs are considered to drive genome rearrangements, including the generation of inversions, translocations, deletions and duplications, we screened Assembly 1 and Assembly 2 combining datasets from Repbase (Jurka 1998) and Repeatmasker (http://repeatmasker.org) for Viridiplantae TEs in search of hallmarks for the duplication of the Bsp*UPG* locus.

In total, 100 and 113 TEs mapped onto Assembly 1 and Assembly 2 respectively, many of them are simple repeats with low sequence complexity. Excluding the simple repeats left 68 and 73 TEs which mapped onto Assembly 1 and Assembly 2 respectively, with about half of them identified in *Arabidopsis* (27 and 35 respectively). The major repeat families are *copia*-like (18 and 16) and *gypsy*-like (8 and 9) LTR retrotransposons, followed by DNA transposons *En-Spm* (9 and 12), *MuDR* (11 and 7), *hAT* (7 and 7) and *Helitrons* (6 and 9). Both Bsp*UPG-1* and the Bsp*UPG-2* are targeted by four TE sequence sites (Fig. 22; Supplemental Tables 19 and 20). Proportions of the TE superfamilies do not vary statistically significant between Assembly 1 and Assembly 2. In contrast, proportions of both loci compared to the distribution across the whole *Arabidopsis* genome partly differ. *Copia*-like LTR retrotransposons are the largest proportions among both *Arabidopsis* TEs (57.70%) and those on the *Boechera* BAC clone assemblies (26.47% and 21.92% respectively). Nonetheless, *copia*-like LTR retrotransposons are more prominent across the whole *Arabidopsis* genome compared to both *Boechera* assemblies (one-tailed Fisher's exact test, LTR/*Copia*: p(Assembly1)=1.83E-07, p(Assembly2)=5.23E-10), in contrast to DNA transposons of the *En-Spm* and *hAT* class and non-LTR retrotransposons of the LINE/L1 class, which are more prominent on both *Boechera* assemblies (one-tailed Fisher's exact test, DNA/*En-Spm*: p(Assembly1)=4.21E-10, p(Assembly2)=3.83E-14, DNA/*hAT*: p(Assembly1)=2.00E-06, p(Assembly2)=3.24E-06, LINE/L1: p(Assembly1)=0.001, p(Assembly2)=0.008, Fig. 22, Supplemental Tables 19 and 20). Interestingely, Fiston-Lavier et al. (2012) observed a greater number of *gypsy*-like LTR retrotransposons and DNA/*En-Spm* transposons in heterochromatic compared to euchromatic regions of *Arabidopsis*, and observation which is congruent with the detection of Bsp*UPG-1* on Assembly 1 in a heterochromatic region of *Boechera*. Although the position of the

duplicated Bsp*UPG-2* in *Boechera* is unknown, the similar prominent occurrence of the DNA/*En-Spm* TE class in both assemblies might point to a heterochromatin insertion position of the candidate transcript, similar to the position of the original locus (Assembly1).

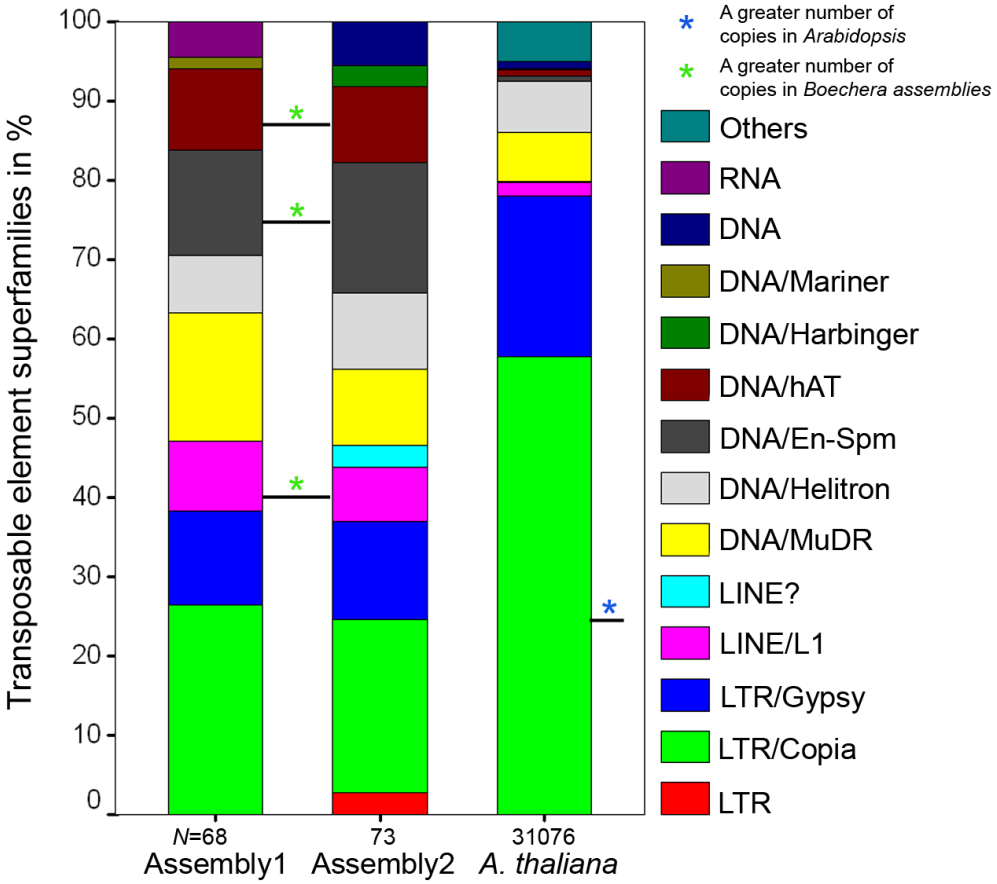

Figure 22. Distribution of the TE superfamilies along the original and duplicated *UPGRADE* locus.

TE superfamily distributions along the original and duplicated *UPGRADE* locus in *Boechera* were examined in comparison with the distribution of TE superfamilies along the complete *A. thaliana* genome.

4.5 The genesis of Bsp*UPG-2*

4.5.1 Sequence divergence of UPGRADE *in sexual and apomictic* Boechera

Although no within-individual (*i.e.* allelic) polymorphisms were identified, Bsp*UPG* is characterized by strong sequence divergence between apomictic and sexual genotypes (63.1% ± 0.26 similarity; Fig. 23A; Supplemental Table 21). Sequence conservation is high between apomicts, with highest nucleotide divergence between genotypes ES 753 (transcriptional outlier) and ES 514 (0.67% ± 0.13), while all other apomicts share complete sequence identity (Fig. 22A). Sequence similarity is also high in sexuals, both

within *B. stricta* (99.69% ± 0.05) and between different species (98.87% ± 0.34). Sequence divergence between apomicts and sexuals is accentuated by four large structural variants (SV>1kb; The-1000-Genomes-Project-Consortium (2010)) based on markedly increased numbers of nucleotide substitutions (39.37% ± 0.17 similarity). Phylogenetic analysis indicates that all apomictic sequences are monophyletic, whereas sexual sequences can be split into two subgroups, one representing *B. stricta* genotypes and the second representing the remaining sexual genotypes (Fig. 23A, Supplemental Table 22).

Together, the lack of intra-individual allelic variation and the absence of a Bsp*UPG-2* at any other locus in both sexuals and apomicts point to a homo- or hemizygous state for Bsp*UPG-2* in apomicts.

Thirty-five indels were identified between Bsp*UPG* in different genotypes (indels, <50 nucleotides, Albers et al. (2011)), 20 of which were located in the highly conserved 3`-end (+1821 to 3153 nt, 82.75% ± 0.86 similarity; Fig. 23A and Supplemental Table 23). Among these 20 indels a single 27 nt sequence was found exclusively in all sexual genotypes (Figs. 23A and B, indel9). Interestingly, Chellappan et al. (2010) identified a new 27-nt small RNA-species that is associated with AGO4 to regulate gene expression at the transcriptional level by directing DNA methylation to some of their target loci *in trans*. Hence, the sex-specific indel9 was tested as a small RNA binding site by northern blot screening of small RNAs from sexual and apomictic pooled flower buds using specific *sense* and *antisense* probes covering the 27 nt site and additional 3`- and 5`-bases (*i.e.* indel9long probes). Expression of the highly conserved plant miRNA167, which was used as positive control, was observed in all flower tissues, whereas no signal was detected for any indel9-specific probe in both apomictic and sexual flower tissues (Fig. 23B).

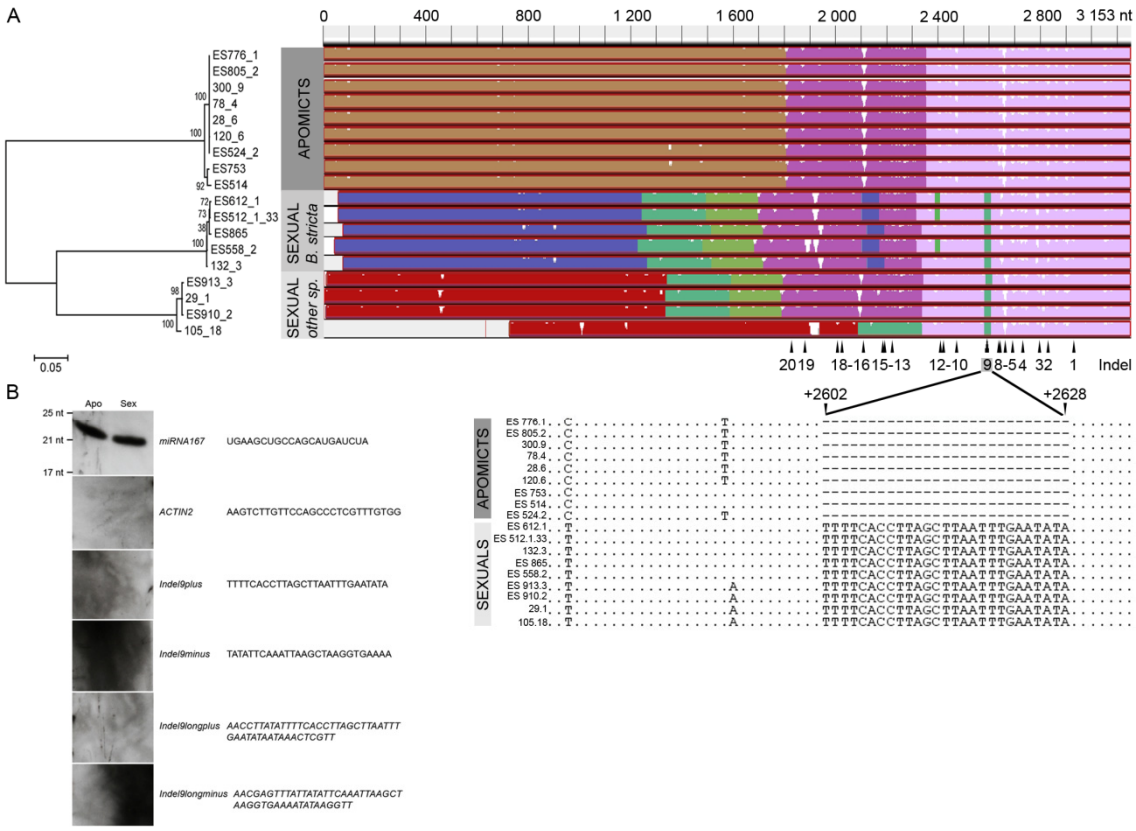

Figure 23. Phylogenetic consensus tree and schematic representation of a CLUSTALW alignment of *UPGRADE* in sexual and apomictic *Boechera*.

Sequences from end-to-end sequenced and cross-match aligned full-length Bsp*UPG-2* from nine diploid sexual and apomictic *Boechera* respectively, were compared. (A) The Maximum Likelihood method using Tamura and Nei (1993) was used to calculate the unrooted phylogenetic consensus tree showing the relatedness of the complete candidate gene between *Boechera* genotypes. The tree with the highest log likelihood (-7579.008) is shown and is drawn to scale, with branch lengths measured in the number of substitutions per site. Cross-match alignment of full-length Bsp*UPG-2* in sexual and apomictic *Boechera*, adapted and modified from the Progressive Mauve algorithm, exhibits locally collinear blocks (LCBs, coded in different colors) between the candidate genes in various genotypes. CLUSTALW multiple pairwise nucleotide sequence alignment of the full-length Bsp*UPG-2* in apomicts and sexuals shows distribution of SNPs and indels between single genotypes compared to a common consensus sequence. (B) Northern blot analysis for putative small RNAs homologous to indel9 at Bsp*UPG-2*. Pooled flower tissues from different developmental stages for each four pooled sexual and four pooled apomictic *Boechera* genotypes were used. The small RNA miRNA167 was used as positive control and *ACTIN2* as negative control.

*4.5.2 A model of Bsp*UPG-2 *genesis*

The structural variants described in chapter *4.5.1* were elucidated using multiple alignments of sexual and apomictic Bsp*UPG* copies with Assembly 1 and Assembly 2 in the presence of rearrangements. Surprisingly, the 3`-end of all gene copies from sexuals and apomicts, here named locally collinear block 1 (LCB1, see definition in methods), shares high similarity with both assemblies, whereas the 5`-ends of apomictic gene copies, here LCB2, were only identified on the duplicated locus of Assembly 2 (Fig. 24A; Supplemental Figure 6 and Supplemental Table 24). Unlike apomictic gene copies (now Bsp*UPG-2*, see nomenclature in chapter *4.4.2*), the copies from sexuals (now Bsp*UPG-1*) share high sequence similarity for the 3`- and 5`-end with the original locus on Assembly 1, but exhibit a mosaic distribution of their LCBs on the duplicated locus of Assembly 2 (Fig. 24B, Supplemental Figure 6). LCB4 of Bsp*UPG-1* in sexual *B. stricta*, which is partially homologous to LCB3 in other sexuals, is translocated 1859 nt upstream from the transcript start of Bsp*UPG-2* in apomictic genomes (Fig. 24B; Supplemental Figure 6 and Supplemental Table 24). LCB5 from sexual *B. stricta* is translocated 10189 nt further downstream of Bsp*UPG-1*, but translocated and additionally inverted 19276 nt further downstream of Bsp*UPG-2* at the duplicated locus (Supplemental Figure 6).

The transcriptional functionality of Bsp*UPG-1* and Bsp*UPG-2* was inferred by mapping sexual and apomictic cDNA (Sharbel et al., 2009) independently onto both, and revealed that only the apo-specific LCB2 is transcribed in both sexuals and apomicts (Fig. 24C). In contrast, apomictic cDNA map to LCB1 and LCB2 including the reads with homology to the candidate microarray 60-mer oligonucleotide probes (Fig. 24C). No cDNA maps to any of the remaining LCBs. PCR on genomic DNA using specific primers for both loci demonstrates the presence of Bsp*UPG-1* in sexuals and apomicts, but absence using cDNA as a PCR template. In contrast genomic copies and transcripts of Bsp*UPG-2* were solely identified in apomicts (Fig. 24E).

Two lines of evidence were used to validate the proposed fusion of the 5`- (LCB1) and 3`-ends (LCB2) of the transcriptionally active Bsp*UPG-2*. A BLASTN search between Bsp*UPG-2* and the complete genomic Sequence Read Archive (SRA) of a sexual *B. stricta* identified a gap at the juncture between all reads mapping to LCB1 and to LCB2, while all reads per LCB overlapped, suggesting their separate genomic origins in the sexual genotype (Fig. 24B, position +1820 nt). We confirmed this result by amplifying a 478 bp-fragment in sexual genotypes using primers corresponding to a fragment of the

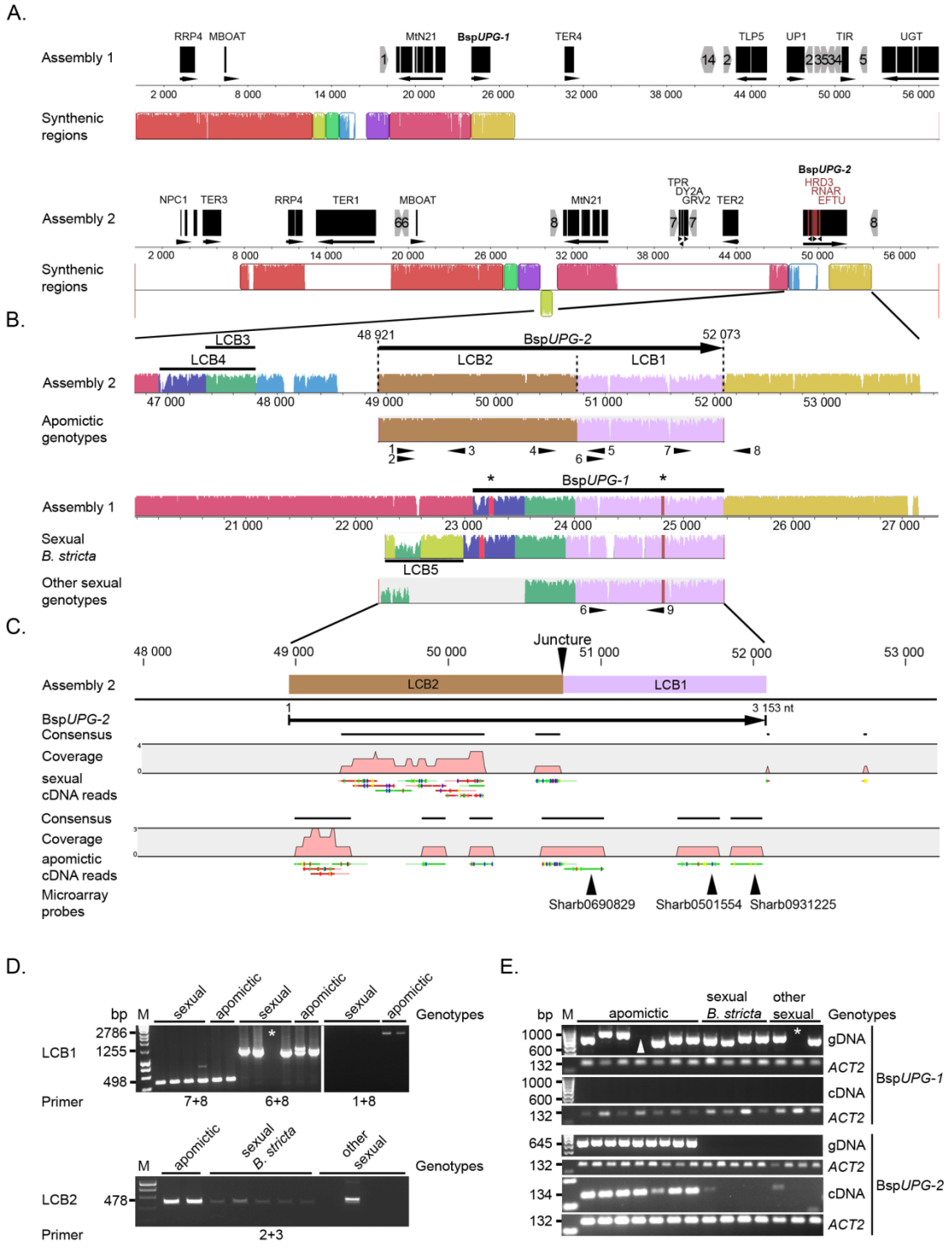

Figure 24. Structure of original and duplicated *UPGRADE* locus.

(A) Schematic representation of the gene annotation of Assembly 1 containing the original Bsp*UPG-1* locus and Assembly 2 containing the duplicated apo-specific Bsp*UPG-2* locus. Synthenic regions of the two assemblies are shown below. Inverted repeats are denoted by grey numbered arrowheads. Insertions of genic fragments are marked in red letters. (B) Mapping of genomic sequences of Bsp*UPG-1* in apomictic (*e.g. B. divaricarpa*, ES 514) and sexual genotypes (*e.g. B. stricta*, ES 612.1) onto both BAC clone assemblies identified rearrangements of locally collinear blocks (LCBs). Black asterisks denoted indels for the sex-specific identity of LCB1 and LCB4. (C) Separate mapping of cDNA reads from sexual and

apomictic genotypes onto Bsp*UPG-2* displays different distribution. Both, sense-oriented (green) and antisense-oriented (red) cDNA reads are displayed. Coloured vertical bars on cDNA reads show SNPs in comparison with genomic DNA of Bsp*UPG-2*. (D) PCRs with independent primer pairs one located on the LCB1 and one on cDNA mapping positions of LCB2 in sexual and apomictic genomes illustrate that LCB1 is highly conserved in both sexuals and apomicts and that LCB2 is also present in sexual genotypes, but at a different genomic position. (E) PCR with primers combining LCB1 and LCB2 shows apo-specific presence of Bsp*UPG-2*, whereas Bsp*UPG-1* is present in both, sexual and apomictic genomes. The white arrow marks a faint band for genotype ES 753 and white asterisk marks missing band for genotype 105.18 which lacks the priming site for primer 6. Black arrows mark PCR primers: 1 - CON234X2L, 2 - CON234X14L, 3 - CON234X10R, 4 - PC1pol1L, 5 - PC1pol1R, 6 - GSP4, 7 - TSP33R, 8 - CON234X5R, 9 - Indel9minus. Refer to Table 4 for gene annotations.

Bsp*UPG-2*-specific LCB2. Additionally, this fragment is mapped by cDNA in sexual and apomictic *Boechera* belonging to the *Boechera* homologue of At*HRD3* (Bsp*HRD3*; Figs. 24A and C, 25A, Table 4).

We tested whether LCB2 derived directly from parental transcripts or from putative duplicated variants by performing RACE-PCR and chromosome walking on Bsp*HRD3* (Fig. 25A; Supplemental Table 11). RACE-PCR amplified Bsp*HRD3* in sexual and apomictic genotypes with high overall sequence similarity to the *AtHRD3* transcript (93.0%), whereas chromosome walking yielded a sequence with only partial similarity (63.5%) to the genomic *AtHRD3* copy (Fig. 25A). To focus on the predicted insertion region of Bsp*HRD3* in Bsp*UPG-2*, a fragment of the 5`-end genomic copy of Bsp*HRD3* extracted from a sexual genome was shown to contain insertions which were not present in the cDNA of Bsp*HRD3* or the genomic copy of *AtHRD3*. Interestingly, comparing the chromosome walking sequence of Bsp*HRD3* with Bsp*UPG-2*, a 550-bp sized fragment was identified as being highly homologous to Bsp*UPG-2,* including these insertions (Fig. 25A; 93.6% similarity, blue boxes denote insertions).

A partial gene duplication as is proposed for Bsp*UPG-1* and Bsp*UPG-2*, should be strongly correlated with insertions and pseudogenization (Ohno 1970; Kaessmann 2010). Hence, we tested Bsp*UPG-2* for duplication by mapping a set of sequence tags from an array-based comparative genome hybridization (aCGH) experiment in 10 sexual and 10 apomictic *Boechera* (Aliyu et al., unpublished results) against Bsp*UPG-2* (Fig. 25B). In total, 79 array probes from the CGH experiment mapped to Bsp*UPG-2*. Whereby most probes mapping onto Bsp*UPG-2* show no copy number variation (CNV)

in apomicts (96.0% with the remaining 4.0% showing depletion), 68% of the tags were depleted in at least one sexual genotype. Approximately half of the depleted sequence tags in sexual genotypes were situated at the 3`-end of Bsp*UPG-2,* representing the start of LCB1. The remaining depleted sequence tags were distributed towards the 5`-end of Bsp*UPG-2* between +0 to +1081 nt (Fig. 25B).

To summarize, both LCB1 and the extreme 5`-end of Bsp*UPG-2* is present in fewer copies in sexual compared to apomictic genomes, whereas the middle part of Bsp*UPG-2* shows no variation in copy number in either reproductive mode. The duplication of LCB1 in apomictic genomes is explained by its presence in both assemblies. The duplication of the 5`-end in apomictic genomes is evidenced by the presence of duplicated source gene insertions covering almost exactly the region represented by depleted sequence tags in sexuals (HRD3: +199 to + 731 nt, RNAR: +732 to 948 nt, EFTU/EF-1A: +1041 to 1191 nt; Figs. 24A and 25A).

Together these results (Figs. 24 and 25) suggested that Bsp*UPG-1* fragments were duplicated from the original locus on Assembly 1 to form the basis of Bsp*UPG-2*, which subsequently underwent sequential insertions of genome fragments derived from at least two unlinked genomic regions. Thereby, LCB1 was inserted between LCB2 and LCB4. The newly formed locus was then the insertion target for the duplicated variant of at least one functional gene (*i.e.* Bsp*HRD3*), with subsequent exonization leading to a gain of Bsp*UPG-2* transciptional activity.

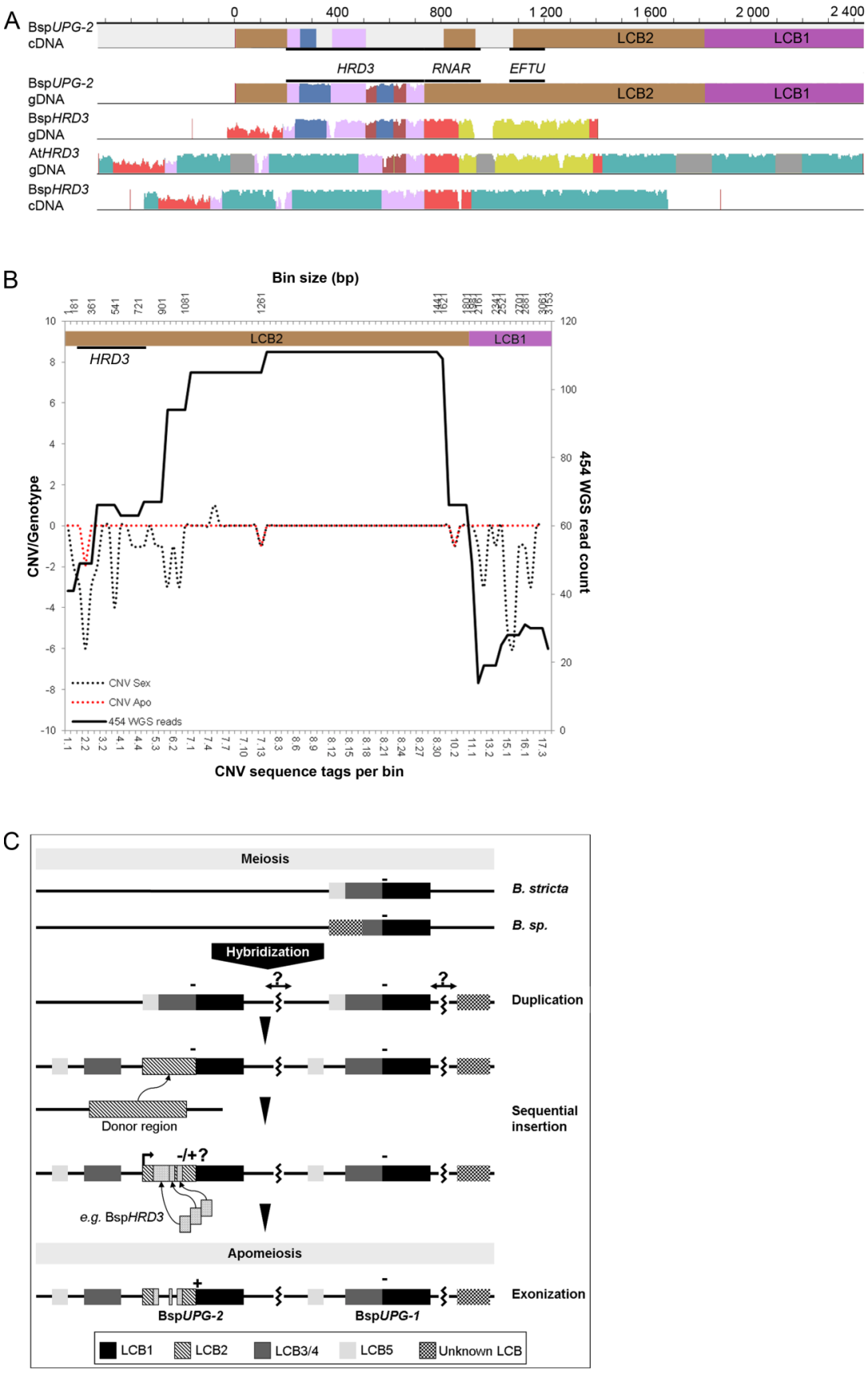

Figure 25. Hypothesized origin of Bsp*UPG-2* through duplication and fragmental insertion.

(A) Locally collinear blocks (LCB) between Bsp*UPG-2* and the source genes from which Bsp*UPG-2* hosts fragments (identical colour display highly similar sequences; only a fragment of Bsp*HRD3* was obtained by chromosome walking) obtained by an modified cross-match alignment using the Progressive Mauve algorithm. (B) 454 whole genome sequencing and CGH sequence tags (10 sexual *versus* 10 apomictic genotypes; Aliyu et al., unpublished results) were mapped onto eighteen 200 bp fragments of the genomic sequence of Bsp*UPG-2* which had 20 bp overlap between each other. The graph shows all existing CNVs for each single sequence tag. Relative CNV frequencies are considered independently for sexual and apomictic genotypes. (C) Schematic representation of parental and chimeric Bsp*UPG* genes. Boxed numbers illustrate different LCBs (Supplemental Table 24, Supplemental Figure 6). (+) Transcription activity and (−) no transcription activity.

4.6 Phylogeography of Bsp*UPG-2*

In this chapter we examined the evolutionary footprint of Bsp*UPG-2* as a candidate factor for unreduced pollen formation on a Boechereae tribe-wide level, by testing its species-specificity and putative origin. A total of 1576 accessions representing 102 taxa of four Boechereae genera (*Boechera, Cusickiella* Rollins, *Sandbergia* Greene (formerly *Halimolobos,* Al-Shehbaz (2007)) and *Polyctenium* Greene) and one taxon of tribe Sisymbrieae (*Schoenocrambe* Greene), which are distributed across 213 cpDNA haplotypes, were examined for presence ('carriers') or absence ('noncarriers') of Bsp*UPG-2 via* a PCR-based screen using a marker fragment which spans the juncture between the fused LCB1 and LCB2.

Bsp*UPG-2* was found in 33.82% of all tested accessions across North America and additionally in a single Greenland accession, while virtually all mid- and south-eastern North American *Boechera* accessions were noncarriers (Figs. 26A and B). Interestingly, these noncarrier accessions belong to cpDNA lineages V and IV which were recently described as closely related to *Borodinia,* and for whom their taxonomic status remains undecided (Kiefer et al., 2009). For each of the remaining lineages the distribution of Bsp*UPG-2* follows approx. a 1:2 presence-to-absence ratio, with the exception of lineages I and III where Bsp*UPG-2* is overrepresented (two-tailed Fisher's exact test, 43.27%, $p>p^0=0.0006$ and 46.47%, $p>p^0<<0.001$, respectively, Fig. 26C). Taken together, except for lineages IV (purely *B. canadensis*) and V (mostly *B. laevigata* and *B. missouriensis*), no pronounced correlation was detected between single lineages and Bsp*UPG-2* levels. As apomictic *Boechera* are pseudogamous and thus strongly depend on central cell fertilization with their own unreduced pollen for induction of a balanced endosperm (Roy 1995; Voigt et al., 2007) and for which unreduced pollen formation is

restricted to apomictic genotypes (Aliyu et al., 2010), a genetic factor causing unreduced pollen formation should be indirectly linked with 'apomixis'. Despite the hypothesized involvement of additional apomeiosis factors, a putative male apomeiosis-

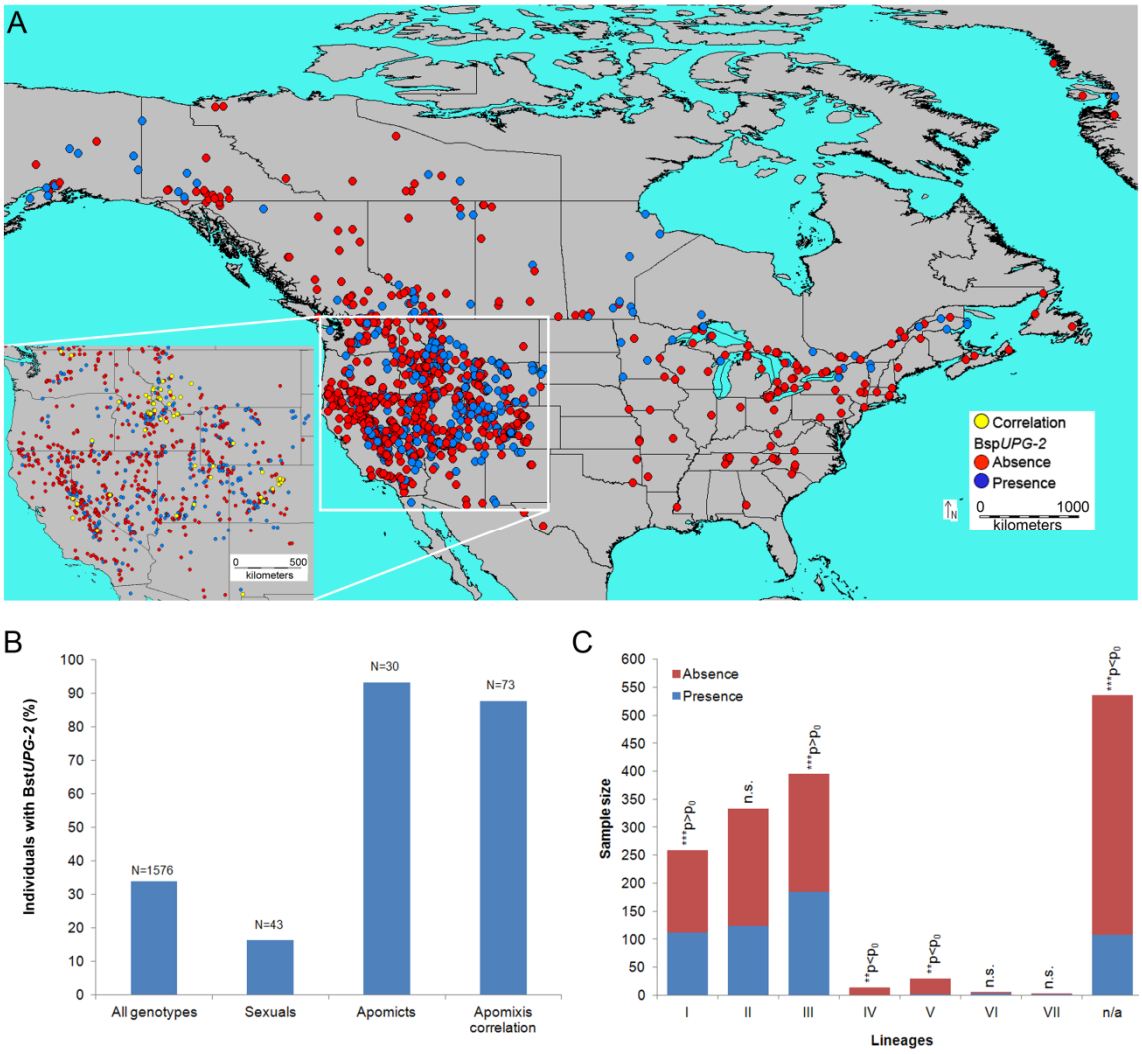

Figure 26. Phylogeographic distribution of Bsp*UPG-2*.
(A) Individuals carrying Bsp*UPG-2* (blue) *versus* individuals lacking Bsp*UPG-2* (red). Subfigure shows distribution of individuals used for correlation of Bsp*UPG-2* appearance with the mode of reproduction (yellow). (B) Penetrance of Bsp*UPG-2* across all genotypes compared to sexuals, and apomicts. (C) Penetrance of Bsp*UPG-2* per cp-haplotype lineage. n/a = not applicable.

associated factor such as Bsp*UPG-2* should be able to separate most of the apomicts from the sexuals within the genus. In order to establish Bsp*UPG-2* as a genetic marker for apomixis in *Boechera* we tested its association with apomixis using a subset of 73 of the 1576 screened *Boechera* accessions from which FCSS data were available genotypes, representing 18 *Boechera* taxa from lineages I, II and III (yellow dots in Fig. 26A, Supplemental Table 25; Mau et. al., unpublished results; Aliyu et al. (2010)). In

93.10% of the apomicts Bsp*UPG-2* is present, whereas in sexual genotypes only 19.51% genotypes carry the candidate gene (Fig. 26B).

In all, Bsp*UPG-2* explains 85.71% of the reproductive modes among the tested *Boechera* genotypes, and together with its ubiquity in all lineages, we thus conclude that Bsp*UPG-2* is a suitable marker for apomixis in the genus *Boechera.*

*4.6.1 Bsp*UPG-2 *is specific to the tribe Boechereae*

According to the haplotype database (Kiefer et al., 2009), 106 of the tested haplotypes are singletons whilst 107 haplotypes are shared by several accessions (hence called suprahaplotypes). Bsp*UPG-2* is present in approx. half of all *Boechera* interconnected haplotypes (*N*=93, 43.66%, Fig. 27A) exept lineage IV. Furthermore, Bsp*UPG-2* was detected in accessions of the oldest haplotype (*i.e.* AB, represented in suprahalotype 8), which is the center of the haplotype network and which has been estimated to be 0.7–2 million years old (Dobeš et al., 2004a). In addition, Bsp*UPG-2* is present in two single accessions which are each assigned to one of two closely related genera of the tribe Boechereae (*i.e.* clade B: *Cusickiella* and *Polyctenium*, refer to Beilstein et al. (2006) and Al-Shehbaz et al. (2006)). Bsp*UPG-2* was not detected in other tested genera of the same clade (clade B, *e.g. Sandbergia whitedii*, *S. perplexa* and *S. perplexa* var. *lemhiensis*) or in neighbouring clades (clade A (Camelineae): *e.g.* genus *Arabidopsis* (*A. thaliana*, *A. lyrata*), clade I (Sisymbrieae): *e.g.* genus *Schoenocrambe* (*S. linifolia*); clade J (Brassicaceae): *e.g.* genus *Brassica* (*B. napus, B. oleracea, B. rapa*); Supplemental Table 26).

All together, the ubiquitous occurence of Bsp*UPG-2* within the genus *Boechera* point to a conserved status of the apomeiosis candidate gene, and places the origin of Bsp*UPG-2* with that of the genus *Boechera* in the middle of the Pleistocene (Dobeš et al., 2004a). As Bsp*UPG-2* was detected in two neighbouring Boechereae genera our working hypothesis is that Bsp*UPG-2* alternatively could have originated earlier at the base of the tribe Boechereae.

*4.6.2 Taxon-specific variation of Bsp*UPG-2

PCR screening data from 28 taxa with statistically representative numbers (*N*≥10) of accessions demonstrate (Fig. 27B, complete list of taxa in Supplemental Table 26) that six taxa do not carry Bsp*UPG-2*, from which two belong to the eastern North American lineages IV and V (*B. breweri* (lineage I, III), *B. platysperma* (III), *B. canadensis* (IV),

Schoenocrambe linifolia, *B. laevigata* (V) and *B. cobrensis* (III)). The mostly sexually reproducing *B. stricta* (Aliyu et al., 2010) which co-dominates lineage II together with *B. divaricarpa*, is consistently correlated with the absence of Bsp*UPG-2*, except for a few accessions for which FCSS confirmed their apomictic mode of reproduction (*N*=183, 5.43%, *e.g.* ES 649; Supplemental Table 26). In other taxa Bsp*UPG-2* is prevalent in a number of accessions (*e.g. B. microphylla* var. *macounii*, *B. crandallii, B. fendleri*, *B. lignifera*, *B. microphylla*, *B. holboellii*, *B. perennans*, *B. holboellii* var. *pinetorum* and *B. divaricarpa*), whereby in accessions of *B. crandallii* Bsp*UPG-2* is virtually omnipresent (*N*=16, 93.75%). In contrast to previous reports (Aliyu et al., 2010; Kiefer and Koch 2012), the two taxa which constitute the largest groups and the largest distribution among the haplotypes, *B. holboellii* var. *retrofracta* which dominates lineage I and III, and *B. divaricarpa* which dominates lineage II, do not correlate purely with apomixis but are characterized by approximately equal proportions of carrier and noncarrier accessions (45.50% and 60.66% carriers respectively, Fig. 27B, Supplemental Tables 26 and 27). In fact, none of the tested taxa with a statistically representative number of accessions (here extended to taxa with $N \geq 5$ accessions) is purely apomictic, but several purely sexual taxa were detected so far (*N*=19), besides the majority of tested taxa which exhibit a mixture of sexual and apomictic accessions (*N*=33, Supplemental Tables 26 and 27).

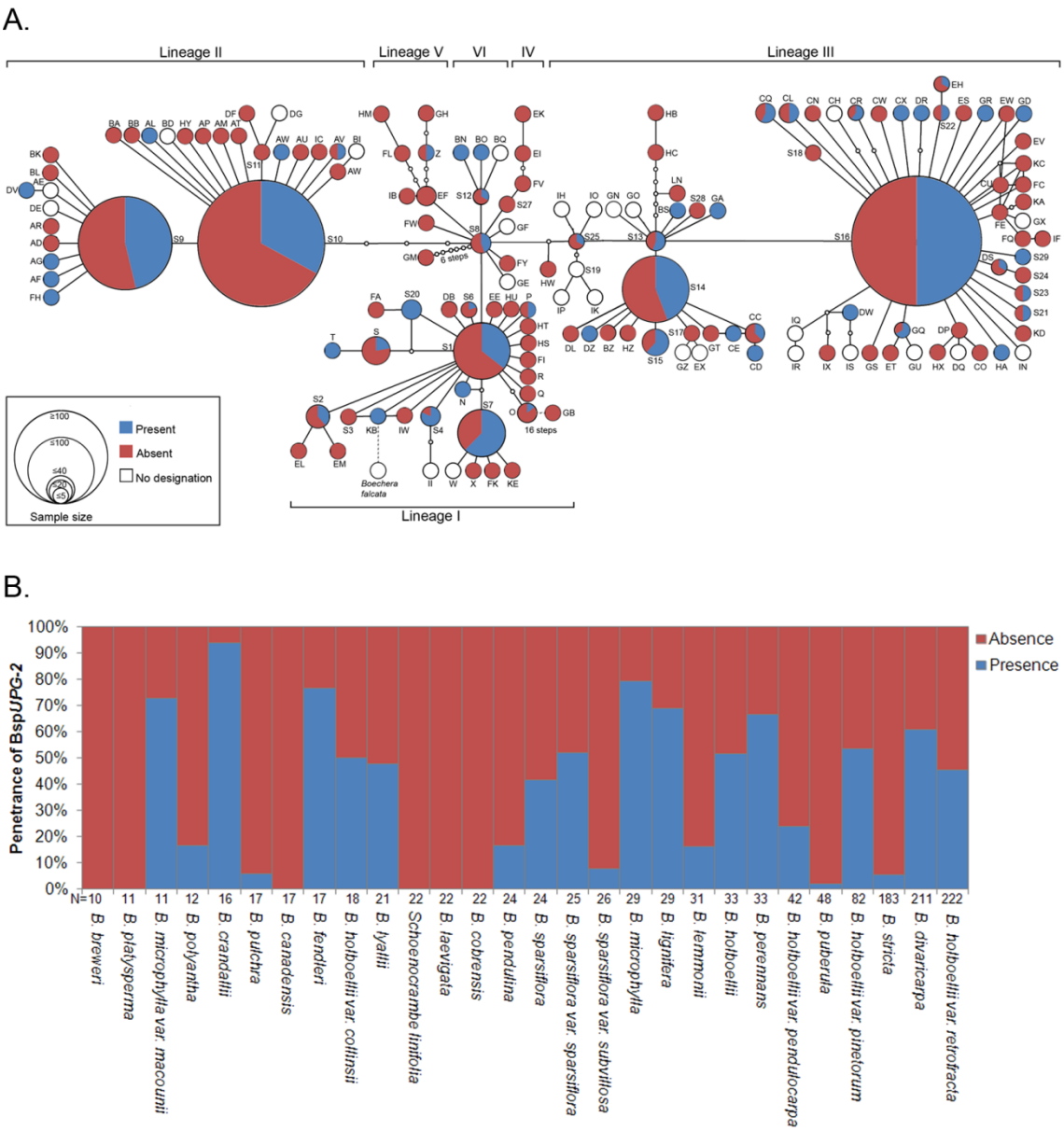

Figure 27. Haplotype- and taxon-specific frequencies of Bsp*UPG-2*.
(A) Network analysis of Bsp*UPG-2* using TCS v1.21, based on the *trnL-F* chloroplast DNA dataset of Kiefer et al. (2009). Node size corresponds to number of accessions carrying the (supra)haplotype (*N* ranges from 1 to 168). Haplotype node sectors indicate the partition of this haplotype between accessions carrying Bsp*UPG-2* and accessions lacking Bsp*UPG-2*. (B) Abundance of Bsp*UPG-2* for 28 *Boechera* taxa (latest nomenclature) which are represented by more than ten accessions.

4.6.3 Indications for geographic parthenogenesis in apomictic Boechera*?*

Previous studies on ecological and geographic distribution patterns of sexual and apomictic higher plants concluded that apomictic populations follow a trend towards larger ranges, in addition to a tendency to higher latitudes and elevations compared to their sexual relatives, a phenomenon which is called “geographical parthenogenesis”

and which was initially described for parthenogenetic animals (Vandel 1928; Bierzychudek 1985). Moreover, growth-chamber experiments with varying environmental conditions for sexual and apomictic *Anntennaria parvifolia* supported the hypothesis that sexual and apomictic genotypes are not ecologically similar, and led to the idea of general purpose genotypes among apomicts (Bierzychudek 1987; Bierzychudek 1989). In this chapter we used the pairwise distances and bioclimate variables to test whether the previously reported differences in geographical and ecological distribution patterns between sexuals and parthenogens are also present between naturally occuring apomictic (*i.e.* accessions which carry the apomeiosis marker candidate Bsp*UPG-2*) and sexual *Boechera* (*i.e.* accessions lacking the marker gene).

Bsp*UPG-2* is geographically widely distributed over North America and Greenland. The most southerly occupied site for noncarriers was Bullen Creek, Jefferson, Mississippi (*B. canadensis*, 31° 34'N, -91° 12'W), and for carriers it was Box Elder County, Utah (*B. perennans*, 31° 59'N, -108° 6'W), while the most northerly occupied site was the Upernivik District, Greenland (*B. holboellii*, 72° 46'N, -56° 10'W) for noncarriers, and for carriers Kangerdluarssuk, Greenland (*B. holboellii*, 69° 40'N, -50° 28'W). The center with the highest abundance differed slightly between carriers and noncarriers. Bsp*UPG-2* carriers were highly abundant at the Colorado Plateau Shrub Steppe and the Central Rockies Forest, Idaho, whereas the highest density of Bsp*UPG-2* noncarriers was detected at the Sierra Nevada/Great Basin Shrub Steppe, Californa and the Central Rockies Forest, Idaho (Figs. 28A and B). In addition, a higher abundance of Bsp*UPG-2* noncarriers compared to Bsp*UPG-2* carriers was detected in Eastern North America.

In search for variation in geographical patterns between carrier and noncarrier accessions we tested for geospatial distribution homogenity using kernel density estimation as a smoothing method (KDE with Epanechnikov kernel, Silverman (1986)) for pairwise distances (PD) between all single accessions per group (N=1502 accessions) in conjunction with Mann-Whitney U tests to evaluate their significant differences per group (*i.e.* apo *versus* sex).

Besides different centers of highest abundance of noncarriers *versus* carriers, no obvious difference in the geospatial pattern of both PD distributions was detected. The same two distance maxima of PDs were identified for carriers and noncarriers (M_1:~750 km, M_2:~2750 km, Fig. 28E) which points to a similar aggregation pattern of both

distributions on a large scale. Interestingely, the frequencies of Bsp*UPG-2* carrier and noncarrier differ at each PD maximum (U=3.01E+10, p<<0.001) exhibiting a higher number of carriers at M_1 and a lower number of carriers at M_2 compared to the distribution of the noncarriers, as reflected by smaller mean PDs in the carrier group (N=140715, PD_{MEAN}=1223.6±1013.69 km, mean rank=284582.98) compared to the mean PDs in the noncarrier group (N=470935, PD_{MEAN}=1414.8±1124.29 km, mean rank=312172.74). Considering lineages IV and V as outliers (see above, Kiefer 2009) the PD analysis was repeated after exclusion of 42 accessions belonging to lineages IV and V, and demonstrated similar results for carriers (N=140185, PD_{MEAN}=1222.12±1014.73 km, mean rank=278686.26) compared to noncarriers (N=434778, PD_{MEAN}=1304.69±1070.20 km, mean rank=290318.00, U=2.92E+10, p<<0.001, Fig. 28E inbox graph).

Ecological habitat modelling of sexual and apomictic accessions, as estimated by presence of Bsp*UPG-2*, was used to provide insights into putative differences in habitat requirements based on primary limiting factors for species distribution such as temperature, moisture, elevation and specific ecoregion profiles (*i.e.* ecologically and geographically defined areas with distinct composition of natural communities and species, after Olson et al. (2001)). Noncarriers were used as a training dataset while carriers composed the test data set. The maximum-entropy-model (Maxent) method (Phillips et al., 2006) was used, which tests separately the worldwide potential distribution (*i.e.* suitability) of both sexual and apomictic *Boechera* accessions based upon their biogeographic abundance. Thereby, the area under the receiver operating characteristic (ROC) curve (AUC) allows the comparison of model performance by plotting the fraction of true positives out of the positives (TPR=true positive rate or sensitivity) *versus* the fraction of false positives out of the negatives (FPR=false positive rate or 1-specificity). An AUC value of 0.5 indicates that the performance of the model meets randomness, while values closer to 1.0 indicate better model performance. The average AUC (N=15 repetitions) for the Bsp*UPG-2* noncarrier model is 0.946±0.0005, and the average AUC for Bsp*UPG-2* carriers (test model) is 0.940±0.0072, which points to a good performance of both models (Figs. 28C, D and F) compared to the real distribution of carriers and noncarriers (Figs. 28A and B). Moreover, both AUC values are similar suggesting a good prediction of the Bsp*UPG*-2 carrier model with the distribution of Bsp*UPG-2* noncarriers. Therefore our prediction

for apomictic *Boechera* demonstrated similar patterns to sexual *Boechera*, but its geogaphic coverage was somewhat less extensive than sexual *Boechera* (Figs. 28A-D).

We predicted the distributions of Bsp*UPG-2* carriers and noncarriers in North America (Fig. 28C and D) using Maxent models driven by the climatic variables that most strongly contributed to the thermal and moisture gradients (Table 5). As discussed earlier, these variables included BIO1 (= Annual Mean Temperature) to BIO11 (= mean temperature of coldest quarter) for the thermal gradient, and BIO12 (= Annual Precipitation) to BIO19 (= precipitation of coldest quarter) for the moisture gradient. Jackknife tests showed that these predictors contributed similarly to the models, but had less impact on the models compared to variable 'altitude', which in both groups had the highest predictive contribution to the models with 31.6% *versus* 21.0% for carriers and noncarriers, respectively (Table 5). Carriers and noncarriers were most responsive to variables BIO11 and BIO2 (= mean diurnal range), respectively, for the thermal gradient and to variable BIO19 for the moisture gradient (Table 5). These four of the nineteen tested variables contributed approx. 50% to the models, in contrast to the remaining variables which contributed relatively little (Table 5).

Boechera inhabits a wide range of elevations (sea level to 3992.88m), and although single noncarriers (*i.e.* sexual accessions) represent the highest elevational record for the genus, we detected a trend towards higher altitudes for carriers (noncarriers: 1831.00±837.70m, carriers: 2024.83±618.86m; Mann-Whitney *U* test, *U*=7.92E+04, *p*=0.0012, Fig. 28G). In order to avoid bias resulting from strong outlier distributions of carriers and noncarriers, we removed accessions of lineages IV and V since virtually all accessions of both lineages lack Bsp*UPG-2* and a previous study placed them into a new genus of the tribe Boechereae (*Borodinia*, see above, Kiefer et al. (2009)). Nonetheless, the corrected data did not change the conclusion that apomictic *Boechera* were characterized by higher elevations (Mann-Whitney *U* test, *U*=7.90E+04, *p*=0.018).

Most bioclimate variables showed no statistically different distributions between noncarriers and carriers (*e.g.* annual precipitation, Fig. 28I, Supplemental Figure 7). On the other hand, small differences were observed, such as the average annual temperature of carriers (4.4±4.06°C, BIO1), which was significantly lower compared to noncarriers (4.9±4.44°C, *U*=1.94E+05, *p*=0.015, Fig. 28H). In addition, carriers tend to occupy locations with higher seasonal variability (856.2±177.12 *versus* 833.7±174.06 100 x standard deviation of mean monthly temp. in °C, BIO4) and exhibit a one degree higher

annual temperature range compared to sexuals (38.5±5.02 *versus* 37.6±4.94 respectively, U=1.86E+05, p<0.001, BIO7; Supplemental Figure 7).

A similar geographic distribution pattern for sexual and apomictic *Boechera* was mirrored by a similar habitat distribution (grouping of the tested ecoregions into biomes according to Olson et al. (2001), http://worldwildlife.org/publications/wildfinder-database; Fig. 29, Supplemental Table 28), which contrasts results for other apomictic species for which habitat differentiation between sexuals and apomicts were observed (*e.g. Ranunculus auricomus*, Hörandl and Paun (2007)). Both sexuals and apomicts prefer temperate conifer forests and desert/xeric shrublands, which together account for 80.8% and 82.9% of the compete distribution, respectively (Fig. 29).

Taken together, with the exception of altitude differences, no habitat differentiation between sexual and apomictic *Boechera* accessions was found according to ecological and geographic distribution analyses. Sexual and apomictic *Boechera* are also similar in terms of their climatic limits. Thus, evidence for geographic parthenogenesis is weak and based on marginal differences in altitude and temperature, which may be interelated as apomicts demonstrated lower annual temperatures and higher altitudes.

Table 5. Contribution of environmental variables to the maximum-entropy-models of Bsp*UPG-2* carriers and noncarriers.

Noncarrier (Bsp*UPG*-2 absent)			Carrier (Bsp*UPG*-2 present)		
Variable[a]	% contribution	% permutation importance	Variable[a]	% contribution	% permutation importance
Altitude	21.0	0.7	Altitude	31.6	1.2
BIO19	11.2	33.9	BIO11	15.0	4.0
BIO2	11.0	19.4	BIO2	11.0	19.0
BIO11	9.1	0.6	BIO19	6.9	23.5
BIO3	8.7	3.3	BIO1	5.7	10.5
BIO1	8.6	5.4	BIO15	5.1	0.3
BIO9	6.6	2.5	BIO17	5.0	10.8
BIO4	6.2	7.7	BIO3	4.1	0.7
BIO6	4.9	7.6	BIO18	3.6	4.8
BIO17	3.9	0.4	BIO4	3.5	8.7
BIO18	3.2	5.8	BIO6	2.4	2.5
BIO10	2.4	1.3	BIO5	2.3	2.9
BIO8	1.4	0.4	BIO8	1.1	1.2
BIO7	0.6	2.4	BIO10	1.0	3.2
BIO15	0.4	0.5	BIO9	0.7	2.3
BIO5	0.4	1.5	BIO7	0.5	1.3
BIO14	0.1	1.1	BIO14	0.3	1.0
BIO16	0.1	0.0	BIO13	0.1	0.4
BIO13	0.1	0.1	BIO16	0.1	0.3
BIO12	0.0	5.3	BIO12	0.0	1.5

[a]BIO1= Annual Mean Temperature, BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)), BIO3 = Isothermality (BIO2/BIO7) (* 100), BIO4 = Temperature Seasonality (standard deviation *100), BIO5 = Max Temperature of Warmest Month, BIO6 = Min Temperature of Coldest Month, BIO7 = Temperature Annual Range (BIO5-BIO6), BIO8 = Mean Temperature of Wettest Quarter, BIO9 = Mean Temperature of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter, BIO11 = Mean Temperature of Coldest Quarter, BIO12 = Annual Precipitation, BIO13 = Precipitation of Wettest Month, BIO14 = Precipitation of Driest Month, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO16 = Precipitation of Wettest Quarter, BIO17 = Precipitation of Driest Quarter, BIO18 = Precipitation of Warmest Quarter, BIO19 = Precipitation of Coldest Quarter; (http://www.worldclim.org/bioclim).

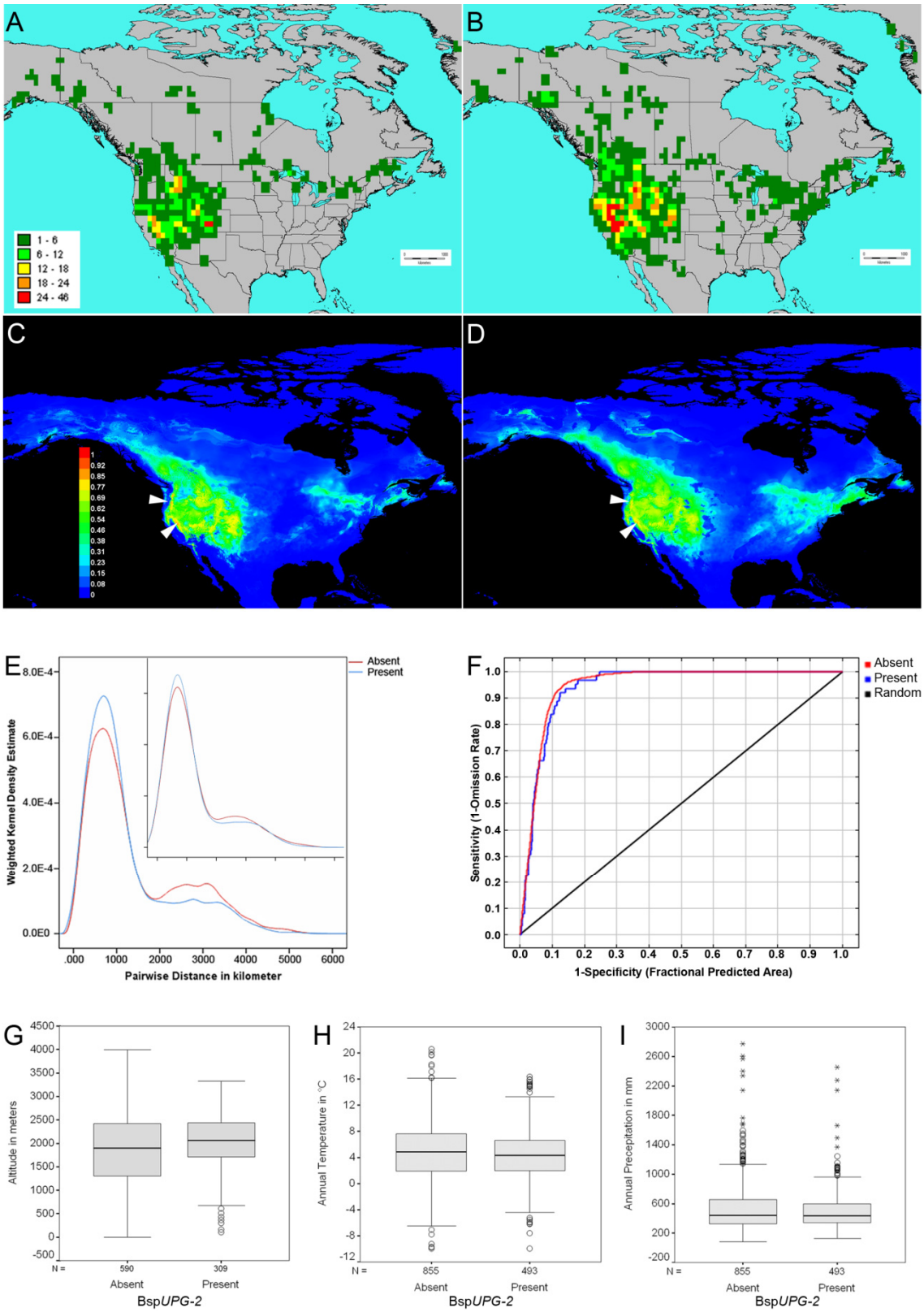

Figure 28. Comparison of the geographic abundance of apomictic and sexual *Boechera* accessions using Maxent models.

Density measures of Bsp*UPG-2* noncarriers (*i.e.* sexuals) (A) and Bsp*UPG-2* carriers (*i.e.* apomicts) (B) per cell (1x1 degree). Bar scale=1000km. Habitat prediction for Bsp*UPG-2* noncarriers (C) and Bsp*UPG-2* carriers (D) shows the center of optimal environmental conditions for both groups (white arrows). The colors indicate the predicted probability that

conditions are suitable to Bsp*UPG-2* noncarriers or Bsp*UPG-2* carriers, with red indicating high probability of suitable conditions, green indicating conditions typical of those where repoductive mode-based group of *Boechera*'s is found, and lighter shades of blue indicating low predicted probability of suitable conditions. (E) Comparance of weighted KDE of pairwise distances between individuals carrying Bsp*UPG-2 versus* individuals lacking Bsp*UPG-2*. The subgraph shows KDE of pairwise distances without accessions of lineages IV and V. (F) Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) enables the comparison of performance of the Bsp*UPG*-2 carrier model with the Bsp*UPG-2* noncarrier model against a random distribution, and is useful in evaluating both maximum entropy models. Abundance of Bsp*UPG-2* carriers and noncarriers per altitude (G), annual temperature (H) and annual precipitation (I) is shown.

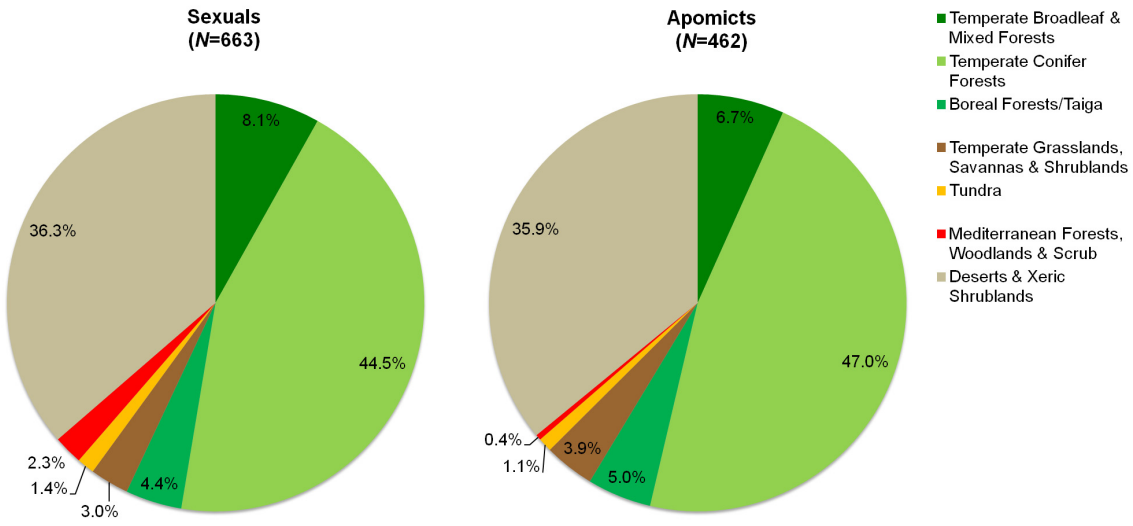

Figure 29. Habitat differentiation of sexual and apomictic *Boechera* in North America. Assignment to types of vegetation was conducted according to a PCR-based screen with the molecular marker Bsp*UPG-2*. Types of vegetation were applied from WWF biomes (http://worldwildlife.org/publications/wildfinder-database) based on Olson et al. (2001).

5 Discussion and Conclusions

Research into the fundamentals of apomixis with the goal of introducing it into crops has become a major issue in the last two decades (Khush et al., 1994; Koltunow et al., 1995; Spillane et al., 2001; Marimuthu et al., 2011) and has an expected high impact in agronomy (Dresselhaus et al., 2001). The hypothesized benefits of introducing apomixis in crops are multifaceted and include drastically decreased costs of hybrid production (hybrids from hybrid seeds through clonal reproduction), clonal propagation of hybrid vigor, the stabilization of locally adapted varieties and reduced costs for farmers.

Besides the inherent potential of apomixis for agriculture, the transfer of apomixis key gene(s) into the desired host crop plants would provide several hurdles such as a proposed high (epi)genetic genetic (*i.e.* mutational) load of apomeiosis-related region(s) (Leblanc et al., 2009), the requirement for adaptation of the host genome to regulate downstream genes following apomixis-induction, and for transferring candidate factors to distantly-related crop species (Polegri et al., 2010).

Several approaches to studying apomixis have been undertaken: (i) interspecific hybridization between crops and wild apomictic varieties, which has so far provided no agronomically useable crop (refer to section *2.3.3*), (ii) mutagenesis in model plants, which has revealed a wide array of genes which when introduced into a sexual background show side effects (*e.g.* polyploidization) or low penetrance of the desired apomixis trait (*e.g.* low rates of unreduced gamete formation, Marimuthu et al., (2011)), (iii) comparative mapping which has identified a number of linkage groups which coseggregate with apomixis but for which functional analyses of robust candidates is still lacking (*e.g.* Nogler (1984), Ozias-Akins et al. (1998), van Dijk and Bakx-Schotman (2004)), and (iv) comparative gene expression studies between natural sexual and apomictic genotypes as presented here (Polegri et al., 2010; Sharbel et al., 2010).

This study represents the first approach to identify components associated with unreduced pollen formation, a trait which is directly linked with pseudogamous apomixis (*sensu stricto* a female trait) in *Boechera*. Using multiple phenotypically characterized genotypes for a customized microarray-based transcriptome analysis of male reproductive tissues, followed by qRT-PCR, RACE and DNA sequencing-based validation steps, identified a single novel transcript associated with unreduced pollen formation, whose structure and origin was subsequently characterized. Furthermore, *Boechera* combines the advantage of being genetically closely related to *Arabidopsis*

which provides an array of tools for molecular genetic analyses, and in addition is closely related to the agronomical important genus *Brassica*, together which could facilitate the introgression of apomixis into a crop species. Considering that apomixis is facultative and represents a derivation of normal sexual development (Koltunow 1993; Grossniklaus 2001), in addition to the fact that diploid sexual and diploid apomictic individuals were compared here, this factor could hypothetically stabilize endosperm development in a sexual plant that has been transformed with apomixis controlling factor(s) ("dosage compensation"; *sensu* Birchler (1993), refer also to section *2.2.2*). Hence, in the following sections the putative role of the candidate factor during unreduced pollen formation and its implications on a genus-wide scale are discussed.

5.1 High quantitative variation for unreduced pollen formation in apomictic *Boechera*

High variability in pollen morphology and unreduced pollen formation, in addition to tolerance to deviations from the sexual endosperm balance number have been described for some *Boechera* (Böcher 1951, 1954; Voigt et al., 2007; Aliyu et al., 2010; Voigt-Zielinski et al., 2012). Despite this variability, castration experiments (Böcher 1951) and extensive flow cytometric analyses of seeds (Aliyu et al., 2010) strongly support selection pressure for the maintenance of unreduced pollen development to fulfill endosperm balance requirements in diploid apomicts.

Here we have performed detailed quantitative analyses of anther growth and microsporogenesis to identify the most optimal developmental stage for comparative expression profiling of reduced (sexual) and unreduced (apomictic) pollen formation in 14 genotypes. A strong correlation between microsporogenesis and anther growth was evident at the pre- and postmeiotic stages, whereas it was difficult to identify antherhead lengths corresponding to the meiotic stage (Fig. 7). Considering spatial and temporal variability of expression profiles in reproductive tissues (Mascarenhas 1989; Honys and Twell 2004), the identification of a specific stage of antherhead development and length characterized by PMCs at the onset of meiosis enabled targeted expression profiling of the meiotic stage which differentiates reduced (sexual) and unreduced (apomeiotic) pollen formation.

Variable levels of apomeiosis penetrance have been reported for *Boechera* (Aliyu et al., 2010), and here both obligate and highly facultative apomicts were analyzed. No evidence for sex- or apomixis-specific flower morphological variation was found,

although meiosis and the appearance of microspores are developmentally-uncoupled in apomicts relative to sexuals (Figs. 7 and 8). This latter observation supports the hybridization-derived floral asynchrony (HFA) theory, which describes the temporal shift (heterochrony) between both reproductive modes (Carman 1997) which has also been identified on the transcriptomic level (Sharbel et al., 2010).

Meiotic chromosome behaviour of apomicts producing unreduced pollen (the same used for expression profiling) was primarily asynaptic (Fig. 9), as has previously been shown (asyndetic, *sensu* Böcher (1951)). The low level of segregation distortions (*e.g.* resulting in the formation of micronuclei, data not shown), point to a fully functioning molecular mechanism which guarantees the equal separation of sister chromatids during meiosis II and subsequent formation of dyads with two unreduced chromosome sets. Sexuals were also able to produce low levels of unreduced pollen (Fig. 9), which is consistent with previous data (Kantama et al., 2007) and might reflect a predisposition for environmentally-induced unreduced gamete formation, such as stimulation by low (Mason et al., 2011) or elevated temperatures (Pécrix et al., 2011).

Although ultimately producing higher levels of dyads compared to sexuals, the same apomicts demonstrated variability in terms of monad, dyad, triad and tetrad formation (Fig. 9), mirroring reported genotype-specific variability for meiotic chromosome synapsis potential in both diploid and polyploid *Boechera* (Böcher 1951; Naumova 2001; Kantama et al., 2007). Interestingly, variation between individuals of the same clonal lineage for dyad and tetrad formation was also apparent in the majority of tested apomictic genotypes, and hence phenotypic variability for pollen formation exists *despite* genetically clonal reproduction. In this light, the observed sexual allopolyploid progeny from both diploid and triploid obligate apomicts (*e.g.* ES 753, *B. divaricarpa*; Schranz et al. (2005)) could be explained by fertilization with reduced self pollen rather than with pollen from another plant (Aliyu et al., 2010), since *Boechera* is a highly selfing system (Roy 1995). Kantama et al. (2007) observed sexual *Boechera* which produce apomeiotic egg cells and postulated that hybridization-associated stress might have induced gene-expression changes required to switch from sex to apomixis. Considering the existing predisposition for environmentally-induced unreduced gamete formation in sexual genotypes (this study and Kantama et al. (2007)), hybridization-derived stress could similarly have induced gene-expression changes required to switch to apomeiosis in the PMCs of *Boechera* to thus lead to enhanced production of unreduced pollen.

These results, and the observed endosperm imbalance tolerance of some taxa, led us to characterize seed and pollen ploidy in every individual plant used for subsequent expression-profiling of antherhead tissues. On a broader scale, our data indicate that fixation of the genotype *via* apomixis does not necessarily lead to phenotypic stability of male meiocytes. While these data are consistent with previous work (Kantama et al., 2007; Voigt et al., 2007), the lack of correlation with apomeiosis expression (*e.g.* egg cell formation as measured by flow cytometric seed screen) suggests separate mechanisms leading to unreduced male *versus* female gametes in *Boechera,* and supports independent genetic control for at least some of the developmental steps required to form apomictic seeds (van Dijk et al., 1999; Noyes and Rieseberg 2000; Matzk et al., 2005).

The analyses of seed production, anther development, meiosis and pollen formation have shown that (1) unreduced pollen formation occurs *via* first division restitution (FDR) in apomictic *Boechera*, (2) this mechanism is not fully penetrant, and (3) that selection pressure for a balanced endosperm in apomictic *Boechera* leads to the almost exclusive contribution of unreduced pollen to endosperm formation.

5.2 Bsp*UPG-2* is highly associated with male apomeiosis

Besides the pioneering work on the identification of apomixis-linked genes in various model apomixis systems using comparative mapping strategies (Leblanc et al., 1995; Ozias-Akins et al., 1998; Noyes and Rieseberg 2000), transcriptional profiling of sexual *versus* apomictic genotypes became more popular in the last years due to reduced costs (*e.g.* microarrays superseeded cDNA-AFLP and differential display techniques), customizing capabilities and an increase in microarray size format which enabled spotting of large genome/transcriptome portions (Sharbel et al., 2009; Polegri et al., 2010). Nonetheless, with rare exceptions (*e.g.* Albertini et al. (2005)) past studies have reported a plethora of genes associated with apomixis for different species in absence of functional analyses to prove their specific involvement in apomixis, neither from mutagenesis approaches, using transposons or t-DNA, (*e.g. A. thaliana*, Chaudhury and Peacock (1994); *H. pilosella*, Ohad et al. (1996)) nor from comparative studies on natural apomictics (Polegri et al., 2010; Sharbel et al., 2010). This might be a consequence of the complexity of the trait considering the hybridization-induced temporal and spatial deregulation of the complete sexual pathway as a potential

apomixis inducer (HFA, Carman (1997)), in addition to the use of single genetically different samples (single genotypes or populations, segregating backcrossing lines) for a molecular-based comparison of apomixis and sexuality (Leblanc et al., 1997; Singh et al., 2007) and the species-specificity of the reported candidate apomixis marker genes (*cp.* Grossniklaus et al. (2001)).

In order to identify genes whose specific expression is associated with single components of the asexual pathway in *Boechera* (*e.g.* apomeiosis for this study) the present study attempted to overcome these hurdles by examining multiple genotypes (*i.e.* biological replicates) for (1) the identification of a single developmental stage showing a significantly different gene expression between sexual and apomictic tissues (*e.g.* PMCs in analogy to MMCs, *sensu* Sharbel et al. (2010)), and (2) detailed phenotypic characterization of the trait. Hence, comparative transcriptome profiling of the identified developmental tissues in multiple genetically and phylographically diverse samples enabled us to filter out most of the non-apomixis related variation.

A global view of our microarray data revealed a group of highly expressed genes with different expression characteristics in apomictic, sexual *B. stricta* and sexual non-*B.stricta* genotypes (Fig. 12; FC≥10, *N*=522 genes). The different relative expression level of highly expressed genes involved in major cellular and metabolic processes between sexual *B. stricta*, other sexuals and apomicts could reflect different contributions of both groups of sexual genotypes to their apomictic derivatives, as has been shown for different degrees of introgession of parental (*i.e.* sexual) chromosome segments in hybrid progenies (L'Hôte et al., 2008). Interestingely, a previous study on apomictic *Boechera* revealed that their genome consisted of variable numbers of *B. stricta* and non-*B. stricta*-like chromosomes (*i.e. B. holboellii*, *sensu* Kantama et al. (2007)). Hence, this phenomenon could apply to the observed gene expression differences of highly expressed genes which exhibit a closer relation between the apomicts and the sexual non-*B. stricta* compared to *B. stricta* genotypes (Fig. 12F), considering that chromosomes from diploid apomicts have by tendency higher numbers of *B. holboellii*-like compared to *B. stricta*-like chromosomes (in 3 of 5 tested apomicts, one has equal contributions, Kantama et al. (2007)). Interestingly, global upregulation of these highly expressed genes in apomictic anther tissues is contrasted by global downregulation of genes in early apomictic ovule development (Sharbel et al., 2010), a phenomenon which could attest to the suggested independent genetic control of male and female gamete formation in apomicts (see *5.1*). Among these genes, which are

averagely but not constantly (*i.e.* outlier) higher expressed in apomicts than in sexuals (Fig. 12), GO classes involved in the serine/threonine protein kinase pathway, especially for mitogen-activated protein (MAP) kinases, were significantly enriched (Fig. 13, Supplemental Table 7). Among other functions (*e.g.* signal transmission), MAP kinases have been implicated in cell cycle and developmental processes and have previously been shown to be differentially expressed in apomictic and sexual ovules of aposporic *Brachiaria* (Rodrigues et al., 2003) and in immature inflorescences of aposporic *Paspalum notatum* (Laspina et al., 2008). In line with these results would be the observation that mutagenesis of cyclin-dependent kinases (*i.e.* the closest relatives of MAP kinases) in sexual reproducing plants (*e.g.* CYCA1;2 in the *tam1* mutant in *Arabidopsis*; Magnard et al. (2001), Wang et al. (2004)) causes asynchronous PMC meiosis, producing elevated levels of unreduced gametes with sister chromatid formation. Therefore, MAP kinases could conceivably act as a part of an apo-specific signal cascade involved in unreduced pollen formation in apomictic *Boechera*, although our data could show only tendencies as none of the corresponding genes were constantly (*i.e.* significantly) differential expressed between all tested apomictic and all sexual genotypes (refer to section *4.3.1*).

The choice of a microarray containing multiple oligonucleotides (*i.e.* technical replicates) of every annotated and non-annotated gene expressed during *Boechera* flower development, together with stringent threshhold criteria such as multiple biological replicates (*i.e.* single genotypes), the collection of a specific antherhead stage which differentiated reduced *versus* unreduced PMCs, and a conservative statistical correction of false positives overcame the problem of genotype and tissue heterogeneity and drastically reduced the number of identified genes being significant differentially expressed between sexual and apomictic antherhead tissues.

The microarray-based analysis of diploid sexual *versus* diploid apomictic *Boechera* genotypes, in conjunction with qRT-PCR-based analyses of additional genotypes, further reduced the field of potential candidates (*e.g.* Sharb0350102; Figs. 15 and 16) and led to the identification of Bsp*UPG-2*. Assuming all other developmental aspects of sexual and apomictic antherheads to be the same, the transcriptional activity of the Bsp*UPG-2* locus is directly correlated with meiotic non-reduction during male sporogenesis. Interestingly, only the apomixis-specific duplicated locus (Bsp*UPG-2*) shows transcriptional activity, while the original locus (Bsp*UPG-1*), which is present in both sexuals and apomicts, does not (Fig. 24E). This is consistent with the “gain in

functions" hypothesis, which proposes a novel function for a putative apomeiosis gene because of the lack of recurrent apomictic mutants in sexual species (Vielle-Calzada et al., 1996). Although the apo-specificity of Bsp*UPG-2* attests to its novelty, the lack of homologs in other species and a missing ORF typical for a protein-coding gene (Fig. 18B) leaves the question open whether Bsp*UPG-2* acts in a complementary fashion to increase the penetrance of an existing predisposition for environmentally-induced unreduced gamete formation in sexual genotypes (see chapter *4.2.2*, sensu Aliyu et al. (2010)), or whether it is responsible for a completely novel pathway.

While the specific function of Bsp*UPG-2* with respect to unreduced pollen formation remains unclear, the fact that it was discovered with 21 biological replicates (*i.e.* 11 sexual and 10 apomictic genotypes) and 3 technical replicates (*i.e.* microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 mapping to different regions of the same locus), in conjunction with qRT-PCR results which show strong apomeiosis-specific upregulation of Bsp*UPG-2* in reproductive tissues, attest to its significance (Figs. 15 and 16). The specificity of Bsp*UPG-2* to male apomeiosis in anthers is additionally supported by a lack of its expression in ovule tissues (J. M. Corral, pers. comm.). In this context, it is unclear why genotype ES 753 was an outlier (Figs. 16B and 17), despite sharing a highly-similar Bsp*UPG-2* sequence with other apomictic genotypes (Fig. 23), although we suspect that genotype-specific shifts in anther development (as observed in other samples) may have led to sampling of pollen formation outside of the developmental window which characterized other genotypes. Besides tremendously higher expression rates in the autosomal tissue low levels of expression of Bsp*UPG-2* were also detected in somatic tissues of all tested apomicts (Fig. 15), an observation which might reflect the 60-90% overlap of genes expressed in pollen with somatic tissues (Willing et al., 1988; Borges et al., 2008).

Another feature of Bsp*UPG-2* is its hemi- or homozygosity (Fig. 23, Supplemental Table 22), which reflects dominant inheritance as a proposed characteristic for an apomeiosis controlling locus (refer to Grossniklaus et al. (2001), Table 1). A prominent example for a hemizygous apomixis-linked molecular marker was previously reported for apomictic *Pennisetum*, which was evidenced by its lack of hybridization in sexual F_1 relatives (Ozias-Akins et al., 1998). A hemizygous status for Bsp*UPG-2* could reflect an extended period of low to no recombination in its genomic region, as has been proposed for the apospory-specific genomic region in *P. squamulatum* (Ozias-Akins et al., 1998).

5.3 Transposable element-driven duplication of the *UPGRADE* locus

Mapping of annotated genes from both BAC assemblies to the genomes of *Arabidopsis* (TAIR database) and a sexual *B. stricta* (Schranz 2007) allocated the original locus (Assembly 1 including Bsp*UPG-1*) to a pericentromeric, highly heterochromatic region (Fig. 21). We could not decipher the location of the duplicated locus containing Bsp*UPG-2* due to an enriched number of gene fragments which are distributed across the *Arabidopsis* genome. One source of the enrichment of these gene fragments (*i.e.* dead-on-arrival (DOA) elements, Petrov et al. (2003)) on Assembly 2 could be transposable elements (TEs) which are known to induce ectopic recombination leading to genome rearrangements (*e.g.* duplications, deletions and inversions). Consequently, their disruptive effect on gene functions ("gene-disruption model", Finnegan (1992)) in addition to induction of strongly deleterious chromosome rearrangements ("ectopic recombination model", Montgomery et al. (1987)) should lead to the elimination of TE sequences in high-recombination rate regions more quickly compared to TEs in regions with low-recombination rates, leading to accumulation of TEs in heterochromatic regions (*e.g.* pericentromers, Petrov et al. (2003)). Interestingly this is consistent with our observation of numerous TEs and TE fragments on both assemblies (Figs. 20B and C, 22), in addition to the allocation of the original locus (Assembly 1) to the pericentromeric region of chromosome At1/linkage group BstLG1 (Schranz et al., 2007). In conjunction with the similar abundance of TEs from *Gypsy*-type and *En-Spm* superfamilies, which insert preferentially into gene poor regions (*i.e.* heterochromatic regions, Fiston-Lavier et al. (2012)) and which are enriched in both assemblies compared to their abundance within the total number of *Arabidopsis* TEs (Fig. 22) we assume that both assemblies, including the original and the duplicated locus of Bsp*UPG*, are co-localized in the same pericentromeric region or are localized on different positions with similar characteristics (*i.e.* interstitial heterochromatic regions comparable to *hk4S* in *Arabidopsis*; Fransz et al. (2000)).

In light of the data presented here, two mechanisms may have contributed to the formation of Bsp*UPG-2*: homeologous recombination and transposable element (TE) activation. Considering homeologous recombination, both nonreciprocal (HNTRs) and reciprocal translocations (HRT) could have led to gene fusions, although we favor the latter considering that HRT drives the rates of both chromosomal segment loss and duplication, as has been shown in *B. napus* (Udall et al., 2005; Nicolas et al., 2007). Duplication, as the fundamental event of Bsp*UPG-2* formation is evidenced by

sequence comparison with both BAC clone assemblies (*cp.* sex-specific indels for the LCB1 and LCB4, Figs. 20 and 24).

Moreover, TE sequences surrounding Bsp*UPG-2*, including three TE-related protein-coding genes (Fig. 20C, Supplemental Table 20), point to hybridization-mediated TE activation and associated genome rearrangements (McClintock 1984) as possible drivers of the Bsp*UPG-2* origin. The increased length of the duplicated *versus* original locus, the putative result of LTR transposons and IR-flanked insertion sequences (*e.g.* IR7; Fig. 20C) points additionally to a TE-induced origin of Bsp*UPG-2*. Evidence for TE-mediated gene creation has been found in plants, where *Mutator*-like DNA elements (*e.g.* Pack-MULES) and *Helitrons* (less abundant than classical LTR retrotransposons) produce chimeric transcripts which can evolve into functional fusion genes (Talbert and Chandler 1988; Yu et al., 2000; Brunner et al., 2005; Gupta et al., 2005). Consistent with this, *Mutator* and *Helitron* repeat families are prominent in both BAC clone assemblies (Fig. 22). Besides *Helitron and Mutator*-like DNA elements, specific TEs of other superfamilies such as *CACTA* and *Harbinger*, have the potential of directly "capturing" genic sequences *via* readthrough events, whereby neighboring TEs and the encircled fragment are joined into one element (*i.e.* TE-driven exon shuffling, Long et al. (2003), Kapitonov and Jurka (2007)). Alternatevely, the Bsp*UPG* locus could have been duplicated *via* the "synthesis-dependent strand annealing" mechanism (SDSA), whereby neighbouring genes are used as filler DNA while invading a foreign DNA strand to reinitiate strand synthesis after a double strand break inside of a TE (*e.g.* in *Helitrons*, Gupta et al. (2005), Kapitonov and Jurka (2007), Wicker et al. (2010)).

However, while a TE-driven duplication scenario for the origin of Bsp*UPG-2* is favored, the specific involvement of TEs in the genesis of BspUPG-2 cannot be asserted because of "footprint" erosion through time (Bennetzen 2005) and limited sequence information from flanking sites.

5.4 High sequence conservation and genus-wide occurence imply a selective advantage of Bsp*UPG-2*

Importantly, Bsp*UPG-2* is chimeric in structure and present together with Bsp*UPG-1* in apomicts, whereas sexuals host the original Bsp*UPG-1* only (Figs. 24 and 25C). This observation is based on BAC sequence comparison in addition to CNV and WGS read distribution analyses which both demonstrate duplication of the 5`-end of LCB2

and the LCB1 of Bsp*UPG-2* in apomicts, an observation which reflects two different processes, and which leads to the following model with respect to its generation.

The initial process was the duplication of a locus (LCB1) which was primary non-genic (*i.e.* non-genic in sexuals) and its insertion into a different genomic region (now referred to as Assembly 2). This event was followed by sequential integration of genes fragments from functional genes of other genomic regions (*e.g. HDR3*, *RNAR*, *EFTU/EF-1A*) into the 5`-end of the LCB1-attatched sequence (now LCB2), which is reflected in the higher copy number genomic sequence reads at the 5`-end of Bsp*UPG-2*. We tested whether LCB2 was derived directly from parental transcripts or from putative duplicated variants of those parental transcripts by performing RACE-PCR and chromosome walking (*e.g.* Bsp*HRD3*, homologous to AT1G18260). Although Bsp*HRD3* cDNA is conserved with *AthHRD3*, variation (*i.e.* insertions) between the transcript and its genomic copy, which are not caused by RNA splicing, were detected (Fig. 25A). The homologous fragment shared between Bsp*UPG-2* and the genomic copy of Bsp*HRD3* includes these insertions, suggesting that a nonfunctional duplicated variant of Bsp*HRD3* (*i.e.* a pseudogene derived through duplication; Ohno (1972)) which is present in sexuals and apomicts, was inserted into the Bsp*UPG-2* locus.

Although the long-term survival of chimeric genes is described to be rare (Bennetzen 2005), Bsp*UPG-2* exhibited an unexpectedly high degree of conservation in DNA sequence (Fig. 23) and expression between apomictic *Boechera* representing different taxa and geographic origins (Fig. 24E). In contrast, higher sequence variation was identified for Bsp*UPG-1*, especially at the 5`-end in sexuals (Fig. 23, Supplemental Tables 21 and 22). Together with the lack of intra-individual allelic variation in any tested *Boechera* genotype (refer to *4.5.1* and *5.2*) these observations have a major implications considering that natural selection acts upon two major sources of genetic variation: mutations and recombination.

They indeed could reflect an extended period of low to no recombination in the genomic region of Bsp*UPG-2*, similar to the apospory-specific genomic region in *P. squamulatum*, which would point to a physical location of Bsp*UPG-2* in a recombinationally suppressed region. This result would also be consistent with the specific composition of TEs at both the original and the duplicated locus (Fig. 22). It is widely known that in regions of low recombination natural selection reduces polymorphism at linked sites because of Hill-Robertson interference (HRi, Hill and Robertson (1966)), which explains the influence of segregation at a second locus on the

chance of fixation at the first locus of a linked neutral site. But the incorporation of deleterious ("background selection", Charlesworth et al. (1993)) and advantageous mutations ("genetic hitchhiking", Maynard-Smith and Haigh (1974)), similarly lead to a positive correlation of genetic invariability at linked neutral sites with a decrease in recombination rate. However, considering the broad phylogeographic distribution of Bsp*UPG-2*, we hypothesize that the canditate gene appears rather to be under selective maintenance (*sensu* Casillas et al. (2007)), which would be consistent with its assumed role in unreduced pollen formation for balanced endosperm in apomictic *Boechera* (Aliyu et al., 2010).

Taken together, we hypothesize that Bsp*UPG-2* arose *via* sequential (and segmental) duplication-insertion events involving at least five loci, from which three originated from transcriptional active genes. Considering its complex chimeric structure, we suggest that the genesis of Bsp*UPG-2* is associated with the recurrent interspecific hybridization which is highly correlated with the origins of apomictic *Boechera* (Schranz et al., 2005; Beck et al., 2011).

5.5 Does Bsp*UPG-2* have a regulatory function?

We have shown that meiocyte production in apomictic *Boechera* is highly variable despite stable Bsp*UPG-2* upregulation in generative tissues at the onset of meiosis and the formation of apomictic seeds (Figs. 9 and 15). The unexpected detection of a single candidate instead of multiple genes may therefore be interpreted in different ways with respect to its association with apomeiosis in anthers: either the locus controlling male apomeiosis contains a key genetic factor (*e.g.* Bsp*UPG-2*) with pleiotropic effects on other putatively essential components of unreduced pollen formation, or it contains several tightly linked factors, one of which is Bsp*UPG-2*, each controlling different aspects of unreduced pollen formation (*i.e.* apomeiosis). Therefore additional factors, including modifier genes (Bicknell et al., 2000), genetic background and/or environmental conditions (Nogler 1984) could explain variation in the level of apomictic trait expression.

The detection of the apo-specific indel9 is one example of a potential co-regulatory target site in Bsp*UPG-2*, although no homologous sRNAs were detected (Fig. 23B). Alternatively, the four detected mRNA isoforms (*i.e.* potentially caused by alternative splicing; Figs. 17 and 18) could regulate transcript abundance (as is predicted for at least 25% of all alternative exons; Stamm et al. (2005)) leading to the observed

variation in terms of reduced and unreduced pollen formation in both facultative and obligate apomicts.

The identification of an apomeiosis-associated short open reading-frame mRNA (sORF mRNA) and/or lncRNA is unique. Usually the identification of lncRNAs is difficult due to microarray designs, missing unannotated genes and the use of standard gene prediction programs that rely on the presence of relatively long ORFs (Zhang 2002). Here, we used detailed bioinformatic analyses of cDNA databases, including coding and miRNA formation potential in conjunction with chromosome walking to identify the chimeric Bsp*UPG-2*. Given its chimeric structure, the novel lncRNA Bsp*UPG-2* could have attained neofunctionalization in the context of apomeiosis for pollen development through a number of mechanisms (Kaessmann 2010), similar to the lncRNAs *BcMF11* from *Brassica campestris* (Song et al., 2007; 2012) or *Zm401* in maize (Ma et al., 2008). In addition, its tissue-specific expression pattern and conservation at the nucleotide level strongly supports a developmental role for the lncRNA Bsp*UPG-2*. Hence, Bsp*UPG-2* could conceivably have a regulatory function *via* a homology-dependent gene silencing mechanism (HDGS, reviewed in Meyer and Saedler (1996)), such as posttranscriptional gene silencing of related genes in *trans* (Eamens et al., 2008), which has been found for chimeric *Helitron* and Pack-MULE RNAs in maize (Jiang et al., 2004; Morgante et al., 2005), or *via* the modulation of DNA methylation patterns, such as reported for lncRNA-like loci which are associated with polycomb components and histone modifications (de Lucia and Dean 2011).

In line with this would be the observation of sequence fragments homologous to three different functional parental genes across the 5`-end of Bsp*UPG-2* which putatively could serve as targets for HDGS (Bsp*HRD3*, *RNAR* and *EFTU/EF-1A*; Figs. 18 and 23). Thereby, Bsp*UPG-2* could belong to a novel class of miRNA-containing lncRNAs which serve as matrices (*i.e.* as precursor RNA) for unidentified miRNAs or endogenous *trans* acting short interfering RNAs (ta-siRNA), as for example were detected in vegetative developmental pathways (Peragine et al., 2004; Vazquez et al., 2004; Hirsch et al., 2006) as well as in generative tissues (*e.g.* in mature pollen; Grant-Downton et al. (2009b)). Unlike other siRNAs in plants, ta-siRNAs silence gene expression by acting in *trans* to cleave mRNAs with sequences only partially complementary to their own (Peragine et al., 2004). The biogenesis of ta-siRNAs comprises their processing from the excised intronic region of a spliced transcript, or from the full-length unspliced transcript with Dicer-like proteins into 21 nt long siRNAs

which are incorporated into a RNA-induced Silencing Complex (RISC) with the appropriate Argonaute protein, which binds to target transcripts which are subsequently converted into dsRNA *via* RNA-dependent RNA polymerases (Peragine et al., 2004; Vazquez et al., 2004; Yoshikawa et al., 2005). One prominent example of this lncRNA-type is the pri-miR162a, whose miRNA product *MIR162a* putatively acts as substrates of DCL1 (Hirsch et al., 2006).

In search for secondary structures from Bsp*UPG-2* similar to those mentioned above, database screens for known *Boechera*-specific and other plant sRNAs for homologous sequences were unfruitful. Despite these negative results, computer analysis of the RNA folding probability for different window slices (*i.e.* 50nt to 300nt with step size=10nt and window delta=10nt) of the complete gene using the ViennaRNA package (Hofacker et al., 1994) detected several sections of Bsp*UPG-2* which form with high probability non-random and stable secondary structures (Fig. 19, Supplemental Figure 2 and Supplemental Tables 13 and 14). According to previous studies the minimal folding free energy index (MFEI) is an appropiate measure to distinguish miRNAs from other non-coding and coding RNAs, whereby 90% of miRNA precursors have a MFEI greater than 0.85 (Zhang et al., 2006). Interestingely, seven of the eight detected secondary structures have MFEIs greater than 1.07 and also fulfill other criterias for miRNAs (*e.g.* elevated A+U content, Supplemental Table 15). Surprisingly, these potential miRNAs were not detected in a previous screen (Amiteye et al., 2011), which might be due to their tissue-specific and short-term upregulation at the onset of male meiosis or due to shortcomings of previous analyses which often refer to homology-based screens, whereas Bsp*UPG-2* demonstrated no homology to any known gene in close relative species.

In how far these conserved, structured and highly expressed RNA domains could be functional elements that play a role in posttranscriptional regulation of target genes in *trans* remains open as most of them demonstrated only minor homologies with known protein-coding genes. However, the detection of a stable secondary structure in a region which is highly homologous to a known protein-coding gene with translation elongation activities during polypeptide synthesis at the ribosome and activities in signal transduction (*i.e.* npcRNA 5 similar to GTP binding Elongation factor Tu/EF-1A family protein; *E*-value=7.00E-24, GO:0003746, AT4G02930), could be a first indication for a HDGS function of Bsp*UPG-2*. The highly conserved eukaryotic EFTU/EF-1A is involved in many cellular processes in plants, and hence modulation of EFTU/EF-1A

activity would have tremendous effects to the translation efficieny of many tRNAs, which it binds in a GTP-dependent reaction to the acceptor site of ribosomes (Fu et al., 2012). Interestingely, such modulation was observed for a homologous factor in *Xenopus* during meiotic progression, were some but not all subunits of EF-1 become phosphorylated by cdc2 kinase (*i.e.* amongst other kinases) - a metaphase promoting factor - during prophase to metaphase transition of meiotic cell division, resulting in an enhancernent of elongation activity (Bellé et al., 1990; Peters et al., 1995). It would now be of great interest to investigate if and how npcRNA 5 could be involved *via* interaction with EFTU/EF-1A in the regulation of protein synthesis during meiotic maturation of male gametes in *Boechera*.

Noteworthy in this context is the observation that one of the other two sequence fragments of the 5`-end of Bsp*UPG-2* is also homologous to a gene with nucleotide binding function (RNAR; GO:0003723), whereas the *Arabidopsis* homolog of Bsp*HRD3* (AT1G18260) is involved in vesicle transport from the endoplasmic reticulum (ER) to the Golgi bodies (GO:0030433) and is associated with salt stress (GO:0042538). Expression of *Arabidopsis* homologs of all three parental genes was found across pollen development (Borges et al., 2008), and interestingly, greater abundance and development of endoplasmic reticulum, Golgi bodies, polysomes and mitochondrial cristae was found in unreduced compared to meiotically derived egg cells in mature aposporous embryo sacs of *P. ciliare* (Naumova and Vielle-Calzada 2001), an observation which was related to early egg cell maturation and the loss or truncation of the quiescent phase of egg cell development in apomicts. In this context an analysis of relative mRNA levels of the parental gene Bsp*HRD3*, which is involved in the ER to Golgi transport and which is actively transcribed in sexual and apomictic genotypes, would give further insights into its regulation and role during unreduced pollen formation in pseudogamous apomicts. Furthermore a functional analysis of the collection of Bsp*UPG-2* npcRNAs should help us to better grasp the role of Bsp*UPG-2* in unreduced pollen formation.

To summarize, despite variability for unreduced pollen formation in apomictic *Boechera*, a single novel transcription unit (Bsp*UPG-2*) is consistently upregulated in apomictic flower tissues at the PMC stage. Bsp*UPG-2* has a chimeric sequence structure which might reflect the interspecific hybridization history of this genus. Whereas many studies have focused on mutation accumulation and deregulation with respect to origins of apomixis elements (*e.g.* Tucker et al. (2003), d'Erfurth et al.

(2008)), the emergence of novel genes in apomicts has not been appreciated, although various identified apomixis-associated loci suggest species-specific inheritance of this trait (Grossniklaus et al., 2001) for which "gains in function" are required (Vielle-Calzada et al., 1996). The identification of the novel apo-specific Bsp*UPG-2* however, supports the HFA theory, which proposes that apomeiosis, and in a broader perspective apomixis, originates from hybrid-specific "genome collisions" and associated induction of gene duplication and TE activation (Carman 1997). How Bsp*UPG-2* has undergone neofunctionalization to develop a hypothesized *trans*-regulatory function remains to be clarified.

5.6 Phylogeographic distribution of Bsp*UPG-2*

*5.6.1 Correlation of Bsp*UPG-2 *with male apomeiosis on a large geographical scale*

Classification of a particular *Boechera* taxon as either sex or apomictic is influenced by sampling and quantitative variation for the trait, as demonstrated for example in earlier experiments where relatively few crosses per plant were assessed (Schranz et al., 2005), or where extensive flow cytometric seed screen data were collected for relatively few genotypes (Aliyu et al., 2010). Furthermore, a classification based on morphological traits (Al-Shehbaz 2010, Kiefer 2012) is not always consistent with molecular phylogenies (Beilstein et al., 2006).

The goal of this study is to shed light on a reproductive-mode based classification of *Boechera* taxa by examining the phylogeographic distribution and abundance of the newly characterized candidate gene (Bsp*UPG-2*) on a genus-wide and continental-wide scale. A correlative analysis with respect to its presence in apomictic *Boechera* can furthermore be used to infer its hypothesized importance for balanced endosperm formation.

The chimeric structure of Bsp*UPG-2*, which is strictly conserved in apomicts, provides an excellent apo-specific polymorphism (Fig. 24) to test the candidate gene's correlation with diploid and polyploid apomicts which were previously screened for quantitative variation for different apomixis components (Aliyu et al., 2010). The expected high correlation between the marker and apomixis was confirmed in diploid (95.24%, *N*=21) and triploid (88.89%, *N*=9) apomictic accessions (Supplemental Table 25). Furthermore, Bsp*UPG-2* was detected in a few sexual accession (30.23%, *N*=43), an observation which is consistent with the potential of some sexual *Boechera* to produce unreduced pollen (Kantama et al., 2007). But considering that no

transcriptional activity of Bsp*UPG-2* in sexuals was detected, our working hypothesis is that Bsp*UPG-2* evolved gradually (*i.e.* duplication, sequential insertion of foreign gene fragments) in a non-recombining region (*e.g.* pericentromere) of sexual genomes under little to no selection pressure. The hypothesis predicts that later recurrent hybridization events between sexuals not only supported the transition to apomixis through gene regulatory changes in the hybrid progeny (Carman 1997; Beck et al., 2011), but additionally activated the transcription of Bsp*UPG-2* in apomicts, which would be consistent with our gene expression data. Therefore, footprints of the candidate gene evolution, such as gene fragments, are likely present in some sexuals and hence may have served as templates for the PCR-based screen and subsequent positive amplifications.

Nonetheless, its high correlation with FCSS-based quantitative measurements across 18 taxa, apomicts of varying ploidy (diploid and triploid) and phylogenetic lineages (I to III) support its use as a marker for analysing genus-wide apomixis frequencies.

*5.6.2 Ubiquitous distribution of Bsp*UPG-2 *across* Boechera *genus attests to its single origin*

A survey of 1576 accessions representing 102 taxa using Bsp*UPG-2* as an apomixis-specific marker demonstrated its ubiquitous abundance in all *Boechera* chloroplast haplotype lineages, including taxa of the ancient AB haplotype in addition to closely-related genera (Figs. 26 and 27, Supplemental Tables 26 and 27). Bsp*UPG-2* is apparently underrepresented in lineages IV and V, which is consistent with previous data showing that these *Boechera* are most closely related to *Borodinia* which exhibits great differences in life form, habitat preferences and morphology (Kiefer et al., 2009). Together, this distribution has several far-reaching implications.

Sharbel et al. (2009) hypothesized that inter-breeding between different *Boechera* taxa led to the independent and convergent expression of apomeiosis in different apomictic lineages. Our results contrast this hypothesis in so far, as at least the genetic base for apomeiosis (*i.e.* for male-apomeiosis) seems not to have evolved convergently as the ubiquitous distribution of the unique apomeiosis-associated Bsp*UPG-2* factor across all *Boechera* chloroplast haplotype lineages point to a single origin. Furthermore, the ubiquitous distribution of Bsp*UPG-2* in all *Boechera* chloroplast haplotype lineages and approx. half of the tested haplotypes supports its hypothesized selective advantage for apomictic *Boechera* in addition to its importance for balanced endosperm formation.

Our marker-based analysis has provided some results which are contradictory to previously-published data (Kiefer and Koch 2012). For example, some taxa that were classified as pure sexual (Schranz et al., 2005) have now been found to contain apomicts (*e.g. B. stricta*, Fig. 27B, Supplemental Table 26), while taxa formerly considered as apomictic (Schranz et al., 2005; Kiefer and Koch 2012) were shown to contain sexual and apomictic members (*e.g. B. divaricarpa*, Fig. 27B, Supplemental Table 26). Furthermore, our marker analysis provides evidence for the existence of both sexual and apomictic *B. microphylla* (Fig. 27B), and could explain contrasting results based on the correlation between microsatellite heterozygozity and apomictic pollen frequencies, whereby highly heterozygous individuals had no apomictic pollen (Beck et al., 2011).

The observation of wide-ranging sympatry between apomicts and sexuals (*i.e.* 63.36% of all taxa with $N \geq 5$ tested accessions, Figs. 27 and 28) is consistent with Mogie's hypothesis (Britton and Mogie 2001; Mogie et al., 2007), whereby the male function can favor long-term coexistence of sexuals and asexuals in contrast to the female function which is associated with the establishment of asexuality (*i.e.* reproductive assurance).

Bsp*UPG-2* was only detected in two single individuals of each of one species from close relative Boechereae genera (*e.g. Polyctenium fremontii*, Supplemental Tables 26 and 27). As these results could be explained by the hypothesized presence of Bsp*UPG-2* gene fragments in sexual genotypes (see above) we favor a genus-specific origin over an origin at the root of the tribe (Fig. 30, 2). The genus-specific distribution of Bsp*UPG-2* could be explained by the observation that certain germline-specific genes, such as has been demonstrated for gamete recognition genes, rapidly diverge as a result of adaptive evolution and are not conserved over species borders (Swanson and Vacquier 2002). The former argument is strengthened by the observation that no apomixis has been reported for any close relatives, and that the apomictic mechanism seems to differ on a species level, which was shown for various apomictic grass species for which different genetic loci were proposed (Grimanelli et al., 2001; Grossniklaus et al., 2001).

The spatial analysis exhibited no global differences in distribution pattern (*e.g.* homogeneous *versus* patchy) between sexual and apomictic *Boechera* (Fig. 28A, B and E). The observed frequency differences for Bsp*UPG-2* carriers and noncarriers at the two common distribution maxima can be explained by a lower abundance of Bsp*UPG-2* in eastern North American accessions, which resulted in decreased frequencies of long

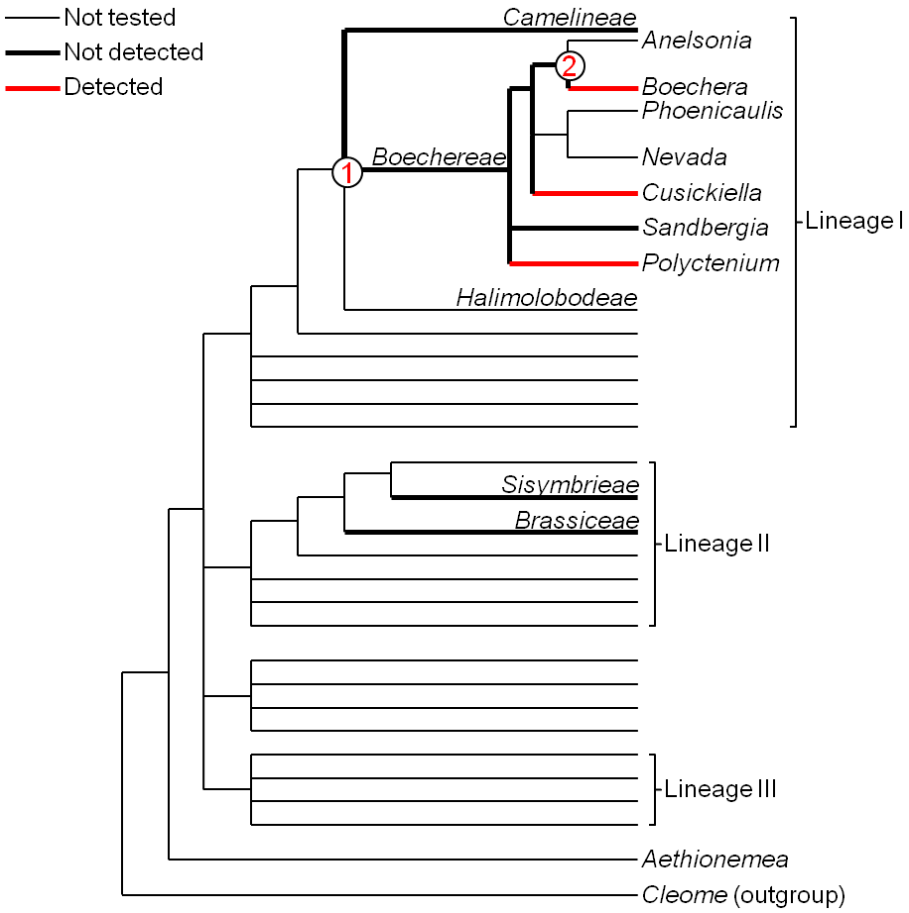

Figure 30. Putative origins of Bsp*UPG-2*.
Phylogenetic relationships among tribes of the Brassicaceae and among genera of the tribe *Boechera* with designation of Bsp*UPG-2* occurrence (modified from Beilstein et al. (2006)). Encircled numbers refer to putative origins of Bsp*UPG-2* at the root of the tribe Boechereae (1) and at the root of the genus *Boechera* (2).

distance pairs. The equality of spatial patterns was mirrored by similar ecological niche models based on similar habitats for sexual and apomictic accessions (Figs. 28A-C).

The majority of tested WWF terestrial ecoregions is shared by apomicts and sexuals with similar proportions (Fig. 29). One explanation for the observed linked distribution of apomicts to sexuals is the growing evidence that apomixis in *Boechera* is the result of a cascade of gene regulatory changes following interspecific hybridization between sexual individuals (Schranz et al., 2005; Kantama et al., 2007; Beck et al., 2011), and hence the distributional linkage is a reflection of their parental relationships (Asker and Jerling 1992; Mogie 1992; Carrillo et al., 2002; Hörandl 2006). Another possible cause for the apparent sympatry (*i.e.* syntopy has not been examined here) between sexual and apomictic *Boechera* could be the impairment of the male function in apomicts, which can reduce the costs of sexuality and enable its invasion into asexual populations

(Mogie et al., 2007; Voigt et al., 2007). The molecular marker-assisted observed sympatry of sexual and apomictic *Boechera* across the North American continent (Figs. 26, 28 and 29) is consistent with the hypothesis that hybridity, rather than polyploidy, is linked to the expression of the apomixis (Beck et al., 2011). In contrast, the present results together with other examples of long-term coexistence of sexuals and apomicts (*e.g. T. officinale*, Verduijn et al. (2004), Hörandl (2006)) are contradictory to previously-published data which demonstrated habitat differentiation between sexual and apomictic populations, such as in the *R. auricomus* complex (Hörandl and Paun 2007) or *Antennaria* (O'Connell and Eckert 1999). As geographic parthenogenesis could also be expressed by contrasting microhabitats as shown for *T. officinale* (Verduijn et al., 2004), one explanation for our results could be that our use of a course-scaled ecological niche model (based on a 5 km^2 grid) prevented the detection of subtle niche differences. But, assuming unbiased sampling, the former explanation is not convincing since only 0.06% of all calculated pairwise distances for sexuals and apomicts together are below the threshold of 5km distance between two accessions. Alternatively, Mogie et al. (2007) stated that for populations at marginal sites, the overall effect of gene flow between well-adapted asexual residents and maladapted sexual immigrants from populations at core sites (*i.e.* asexual pollen fertilizes sexual egg cell) increases the level of adaptedness of the sexual component and lowers the level of the asexual component to potentially result in an increase in period of coexistence. In addition, Mogie et al. (2007) assumed a similar effect for the opposite scenario, were asexual immigration into core population sites is observed and predicts an increased level of adaptedness of the asexual component while lowering the level of the sexual component of the same population. Both trends in gene flow together could also explain extented periods of sympatry of sexual and apomictic *Boechera* accessions. Finally, the diverging results could be explained by parental sexual *Boechera* accessions which continuously generate new asexual accessions, which would lead to younger asexual lineages relative to sexual lineages through neutral clonal turnover as has been shown in a model of dynamic equilibrium between asexual lineage generation and neutral loss (Janko et al., 2008).

Besides the generally observed sympatry between sexuals and apomicts, the latter tend towards higher - although on low level - tolerance of annual temperature ranges, higher alitudes and lower temperatures (Figs. 28G and H, Supplemental Figure 7). In total, only in 9 of 20 tested environmental variables significantly differed between both groups (Supplemental Figure 7). The observed marginal differences in altitude and

prevalently in temperature-related bioclimate variables cannot be discarded, but could rather reflect the emergence of local niche differentiation or contrasting microhabitats between apomictic and sexual accessions according to a broadening of ecological tolerances (*sensu* Hörandl (2006) and (Mogie et al. (2007)), rather than being historical carry-overs of an established geographical parthenogenesis as proposed for other species (Bierzychudek 1985; de Kovel and de Jong 2000; Hörandl and Paun 2007), which is also consistent with the relatively young age of the genus (Koch et al., 2001; Dobeš et al., 2004a; Beilstein et al., 2006; Kiefer et al., 2009; Kiefer and Koch 2012). Considering the latter assumption in conjunction with polyploidization and its disruptive consequences for the male function in apomicts, both would explain the observed distribution pattern in which the ranges of sexuals coincides for the most part with that of apomicts (Mogie et al., 2007).

An interesting aspect thereby is that sexuals occupy extreme altitudes whereas apomicts tend to higher altitudes in general (Fig. 28G). Similar observations were made with subnival to nival plants of the European Alps (Hörandl et al., 2011), in addition to hypotheses which revealed that apomixis turns out to be very rare in plants at extremely high elevations declining from high to very high altitudes (~1800 to 2900m; Gustafsson (1952; 1953)). It is speculated that the evolutionary advantages of gametophytic apomixis for colonization, such as pollinator-independency (*i.e.* reduced pollinator activities caused by unfavorable weather conditions in high altitudes, McCall and Primack (1992)) and rapid development, are not strong enough to establish apomixis frequently at extreme elevations and hence other strategies, such as facultative selfing, are functionally easier to establish (Hörandl 2006; Hörandl et al., 2011).

Nonetheless, our results based on ecological niche modeling should be interpreted with caution as they often miss environmental variables (*e.g.* solar radiation, soil characteristics), and are prone to putative sampling bias (*i.e.* bias through missing 100% penetrance of the molecular marker Bsp*UPG-2* in apomicts in addition to low frequencies in sexuals) or low spatial and temporal resolution of the datasets.

5.7 Final remarks

This study provides a detailed analysis of the first candidate gene reported for unreduced pollen formation in *Boechera* using a broad variety of currently available procedures and resources which will also enable the rapid discovery of more candidate genes towards the improvement of crop plants.

We present an apomeiosis candidate gene which fulfills most of the previously proposed criteria for the genetic control of apomixis components. Bsp*UPG-2* is only present in apomictic *Boechera*, and its transcript is highly abundant at the onset of male meiosis in apomicts, which together imply that it is a "gain in function" factor (Vielle-Calzada et al., 1996; Grimanelli et al., 2001). Bsp*UPG-2* is putatively located in a non-recombining region were it evolved in a stepwise fashion (*e.g.* duplication, sequential insertion of foreign gene fragments, exonization), an observation which was also made for apomeiotic loci in other aposporous (Ozias-Akins et al., 1998) and diplosporous apomicts (Noyes and Rieseberg 2000). The structure and putative genesis of Bsp*UPG-2* reflect molecular signatures of apomictic ovule transcriptomes which implicated both hybridization and gene duplication in the switch from sexuality to apomixis (Sharbel et al., 2009). Comparable to other known apomixis candidate factors, Bsp*UPG-2* is likely genus-specific, and its homo- or hemizygous status in addition to its presence in obligate and facultative apomicts attests to its proposed dominant genetic control of apomixis (Grossniklaus et al., 2001). Furthermore, the chimeric appearance of Bsp*UPG-2*, its high abundance in both diploid and tripoid apomicts and the sympatric occurrence of most of sexual and apomictic *Boechera* is in agreement with the hypothesis that the hybrid constitution of apomictic lineages, rather than polyploidy causes the expression of the apomixis (Bicknell & Koltunow, 2004).

The detailed analyses of Bsp*UPG-2* point to a complex regulatory network for unreduced pollen formation, which is consistent with previous reports which emphasize the role of long non-coding mRNAs (Ansaldi et al., 2000) and miRNAs (Olmedo-Monfil et al., 2010; Amiteye et al., 2011) in the control of gamete formation, and illustrates the importance of novel pathways for regulation of apomixis components. Nonetheless, its detailed contribution to apomixis remains to be elucidated. In future studies we will characterize the molecular function and putative interaction of Bsp*UPG-2* with target transcripts or proteins using biochemical assays, such as the electromobility shift assay (EMSA) and *in vitro*-expression studies (*e.g.* in wheat germline systems). To further advance the study of the contribution of Bsp*UPG-2* to

unreduced pollen formation transformation into sexual *Boechera* and crop species in parallel to knockdown experiments in apomictic *Boechera* will be assessed.

A better understanding of the molecular mechanism for unreduced pollen formation will be vital in our attempt to engineer apomictic crops, by stably expressing all components of apomixis in crop plants, including a stable endosperm formation (*c.f.* Birchler (1993)). In this regard the identification of Bsp*UPG-2* is a major step forward to learn more about the genomic history of apomicts and the putative important regulatory roles of mRNA-like non-protein-coding RNAs in gametophytic apomixis.

6 References

Ahmed, I., Sarazin, A., Bowler, C., Colot, V. and Quesneville, H. (2011). "Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*." Nucleic Acids Research **39**(16): 6919–6931.

Al-Shebaz, I.A. and Windham, M. (1993+). *In:* Flora of North America Editorial Committee [eds.], Flora of North America North of Mexico, 16+ vols. . New York and Oxford, Oxford University Press. **7:** 348–367.

Al-Shehbaz, I.A. (2007). "Brassicaceae *Sandbergia perplexa* (L.F.Hend.)." Harvard Papers in Botany **12**(2): 426.

Al-Shehbaz, I.A., Beilstein, M.A. and Kellogg, E.A. (2006). "Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview." Plant Systematics and Evolution **259**: 89–120.

Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H. and Durbin, R. (2011). "Dindel: Accurate indel calls from short-read data." Genome Research **21**: 961-973.

Albertini, E., Marconi, G., Barcaccia, G., Raggi, L. and Falcinelli, M. (2004). "Isolation of candidate genes for apomixis in *Poa pratensis* L." Plant Molecular Biology **56**: 879–894.

Albertini, E., Marconi, G., Reale, L., Barcaccia, G., Porceddu, A., Ferranti, F. and Falcinelli, M. (2005). "SERK and APOSTART. Candidate genes for apomixis in *Poa pratensis*." Plant Physiology **138**: 2185–2199.

Aliyu, O.M., Schranz, M.E. and Sharbel, T.F. (2010). "Quantitative variation for apomictic reproduction in the genus *Boechera* (Brassicaceae)." American Journal of Botany **97**(10): 1719-1731.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389–3402.

Amien, S., Kliwer, I., Marton, M.L., Debener, T., Geiger, D., Becker, D. and Dresselhaus, T. (2010). "Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1." PLoS Biology **8**.

Amiteye, S., Corral, J.M., Vogel, H. and Sharbel, T.F. (2011). "Analysis of conserved microRNAs in floral tissues of sexual and apomictic *Boechera* species." BMC Genomics **12**(500).

Ansaldi, R., Chaboud, A. and Dumas, C. (2000). "Multiple *S* gene family members including natural antisense transcripts are differentially expressed during development of maize flowers." The Journal Of Biological Chemistry **275**(31): 24146–24155.

Arabidopsis Genome Initiative (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*." Nature **408**(6814): 796-815.

Armbruster, W.S. and Gobeille Rogers, D. (2004). "Does pollen competition reduce the cost of inbreeding?" American journal of botany **91**(11): 1939–1943.

Armstrong, S.J., Caryl, A.P., Jones, G.H. and Franklin, F.C.H. (2002). "Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in *Arabidopsis* and *Brassica*." Journal of Cell Science **115**: 3645-3655.

Armstrong, S.J. and Jones, G.H. (2001). "Female meiosis in wild-type Arabidopsis thaliana and in two meiotic mutants." Sexual Plant Reproduction **13**: 177–183.

Asker, S.E. and Jerling, L. (1992). Apomixis in plants. Boca Raton, CRC Press.

Au, P.C.K., Zhu, Q.-H., Dennis, E.S. and Wang, M.-B. (2011). "Long non-coding RNA-mediated mechanisms independent of the RNAi pathway in animals and plants." RNA Biology **8**(3): 404-414.

Azumi, Y., Liu, D., Zhao, D., Li, W. and Wang, G. (2002). "Homolog interaction during meiotic prophase I in *Arabidopsis* requires the *SOLO DANCERS* gene encoding a novel cyclin-like protein." EMBO **21**: 3081–95.

Baack, E.J., Whitney, K.D. and Rieseberg, L.H. (2005). "Hybridization and genome size evolution: timing and magnitude of nuclearDNAcontent increases in *Helianthus* homoploid hybrid species." New Phytologist **167**: 623–30.

Baker, B.S., Carpenter, A.T.C., Esposito, M.S., Esposito, R.E. and Sandler, L. (1976). "The genetic control of meiosis." Annual Review of Genetics **10**: 53-134.

Baker, H.G. (1955). "Self-compatibility and establishment after 'long distance' dispersal." Evolution **9**: 347–348.

Barcaccia, G., Mazzucato, A., Falcinelli, M. and Veronesi, F. (1996). "Callose localization in cell walls during meiotic and apomeiotic megasporogenesis in diploid alfalfa (*Medicago* spp.)." Caryologia **49**: 45-56.

Bashaw, E.C. and Hanna, W.W. (1990). In reproductive versatility in the grasses. Apomictic reproduction. G.P. Chapman. Cambridge, Cambridge University Press**:** 100-130.

Bechsgaard, J.S., Castric, V., Charlesworth, D., Vekemans, X. and Schierup, M.H. (2006). "The transition to self-compatibility in *Arabidopsis thaliana* and evolution within *S*-haplotypes over 10 myr." Molecular Biology and Evolution **23**(9): 1741–1750.

Beck, J.B., Alexander, P.J., Allphin, L., Al-Shehbaz, I.A., Rushworth, C., Bailey, C.D. and Windham, M.D. (2011). "Does hybridization drive the transition to asexuality in diploid *Boechera*? ." Evolution **(in press)**.

Beilstein, M.A., Al-Shehbaz, I.A. and Kellogg, E.A. (2006). "Brassicaceae phylogeny and trichome evolution." American Journal of Botany **93**(4): 607–619.

Bell, G. (1982). The Masterpiece of Nature. The Evolution and Genetics of Sexuality. Los Angeles, USA, University of California Press.

Bellé, R., Cormier, P., Poulhe, R., Morales, J., Huchon, D. and Mulner-Lorillon, O. (1990). "Protein phosphorylation during meiotic maturation of *Xenopus* oocytes: cdc2 protein kinase targets." International Journal of Developmental Biology **34**: 111-115.

Bennetzen, J.L. (2005). "Transposable elements, gene creation and genome rearrangement in flowering plants." Current Opinion in Genetics & Development **15**: 621–627.

Berger, F. (1999). "Endosperm development." Current Opinion in Plant Biology **2**: 28–32.

Berger, F. and Twell, D. (2011). "Germline specification and function in plants." Annual Review of Plant Biology **62**: 461–84.

Bernstein, H. (1977). "Germ line recombination may be primarily a manifestation of DNA repair processes." Journal of Theoretical Biology **69**: 371-380.

Bicknell, R.A., Borst, N.K. and Koltunow, A.M. (2000). "Monogenic inheritance of apomixis in two Hieracium species with distinct developmental mechanisms." Heredity **84**(2): 228-237.

Bicknell, R.A. and Koltunow, A.M. (2004). "Understanding apomixis: recent advances and remaining conundrums." The Plant Cell **16**: 228–245.

Bierzychudek, P. (1985). "Patterns in plant parthenogenesis." Experientia **41**: 1255–1264.

Bierzychudek, P. (1987). Patterns in plant parthenogenesis. The evolution of sex and its consequences. S.C. Stearns. Basel, Birkhäuser Verlag**:** 197-217.

Bierzychudek, P. (1989). "Environmental sensitivity of sexual and apomictic *Antennaria*: do apomicts have general-purpose genotypes?" Evolution **43**(7): 1456-1466.

Birchler, J.A. (1993). "Dosage analysis of maize endosperm development." Annual Review of Genetics **27**: 181-204.

Blázquez, M.A., Green, R., Nilsson, O., Sussman, M.R. and Weigel, D. (1998). "Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter." The Plant Cell **10**: 791–800.

Böcher, T.W. (1951). "Cytological and embryologal studies in the amphiapomictic *Arabis holboellii* complex." Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab **6**(7): 1-59.

Böcher, T.W. (1954). "Experimental taxonomical studies in the *Arabis holboellii* complex." Svensk Botanisk Tidskrift **48**: 31-44.

Bonnet, E., Van de Peer, Y. and Rouzé, P. (2006). "The small RNA world of plants." New Phytologist **171**: 451–468.

Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. (2004). "Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences." Bioinformatics **20**: 2911–2917.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J.A. and Becker, J.D. (2008). "Comparative transcriptomics of *Arabidopsis* sperm cells." Plant Physiology **148**: 1168–1181.

Borges, F., Pereira, A.P., Keith Slotkin, R., Martienssen, R.A. and Becker, J.D. (2011). "MicroRNA activity in the *Arabidopsis* male germline." Journal of Experimental Botany **62**(5): 1611–1620.

Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M. and Simon, R. (2000). "Dependence of stem cell fate in *Arabidopsis* on a feedback loop regulated by *CLV3* activity." Science **289**(617-619).

Bretagnolle, F. and Thompson, J.D. (1995). "Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants." New Phytologist **129**: 1-22.

Brewbaker, J.L. (1967). "The distribution and phylogenetic significance of binucleate and trinucleate pollen grains in the angiosperms." American Journal of Botany **54**(9): 1069-1083.

Britton, N.F. and Mogie, M. (2001). "Poor male function favours the coexistence of sexual and asexual relatives." Ecology Letters **4**(2): 116-121.

Brownfield, L. and Köhler, C. (2010). "Unreduced gamete formation in plants: mechanisms and prospects." Journal of Experimental Botany **62**(5): 1659-1668.

Brunner, S., Pea, G. and Rafalski, A. (2005). "Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize." The Plant Journal **43**: 799-810.

Buxton, B.H. and Darlington, C.D. (1932). "Crosses between *Digitalis purpurea* and *Digitalis ambigua*." The New Phytologist **31**(4): 225-240.

Capron, A.Gourgues, M.Neiva, L.S.Faure, J.-E.Berger, F.Pagnussat, G.Krishnan, A.Alvarez-Mejia, C.Vielle-Calzada, J.-P.Lee, Y.-R.Liu, B.Sundaresana, V. (2008). "Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene." The Plant Cell **20**: 3038–3049.

Carman, J. (1997). The gene effect - genome collisions and apomixis. The flowering of apomixis: from mechanisms to genetics. Y. Savidan, J.G. Carmanet al.,. Mexico, CIMMYT, IRD, European Commission DC VI (FAIR). **15**.

Carman, J.G. (1997). "Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony." Biological Journal of the Linnean Society **61**(1): 51–94.

Carman, J.G. (2001). The gene effect: genome collision and apomixis. The flowering of apomixis: from mechanisms to genetic engineering. Y. Savidan, J.G. Carmanet al.,. Mexico, DF, CIMMYT, IRD, European Commission DG VI (FAIR)**:** 95–110.

Carrillo, C., Britton, N.F. and Mogie, M. (2002). "Coexistence of sexual and asexual conspecifics: a cellular automaton model." Journal of Theoretical Biology **217**: 275–285.

Casillas, S., Barbadilla, A. and Bergman, C.M. (2007). "Purifying selection maintains highly conserved noncoding sequences in *Drosophila*." Molecular Biology and Evolution **24**(10): 2222–2234.

Chang, W.C., Lee, T.Y., Huang, H.D., Huang, H.Y. and Pan, R.L. (2008). "PlantPAN: Plant Promoter Analysis Navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene group." BMC Genomics **9**(561).

Charlesworth, B. (1980). "The cost of sex in relation to mating system." Journal of Theoretical Biology **84**: 655-671.

Charlesworth, B. and Charlesworth, D. (1978). "A model for the evolution of dioecy and gynodioecy." American Naturalist **112**: 975–997.

Charlesworth, B., Morgan, M.T. and Charlesworth, D. (1993). "The effect of deleterious mutations on neutral molecular evolution." Genetics **134**: 1289-1303.

Chaudhury, A.M. and Peacock, J.W. (1994). Approaches towards isolating apomictic mutants in Arabidopsis thaliana: Prospects and progress. Apomixis: Exploiting Hybrid Vigor in Rice. G. Kush. Manila, IRRI**:** 66–71.

Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W. and Jin, H. (2010). "siRNAs from miRNA sites mediate DNA methylation of target genes." Nucleic Acids Research **38**(20): 6883–6894.

Chen, L.Z., Miyazaki, C., Kojima, A., Saito, A. and Adachi, T. (1999). "Isolation and characterization of a gene expressed during early embryo sac development in apomictic guinea grass (*Panicum maximum*)." Journal of Plant Physiology **154**: 55–62.

Chen, M., Sanmiguel, P., De Oliveira, A.C., Woo, S.-S., hang, H.Z., Wing, R.A. and Bennetzen, J.L. (1997). "Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes." Proceedings of the National Academy of Sciences **94**: 3431–3435.

Chen, Y.-C.S. and McCormick, S. (1996). "*Sidecar pollen*, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development." Development **122**: 3243-3253

Chen, Z.J., Comai, L. and Pikaard, C.S. (1998). "Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids." Proceedings of the National Academy of Sciences **95**: 14891–14896.

Choulet, F., Wicker, T. and Rustenholz, C. (2010). "Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces." Plant Cell **22**: 1686-1701.

Clement, M., Posada, D. and Crandall, K.A. (2000). "TCS: a computer program to estimate gene genealogies." Molecular Ecology **9**: 1657–1660.

Cock, J.M., Swarup, R. and Dumas, C. (1997). "Natural antisense transcripts of the *S* locus receptor kinase gene and related sequences in *Brassica oleracea*." Molecular Genetics and Genomics **255**: 514-524.

Coen, E.S. and Meyerowitz, E.M. (1991). "The war of the whorls: Genetic interactions controlling flower development." Nature **353**: 31-37.

Comai, L. (2000). "Genetic and epigenetic interactions in allopolyploid plants." Plant Molecular Biology **43**: 387–399.

Connelly, J.C. and Leach, D.R. (2002). "Tethering on the brink: the evolutionarily conserved Mre11-Rad50 complex." Trends in Biochemical Sciences **27**: 410–418.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D. and Doutriaux, M.P. (1999). "Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of *Arabidopsis*." The Plant Cell **11**: 1623–1634.

Crespi, M.D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E. and Kondorosi, A. (1994). "*enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth." European Molecular Biology Organization reports **13**: 5099–5112.

Cresti, M., Blackmore, S. and van Went, J.L. (1992). Atlas of sexual reproduction in flowering plants. Berlin - Heidelberg, Springer - Verlag.

Crick, F. (1979). "Split genes and RNA splicing." Science **204**: 264–271.

d'Erfurth, I., Cromer, L., Jolivet, S., Girard, C., Horlow, C., Sun, Y., To, J.P.C., Berchowitz, L., Copenhaver, G.P. and Mercier, R. (2010). "The CYCLIN-A CYCA1;2/TAM is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition." PLoS Genetics **6**(6).

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., Simon, M., Jenczewski, E. and Mercier, R. (2008). "Mutations in *AtPS1* (*Arabidopsis thaliana Parallel Spindle 1*) lead to the production of diploid pollen grains." PLoS Genetics **4** (11).

Darling, A.C.E., Mau, B., Blattner, F.R. and Perna, N.T. (2004). "Mauve: multiple alignment of conserved genomic sequence with rearrangements." Genome Research **14**: 1394–1403.

Darwin, C. (1877). The different forms of flowers on plants of the same species. London.

de Kovel, C.G.F. and de Jong, G. (2000). "Selection on apomictic lineages of *Taraxacum* at establishment in a mixed sexual-apomictic population." Journal of Evolutionary Biology **13**: 561-568.

de Lucia, F. and Dean, C. (2011). "Long non-coding RNAs and chromatin regulation." Current opinion in plant biology **14**(2): 168-173.

de Storme, N. and Geelen, D. (2011). "The *Arabidopsis* mutant jason produces unreduced FDR male gametes through a parallel-fused spindle mechanism in meiosis II." Plant Physiology **55**(3): 1403-1415.

Deaton, A.M. and Bird, A. (2011). "CpG islands and the regulation of transcription." Genes and Development **25**(10): 1010–1022.

Derrien, T., Guigó, R. and Johnson, R. (2012). "The long non-coding RNAs: a new (p)layer in the "dark matter"." Frontiers in Genetics **10.3389/fgene.2011.00107**

Dickinson, H. and Grant-Downton, R. (2011). Small RNA pathways and their function in the male gametophyte. Non coding RNAs in plants. V.A.E.a.J.B. (eds.). Oxford, Springer-Verlag Berlin Heidelberg**:** 175-191.

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., Xu, C., Li, X., Xiao, J. and Zhang, Q. (2012). "A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice." Proceedings of the National Academy of Sciences **109**(7): 2654-2659

Dobeš, C., Koch, M. and Sharbel, T.F. (2006). "Embryology, karyology and modes of reproduction in the north american genus *Boechera* (Brassicaceae): a compilation of seven decades of research." Annals of the Missouri Botanical Garden **93**: 517–533.

Dobeš, C.H., Mitchell-Olds, T. and Koch, M.A. (2004a). "Extensive chloroplast haplotype variation indicates pleistocene hybridization and radiation of north american *Arabis drummondii*, *A. divaricarpa*, and *A. holboellii* (Brassicaceae)." Molecular Ecology **13**: 349–370.

Doutriaux, M.P., Couteau, F., Bergounioux, C. and White, C. (1998). "Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*." Molecular Genetics and Genomics **257**: 283–91.

Dowd, P.E., Coursol, S., Skirpan, A.L., Kao, T.-H. and Gilroy, S. (2006). "Petunia phospholipase C1 is involved in pollen tube growth." The Plant Cell **18**: 1438-1453.

Dresselhaus, T., Carman, J.G. and Savidan, Y. (2001). Genetic engineering of apomixis in sexual crops: a critical assessment of the apomixis technology. The flowering of apomixis: from mechanisms to genetic engineering. Y. Savidan, J.G. Carmanet al.,. Mexico, D.F., CIMMYT,IRD, European Commission DG VI (FAIR).

Eamens, A., Wang, M.-B., Smith, N.A. and Waterhouse, P.M. (2008). "RNA silencing in plants: yesterday, today, and tomorrow." Plant Physiology **147**(2): 456-468.

East, E.M. (1940). "The distribution of self-sterility in the flowering plants." Proceedings of the American Philosophical Society **82**(4).

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

Eddy, S. (2001). "Non-coding RNA genes and the modern rna world." Nature Reviews **2**: 919-929.

Erdtman, G. (1947). "Suggestions for the classification of fossil and recent pollen grains and spores." Svensk Botanisk Tidskrift **41**: 104-114.

Erdtman, G. (1952). Pollen Morphology and Plant Taxonomy. Angiosperms. Stockholm, Almqvist and Wiksell.

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Mittelsten Scheid, O., Hennig, L. and Köhler, C. (2009). "Imprinting of the Polycomb proup gene *MEDEA* serves as a ploidy sensor in *Arabidopsis*." PLoS Genetics **5**(9): e1000663.

Escobar-Restrepo, J.M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W.C. and Grossniklaus, U. (2007). "The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception." Science **317**: 656–660.

Esteller, M. (2011). "Non-coding RNAs in human disease." Nature Reviews **12**: 861-874.

Estruch, J.J., Kadwell, S., Merlin, E. and Crossland, L. (1994). "Cloning and characterization of a maize pollen-specific calcium-dependent calmodulin-independent protein kinase." Proceedings of the National Academy of Sciences **91**: 8837-8841.

Fan, L.M., Fei Wang, Y., Wang, H. and Wu, W.H. (2001). "In vitro *Arabidopsis* pollen germination and characterization of the inward potassium currents in *Arabidopsis* pollen grain protoplasts." Journal of Experimental Botany **52** (361): 1603-1614.

FANTOM-Consortium (2005) "The transcriptional landscape of the mammalian genome." Science **309**, 1559-1563

Feijó, J.A., Malhó, R. and Pais, M.S. (1995). "Electrical currents, ion channels and ion pumps during germination and growth of pollen tubes." Protoplasma **187**: 155-167.

Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A.A. and Vega, J.M. (1997). "Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes." Genetics **147**: 1381–1387.

Fielding, A.H. and Bell, J.F. (1997). "A review of methods for the assessment of prediction errors in conservation presence/absence models." Environmental Conservation **24**: 38–49.

Finnegan, D.J. (1992). "Transposable elements." Current Opinion in Genetics & Development **2**: 861–867.

Fiston-Lavier, A.-S., Vejnar, C.E. and Quesneville, H. (2012). "Transposable sequence evolution is driven by gene context." arXiv:1209.0176

Fletcher, J.C., Brand, U., Running, M.P., Simon, R. and Meyerowitz, E.M. (1999). "Signaling of cell fate decisions by *CLAVATA3* in *Arabidopsis* shoot meristems." Science **283**: 1911-1914.

Fransz, P.F., Armstrong, S.J., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. and Jones, G.H. (2000). "Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region." Cell **100**: 367-376.

Fryxell, P. (1957). "Mode of reproduction of higher plants." The Botanical Review **23**(3): 135-213.

Fu, J., Momcilovic, I. and Prasad, P.V.V. (2012). "Roles of protein synthesis elongation factor EF-Tu in Heat tolerance in plants." Journal of Botany.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P. and Firoozabad, E. (1983). "Rapid flow cytometric analysis of the cell cycle in intact plant tissues." Science **220**(4601): 1049-1051.

Gallo, P. (2007). "2n gamete formation in the genus *Brachiaria* (Poaceae: Paniceae)." Euphytica **2**: 255-260.

Gaut, B.S. and Doebley, J.F. (1997). "DNA sequence evidence for the segmental allopolyploid origin of maize." Proceedings of the National Academy of Sciences **94**: 6809–6814.

Geitmann, A. and Palanivelu, R. (2007). "Fertilization requires communication: signal generation and perception during pollen tube guidance." Floriculture and Ornamental Biotechnology **1**(2): 77-89.

Gladyshev, E. and Meselson, M. (2008). "Extreme resistance of bdelloid rotifers to ionizing radiation." Proceedings of the National Academy of Sciences **105**(13): 5139-5144

Glover, J., Grelon, M., Craig, S., Chaudhury, A. and Dennis, E. (1998). "Cloning and characterization of *MS5* from *Arabidopsis*: a gene critical in male meiosis." The Plant Journal **15**: 345–356.

Goldberg, R.B. (1993). "Anther development: basic principles and practical applications." The Plant Cell **5**: 1217-1229.

Grant-Downton, R., Hafidh, S., Twell, D. and Dickinson, H.G. (2009a). "Small RNA pathways are present and functional in the angiosperm male gametophyte." Molecular Plant **2**: 500-512.

Grant-Downton, R., Le Trionnaire, G., Schmid, R., Rodriguez-Enriquez, J., Hafidh, S., Mehdi, S., Twell, D. and Dickinson, H. (2009b). "MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*." BMC Genomics **10**(643).

Grelon, M., Vezon, D., Gendrot, G. and Pelletier, G. (2001). "AtSPO11-1 is necessary for efficient meiotic recombination in plants." EMBO **20**: 589–600.

Grimanelli, D. (2003). "Heterochronic expression of sexual reproductive programs during apomictic development in Tripsacum." Genetics **165**: 1521–1531.

Grimanelli, D. (2011). "Epigenetic regulation of reproductive development and the emergence of apomixis in angiosperms." Current Opinion in Plant Biology **15**: 1–6.

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

Grimanelli, D., Leblanc, O., Perotti, E. and Grossniklaus, U. (2001). "Developmental genetics of gametophytic apomixis." Trends in Genetics **17** (10): 597-604.

Grossniklaus, U. (2001). From sexuality to apomixis: Molecular and genetic approaches. The Flowering of Apomixis: From Mechanisms to Genetic Engineering. Y. Savidan, J.G. Carmanet al.,. Mexico, CIMMYT, IRD European Commission DG VI (FAIR)**:** 168-211.

Grossniklaus, U. (2001). From sexuality to apomixis: molecular and genetic approaches. The flowering of apomixis: from mechanisms to genetic engineering. Y. Savidan, J.G. Carman et al.,. Mexico, D.F., CIMMYT,IRD, European Commission DG VI (FAIR).

Grossniklaus, U., Nogler, G.A. and van Dijk, P.J. (2001). "How to avoid sex: the genetic control of gametophytic apomixis." The Plant Cell: 1491-1497.

Grossniklaus, U., Spillane, C., Page, D.R. and Koehler, C. (2001). "Genomic imprinting and seed development: endosperm formation with and without sex." Current Opinion in Plant Biology **4**: 21–27.

Guerin, J., Rossel, J.B., Robert, S., Tsuchiya, T. and Koltunow, A. (2000). "A *DEFICIENS* homologue is down-regulated during apomictic initiation in ovules of *Hieracium*." Planta **210**: 914–920.

Guitton, A.-E. and Berger, F. (2005). "Control of reproduction by Polycomb Group complexes in animals and plants." International Journal of Developmental Biology **49**: 707-716.

Gupta, P., Shivanna, K.R. and Mohan Ram, H.Y. (1998). "Pollen-pistil interaction in a non-pseudogamous apomict, *Commiphora wightii*." Annals of Botany **81**: 589-594.

Gupta, S., Gallavotti, A., Stryker, G., Schmidt, R.J. and Lal, S.K. (2005). "A novel class of *Helitron*-related transposable elements in maize contain portions of multiple pseudogenes." Plant Molecular Biology **57**: 115-127.

Gustafsson, A. (1946). "Apomixis in higher plants. Part I. The mechanisms of apomixis." Lunds Univ. Arsskr. **2**(42): 1-66.

Gustafsson, A. (1947a). "Apomixis in higher plants. Part 2. The causal aspect of apomixis." Lunds Univ. Arsskr. **2**(43): 71-178

Gustafsson, A. (1947b). "Apomixis in higher plants. Part 3. Biotype and species formation." Lunds Univ. Arsskr. **2**(43): 183-370.

Gustafsson, A. (1952). "Apomixis in higher plants. Part I. The mechanisms of apomixis." Lunds Univ. Arsskr. **42**: 1-66.

Gustafsson, A. (1953). "Apomixis in higher plants. Part II. The causal aspects of apomixis." Lunds Univ. Arsskr. **43**: 71–178.

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

Haldane, J.B.S. (1932). The causes of evolution. London.

Hamamura, Y.Saito, C.Awai, C.Kurihara, D.Miyawaki, A.Nakagawa, T.Kanaoka, M.M.Sasaki, N.Nakano, A.Berger, F.Higashiyama, T. (2011). "Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in *Arabidopsis thaliana*." Current Biology **21**(6): 497–502.

Hanna, W.W. and Bashaw, E.C. (1987). "Apomixis: its identification and use in plant breeding." Crop Science **27**(6): 1136-1139

Harlan, J.R., Brooks, M.H., Borgaonkar, D.S. and DeWet, J.M.J. (1964). "Nature and inheritance of apomixis in *Bothriochloa* and *Dichantium*." Botanical Gazette **125**: 41–46.

Harlan, J.R. and deWet, J.M.J. (1975). "On Ö. Winge and a prayer: the origins of polyploidy." The Botanical Review **41**: 361-390.

Harrison, C.J., Alvey Elizabeth and Henderson, I.R. (2010). "Meiosis in flowering plants and other green organisms." Journal of Experimental Botany **61**(11): 2863-2875.

Haughn, G.W. and Sommerville, C.R. (1988). "Genetic control of morphogenesis in Arabidopsis." Developmental Genetics **9**: 73-89.

Hazzouri, R.M., Mohajer, A., Dejak, S.I., Otto, S.P. and Wright, S.I. (2008). "Contrasting patterns of transposable-element insertion polymorphism and nucleotide diversity in autotetraploid and allotetraploid *Arabidopsis* species." Genetics **179**: 581-592.

Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W. and Dennis, E.S. (2006). "The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex." The Plant Journal **46**: 183-192.

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

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. (2005). "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology **25**: 1965-1978.

Hill, W.G. and Robertson, A. (1966). "The effect of linkage on limits to artificial selection." Genetical Research **8**: 269–294.

Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C., d'Aubenton-Carafa, Y. and Crespi, M. (2006). "Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts." Plant Physiology **140**(4): 1192-1204.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S.L., Tacker, M. and Schuster, P. (1994). "Fast folding and comparison of RNA secondary structures." Chemical Monthly **125**: 167-188.

Honys, D. and Twell, D. (2004). "Transcriptome analysis of haploid male gametophyte development in Arabidopsis." Genome Biology **5**(R85).

Hörandl, E. (2006). "The complex causality of geographical parthenogenesis." New Phytologist **171**: 525–538.

Hörandl, E., Dobeš, C., Suda, J., Vít, P., Urfus, T., Temsch, E.M., Cosendai, A.-C., Wagner, J. and Ladinig, U. (2011). "Apomixis is not prevalent in subnival to nival plants of the European Alps." Annals of Botany **108**(2): 381–390.

Hörandl, E. and Paun, O. (2007). Patterns and sources of genetic diversity in apomictic plants: implications for evolutionary potentials. Apomixis: evolution, mechanisms and perspectives. E. Hörandl, U. Grossniklauset al.,. Rugell, Liechtenstein, A. R. G. Gantner Verlag**:** 169-194.

Hörandl, E. and Temsch, E.M. (2009). "Introgression of apomixis into sexual species is inhibited by mentor effects and ploidy barriers in the *Ranunculus auricomus* complex." Annals of Botany **104**: 81–89.

Hosouchi, T., Kumekawa, N., Tsuruoka, H. and Kotani, H. (2002). "Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3." DNA Research **9**: 117–121.

Huck, N., Moore, J.M., Federer, M. and Grossniklaus, U. (2003). "The *Arabidopsis* mutant feronia disrupts the female gametophytic control of pollen tube reception." Development **130**: 2149–2159.

Hülskamp, M., Kopczak, S.D., Horejsi, T.F., Kihl, B.K. and Pruitt, R.E. (1995). "Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*." The Plant Journal **8**(5): 703-714.

Hulskamp, M., Parekh, N.S., Grini, P., Schneitz, K., Zimmermann, I., Lolle, S.J. and Pruitt, R.E. (1997). "The *STUD* gene is required for male-specific cytokinesis after telophase II of meiosis in *Arabidopsis thaliana*." Developmental Biology **187**: 114–124.

Hunt, A.G. (1994). "Messenger RNA 3' end formation in plants." Annual Reviews in Plant Physiology and Plant Molecular BioIogy **45**: 47-60.

Indriolo, E. and Goring, D.R. (2010). "Pollen gets more complex." Science **330**: 767.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T. and Berger, F. (2007). "Distinct dynamics of HISTONE3 variants between the two fertilization products in plants." Current Biology **17**(12): 1032–1037.

Ingouff, M., Sakata, T., Li, J., Sprunck, S., Dresselhaus, T. and Berger, F. (2009). "The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*." Current Biology **19**: R19–R20.

Ishiguro, S., Kawai-Oda, A., Ueda, J., Nishida, I. and Okada, K. (2001). "The *DEFECTIVE IN ANTHER DEHISCENCE1* gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in *Arabidopsis*." Plant Cell **13**: 2191–2209.

Jacobsen, S.E. and Meyerowitz, E.M. (1997). "Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*." Science **277**: 1100–1103.

Jain, S.K. (1976). "The evolution of inbreeding in plants." Annual Review of Ecology and Systematics **7**: 469–495.

Janko, K., Drozd, P., Flegr, J. and Pannell, J.R. (2008). "Clonal turnover versus clonal decay: a null model for observed patterns of asexual longevity, diversity and distribution." Evolution **62**: 1264–1270.

Jankun, A. and Kovanda, M. (1987). "Apomixis and origin in *Sorbus bohemica* (Embryological studies in *Sorbus* 2)." Preslia **59**: 97–116.

Jefferson, R.A. (1993). Strategic development of apomixis as a general tool for agriculture. Proceeding international workshop on apomixis in rice. K.J. Wilson. Changsa, China, Canberra, Australia, CAMBIA**:** 206-217.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R. and R., W.S. (2004). "Pack-MULE transposable elements mediate gene evolution in plants." Nature **431**(7008): 569-573.

Johnston, S.A., den Nijs, T.P.M., Peloquin, S.J. and Hanneman, J., R.E. (1980). "The significance of genic balance to endosperm development in interspecific crosses." Theoretical and Applied Genetics **57**: 5-9

Jurka, J. (1998). "Repeats in genomic DNA: mining and meaning." Current Opinion in Structural Biology **8**: 333-337

Kaessmann, H. (2010). "Origins, evolution, and phenotypic impact of new genes." Genome Research **20**: 1313–1326.

Kalla, S.E. and Ashman, T.-L. (2002). "The effects of pollen competition on progeny vigor in *Fragaria virginiana* (Rosaceae) depend on progeny growth environment." International Journal of Plant Sciences **163**(2): 335–340.

Kamalay, J.C. and Goldberg, R. (1980). "Regulation of structural gene expression in tobacco." Cell **19**: 935-946.

Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., Vries, S.d. and de Jong, H. (2007). "Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes." Proceedings of the National Academy of Sciences **104** (35): 14026–14031

Kapitonov, V.V. and Jurka, J. (2007). "Helitrons on a roll: Eukaryotic rolling-circle transposons." Trends in Genetics **23**: 521–529.

Kashkush, K., Feldman, M. and Levy, A.A. (2003). "Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat." Nature Genetics **33**: 102–106.

Kaul, M.L.H. and Murthy, T.G.K. (1985). "Mutant genes affecting higher plant meiosis." Theoretical and Applied Genetics **70**: 449-466.

Kavanaugh, L.A. and Dietrich, F.S. (2009). "Non-Coding RNA Prediction and Verification in *Saccharomyces cerevisiae*." PLoS Genetics **5**(1).

Kazazian Jr, H.H. (2004). "Mobile elements: drivers of genome evolution." Science **303**: 1626-1632.

Kentner, E.K., Arnold, M.L. and Wessler, S.R. (2003). "Characterization of high-copy-number retrotransposons from the large genomes of the louisiana iris species and their use as molecular markers.." Genetics **164**: 685-697.

Khush, G., Brar, D.S., Bennett, J. and Virmani, S.S. (1994). Apomixis for rice improvement. Apomixis: exploiting hybrid vigor in rice. G. Khush. Manila, IRRI**:** 15–21.

Kiefer, C., Dobeš, C. and Koch, M.A. (2009). "*Boechera* or not? Phylogeny and phylogeography of eastern North American *Boechera* species (Brassicaceae)." Taxon **58**(4): 1109-1121.

Kiefer, C., Dobeš, C., Sharbel, T.F. and Koch, M.A. (2009). "Phylogeographic structure of the chloroplast DNA gene pool in North American *Boechera* – A genus and continental-wide perspective." Molecular Phylogenetics and Evolution **52** 303–311.

Kiefer, C. and Koch, M.A. (2012). "A continental-wide perspective: the genepool of nuclear encoded ribosomal DNA and single-copy gene sequences in North American *Boechera* (Brassicaceae)." PLoS ONE **7**(5).

Kim, E.-D. and Sung, S. (2012). "Long noncoding RNA: unveiling hidden layer of gene regulatory networks." Trends in Plant Science **17**(1).

Kim, H.U., Cotter, R., Johnson, S., Senda, M., Dodds, P., Kulikauska, R., Tang, W., Ezcurra, I., Herzmark, P. and McCormick, S. (2002). "New pollen-specific receptor kinases identified in tomato, maize and *Arabidopsis*: the tomato kinases show overlaping but distinct localization patterns on pollen tubes." Plant Molecular Biology **50**: 1-16.

Kimura, M. (1967). "On the evolutionary adjustment of spontaneousmutation rates." Genetics Research **9**: 23–24.

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. (2001). "Molecular systematics of the Brassicaceae: evidence from coding plastidic *matk* and nuclear *chs* sequences." American Journal of Botany **88**(2): 534–544.

Koch, M.A. (2010). "Colonizing the american continent: systematics of the genus *Arabis* in North America (Brassicaceae)." American Journal of Botany **97**(6): 1040–1057.

Koch, M.A., Dobes, 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**(3): 338–350.

Kohany, O. (2006). "Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor." BMC Bioinformatics **7**(474).

Koltunow, A.M. (1990). "Different temporal and spatial gene expression patterns occur during anther development." The Plant Cell **2**: 1201 -1 224.

Koltunow, A.M. (1993). "Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules." The Plant Cell **5**: 1425-1437.

Koltunow, A.M. (1993). "Apomixis: Embryo Sacs and Embryos Formed without Meiosis or Fertilization in Ovules." Plant Cell **5**(10): 1425-1437.

Koltunow, A.M., Bicknell, R.A. and Chaudhury, A.M. (1995). "Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization." Plant Physiology **108**: 1345-1 352.

Koltunow, A.M. and Grossniklaus, U. (2003). "Apomixis: A Developmental Perspective." Annual Review of Plant Biology **54**: 547–74.

Kondo, T. (2007). "Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA." Nature Letters **9**(6): 660-665.

Kondrashov, A.S. (1985). "Deleterious mutations as an evolutionary factor: facultative apomixis and selfing." Genetics **111**: 635-653.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L. and Gao, G. (2007). "CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine." Nucleic Acids Research **36**: 345-349.

Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B. and Nasrallah, M.E. (2001). "Self-Incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*." The Plant Cell **13**: 627–643.

L'Hôte, D., Serres, C., Veitia, R.A., Montagutelli, X., Oulmouden, A. and Vaiman, D. (2008). "Gene expression regulation in the context of mouse interspecific mosaic genomes." Genome Biology **9: R133**.

Lagercrantz, U. (1998). "Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements." Genetics **150**: 1217–1228.

Laspina, N.V.Vega, T.Seijo, J.G.González, A.M.Martelotto, L.G.Stein, J.Podio, M.Ortiz, J.P.Echenique, V.C.Quarin, C.L.Pessino, S.C. (2008). "Gene expression analysis at the onset of aposporous apomixis in Paspalum notatum." Plant Molecular Biology **67**(6): 615-628.

Laux, T., Mayer, K.F.X., Berger, J. and Jürgens, G. (1996). "The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis*." Development **122**: 87-96

Leblanc, O., Armstead, I., Pessino, S., Ortiz, J.P.A., Evans, C., do Valle, C. and Hayward, M.D. (1997). "Non-radioactive mRNA fingerprinting to visualise gene expression in mature ovaries of *Brachiaria* hybrids derived from *B. brizantha*, an apomictic tropical forage." Plant Science **126**: 49–58.

Leblanc, O., Grimanell, D., Hernandez-Rodriguez, M., Galindo, P.A., Soriano-Martinez, A.M. and Perotti, E. (2009). "Seed development and inheritance studies in apomictic maize-Tripsacum hybrids reveal barriers for the transfer of apomixis into sexual crops." International Journal of Developmental Biology **53**(4): 585-596.

Leblanc, O., Grimanelli, D., Gonzfilez-de-Leon, D. and Savidan, Y. (1995). "Detection of the apomictic mode of reproduction in maize-*Tripsacum* hybrids using maize RFLP markers." Theoretical and applied genetics **90**(7/8): 1198-1203.

Lenhard, M., Jürgens, G. and Laux, T. (2002). "The WUSCHEL and SHOOTMERISTEMLESS genes fulfil complementary roles in *Arabidopsis* shoot meristem regulation." Development **129**: 3195-3206.

Leshem, Y., Johnson, C., Wuest, S.E., Song, X., Ngo, Q.A., Grossniklaus, U. and Sundaresana, V. (2012). "Molecular characterization of the *glauce* mutant: a central cell–specific function is required for double fertilization in *Arabidopsis*." The Plant Cell **24**: 3264–3277.

Levins, R. (1967). "Theory of fitness in a heterogeneous environment. VI. The adaptive significance of mutation." Genetics **56**: 163–178.

Li, H., Wu, G., Ware, D., Davis, K.R. and Yang, Z. (1998). "*Arabidopsis* Rho-Related GTPases: Differential gene expression in pollen and polar localization in fission yeast." Plant Physiology **118**: 407–417.

Lin, Y. and Yang, Z. (1997). "Inhibition of pollen tube elongation by microinjected anti-Rop1Ps antibodies suggests a crucial role for Rho-type GTPases in the control of tip growth." The Plant Cell **9**: 1647-1659.

Lisch, D. (2009). "Epigenetic regulation of transposable elements in plants." Annual Review of Plant Biology **60**: 43-66.

Liu, J. and Qu, L.-J. (2008). "Meiotic and mitotic cell cycle mutants involved in gametophyte development in *Arabidopsis*." Molecular Plant **1**(4): 564–574.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E. and Snell, W.J. (2008). "The conserved plant sterility gene *HAP2* functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes." Genes & Development **22**: 1051–1068.

Liu, Z.W., Wang, R.R.C. and Carman, J.G. (1994). " Hybrids and backcross progenies between wheat (*Triticum aestivum* L.) and apomictic Australian wheatgrass (*Elymus rectisetus* (Nees in Lehm.) A. Loeve and Connor): karyotypic and genomic analyses." Theoretical and Applied Genetics **89** (5): 599-605.

Long, M., Betran, E., Thornton, K. and Wang, W. (2003). "The origin of new genes: glimpses from the young and old." Nature Reviews Genetics **4**(11): 865-875.

Löve, A. and Löve, D. (1975). "Nomenclatural notes on arctic plants." Botaniska Notiser **128**: 497–523.

Lu, X., Liu, X., An, L., Zhang, W. and Zhang, C. (2008). "The *Arabidopsis MutS* homolog *AtMSH5* is required for normal meiosis." Nature **18**: 589-599.

Lynch, M. (1984). "Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis." The quarterly review of biology **59**(3): 257-290.

Lynch, M. (2010). "Evolution of the mutation rate." Trends in Genetics **26**: 345–352.

Lynch, M., O'Hely, M., Walsh, B. and Force, A. (2001). "The probability of preservation of a newly arisen gene duplicate." Genetics **159**(4): 1789-1804

Ma, H. (2005). "Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants." Annual Review of Plant Biology **56**: 393–434.

Ma, J., Yan, B., Qu, Y., Qin, F., Yang, Y., Hao, X., Yu, J., Zhao, Q., Zhu, D. and Ao, G. (2008). "*Zm401*, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize." Journal of cellular biochemistry **105**(1): 136-146

Magnard, J.-L., Yang, M., Chen, Y.-C.S., Leary, M. and McCormick, S. (2001). "The Arabidopsis gene *Tardy Asynchronous Meiosis* is required for the normal pace and synchrony of cell division during male meiosis." Plant Physiology **127**: 1157–1166.

Malhó, R., Read, N.D., Salomé Pais, M. and Trewavas, A.J. (1994). "Role of cytosolic free calcium in the reorientation of pollen tube growth." The Plant Journal(5): 331-341.

Malhó, R., Read, N.D., Trewavas, A.J. and Salomé Pais, M. (1995). "Calcium channel activity during pollen tube growth and reorientation." The Plant Cell **7**: 1173-1184.

Manly, B.F.J. (1994). Multivariate Statistical Methods: A Primer. Boca Raton, Chapman and Hall.

Marimuthu, M.P.A.Jolivet, S.Ravi, M.Pereira, L.Davda, J.N.Cromer, L.Wang, L.Nogué, F.Chan, S.W.L.Siddiqi, I.Mercier, R. (2011). "Synthetic clonal reproduction through seeds." Science **331**: 876.

Mark Welch, D.B., Mark Welch, J.L. and Meselson, M. (2008). "Evidence for degenerate tetraploidy in bdelloid rotifers." Proceedings of the National Academy of Sciences **105**(13): 5145-5149

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

Martincorena, I., Seshasayee, A.S.N. and Luscombe, N.M. (2012). "Evidence of non-random mutation rates suggests an evolutionary risk management strategy." Nature **485**: 95-98.

Martínez, E.J., Urbani, M.H., Quarin, C.L. and Ortiz, J.P.A. (2001). "Inheritance of apospory in bahiagrass, *Paspalum notatum*." Hereditas **135**: 19–25.

Márton, M.L., Cordts, S., Broadhvest, J. and Dresselhaus, T. (2005). "Micropylar pollen tube guidance by Egg Apparatus 1 of maize." Science **307**: 573–576.

Márton, M.L. and Dresselhaus, T. (2008). "A comparison of early molecular fertilization mechanisms in animals and flowering plants." Sexual Plant Reproduction **21**: 37-52.

Mascarenhas, J.P. (1989). "The Male Gametophyte of Flowering Plants." The Plant Cell **1**: 657-664.

Mason, A.S., Nelson, M.N., Yan, G. and Cowling, W.A. (2011). "Production of viable male unreduced gametes in *Brassica* interspecific hybrids is genotype specific and stimulated by cold temperatures." BMC Plant Biology **11**(103).

Matzk, F., Meister, A. and Schubert, I. (2000). "FC analysis on seeds of mono & dicots." The Plant Journal **21**(1): 97-108.

Matzk, F., Prodanovic, S., Bäumlein, H. and Schubert, I. (2005). "The inheritance of apomixis in *Poa pratensis* confirms a five locus model with differences in gene expressivity and penetrance." The Plant Cell **17**: 13-24.

Maynard-Smith, J. and Haigh, J. (1974). "The hitch-hiking effect of a favorable gene." Genetical Research **23**: 23-35.

McCall, C. and Primack, R.B. (1992). "Influence of flower characteristics, weather, time of day, and season on insect visitation rates in three plant communities." American Journal of Botany **79**: 434–442.

McClintock, B. (1984). "The significance of responses of the genome to challenge." Science **226**: 792–801.

McCormick, S. (1993). "Male gametophyte development." The Plant Cell **5**: 1265–1275.

McDonald, J.F. (1998). "Transposable elements, gene silencing, and macroevolution." Trends in Ecology and Evolution **13**: 94–95.

McKenna, M.A. and Mulcahy, D.L. (1993). Ecological aspects of gametophytic competition in *Dianthus chinensis*. Pollen: Biology and implications for plant breeding. D.L. Mulcahy and E. Ottaviano. New York, Elsevier**:** 419–424.

Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009). "Long non-coding RNAs: insights into functions." Nature Reviews **10**: 155-159.

Mercier, R., Vezon, D., Bullier, E., Motamayor, J.C., Sellier, A., Lefevre, F., Pelletier, G. and Horlow, C. (2001). "SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis." Genes & Development **15**: 1859–1871.

Meyer, P. and Saedler, H. (1996). "Homology-dependent gene silencing in plants." Annual Review of Plant Physiology and Plant Molecular Biology **47**: 23–48.

Meyer, S. and Scholten, S. (2007). "Equivalent parental contribution to early plant zygotic development." Current Biology **17**: 1686–1691.

Michurin, L. (1950). Selected works. Moscow.

Mogie, M. (1988). "A model for the evolution and control of generative apomixis." Biological Journal of the Linnean Society **35**: 127–153.

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

Mogie, M., Britton, N.F. and Stewart-Cox, J.A. (2007). Asexuality, polyploidy and the male function. Apomixis: evolution, mechanisms and perspectives. E. Hörandl, U. Grossniklauset al.,. Rugell, Liechtenstein, A. R. G. Gantner Verlag.

Montgomery, E., Charlesworth, B. and Langley, C.H. (1987). "A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*." Genetics Research **49**: 31–41.

Montgomery, T.A., Howell, M.D. and Cuperus, J.T. (2008). "Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation." Cell **133** (1): 128–141.

Morgan, D. (2006). The Cell Cycle: Principles of Control. Oxford, Oxford University Press.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A. and Rafalsk, i.A. (2005). "Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize." Nature Genetics **37**(9): 997-1002.

Mulcahy, D.L. and Mulcahy, G.B. (1975). "The influence of gametophytic competition on sporophytic quality in *Dianthus chinensis*." Theoretical and Applied Genetics **46**: 277–280.

Mulcahy, D.L. and Mulcahy, G.B. (1987). "The effects of pollen competition." American Scientist **75**(1): 44–50.

Müller-Stoll, W.R. and Lerch, G. (1957). "Über Nachweis, Entstehung und Eigenschaften der Kallosebildungen in Pollenschläuchen." Flora **144**: 297-334.

Muller, H.J. (1964). "The relation of recombination to mutational advance." Mutation Research **106**: 2–9.

Müntzing, A. (1940). "Further studies on apomixis and sexuality in *Poa*." Hereditas **26**: 115-190.

Muschietti, J., Eyal, Y. and McCormick, S. (1998). "Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases *LePRK1* and *LePRK2*." The Plant Cell **10**: 319-330.

Naumova, T.N. (2001). "Reproductive development in apomictic populations of *Arabis holboellii* (Brassicaceae)." Sexual Plant Reproduction **14**: 195–200.

Naumova, T.N. and Vielle-Calzada, J.-P. (2001). Ultrastructural analysis of apomictic development. Flowering of apomixis: from mechanisms to genetic engineering. Y. Savidan, J.G. Carmanet al.,, Mexico: CIMMYT, IRD, European Commission DG VI**:** 44–63.

Nicolas, S.D.Le Mignon, G.Eber, F.Coriton, O.Monod, H.Clouet, V.Huteau, V.Lostanlen, A.Delourme, R.Chalhoub, B.Ryder, C.D.Chevre, A.M.Jenczewski, E. (2007). "Homeologous

recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids." Genetics **175**: 487–503

Nogler, G.A. (1984). Gametophytic apomixis. In Embryology of Angiosperms. B.M.e. Johri. Berlin, Springer-Verlag**:** 475-518.

Noirot, M., Couvet, D. and Hamon, S. (1997). "Main role of self-polination rate on reproductive allocations in pseudogamous apomicts." Theoretical and Applied Genetics **95**: 479-483.

Nowak, M., Grini, P.E., Jakoby, M.J., 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**(1): 63-67.

Noyes, R.D. and Rieseberg, L.H. (2000). "Two independent loci control agamospermy (apomixis) in the triploid flowering plant *Erigeron annuus*." Genetics **155**: 379–390.

O'Connell, L.M. and Eckert, C.G. (1999). "Differentiation in sexuality among populations of *Antennaria parlinii* (Asteraceae)." International Journal of Plant Sciences **160**(3): 567-575.

Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P. and Fisher, R.L. (1996). "A mutation that allows endosperm development without fertilization." PNAS **93**(11): 5319–5324.

Ohno, S. (1970). Evolution by gene duplication. Berlin, Springer Verlag.

Ohno, S. (1972). "So much ‘‘junk’’ DNA in our genome." Brookhaven Symposia in Biology **23**: 366–370.

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A. and Vielle-Calzada, J.P. (2010). "Control of female gamete formation by a small RNA pathway in *Arabidopsis*." Nature **464**(7288): 628-632.

Olson, D.M.Dinerstein, E.Wikramanayke, E.D.Burgess, N.D.Powell, G.V.N.Underwood, E.C.D'Amico, J.Itoua, I.Strand, H.E.Morrison, J.C.Loucks, C.J.Allnutt, T.F.Ricketts, T.H.Kura, Y.Lamoreux, J.F.Wettengel, W.W.Hedao, P.Kassem, K.R. (2001). "Terrestrial ecoregions of the world: new map of life on earth." Bioscience **51**: 933–938.

Ostenfeld, C.H. (1910). "Further studies on the apogamy and hybridization of the *Hieracia*." Z. indukt. Abstammungs- u. Vererbungslehre **3**(241-285).

Ozias-Akins, P., Akiyama, Y. and Hanna, W.W. (2003). "Molecular characterization of the genomic region linked with apomixis in *Pennisetum/Cenchrus*." Functional & Integrative Genomics **3**: 94–104.

Ozias-Akins, P., Roche, D. and Hanna, W.W. (1998). "Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes." Proceedings of the National Academy of Sciences **95**: 5127–5132.

Pace, L. (1913). "Apogamy in Atamosco." Botanical Gazette **56**: 376–394.

Pang, K.C., Frith, M.C. and Mattick, J.S. (2006). "Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function." Trends in Genetics **22**(1).

Parisi, S., McKay, M.J., Molnar, M., Thompson, M.A. and van der Spek, P.J. (1999). "Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family conserved from fission yeast to humans." Molecular and Cellular Biology **19**: 3515–3528.

Park, S.K., Howden, R. and Twell, D. (1998). "The *Arabidopsis thaliana* gametophytic mutation *gemini pollen1* disrupts microspore polarity, division asymmetry and pollen cell fate." Development **125**: 3789-3799.

Peakall, R. and Smouse, P.E. (2012). "GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update." Bioinformatics **28**: 2537-2539.

Pécrix, Y., Rallo, G., Folzer, H., Cigna, M., Gudin, S. and Le Bris, M. (2011). "Polyploidization mechanisms: temperature environment can induce diploid gamete formation in *Rosa* sp.." Journal of Experimental Botany.

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

Peloquin, S.J., Boiteux, L.S. and Carputo, D. (1999). "Meiotic mutants in potato: valuable variants." Genetics **153**(4): 1493-1499

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L. and Poethig, R.S. (2004). "SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*." Genes and Development **18**: 2368–2379.

Pessino, S.C., Espinoza, F., Martinez, E.J., Ortiz, J.P.A., Valle, E.M. and Quarin, C.L. (2001). "Isolation of cDNA clones differentially expressed in flowers of apomictic and sexual *Paspalum notatum*." Hereditas **134**: 35–42.

Peters, H.I., Chang, Y.-W.E. and Traugh, J.A. (1995). "Phosphorylation of elongation factor 1 (EF-1) by protein kinase C stimulates GDP/GTP-exchange activity." European Journal of Biochemistry **233**: 550-556.

Petrov, D.A., Aminetzach, Y.T., Davis, J.C., Bensasson, D. and Hirsh, A.E. (2003). "Size matters: Non-LTR retrotransposable elements and ectopic recombination in *Drosophila*." Molecular Biology and Evolution **20**(6): 880–892.

Pfaffl, M.W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acid Research **29**(9): 2002-2007.

Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006). "Maximum entropy modeling of species geographic distributions." Ecological Modelling **190**(3-4): 231-259.

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

Pikaard, C.S., Haag, J.R., Ream, T. and Wierzbicki, A.T. (2008). "Roles of RNA polymerase IV in gene silencing." Trends in Plant Science **13**(7): 390-397.

Pina, C., Pinto, F., Feijo, J.A. and D., B.J. (2005). "Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications and gene expression regulation." Plant Physiology **138**: 744–756,.

Polegri, L., Calderini, O., Arcioni, S. and Pupilli, F. (2010). "Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic *Paspalum simplex* Morong flowers." Journal of Experimental Botany **61**(6): 1869–1883.

Powers, L. (1945). "Fertilization without reduction in guayule (*Parthenium argentatum* Gray) and a hypothesis as to the evolution of apomixis and polyploidy." Genetics **30**(323-346).

Procissi, A., Guyon, A., Pierson, E.S., Giritch, A., Knuiman, B., Grandjean, O., Tonelli, C., Derksen, J., Pelletier, G. and Bonhomme, S. (2003). "*KINKY POLLEN* encodes a SABRE-like protein required for tip growth in *Arabidopsis* and conserved among eukaryotes." The Plant Journal **36**(6): 894-904.

Quarin, C.L. (1999). "Effect of pollen source and pollen ploidy on endosperm formation and seed set in pseudogamous apomictic *Paspalum notatum*." Sexual Plant Reproduction **11**: 331–335.

Rambuda, T.D. and Johnson, S.D. (2004). "Breeding systems of invasive alien plants in South Africa: does Baker's rule apply?" Diversity and Distributions **10**(5-6): 409-416.

Reagin, M.J. (2003). "TempliPhi - a sequencing template preparation procedure that eliminates ov and DNA purification." Journal of Biomolecular Techniques **14**(2): 143–148.

Reddy, T.V., Kaur, J., Agashe, B., Sundaresan, V. and Siddiqi, I. (2003). "The *DUET* gene is necessary for chromosome organization and progression during male meiosis in *Arabidopsis* and encodes a PHD finger protein." Development **130**: 5975–5987.

Regan, S.M. and Moffatt, B.A. (1990). "Cytochemical analysis of pollen development in wild-type *Arabidopsis* and a male-sterile mutant." The Plant Cell **2**: 877-889.

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

Renaudin, J.-P., Doonan, J.H., Freeman, D., Hashimoto, J., Hirt, H. and al., e. (1996). "Plant cyclins: a unified nomenclature for plant A-, B- and D-type cyclins based on sequence organization." Plant Molecular Biology **32**: 1003-1018.

Rice, P., Longden, L. and Bleasby, A. (2000). "EMBOSS: the european molecular biology open software suite." Trends in Genetics **16**(6): 276–277.

Riggs, C.D. and Hasenkampf, C.A. (1991). "Antibodies directed against a meiosisspecific, chromatin-associated protein identify conserved meiotic epitopes." Chromosoma **101**: 92-98.

Rodrigues, J.C., Cabral, G.B., Dusi, D.M., de Mello, L.V., Rigden, D.J. and Carneiro, V.T. (2003). "Identification of differentially expressed cDNA sequences in ovaries of sexual and apomictic plants of *Brachiaria brizantha*." Plant Molecular Biology **53**(6): 745-757.

Rollins, R.C. (1941). "Monographic study of *Arabis* in western North America." Rhodora **43**: 289-483.

Rollins, R.C. (1993). *Cruciferae* of continental North America – Systematics of the mustard family from the Arctic to Panama. Stanford, CA., Stanford University Press.

Ross, K.J., Fransz, P., Armstrong, S.J., Vizir, I., Mulligan, B., Franklin, F.C.H. and Jones, G.H. (1997). "Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines." Chromosome Research **5**: 551-559.

Rowley, J.R., Skvarla, J.J. and Walles, B. (2000). "Microsporogenesis in *Pinus sylvestris* L. VIII. Tapetal and late pollen grain development." Plant systematics and evolution **225**: 201-224.

Roy, B.A. (1995). "The breeding systems of six species of *Arabis* (Brassicaceae)." American Journal of Botany **82**: 869–877.

Rutishauser, A. (1954). "Die Entwicklungserregung des Endosperms bei pseudogamen Ranunculusarten." Mitteilungen der Naturforschenden Gesellschaft Schaffhausen **25**: 1-45.

Ryan, E., Grierson, C.S., Cavell, A., Steer, M. and Dolan, L. (1998). "*TIP1* is required for both tip growth and non-tip growth in *Arabidopsis*." New Phytologist **138**(1): 49–58.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.-C., Lee, P.Y., Truong, M.T., Beals, T.P. and Goldberg, R.B. (1999). "Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants." Sexual Plant Reproduction **11**: 297–322.

Sasaki, Y., Yasuda, H., Ohba, Y. and Harada, H. (1990). "lsolation and characterization of a nove1 nuclear protein from pollen mother cells of lily." Plant Physiology **94**: 1467-1471.

Savidan, Y. (2000). "Apomixis: genetics and breeding." Plant breeding reviews **18**: 13–86.

Savidan, Y. (2001). Transfer of apomixis through wide crosses. The flowering of apomixis: from mechanisms to genetic engineering. Y. Savidan, J.G. Carmanet al.,. Mexico, D.F., CIMMYT,IRD, European Commission DG VI (FAIR).

Savidan, Y. (2007). Apomixis in higher plants. Apomixis: evolution, mechanisms and perspectives. E. Hörandl, Grossniklaus, U., Van Dijk, P. J., Sharbel, T., editors. Ruggell, ARG-Gantner**:** 15–22.

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

Schiefelbein, J., Galway, M., Masucci, J. and Ford, S. (1993). "Pollen tube and root-hair tip growth is disrupted in a mutant of *Arabidopsis thaliana*." Plant Physiology **103**(3): 979–985.

Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E. and Schneitz, K. (1999). "Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*." Proceedings of the National Academy of Sciences **96**: 11664–11669.

Schierup, M.H., Vekemans, X. and Christiansen, F.B. (1998). "Allelic genealogies in sporophytic self-incompatibility systems in plants." Genetics **150**: 1187–1198.

Schiott, M., Romanowsky, S.M., Baekgaard, L., Jakobsen, M.K., Palmgren, M.G. and Harper, J.F. (2004). "A plant plasma membrane Ca2+ pump is required for normal pollen tube growth and fertilization." Proceedings of the National Academy of Sciences **101**: 9502–9507.

Schommer, C., Beven, A., Lawrenson, T., Shaw, P. and Sablowski, R. (2003). "*AHP2* is required for bivalent formation and for segregation of homologous chromosomes in *Arabidopsis* meiosis." The Plant Journal **36**: 1–11.

Schopfer, C.R., Nasrallah, M.E. and Nasrallah, J.B. (1999). "The male determinant of self-incompatibility in *Brassica*." Science **286**: 1697-1700.

Schranz, E.M., Dobeš, C., Koch, M.A. and Mitchell-Olds, T. (2005). "Sexual reproduction, hybridization, apomixis and polyploidization in the genus *Boechera* (Brassicaceae)." American Journal of Botany **92**(11): 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)." New Phytologist **171**: 425–438.

Schranz, M.E., Windsor, A.J., Song, B.-h., Lawton-Rauh, A. and Mitchell-Olds, T. (2007). "Comparative genetic mapping in *Boechera stricta* - a close relative of *Arabidopsis*." Plant Physiology **144**: 286–298.

Schwartz, S., Elnitski, L., Li, M., Weirauch, M., Riemer, C., Smit, A., NISC Comparative Sequencing Program, Green, E.D., Hardison, R.C. and Miller, W. (2003). "MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences." Nucleic Acids Research **31**(13): 3518–3524.

Schwartz, Y.B. and Pirrotta, V. (2008). "Review Polycomb complexes and epigenetic states." Current Opinion in Cell Biology **20**(3): 266-273.

Scott, R., Dagless, E., Hodge, R., Paul, W., Soufleri, I. and Draper, J. (1991). "Patterns of gene expression in developing anthers of *Brassica napus*." Plant Molecular Biology **17**: 195-207.

Scott, R.J. (2008). "Double Fertilization in Arabidopsis thaliana Involves a Polyspermy Block on the Egg but Not the Central Cell." Molecular Plant 1–9.

Scott, R.J., Spielman, M. and Dickinson, H.G. (2004). "Stamen structure and function." The Plant Cell **16**: 46–60.

Sears, E.R. (1977). "An induced mutant with homoeologous pairing in common wheat." Canadian journal of genetics and cytology **19**: 585–593.

Seffens, W. and Digby, D. (1999). "mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences." Nucleic Acids Research **27**: 1578–1584.

Sharbel, T.F., 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, T.F., Voigt, M.-L., Corral, J.M., Galla, G., Kumlehn, J., Klukas, C., Schreiber, F., Vogel, H. and Rotter, B. (2010). "Apomictic and sexual ovules of *Boechera* display heterochronic global gene expression patterns." The Plant Cell **22** (3): 655-671

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

Sharbel, T.F., Voigt, M.L., 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.

Shinohara, A., Ogawa, H. and Ogawa, T. (1992). "Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein." Cell **69**: 457–70.

Siddiqi, I., Ganesh, G., Grossniklaus, U. and Subbiah, V. (2000). "The *dyad* gene is required for progression through female meiosis in *Arabidopsis*." Development **127**: 197–207.

Siena, L.A., Sartor, M.E., Espinoza, F., Quarin, C.L. and Ortiz, J.P.A. (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.

Silverman, B.W. (1986). Density estimation for statistics and data analysis. London, Chapman and Hall.

Simpson, G.G. and Filipowicz, W. (1996.). "Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery." Plant Molecular Biology **32**: 1-41.

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

Smit, A.F.A., Hubley, R. and Green, P. (1996-2004). "RepeatMasker Open-3.0."

Smyth, D.R., Bowman, J.L. and Meyerowitz, E.M. (1990). "Early flower development in *Arabidopsis*." The Plant Cell **2**: 755-767.

Soltis, P.S. and Soltis, D.E. (2009). "The role of hybridization in plant speciation." Annual Review of Plant Biology **60**: 561–88.

Song, J.H., Cao, J.S. and Wang, C.G. (2012). "*BcMF11*, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility." Plant Cell Reports.

Song, J.H., Cao, J.S., Yu, X.L. and Xiang, X. (2007). "*BcMF11*, a putative pollen-specific non-coding RNA from *Brassica campestris* ssp. *chinensis*." Journal of Plant Physiology **164**(8): 1097-1100.

Spielman, M., Preuss, D., Li, F.-L., Browne, W.E., Scott, R.J. and Dickinson, H.G. (1997). "TETRASPORE is required for male meiotic cytokinesis in *Arabidopsis thaliana*." Development **124**: 2645–2657.

Spielman, M. and Scott, R.J. (2008). "Polyspermy barriers in plants: from preventing to promoting fertilization." Sexual Plant Reproduction **21**: 53–65.

Spillane, C., Steimer, A. and Grossniklaus, U. (2001). "Apomixis in agriculture: the quest for clonal seeds." Sexual Plant Reproduction **14**: 179–187.

Spurr, A.R. (1969). "A low-viscosity epoxy resin embedding medium for electron microscopy." Journal of Ultrastructure Research **26**(1-2): 31-43.

Staiger, C.J., Gibbon, B.C., Kovar, D.R. and Zonia, L.E. (1997). "Profilin and actindepolymerizing factor: modulators of actin organization in plants." Trends in Plant Science **2**: 275-281.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. (2005). "Function of alternative splicing." Gene **344**: Gene 1–20.

Stebbins, G.L. (1950). Variation and evolution in plants. New York, NY, USA, Columbia University Press.

Steeves, T.A. and Sussex, I.M. (1989). Patterns in plant development. Cambridge, UK, Cambridge University Press.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., Ma, H. and Martienssen, R. (1995). "Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements." Genes & Development **9**: 1797-1810.

Suzuki, T., Masaoka, K., Nishi, M., Nakamura, K. and Ishiguro, S. (2008). "Identification of kaonashi mutants showing abnormal pollen exine structure in *Arabidopsis thaliana*." Plant and Cell Physiology **49**(10): 1465–1477.

Swanson, W.J. and Vacquier, V.D. (2002). "The rapid evolution of reproductive proteins." Nat Rev Genet **3**(2): 137-144.

Talbert, L.E. and Chandler, V.L. (1988). "Characterization of a highly conserved sequence related to *Mutator* transposable elements in maize." Molecular Biology and Evolution **5**: 519-529.

Tamura and Nei (1993). "Estimation of the number of nucleotide substitutions in the control region of mtDNA in humans and chimpanzees." Molecular Biology and Evolution **10**(3): 512-526.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). "MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Molecular Biology and Evolution: (In Press).

Tanaka, H., Ishikawa, M., Kitamura, S., Takahashi, Y., Soyano, T., Machida, C. and Machida, Y. (2004). "The *AtNACK1/HINKEL* and *STUD/TETRASPORE/AtNACK2* genes, which encode functionally redundant kinesins, are essential for cytokinesis in *Arabidopsis*." Genes to Cells **9**(12): 1199-1211.

Taskin, K.M., Turgut, K. and Scott, R.J. (2009). "Apomeiotic pollen mother cell development in the apomictic *Boechera* species." Biologia Plantarum **53** (3): 468-474.

Teppner, H. (1996). "Adventitious embryony in *Nigritella* (Orchidaceae)." Folia Geobotanica **31**(3): 323-331.

Terziiski, D., Jurukova-Grancharova, P., Robeva, P. and Daskalova, T. (1995). "Contribution to the embryology of genus *Achillea* L. *(Asteraceae*)." Nauka-za-Gorata **32**: 36-42.

The-1000-Genomes-Project-Consortium (2010). "A map of human genome variation from population scale sequencing." Nature **467**(7319): 1061–1073.

The ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M. and Dermitzakis, E.T. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**: 799–816.

Theissen, G. and Saedler, H. (2001). "Floral quartets." Nature **409**: 469–471.

Thomson, J.D. (1989). "Germination schedules of pollen grains: implications for pollen selection." Evolution **43**(1): 220-223.

Tucker, M.R., Araujo, A.-C.G., Paech, N.A., Hecht, V., Schmidt, E.D.L., Rossell, J.-B., de Vries, S.C. and Koltunow, A.M.G. (2003). "Sexual and apomictic reproduction in *Hieracium* subgenus *Pilosella* are closely interrelated developmental pathways." The Plant Cell **15**(7): 1524-1537.

Tucker, M.R., Paech, N.A., Willemse, M.T.M. and Koltunow, A.M.G. (2001). "Dynamics of callose deposition and β-1,3-glucanase expression during reproductive events in sexual and apomictic *Hieracium*." Planta **212**(4): 487-498.

Twell, D. (2010). "Male gametogenesis and germline specification in flowering plants." Sexual Plant Reproduction **24**: 149-160.

Twell, D., Oh, S.-A. and Honys, D. (2006). "Pollen development, a genetic and transcriptomic view." Plant Cell Monographs **3**(15).

Twell, D., Park, S.K. and Lalanne, E. (1998). "Asymmetric division and cell-fate determination in developing pollen." Trends in Plant Science **3**(8).

Udall, J.A., Quijada, P.A. and Osborn, T.C. (2005). "Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L." Genetics **169**(2): 967–979.

van Dijk, P.J. and Bakx-Schotman, J.M.T. (2004). "Formation of unreduced megaspores (diplospory) in apomictic Dandelions (*Taraxacum officinale, s.l.*) is controlled by a sex-specific dominant locus." Genetics **166**(1): 483-492

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

Vandel, A. (1928). "La parthénogenese geographique. Contribution a l'étude biologique et cytologique de la parthénogenese naturelle." Bulletin Biologique de la France et de la Belgique **62**: 164–182.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P. and Crete, P. (2004). "Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs." Molecular Cell **16**: 69–79.

Veilleux, R.E. (1985). "Diploid and polyploid gametes in crop plants: mechanisms of formation and utilization in plant breeding." In Janick, J. (ed.). Plant Breeding Reviews Avi Publish. Co., Wesport, CT, **3**: 253-288.

Verduijn, M.H., Dijk Van, P.J. and Van Damme, J.M.M. (2004). "Distribution, phenology and demography of sympatric sexual and asexual dandelions (*Taraxacum officinale s.l.*): geographic parthenogenesis on a small scale." Biological Journal of the Linnean Society **82**: 205-218.

Vielle-Calzada, J.-P., Crane, C.F. and Stelly, D.M. (1996). "Apomixis: The asexual revolution." Science **274**: 1322–1323.

Vielle-Calzada, J.P., Baskar, R. and Grossniklaus, U. (2000). "Delayed activation of the paternal genome during seed development." Nature **404**: 91–94.

Vogel, H.J. (1994). "Calmodulin - a versatile calcium mediator protein." Biochemistry and Cell Biology **72**(9-10): 357-376.

Voigt-Zielinski, M.-L., Piwczynski, M. and Sharbel, T.F. (2012). "Differential effects of polyploidy and diploidy on fitness of apomictic *Boechera*." Sexual Plant Reproduction **25**: 97–109.

Voigt, M.-L., Melzer, M., Rutten, T., Mitchell-Olds, T. and Sharbel, T.F. (2007). Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. In Apomixis: Evolution, Mechanisms and Perspectives E. Hörandl, U. Grossniklauset al.,, Ruggell: A.R.G. Gantner Verlag K.G.,**:** 235-257.

Völz, R., von Lyncker, L., Baumann, N., Dresselhaus, T., Sprunck, S. and Groß-Hardt, R. (2012). "*LACHESIS*-dependent egg-cell signaling regulates the development of female gametophytic cells." Development **139**: 498-502.

Vorsa, N. and Rowland, L.J. (1997). "Estimation of 2n megagametophyte heterozygosity in a diploid blueberry (Vaccinium darrowi Camp) clone using RAPDs." Journal of Heredity **8**: 423-426.

Walsh, B. (2003). "Population-genetic models of the fates of duplicate genes." Genetica **118**: 279–294.

Wang, C., Rathore, K.S. and Robinson, K.R. (1989). "The response of pollen to applied electrical fields." Developmental Biology **136**: 405-410.

Wang, J., Zhang, J., Zheng, H., Li, J. and Liu, D. (2004). "Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs." Nature **431**.

Wang, Y., Magnard, J.L., McCormick, S. and Yang, M. (2004). "Progression though meiosis I and meiosis II in *Arabidopsis* anthers is regulated by an A-type cyclin predominately expressed in prophase I." Plant Physiology **136**: 4127–4135.

Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005). "Fast and reliable prediction of noncoding RNAs." Proceedings of the National Academy of Sciences **102** 2454-2459.

Weijers, D., Geldner, N., Offringa, R. and Juergens, G. (2001). "Seed development: early paternal gene activity in *Arabidopsis*." Nature **414**: 709–710.

Wicker, T., Buchmann, J.P. and Keller, B. (2010). "Patching gaps in plant genomes results in gene movement and erosion of colinearity." Genome Research **20**: 1229–1237.

Wierzbicki, A.T., Haag, J.R. and Pikaard, C.S. (2008). "Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes." Cell 135 **135**: 635–648.

Wilkins, A.S. and Holliday, R. (2009). "The Evolution of Meiosis From Mitosis." Genetics **181**(1): 3-12

Willing, R.P., Bashe, D. and Mascarenhas, J.P. (1988). "An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*." Theoretical and Applied Genetics **75**: 751-753.

Windham, M. and Al-Shehbaz, I.A. (2007). "New and noteworthy species of *Boechera* (Brassicaceae) III: additional sexual diploids and apomictic hybrids." Harvard Papers in Botany **12**(1): 235–257.

Windham, M.D., AL-SHEHBAZ, I.A., BAILEY, C.D. and LOREEN, A. (2004). "A taxonomist's worst nightmare: a preliminary glimpse into the systematics of *Boechera* (Brassicaceae)." Botany 2004: Annual meeting of The Botanical Society of America in Salt Lake City, Utah, USA (Abstract).

Windsor, A.J., Schranz, M.E., Formanová, N., Gebauer-Jung, S., Bishop, J.G., Schnabelrauch, D., Kroymann, J. and Mitchell-Olds, T. (2006). "Partial shotgun sequencing of the *Boechera stricta* genome reveals extensive microsynteny and promoter conservation with *Arabidopsis*." Plant Physiology **140**: 1169–1182.

Wing, R.A., Yamaguchi, J., Larabell, S.K., Ursin, V.M. and McCormick, S. (1989). "Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia.*" Plant Molecular Biology **14**: 17-28.

Winge, Ö. (1917). "The chromosomes. Their numbers and general importance." Compt. Rend. Trav. du Lab. de Carlsberg **13**: 131-275.

Winkler, H.K.A. (1908). "Über Parthenogenesis und Apogamie im Pflanzenreich " Progressus Rei Botanicae **4**: 293–454.

Workman, C. and Krogh, A. (1999). "No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution." Nucleic Acid Research **27**: 4816–4822.

Xu, Z. and Wang, H. (2007). "LTR FINDER: an efficient tool for the prediction of full-length LTR retrotransposons." Nucleic Acids Research **35**: 265–268.

Yamada, K., Lim, J. and Dale, J.M. (2003). "Empirical analysis of transcriptional activity in the *Arabidopsis* genome." Science **302**: 842–846.

Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S. and Yang, W.C. (2003). "*TAPETUM DETERMINANT1* is required for cell specialization in the *Arabidopsis* anther." Plant Cell **15**: 2792–2804.

Yang, W.-C., Ye, D., Xu, J. and Sundaresan, V. (1999). "The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein " Genes & Development **13**: 2108–2117

Yang, W.-C., Ye, D., Xu, J. and Sundaresan, V. (1999). "The *SPOROCYTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein." Genes & Development **13**: 2108-2117.

Yang, X., Makaroff, C.A. and Ma, H. (2003). "The *Arabidopsis MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis " The Plant Cell **15**: 1281-1295.

Yanofsky, M.F. (1995). "Floral meristems to floral organs: genes controlling early events in *Arabidopsis* flower development." Annual Review of Plant Physiology and Plant Molecular Biology **46**: 167-88.

Yoshikawa, M., Peragine, A., Park, M.Y. and Poethig, R.S. (2005). "A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*." Genes and Development **19**(18): 2164–2175.

Yu, Z., Wright, S.I. and Bureau, T.E. (2000). "*Mutator*-like elements in *Arabidopsis thaliana*: structure, diversity and evolution." Genetics **156**: 2019-2031.

Zaratiegui, M., Irvine, D.V. and Martienssen, R.A. (2007). "Noncoding RNAs and gene silencing." Cell **128**: 763-776.

Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P. and And, T.A. (2006). "Evidence that miRNAs are different from other RNAs." Cellular and Molecular Life Science **63**: 246-254.

Zhang, M.Q. (2002). "Computational prediction of eukaryotic protein-coding genes." Nature reviews **3**: 698-709.

Zhao, D. and Ma, H. (2000). "Male fertility: a case of enzyme identity." Current Biology **10**: R904–R907.

Zhao, D.Z., Wang, G.F., Speal, B. and Ma, H. (2002). "The *EXCESS MICROSPOROCYTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther." Genes & Development **16**: 2021–2031.

Zhao, S. and Fernald, R.D. (2005). "Comprehensive algorithm for quantitative real-time polymerase chain reaction." Journal of Computational Biology **12**(8): 1047–1064.

Zhu, Q.-H. and Wang, M.-B. (2012). "Molecular functions of long non-coding RNAs in plants." Genes **3**: 176-190.

Zinkl, G.M., Zwiebel, B.I., Grier, D.G. and Preuss, D. (1999). "Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine." Development **126**(23): 5431-5440.

Zubko, E. and Meyer, P. (2007). "A natural antisense transcript of the *Petunia hybrida Sho* gene suggests a role for an antisense mechanism in cytokinin regulation." The Plant Journal **52**: 1131–1139.

Zweig, M.H. and Campbell, G. (1993). "Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine." Clinical Chemistry **39**: 561–577.

7 Acknowledgements

First of all, I would like to thank the Institute of Plant Genetics and Crop Plant research (IPK) for funding my research and for providing the facilities for generating most of the data presented in this study.

Exceedingly, I would like to express my appreciation to my supervisor Dr. Timothy F. Sharbel who convinced me to start this highly interesting PhD project in his research group and who always provided hints and new directions in a friendly and openhearted atmosphere. He also enabled my participation in the International Max Planck Research School (IMPRS) Jena, Germany, for which I send him my sincere gratitude. In this regard I extend my acknowledgment to Dr. Karin Groten for her assistance with all the administrative issues, which finally made my life much easier. Next, I would like to acknowledge my co-supervisor Prof. Dr. Marcus Koch (HIP, University of Heidelberg), who shared not only ideas on my project but also provided material and unpublished results which pushed this study forward. Thank you also for giving me the opportunity to defend my work at the Heidelberg University.

I would like to express my appreciation to Prof. Dr. Ingo Schubert for the constructive comments on the manuscript and interesting and productive discussions.

A special 'Thank You' goes to Dr. José María Corral García, who is a great character and scientist. I cannot sufficiently express my admiration and gratitude for all your small scientific hints, for your help in strategically planning my thesis and for all the small emotional uplifts which ensured that I kept on track with the dissertation.

I am especially grateful to Dr. Heiko Vogel (MPI-CE, Jena) for sharing his immense knowledge on microarrays and his participation in the manuscript preparation, which also applies to Dr. Jörg Fuchs whom I additionally thank for letting me use his well equipped flow-cytometry laboratory.

It is my big pleasure to thank all members of the Apomixis research group, IPK (ARG-IPK) for the great atmosphere in and outside of the lab, especially the 'good soul' of the lab a.k.a. Heidi who always offers a helping hand in the lab and keeps the lab running, although having sometimes a hard time with our occasionally 'rudimentary' sense for orderliness. Also I greatly appreciate the technical support of Leane, Vera, Sheila and Robert; for assiduously taking care of my plants and for endless PCR sessions. Thank you, Marco and Javier for your outstanding performance and for making the lab a special place to work. Thanks for being a wonderful team.

In addition, I also send my regards to all the past members of the ARG-IPK, Marta and Thomas who hopefully enjoy their new stage of life, and especially Marie who helped me in the beginning of my thesis introducing me into the 'world of pollen'. Above all, I lively remember all the vivid, extended and prolific discussions with Olawale and Samuel about politics, economics and believe with special focus on the present situation of the African continent. Thank you for sharing new insights and your thoughts with me!

Hello to all lab visitors (Dorota, Ana, Alex, John…) from all over the world, you have been the cream on top of an international and collegial time at the IPK; thank you for all the joyful moments.

My great thanks go to all the staff for making scientific life easier at the IPK, especially I am referring to you Dr. Britt Leps; you are doing a great job in assisting with all administrative issues. To all friends and students of IPK, I wish you all the best.

Sometimes life is a miracle. In this sense I met you Tiina, the woman of my life, during this study. You give meaning to this work and to my life. Hopefully you will never stop being what you are: a joyful, buoyant, inquisitive and openhearted woman.

Finally, I would like to express my deepest gratitude to my parents, my brothers Achim and Max, the whole family and all the friends who supported me in any respect throughout the years during good and hard times.

8 Appendix

8.1 Content of supplemental material (see attached CD)

Supplemental Figures

Supplemental Figure 1. Quality control report for *Boechera* microarray hybridizations with *Boechera* antherhead cRNA.

Supplemental Figure 2. Thermodynamically stable non-random secondary structures of BspUPG-2.

Supplemental Figure 3. Overlap of npcRNA 5 with exon 3 of a GTP binding Elongation factor Tu family protein (EFTU/EF-1A; AT4G02930).

Supplemental Figure 4. Mapping of the twelve positive BAC clones on a high-density *Boechera* BAC filter membrane.

Supplemental Figure 5. Fingerprinting of positive BAC clones with *Hind*III, *BamH*I and *Bgl*II.

Supplemental Figure 6. Distribution of locally collinear sequence blocks between Bsp*UPG-1*, Bsp*UPG-2*, the original locus (Assembly 1) and the duplicated locus (Assembly 2).

Supplemental Figure 7. Bioclimate variables used for ecological niche modeling of *Boechera* accessions which carry the male apomeiosis marker gene Bsp*UPG-2 versus* accessions which lack Bsp*UPG-2*.

Supplemental Tables

Supplemental Table 1. Key regulators and their regulatory effects in male germline malfunction mutants in various species of flowering plants.

Supplemental Table 2. List of *Boechera* genotypes with population information used for histological examination of microsporogenesis and candidate gene analyses.

Supplemental Table 3. Microarray probe corresponding cDNAs from a local BLAST search against the *Boechera* 454 cDNA database.

Supplemental Table 4. Relative mRNA expression values for six candidate microarray probes in somatic and reproductive tissues.

Supplemental Table 5. Frequencies of microspores, meiotic cells and pollen mother cells relative to antherhead length.

Supplemental Table 6. Analysis of meiocyte constitution at the tetrad stage in diploid sexual and high facultative and obligate apomictic *Boechera* genotypes.

Supplemental Table 7. Significantly enriched GO classes among 522 highly differential expressed microarray probes in apomictic *Boechera.*

Supplemental Table 8. Relative mRNA expression of seven microarray candidate probes for somatic and reproductive tissues.

Supplemental Table 9. Expression values of differentially-expressed microarray probes from qRT-PCR in antherhead tissue in seven microarray and seven additional *Boechera* genotypes.

Supplemental Table 10. Genomic DNA sequence of the unreduced pollen grain development gene (*Boechera* species *UPGRADE*, Bsp*UPG-2*) in *B. divaricarpa* ES 524.

Supplemental Table 11. Full length cDNA sequence of the HMG-coA Reductase Degradation 3 (AT1G18260) *Arabidopsis* homolog in *Boechera* (*Boechera* species *HRD3*, Bsp*HRD3*).

Supplemental Table 12. Transcription factor binding sites upstream of Bsp*UPG-2*.

Supplemental Table 13. Detection of candidate regions for stable non-protein-coding RNA secondary structures of Bsp*UPG-2*.

Supplemental Table 14. All Z-score values for forward and reverse strand of Bsp*UPG-2*.

Supplemental Table 15. Characteristics of thermodynamically stable secondary non-protein-coding RNA structures from Bsp*UPG-2*.

Supplemental Table 16. Assembly of *Boechera* bacterial artificial chromosome (BAC) clone C8B11.

Supplemental Table 17. Assembly of *Boechera* bacterial artificial chromosome (BAC) clones A4O22, F8G11 and E7K5.

Supplemental Table 18. Distribution of inverted repeats on Assembly 1 and Assembly 2.

Supplemental Table 19. Green plant (Viridiplantae) homologous repetitive DNA element sequences mapping on Assembly 1.

Supplemental Table 20. Green plant (Viridiplantae) homologous repetitive DNA element sequences mapping on Assembly 2.

Supplemental Table 21. Sequence divergence of Bsp*UPG-1* and Bsp*UPG-2*.

Supplemental Table 22. Alignment of genomic DNA of the original (Bsp*UPG-1*) and the duplicated locus (Bsp*UPG-2*) of *UPGRADE*.

Supplemental Table 23. Identified indels on Bsp*UPG-2* isolates of sexual and apomictic *Boechera* genotypes.

Supplemental Table 24. Structural variants of the 5`-end of Bsp*UPG-2* according to the presence of different locally collinear blocks in sexual and apomictic *Boechera* genotypes compared to Assembly 2.

Supplemental Table 25. Correlation of FCSS data from 73 *Boechera* genotypes with the abundance of Bsp*UPG-2*.

Supplemental Table 26. Bsp*UPG-2* abundance in taxa of *Brassicaceae*.

Supplemental Table 27. Distribution of BspUPG-2 across 264 *Boechera* ITS types.

Supplemental Table 28. Abundance of Bsp*UPG-2* per ecoregion.

8.2 Curriculum vitae

Personal data

Name:	Martin Mau
Permanent address:	Lange Strasse 2, D-06466 Gatersleben
Date of birth:	19.07.1982, Pasewalk
Family status:	Unmarried
Nationality:	German

Academical carreer

11/2008	Research stay, Faculty of Bioscience Engineering, University Ghent/Belgium
Since 09/2008	PhD thesis, Faculty of Biosciences, Heidelberg and Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben Title: "Identification, molecular analysis and phylogeographic distribution of the chimeric *UPGRADE-2* gene, a candidate for the initiation of unreduced pollen formation in apomictic *Boechera* (Brassicaceae)."
04/2008 - 08/2008	Scientific research staff, Department of Biotechnology, AG Benz, Bavarian Julius-Maximilians-University, Würzburg
07/2007 - 04/2008	Diploma thesis, Department of Biotechnology, AG Benz, Bavarian Julius-Maximilians-University, Würzburg Title: „Purification and biophysical characterization of different mutants of the Outer Membrane Protein OprP from *Pseudomonas aeruginosa*"
10/2002 - 06/2007	University study of biology, Bavarian Julius-Maximilians-University, Würzburg Major subject: Biotechnology Minor subjects: Molecular plant-physiology and biophysics, Plant physiology, morphology and anatomy, Genetics

Schooling

06/2002	General qualification for university entrance (Abitur)
06/2002 - 08/1993	Secondary school, Oskar-Picht-Gymnasium, Pasewalk
08/1993 - 09/1989	Elementary school, Strasburg/Uckermark

9 Declaration

"I confirm that this dissertation was solely undertaken by myself at the Apomixis Research Group of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany and that no help was provided from other sources as those allowed. All sections of the dissertation that use quotes or describe an argument or concept developed by another author have been referenced, including all secondary literature used, to show that this material has been adopted to support my thesis. The content of this work has not been previously submitted for a degree in any educational institution."

Gatersleben, 24. April 2013

Martin Mau