RUNNING TITLE
Unreduced pollen formation in apomicts

CORRESPONDING AUTHOR
Timothy F. Sharbel
Department of Cytogenetics and Genome Analyses, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany
Telephone number: +49 (0)39482 5608
Email address: sharbel@ipk-gatersleben.de
Research area: Genes, Development and Evolution
TITLE
The conserved chimeric transcript *UPGRADE-2* is associated with unreduced pollen formation and is exclusively found in apomictic *Boechera*

AUTHORS
Martin Mau¹, José M. Corral¹, Heiko Vogel², Michael Melzer¹, Jörg Fuchs¹, Markus Kuhlmann¹, Nico de Storme³, Danny Geelen³ and Timothy F. Sharbel¹,⁴

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben, Germany
²Genomics Research Group, Department of Entomology, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany
³Plant Production, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Gent, Belgium

ONE-SENTENCE SUMMARY
A conserved chimeric genomic sequence found exclusively in apomictic *Boechera* generates a putative long non-protein coding transcript with secondary structure folding capability, and is highly expressed in antherheads characterized by meiotically unreduced pollen formation.
FINANCIAL SOURCE
(1) Apomixis research group by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
(2) EU COST action “Harnessing Plant Reproduction for Crop Improvement” (FA0903)

CORRESPONDING AUTHOR
Timothy F. Sharbel
Email adress: sharbel@ipk-gatersleben.de
ABSTRACT

In apomictic Boechera meiotic diplospory leads to the circumvention of meiosis and suppression of recombination to produce unreduced male and female gametes (i.e. apomeiosis). Here we have established an early flower developmental staging system and have performed microarray-based comparative gene expression analyses of the pollen mother cell (PMC) stage in seven diploid sexual and seven diploid apomictic genotypes to identify candidate factors for unreduced pollen formation. We identified a transcript unique to apomictic Boechera called UPGRADE-2 (BspUPG-2), which is highly-upregulated in their pollen mother cells. BspUPG-2 is highly conserved among apomictic Boechera genotypes but has no homologue in sexual Boechera or in any other taxa. BspUPG-2 undergoes post-transcriptional processing but lacks a prominent open reading frame. Together with the potential of stably forming miRNA-like secondary structures, we hypothesize that BspUPG-2 functions as long regulatory non-coding mRNA-like RNA. BspUPG-2 has apparently arisen through a three-step process initiated by ancestral gene duplication of the original BspUPG-1 locus, followed by sequential insertions of segmentally duplicated gene fragments, with final exonization of its sequence structure. Its genesis reflects the hybridization history which characterizes the genus Boechera.
INTRODUCTION

Haploid gamete formation is an important step during the diplohaplontic plant life cycle which marks the switch from sporophyte to gametophyte. During female and male sporogenesis, the diploid megaspore mother cell (MMC) and pollen mother cell (PMC) respectively undergo one reductional and equational mitotic division to form four haploid daughter cells. During subsequent male spore development (microgametogenesis), each of the four microspores undergoes two mitotic divisions to form a tricellular pollen grain containing two mature reduced gametes (generative nuclei) and one vegetative nucleus.

Despite a relatively simple structure, the male gametophyte is regulated by complex gamete-specific molecular programs which are characterized by (1) enrichment of transcripts 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) developmental stage-specific transcript sets whose activities vary temporally from the uninuclear microspore towards the mature pollen grain (Mascarenhas, 1989; Honys and Twell, 2004).

Loss of function studies of male and female meiosis have revealed cell cycle defects (Yang et al., 2003), effects on homeologous chromosome recombination (Lu et al., 2008), and defects in meiotic cell fate (Sundaresan et al., 1995; Yang et al., 1999), all of which lead to reduced meiocyte fertility or death. Other mutations lead to unreduced (2C) gametes (Brownfield and Köhler, 2010).

Apomeiosis is a naturally occurring form of sensu stricto unreduced egg formation, which is the first step of gametophytic apomixis, or asexual (clonal) seed formation in plants (Nogler, 1984; Grimanelli et al., 2001). During apomixis the apomeiotically-derived egg-cell undergoes parthenogenetic (i.e. without fertilization) development, while endosperm develops with (pseudogamous) or without (autonomous) fertilization (Nogler, 1984; Grimanelli et al., 2001). Apomixis bears the potential to fix hybrid genotypes over generations and is therefore of high interest to agronomy (Grimanelli et al., 2001; Grossniklaus, 2001).

Apomixis research can be divided into studies of de novo induction of apomixis-like processes in sexual plants, or naturally-occurring apomixis (Savidan et al., 2001). Mutant screening and attempts to introgress apomixis into sexual taxa have been hindered by its molecular genetic complexity (d’Erfurth et al., 2008; Hörandl and Temsch, 2009; Marimuthu et al., 2011). Comparative mapping strategies in natural apomicts, with an emphasis on aposporic species, have identified candidate factors, although proof of function in crop plants is still pending (Guerin et al., 2000; Albertini et al., 2004; Matzk et al., 2005; Schallau et al.,...
6

2010). Beside theories of apomixis inheritance through a single dominant locus (Mogie, 1988) or via a complex of physically-linked coadapted genes (van Dijk et al., 1999), the hybridization-derived floral asynchrony (HFA) theory proposes deregulation of sexual genetic pathways via asynchronous expression of duplicated gene sets as a mechanism for apomixis induction (Carman, 1997). Reproductive deregulation was recently supported by heterochronic expression patterns during ovule development in sexual and apomictic Boechera (Sharbel et al., 2010), pointing to a network of epigenetic and post-transcriptional regulation during germline specification (Twell, 2010).

The approximately 110 species comprising the North-American genus Boechera Á. Löve & D. Löve (Brassicaceae) exist as diploid sexual or diploid and allopolyploid apomictic forms (Böcher, 1951; Al-Shehbaz and Windham, 1993+), the latter of which are highly correlated with interspecific hybridization (Sharbel et al., 2009; Sharbel et al., 2010; Beck et al., 2011). Diploid apomixis is rare, and facilitates comparisons between sex and apomixis without the added complexity of differing ploidy (Sharbel et al., 2009; 2010). Apomictic Boechera taxa are characterized by Taraxacum-type pseudogamous diplospory (Rollins, 1941; Böcher, 1951). Besides the formation of unreduced female gametes, apomictic Boechera also produce unreduced pollen, as is demonstrated by the fact that diploids and triploids produce seeds almost exclusively with hexaploid (6C = [4C\text{maternal}] + [2C\text{paternal}]) and nonaploid (9C = [6C_m] + [3C_p]) endosperm (Voigt et al., 2007; Aliyu et al., 2010; Voigt-Zielinski et al., 2012). Hence, strong selection pressure to maintain a balanced 2 maternal : 1 paternal genome ratio (i.e. endosperm balance number, EBN, Johnston et al. (1980)) during endosperm formation apparently characterizes apomictic Boechera (Johnston et al., 1980; Voigt et al., 2007; Aliyu et al., 2010). While we have thus far focused our transcriptome comparisons on sexual and apomictic ovules (Sharbel et al., 2009; Sharbel et al., 2010), strong selection for genetically balanced endosperm during apomictic seed formation has led us to study unreduced pollen formation.

Hence, we have undertaken a microarray-based transcriptome comparison of a single stage of antherhead development differentiating reduced and unreduced pollen formation between diploid sexual and diploid apomictic Boechera. Here, we describe the isolation and characterization of a differentially-expressed candidate gene, BspUPG-2 (Boechera species UPGRADE-2, unreduced pollen grain development), which is associated with unreduced pollen formation and is potentially part of a homology dependent gene silencing mechanism. BspUPG-2 has apparently arisen through gene duplication and subsequent insertions of genic
and non-genic fragments in the hybrid apomictic genome. Its putative molecular function in apomictic genomes of *Boechera* is discussed.

**RESULTS**

**Apomictic *Boechera* produce (FDR type) unreduced gametes**

One major mechanism which leads to unreduced pollen formation is the meiotic nuclear restitution which comprises a failure during the first (FDR) or second meiotic division (SDR). Meiotic chromosome behavior (Ross et al., 1996) was examined to determine the exact mechanism by which unreduced pollen is formed in facultative and obligate apomictic *Boechera*, and showed that pollen formation differed between apomicts and sexuals (Fig. 1, Table S1). The sexual *B. stricta* ES 612.1 showed expected homologous chromosome pairing, with juncture of non-sister chromatids leading to seven bivalents (Fig. 1A). Subsequent two nuclear divisions lead to a tetrad with four haploid nuclei (Figs. 1B-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 univalents at metaphase I which do not (or only partially) segregate during meiosis I (Figs. 1I, J and P, Q). As a result, metaphase II plates were frequently fused (Figs. 1K and R) in contrast to their perpendicular orientation in the sexual reference accession (Fig. 1D), leading to the generation of dyads with balanced and unreduced chromosome numbers (Figs. 1L, M and S, T). Low levels of chromosomal synapsis leading to tetrad formation were also observed in all examined diploid obligate apomicts (Figs. 1N and U). Compared to obligate apomicts the facultative apomictic *B. divaricarpa* ES 514 exhibited higher levels of nuclei with bivalents (Figs. 1W and Y) which proceeded through both meiotic cell divisions (Figs. 1A’, C’ and E’), generating tetrads with four haploid cells (Fig. 1G’). Nonetheless, nuclei with univalents were frequently observed (Figs. 1V, X and Z) whose sister chromatids were equally separate at metaphase II (Fig. 1B’) to develop balanced dyads (Figs. 1D’ and F’). Meiotic chromosome counts of all apomicts strongly support first division restitution (FDR) 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 diploid chromosomes sets in dyads.

Approximately 28 000 meiocytes from single antherheads of six apomictic and a sexual reference genotype were counted to quantify meiotic product frequencies. The sexual
reference genotype produced the expected high frequency of tetrads (88%), whereas the tetrad frequency varied among all tested apomicts (0 to 87%) between individuals of one genotype and even between different flower buds from the same individual (Fig. 2A, Table S2). Low levels of triads and tetrads were observed for apomictic genotypes ES 805.2 (0.55% and 0.35%, respectively) and ES 776.2 (0.00%, respectively), both of which showed high levels of monad and dyad formation. However, surprisingly high levels of triads and tetrads were found for other individuals of both, obligate (300.9 – 24.98% and 57.98%; ES 753 - 32.78% and 62.16%) and facultative apomicts (ES 514 - 11.22% and 87.00%; ES 524.2 - 15.67% and 64.53%, Fig. 2A). The formation of high levels of balanced dyads confirmed that unreduced pollen in apomictic Boechera is produced by a defect in chromosome segregation during meiosis I. However, unequal sister chromatid segregation during meiosis II, resulting in the formation of extra nuclei were low (in total 0.19% meiocytes with micronuclei and 0.02% polyads) and mainly detected in the apomict ES 753 (0.88% meiocytes with micronuclei, Fig. S1).

Considering the highly variable pattern of meiocyte frequencies between individuals and between flowers of single specific apomictic genotypes (Fig. 2A), only plants producing exclusively unreduced 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. S1). Interestingly, flow cytometric seed screen (FCSS) data of all tested obligate apomictic individuals suggests successful fertilization with only unreduced pollen, although also some individuals produce high levels of reduced pollen (embryo to endosperm ratio of 2C [1Cm + 1Cp]: 6C [4Cm + 2Cp]; Fig. 2A, Fig. S1; Table S3).

**Staging of meiosis in Boechera**

To optimize for comparative gene expression analyses between reduced (sex) and unreduced (apomeiosis) pollen, the morphological development of antherheads was quantified through staging (*sensu* Sharbel et al., 2010). Analyses of early flower development from about 780 flower buds in six diploid Boechera genotypes (Figs. 2B and C, Fig. 3, Tables 1 and S3) showed a linear relationship between bud length and flower organ size (*e.g.* stage S4 to S12 antherheads: $R^2_{apo} = 0.87$, $F = 1273.50$, $P < 0.001$; $R^2_{sex} = 0.95$, $F = 4237.23$, $P < 0.001$). Minor variations among genotypes, but no overall differences between sex and apomixis were detected (*i.e.* stages predict similar organ size classes; 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. S2).

We focused on four major histodifferentiation steps of male gametophyte development which are characterized by their pre-meiotic, meiotic and post-meiotic appearance (Fig. 1, Table 1). Despite some marker-specific variations in correlation strength, flower bud length predicts the gametophytic stage of anthers with relatively high accuracy \((N = 455; \text{Sp - 100\%, PMC - 100\%, Me - 52.27\% and Msp - 58.24\%}, \) Fig. 2B). Antherhead length predicts the gametophytic stage of anther development, except for meiotic and microspore stages in sexuals, which overlapped in antherhead size class (one-way ANOVA with Tukey-HSD post hoc test, between \(*P < 10^{-2}\) and \(***P < 10^{-9}\), Fig. 2C). The correlation between antherhead length and gametophyte stage led to the selection of \(400 \pm 30 \mu\text{m}\) antherheads, a stage which is significantly enriched for PMCs being close to meiosis in both sexual and apomictic genotypes, for collection of total RNA \((**P < 0.001, \) Fisher’s exact test, Table S4; decision to take the PMC stage detailed in Materials and Methods section).

Identification of the apo-specific \textit{UPGRADE} transcript

In search for candidate transcripts which are differentially expressed between the sexual and the apomictic microsporogenesis program, we performed pairwise comparisons of RNA expression levels in microdissected antherheads from 7 sexual and 7 apomictic genotypes using a custom Agilent microarray (detailed in “Material and Methods”). Using a foldchange threshold \(\geq 2\) and corrected \(P\)-value \(\leq 0.05\), while allowing one outlier sample, five microarray probes were identified to be highly upregulated in all apomicts, except in genotype ES 753 \((B. lignifera)\), exhibiting absolute fold change (AFC) levels between 48.74 and 849.31 compared to sexuals (Fig. S3; Table S5).

Homologous cDNAs corresponding to the five candidate array probes were identified from a flower-specific cDNA library (Table S5; Sharbel et al. (2009)). A BLASTN (Altschul et al., 1997) search of these cDNAs against the whole GenBank nucleotide collection (http://www.ncbi.nlm.nih.gov/genbank/) revealed hits for only two of the candidate probes; probe Sharb1199059 \((\text{EMBL/GenBank/DDBJ No. ERS317552})\) was homologous to an \textit{Arabidopsis thaliana S-locus} lectin protein kinase involved in pollen recognition, and probe Sharb0501554 \((\text{EMBL/GenBank/DDBJ No. ERS317556})\) matched two \textit{SRK} genes (S-12 and S-15 type) encoding a \textit{Brassica oleracea S-locus} receptor kinase involved in pollen self-recognition specificity (Table S5).
Quantitative real-time RT-PCR confirmed that 4 of the 5 microarray probes had expression
profiles corresponding to the microarray data (Fig. S4; Tables S6 and S7). The single probe
which could not be qRT-PCR validated (Sharb1199059; lectin homolog) was furthermore
characterized by a negative BAC screen result (Table S8), and was not considered for further
analysis. With exception of the apomictic outlier sample (ES 753), the four microarray probes
were highly upregulated in apomictic antherheads. 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 and reproductive tissues. Moreover, for microarray probes
Sharb0931225, Sharb0501554 and Sharb0690829 expression was substantially higher in
antherheads compared to levels in leaf tissue (one-way ANOVA with Tukey-HSD post hoc
test, $P = 0.022$, $P = 0.008$ and $P = 0.027$, respectively; Fig. S4; Table S7).

Rapid amplification of cDNA ends (RACE) was performed using gene specific primers
from microarray probe Sharb0690829 (Table S5) to obtain full-length cDNA including the 3’-
and 5’-cDNA ends. Although total RNA of antherheads and whole flower tissue from two
sexuals and eight apomicts was used, 5’- and 3’-RACE generated similar DNA fragments
exclusively from apomicts, except for the apomictic outlier ES 753 (Fig. S5).

All 3’- and 5’-RACE fragments from five apomicts were cloned and sequenced, and led to
the identification of a single polyadenylated full-length transcript, BspUPG (Boechera species
unreduced pollen grain development (UPGRADE); EMBL/GenBank/DDBJ No. HF930769;
Fig. 4) with about 2648 nt length (for B. divaricarpa ES 524), excluding the poly(A)
sequence. The microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 were
found in the most 3’-exon of the BspUPG transcript (Fig. 4). A comparison of different
BspUPG full-length cDNA variants with genomic DNA by isolation of 3’-RACE and the
various 5’-RACE fragments in different Boechera genotypes (Fig. S5, white arrows) showed
that the gene has an overall length of 3156 nt (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, Figs. 4 and S5). The U2-dependent classified
introns (5’-GT/3’-AG splicing site; Simpson and Filipowicz (1996)) are located towards the
5’-end of the transcript.

**UPGRADE is a chimeric gene encoding a transcript without translational capacity**

BspUPG fragments are homologous to sequences in *Arabidopsis* located on chromosomes
1, 4 and 5, which encode respectively an Elongation factor TU/EF1-A protein (EFTU/EF1-A;
AT4G02930), a RNA recognition motif-containing protein (RNAR; AT5G19960), and the
HRD3 protein (HMG-coA Reductase Degradation) which is homologous to components of the yeast HRD1 complex (AT1G18260; Fig. 4). From these candidates, AtHRD3, which plays a central role in endoplasmic reticulum (ER)-associated protein degradation, exhibits the highest similarity to BspUPG, with fragments of 145 nt (83% identity, P-value < 5E-31), 118 nt (87% identity, P-value < 6E-30) and 47 nt in length (89% identity, P-value < 0.0002).

The transcript carries a poly(A)-tail with some variation in length and initiation position. In search for regulatory motifs on BspUPG, a putative near upstream element (NUE, AATAAA) of a polyadenylation signal, which is common in plants (Hunt, 1994), was identified 31 nt upstream of the polyadenylation site. Furthermore, a single CpG island was detected at the 5' end of the transcript locus (+140 to 312 nt, C+G >60%; Figs. 4 and S6C). Sequences that are rich in the CpG/CpNpG patterns are predominantly nonmethylated and tend to be associated with genes which are frequently switched on (Deaton and Bird, 2011). A screen for putative regulatory features in the upstream sequences (between +47809 and +48921 nt on Assembly 2, PlantPAN database, Chang et al., (2008); http://plantpan.mbc.nctu.edu.tw/) detected a single tandem repeat (-726 to -985 nt, 82 nt consensus size, 3.2 copies; Fig. 4) and multiple transcription factor binding site (TFBS; Table S9).

Coding Potential Calculation of BspUPG (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, +895 to +1090 nt, coding potential score (cps) = -0.0298) and the reverse complement strand revealed a 202 nt sized ORF (68 aa, +1958 to +2159 nt, cps = -1.180). The lack of an ORF length typical for a protein encoding mRNA suggested a long non-protein-coding mRNA-like RNA (lncRNA) product of BspUPG (ORF typically >100 aa; Kondo (2007)).

The genomic sequence information of the candidate locus was augmented by a screen of a B. divaricarpa bacterial artificial chromosome (BAC) library. Colony-PCR using target-specific primers from the chromosome walking approach (detailed in “Material and Methods”; Table S10) led to the identification of 12 Boechera BAC clones which hybridized with one or several probes simultaneously (Table S8). None of the clones were positive for probe Sharb1199059, whereas three single hits for probe Sharb0425060 and nine triple hits for probes Sharb0931225, Sharb0501554 and Sharb0690829 were detected, which confirmed the results from the RACE experiment (Fig. 4). Restriction digests of the twelve BACs suggested partial overlap of their DNA inserts. 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, EMBL/GenBank/DDBJ No. HF954100), 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 which increased to 40.1% in genic regions. BAC
sequence contigs from clones A4O22, F8G11 and E7K5 overlapped and were assembled into
the 58 769 bp Assembly 2 (EMBL/GenBank/DDBJ No. HF954101), 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 which (except for TER4) were
homologous to genes located on Arabidopsis chromosome 1 (Figs. 5A, S6A and B; Table
S11). 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
chromosome 3 (Figs. 5A, S6A and C; Table S11). Comparison of both assemblies showed
that both fragments share several highly similar 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 BspUPG. 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. 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. 5A and S6C; Table S12), suggesting that parts of Assembly 2 arose
via partial duplications of Assembly 1. Hence, considering the partial presence of the
candidate gene at both loci, we subsequently labeled the original locus in Assembly 1, BspUPG-1, and the duplicated variant on Assembly 2, BspUPG-2 (Figs. 5A and S6).

Both apomictic and sexual gene variants were aligned with Assembly 1 and Assembly 2
(i.e. BAC DNA sequences), allowing for rearrangements. The 3`-end of all gene variants from
sexuals and apomicts, here named locally collinear block 1 (LCB1 in violet, see definition in
“Materials and Methods”), shares high similarity with both assemblies. The 5`-ends of
apomictic gene copies (LCB2 in brown) were only identified on Assembly 2 (Figs. 5A, B and
S6). Unlike apomictic gene variants (now BspUPG-2, see nomenclature above), the 5`-ends
(LCBs 3 to 5) of variants from sexuals (now BspUPG-1) share high sequence similarity with
Assembly 1, but also exhibit a mosaic distribution of their LCBs relative to the duplicated
locus of Assembly 2 (Figs. S6 and S7).

Transcriptional activity of BspUPG-2 is restricted to apomictic Boechera
The transcriptional functionality of *BspUPG-1* and *BspUPG-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 (see sexual/apomictic cDNA reads, Fig. 5C). In contrast, apomictic cDNA maps to LCB1 and LCB2, including the reads with homology to the candidate microarray 60-mer oligonucleotide probes (Sharb931225, Sharb501554 and Sharb690829, Fig. 5C). No cDNA maps to any of the remaining LCBs. PCR on genomic DNA using specific primers for both loci demonstrates the presence of *BspUPG-1* in sexuals and apomicts, but absence using cDNA as a PCR template. In contrast genomic copies and transcripts of *BspUPG-2* were solely identified in apomicts (Figs. 5D and E).

*BspUPG-1* is flanked by four genes centred in the centromeric region on the lower arm of *Arabidopsis* chromosome 1 (Figs. 5A and S8, Table S11; ~0.57 megabases distance to T3P8-sp6, Hosouchi et al. (2002)). Interestingly, two genetic markers (Bst006701 and BSTES0032) from the synthetic F2 linkage map of sexual *B. stricta* (Schranz et al., 2007), both of which are homologous to the *Arabidopsis* locus identifiers At1G51310 and At1G43245, flank *BspUPG-1*. Both markers span the interval of genomic block C1, which localizes *BspUPG-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 *BspUPG-2* is unknown.

*BspUPG-2* has high secondary structure folding potential

Apart from other lncRNA types in plants, some lncRNA act as precursors for small non-protein-coding RNAs (ncRNAs or npcRNAs) such as microRNAs (miRNAs) or endogenous trans-acting small interfering RNAs (tasiRNAs), 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; Amiteye et al., 2013) revealed no mature miRNAs matching the candidate transcript. Considering this, in addition to the obvious lack of a long ORF on *BspUPG-2*, a bioinformatics approach was used to examine whether thermodynamically stable npcRNA structural elements could be predicted from its genomic sequence (detailed in “Protocol S1”).

As shown in Figure 6A, the *BspUPG* forward and reverse strand contains eight regions (i.e. overlapping windows are combined to determine the total length of a candidate region, Table S13) characterized by significant Z-score values, suggesting that parts of *BspUPG* are able to form stable secondary structures (Figs. 6A and S9). These candidate regions span only
exons, comprising together 25% and 27% of the total length of each strand of *BspUPG* respectively, whereby npcRNAs 1/6, 2/7 and 4/8 cover similar regions on both strands of the genomic sequence. The predicted secondary structures were classified based upon their (1) minimal folding free energy, (2) adjusted minimal folding free energy (AMFE), (3) minimal folding free energy index (MFEI), and (4) A+U content (Seffens and Digby, 1999; Bonnet et al., 2004; Zhang et al., 2006). Interestingly, all detected secondary structures met most of the criteria for miRNA precursors (Figs. 6B and S9, Table S14). Comparable with known pri-miRNAs in *Boechera* (Amiteye et al., 2011; Amiteye et al., 2013) and other plant species (Zhang et al., 2006), npcRNAs 1 to 8 demonstrate a lower minimal folding 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, with exception for npcRNA 5 (0.70), have a MFEI greater than 1.07 which is much higher compared to transfer RNAs (tRNA, 0.64), ribosomal RNAs (rRNA, 0.59), or random messenger RNAs (mRNA, 0.65).

A BLASTN search of the GenBank nucleotide and TAIR10 Genes databases showed that three npcRNAs are significantly similar to known protein-coding genes (*i.e.* potential regulatory targets). Two of these potential targets, npcRNA 5 and 8, encode for proteins with ribosomal functions. The 70 nt-sized npcRNA 5 (position +1090 nt and +1160 nt on the reverse strand of *BspUPG*) covers 75% of the third exon of the elongation factor EFTU/EF-1A protein homolog in *Arabidopsis*, which binds tRNAs in a GTP-dependent reaction to the acceptor site of ribosomes (AT1G18260, 92.86% similarity, \( P = 7.00E-24 \), Figs. 6B and C, Table S14). A fragment of the npcRNA 8 maps to a homolog in *Arabidopsis lyrata* and *A. thaliana* which is associated with a U3 small nucleolar RNA involved in the processing of pre-rRNA (AT3G06530, \( P = 0.008 \)). The third, npcRNA 3, maps to exon 25 of the myosin XI B protein-coding gene (AT1G04160, \( P = 0.01 \)), which belongs to a class of myosins potentially involved in the process of spindle/phragmoplast alignment during cytokinesis (Hepler et al., 2002), and whose mutation led to abnormal pollen development in *Oryza sativa*, suggesting a crucial role of myosin XI B in pollen formation (Jiang et al., 2007).

**BspUPG-2 is highly conserved in apomictic Boechera**

Within-individual (*i.e.* allelic) polymorphisms were not identified for *BspUPG-2* among nine tested apomicts. Furthermore, both the original locus (*BspUPG-1*) and its duplicated variant (*BspUPG-2*) are highly conserved at the 3'-end between all tested genotypes, which decreases markedly towards the 5'-end (Fig. 7A, Table S15). Sequence conservation is high between all apomicts (*i.e.* *BspUPG-2*), including the transcriptional outlier ES 753, with the
highest nucleotide divergence between genotypes ES 753 and ES 514 (99.33% ± 0.13 similarity), while BspUPG-2 in all other apomicts shares complete sequence identity (Fig. 7A). In contrast, sexual sequences can be split into two subgroups based on low sequence conservation at the 5′-end of BspUPG-1 (39.95% ± 0.02 similarity), one representing B. stricta genotypes (99.69% ± 0.05 similarity between B. stricta genotypes), and the second representing the remaining sexual genotypes (98.87% ± 0.30 similarity between non-B. stricta genotypes; Fig. 7A, Table S15).

Among all indels which were detected between BspUPG-1 and BspUPG-2 (indels, <50 nucleotides, Albers et al. (2011), Table S16), a single 27 nt sequence was exclusively present in BspUPG-1 but not in BspUPG-2 (Fig. 7A, indel 9). 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, indel 9 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 (detailed in “Protocol S2”; Fig. S10). 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 indel 9-specific probe in both apomictic and sexual flower tissues (Fig. S10).

Together, the lack of intra-individual allelic variation and the absence of BspUPG-2 at any other locus in both sexuals and apomicts (Fig. 5) point to a homo- or hemizygous state for BspUPG-2 in apomicts.

**BspUPG-2 arose via genome rearrangements**

Two lines of evidence were used to validate the proposed fusion of the 5′- (LCB1) and 3′-ends (LCB2) of the transcriptionally active BspUPG-2. A BLASTN search between BspUPG-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 (position +1820 nt; Fig. 7C). We confirmed this result by amplifying a 478 bp-fragment in sexual genotypes using primers corresponding to a fragment of the BspUPG-2-specific LCB2. Additionally, this fragment is mapped in sexual and apomictic Boechera by cDNA belonging to the Boechera homologue of AtHRD3 (BspHRD3; Figs. 4 and 7B).

BspHRD3 was used to test whether the 5′-end of BspUPG-2 derived directly from parental genes or from putative duplicated variants. Although BspHRD3 cDNA is conserved with
Arabidopsis HRD3 (93.0%), variation (i.e. insertions) between the transcript and the detected genomic copy, which are not caused by RNA splicing, was identified (Fig. 7B). This would point to two different genomic variants of BspHRD3: (1) one from which the transcript could be detected but not its genomic copy and which lacks the insertions, and (2) a second variant from which the genomic copy could be sequenced but not the transcript and which includes the inserts (Fig. 7B). Interestingly, comparing the genomic sequence of BspHRD3 with BspUPG-2, a 550-bp sized fragment was identified as being homologous to BspUPG-2, including these insertions (Fig. 7B; 93.6% similarity, blue boxes denote insertions), thus pointing to the integration of a fragment of one of the two genomic variants of BspHRD3 into BspUPG-2.

As gene duplication is strongly correlated with insertions and pseudogenization (Ohno, 1970; Kaessmann, 2010), we tested for duplication by mapping a set of sequence tags from an array-based comparative genome hybridization (aCGH) experiment in sexual and apomictic Boechera (Aliyu et al., in review) against BspUPG-2 (Fig. 7C). In total, 79 array probes from the CGH experiment mapped to BspUPG-2. Whereby most probes mapping onto BspUPG-2 show no copy number variations (CNV) in apomicts (96% with the remaining 4% 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 BspUPG-2, representing the start of LCB1. The remaining depleted sequence tags were distributed towards the 5'-end of BspUPG-2 between +0 to +1081 nt. To summarize, both LCB1 and the extreme 5'-end of BspUPG-2 is present in fewer copies in sexual compared to apomictic genomes, whereas the middle part of BspUPG-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: +1041 to +1191 nt; Figs. 5 and 7C).

Since transposable elements (TEs) are considered to drive genome rearrangements, including duplications, after interspecific hybridization, Assembly 1 and Assembly 2 were screened for Viridiplantae transposable elements. In total, 100 and 113 TEs mapped onto Assembly 1 and Assembly 2 respectively, many of which are simple repeats with low sequence complexity. Excluding all 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; Fig. S11). 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). Proportions of the TE superfamilies do not vary significantly between Assembly 1 and Assembly 2, but do when compared to their distribution across the whole Arabidopsis genome. Copia-like LTR retrotransposons represent the largest proportion of 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. S11, Tables S17 and S18).

Together these results (Figs. 5 and 7) indicate a duplication of BspUPG-1 fragments from the original locus on Assembly 1 to form the basis of BspUPG-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 (e.g. BspHRD3), with subsequent exonization leading to a gain of BspUPG-2 transcriptional activity (Fig. 7D).

DISCUSSION

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 taxa (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 optimal developmental stage for comparative expression
profiling 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. 2B). 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 (Table S4). Variable levels of apomeiosis frequencies 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. 2A and S1). 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 behavior of apomicts producing unreduced pollen (the same used for expression profiling) was primarily asynaptic (Fig. 1), as has previously been shown (asynaptic, sensu Böcher (1951)). 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. 2A), 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 (Fig. 2A), 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). 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 FCSS; Fig. S1, Table S1) 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.

**BspUPG-2, unreduced pollen and stable apomixis in Boechera**

The detailed phenotypic analyses and selection of a specific antherhead stage which differentiated reduced versus unreduced PMCs, in conjunction with a microarray-based analysis of diploid sexual versus diploid apomictic Boechera genotypes, led to the identification of BspUPG-2. Assuming all other developmental aspects of sexual and apomictic antherheads to be the same, the transcriptional activity of this locus is directly correlated with meiotic non-reduction during male sporogenesis, and was furthermore qRT-PCR validated in 14 biological replicates (i.e. 7 sexual and 7 apomictic genotypes) and 3 technical replicates (i.e. microarray probes Sharb0931225, Sharb0501554 and Sharb0690829 mapping to different regions of the same locus) to show strong apomeiosis-specific upregulation of BspUPG-2 in reproductive tissues of apomictic Boechera (Fig. S4, Table S7). It is unclear why genotype ES 753 was an outlier (Fig. S5), despite sharing a highly-similar BspUPG-2 sequence with other apomictic genotypes. 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.

Flow cytometric analyses of over 20 000 single seeds have demonstrated that virtually all apomictic Boechera are characterized by balanced endosperm (6C = [4C_{maternal}] + [2C_{paternal}]) in diploids and 9C = [6C_m] + [3C_p] in triploids), leading to the hypothesis that both meiotically-unreduced egg and pollen are required for stable apomictic seed production (Aliyu et al., 2010). Similar to the analyses here which lead to the identification of BspUPG-2, a second approach identified APOLLO, a similarly-conserved factor which is correlated with unreduced egg cell formation in apomictic Boechera (Corral et al., submitted in parallel to this manuscript). We tested the hypothesis that both factors should be required to form stable apomictic seed by analyzing for the presence of both factors in the same 73 genotypes examined in Aliyu et al. (2010). Both factors are highly co-correlated (BspUPG-2 - 93.33%; APOLLO - 100%; Corral et al, submitted to Plant physiology in parallel this manuscript, Fig.
BspUPG-2 is a long pri-miRNA with potential regulatory functions in trans

Although the apo-specificity of BspUPG-2 attests to its novelty, the lack of homologs in other species and a missing ORF typical for a protein-coding gene leaves the question open whether BspUPG-2 acts in a complementary fashion to increase the penetrance of an existing predisposition for environmentally-induced unreduced gamete formation in sexual genotypes (sensu Aliyu et al. (2010)), or whether it is part of a completely novel pathway.

It is unclear whether BspUPG-2 is a key genetic factor or one of several tightly linked factors which control different aspects of unreduced pollen formation (i.e. apomeiosis). Furthermore, modifier genes (Bicknell et al., 2000), genetic background and/or environmental conditions (Nogler, 1984) could explain variation in the level of apomictic trait expression (refer also to Fig. 2A and 8). The detection of the apo-specific indel 9 is one example of a potential co-regulatory target site in BspUPG-2, although no homologous sRNAs were detected (Figs. 7A and S10). Alternatively, the four detected mRNA isoforms (i.e. potentially caused by alternative splicing; Figs. 4 and S5) 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 (Fig. 2A).

Nevertheless, its tissue-specific expression pattern in addition to conservation at the nucleotide level strongly supports a developmental role for the lncRNA BspUPG-2. Given its chimeric structure, the novel lncRNA BspUPG-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; Song et al., 2012) or Zm401 in maize (Ma et al., 2008). BspUPG-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).

Consistent with this are the sequence fragments homologous to three different functional parental genes across the 5'-end of BspUPG-2, which putatively could serve as targets for
HDGS ($BspHRD3$, $RNAR$ and $EFTU/EF-1A$; Figs. 4, 5A and 7B). Thereby, $BspUPG-2$ could belong to a novel class of miRNA-containing lncRNAs which serve as a matrix (i.e. as long primary miRNA (pri-miRNA)) for several unidentified precursor miRNAs (pre-miRNA, Altuvia et al., (2005)) or endogenous trans acting short interfering RNAs (ta-siRNA). Such regulatory elements are processed from the excised intronic region of a spliced transcript, or from the full-length unspliced transcript, and have been 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)).

Computational analysis of RNA folding detected several sections of $BspUPG-2$ which were predicted to form with high probability non-random and stable secondary structures (Fig. 6A, B and S8, Table S13), and which fulfill most criteria for pre-miRNAs (e.g. elevated A+U content, Table S14). Interestingly, these potential pre-miRNAs were not detected in previous screens (Amiteye et al., 2011; Amiteye et al., 2013), which may reflect (1) its tissue-specific and short-term upregulation at the onset of male meiosis, or (2) limitations of homology-based analyses considering that $BspUPG-2$ is unique to apomictic Boechera. The detection of a stable secondary structure in a region which is 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 $BspUPG-2$. For example, a homologous factor to EFTU/EF-1A has been implicated with meiosis progression in Xenopus (Bellé et al., 1990; Peters et al., 1995), while one of the other two sequence fragments of the 5′-end of $BspUPG-2$ is also homologous to a gene with nucleotide binding function ($RNAR$; GO:0003723).

High sequence conservation and hemi-or homozygous status imply a selective advantage of $BspUPG-2$

Importantly, $BspUPG-2$ is a chimera of non-genic (e.g. LCB1) and genic regions (e.g. $HDR3$, $RNAR$, $EFTU$), in which genic regions were derived from partial duplicated variants of their parental genes (e.g. $BspHRD3$, homologous to AT1G18260, i.e. a pseudogene derived through duplication; Ohno (1972), Figs. 7C and D). Although the long-term survival of chimeric genes is described to be rare (Bennetzen, 2005), $BspUPG-2$ exhibited an unexpectedly high degree of conservation in DNA sequence (Fig. 7A, Table S15) and expression (Fig. S4) between apomictic Boechera representing different taxa and geographic
origins. In contrast, higher sequence variation was identified for BspUPG-1 in sexuals (e.g. 5'-end; Fig. 7A, Table S15). Thus BspUPG-2 appears 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 apomicts (Aliyu et al., 2010).

The hemi- or homozygous status of BspUPG-2 reflects dominant inheritance as a proposed characteristic for an apomeiosis controlling locus (refer to Grossniklaus et al. (2001), Table 1), as has been found for example in Pennisetum (Ozias-Akins et al., 1998). Hemizygosity would reflect an extended period of low to no recombination in its genomic location, and would be consistent with the distribution of TEs at both the original and duplicated locus (Fig. S11). 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. S11) points to a co-localization of the original and duplicated locus of BspUPG in the same pericentromeric region (Fig. S9), or to regions with similar characteristics (i.e. interstitial heterochromatic regions comparable to hk4S in Arabidopsis; Fransz et al. (2000)).

We hypothesize that the genesis of BspUPG-2 is associated with the recurrent interspecific hybridization which is highly correlated with origins of apomictic Boechera (Schranz et al., 2005; Beck et al., 2011). Assuming co-localization of BspUPG-2 with BspUPG-1 and the formation of highly heterochromatic regions in apomicts as protected zones (e.g. Het- and Del-chromosomes, Kantama et al. (2007)), the genome-wide effects of hybridization (Comai, 2000) could have provided the mechanism through which BspUPG-2 originated and gained regulatory function.

To summarize, despite variability for unreduced pollen formation in apomictic Boechera, a single novel transcription unit (BspUPG-2) is consistently upregulated in apomictic flower tissues at the PMC stage, and its chimeric sequence structure might reflect the interspecific hybridization history of this genus. We hypothesize that BspUPG-2 arose via sequential (segmental) duplication-insertion events involving at least five loci, from which three originated from transcribed genes (Fig. 7D). In this regard, it is interesting that only the apomixis-specific duplicated locus (BspUPG-2) shows transcriptional activity, while the original locus (BspUPG-1), which is present in both sexuals and apomicts, does not (Fig. 5E). 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 apomixis factors (Grossniklaus et al., 2001) for which gains in function are hypothesized (Vielle-Calzada et al., 1996).

**CONCLUSION**

The identification of the novel apo-specific BspUPG-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). BspUPG-2 could have a potential application in stabilizing the endosperm balance in the course of the implementation of apomixis into crop plants. How BspUPG-2 has undergone neofunctionalization to develop a trans-regulatory function remains to be clarified.

**MATERIALS AND METHODS**

**Plant material and cultivation conditions**

Several sets of the same diploids, ten apomictic and twelve sexual Boechera, were used for all experiments (Table S1; 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 grown in a phytotron without insecticides and herbicides under long-day conditions (16 h light and 8 h dark, 21°C).

**Measurement of relative nuclear DNA content and reproductive mode**

Relative nuclear DNA content (referred to as ploidy) in leaf, seed (Matzk et al., 2000) and pollen (De Storme et al., 2013) was quantified using leaf tissue from a diploid sexual B. stricta (ES 558.2; Table S1) as an external control. Leaf and seed material (20 seeds per individual plant) were chopped with a razor blade in a drop of Galbraith’s buffer containing 4 μg/ml DAPI (Galbraith et al., 1983). 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 (The Scripps Research Institute, http://facs.scripps.edu/software.html); “C” refers to DNA content of a haploid anaphase cell, and “x” the basic chromosome number.

**Microscopy analysis of gametophyte stages**

Pollen developmental stages were defined (Fig. 3; Regan and Moffatt (1990)) from each of twelve flower bud size stages (S1-S12) which differed in 100 µm length increments (*sensu* Smyth et al. (1990) and Sanders (1999)). For dissection of flower buds at stage S3 and antherheads at flower developmental stages S8 – S12 under a Zeiss Discovery V20 (Carl Zeiss, Jena, Germany), stereomicroscope sterile glass needles bent to an angle of appr. 100-120° and tip size of 50-100 µm using a Narishige PC-10 puller (Narishige Group, Kasuya, Japan) were used. Sample preparation of size-staged buds and antherheads for histology and light microscopy was carried out according to Grasser et al. (2009).

The decision to take the PMC stage for comparative gene expression profiling is based on two criteria: (1) apomeiosis candidate genes should affect early meiotic stages (Grimanelli et al., 2001), and (2) in previous analyses the spike of gene expression change between sexual and apomictic ovules was detected in the microspore mother cell (Sharbel et al., 2010), a stage which corresponds to the PMC in antherheads.

**Cytochemical analysis of fixated Boechera anthers**

Cytological observations of meiosis and PMCs were made from sexuals and apomicts at flower bud stages S9 to S10. Meiotic chromosome spreads of anthers were prepared according to Ross et al. (1996) with minor changes, and observed under a Zeiss Axioplan 2 imaging microscope (Carl Zeiss). Photographs were taken with an AxioCam HRC Rev. 2 camera 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. 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).

**RNA isolation, microarray hybridization and qRT-PCR of antherhead tissue**
Microdissections and total RNA extractions were prepared using tools and a dissection area 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 collected in 500 µl RNAlater (Qiagen, Hilden, Germany). 18 µl of RNase-free DNase I (Qiagen) digested total RNA extracts (RNeasy® Micro kit, Qiagen) 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.

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 18 hrs 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≤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.

The real-time PCR reactions, using primers which were designed on the cDNA reads homologous to the candidate microarray probes Sharb0690829 (EMBL/GenBank/DDBJ No. ERS317557), Sharb0501554 (EMBL/GenBank/DDBJ No. ERS317556) and Sharb0931225 (EMBL/GenBank/DDBJ No. ERS317555; Primer3 v0.4.0, http://frodo.wi.mit.edu/primer3/) were performed according to Sharbel et al. (2010). 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 ELONGATION FACTOR α1 (EFα1), Pellino et al. (2011)), negative template and reverse
transcriptase controls. PCR efficiencies and normalized C\textsubscript{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\textsubscript{t}) values of the amplified targets were calculated separately with reference to the expression levels of each of the two housekeeping genes employing the \(\Delta\Delta\text{C}_t\) method (Pfaffl, 2001) using a calibrator sample (ES910.2; Table S7). 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.

**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 BspUPG-2 and BspHRD3, an Arabidopsis homologous of HRD3 in Boechera. Primers (Primer3 v0.4.0) for BspUPG-2 RACE were derived from microarray probe homologous cDNA read ERS317557 (5\'-end primer GSP3: 5\'-TCTTCGCCATCGTTCATGTTTACTTCCG-3\'; 3\'-end primer GSP4: 5\'-TCATCATGTCTTCTTGCCATCGTTCA-3\'), and those for BspHRD3 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 previous section).

**Computational analysis of RNA folding probabilities of BspUPG**

Structural RNAs are usually 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 (\textit{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). The presence of statistically significant secondary structures of BspUPG-2 was monitored using Z-score values as described by Crespi et al. (1994) and Bonnet et al. (Bonnet et al., 2004) using similar thresholds to Kavanaugh and Dietrich (2009). All significant secondary structures of BspUPG-2 were classified according to Zhang et al. (2006). The “npcRNA
[number]“-names were chosen to follow the naming convention established by previous
investigators (Hirsch et al., 2006).

**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). Flanking regions to the mapped microarray probes were used
to design specific primers (Primer3 v0.4.0) for chromosome walking in sexuals and apomicts
using a DNA Walking SpeedUp Premix Kit I from Seegene (Seegene, Seoul, Korea, Table
S10). Resultant products were transformed into *E. coli* TOP10 cells using a TOPO TA
Cloning Kit for Sequencing (Invitrogen), individual clones were amplified using the
TempliPhi™ DNA Sequencing Template Amplification Kit (Reagin, 2003) and Sanger
sequenced on an ABI 3730 XL sequencing system. Sequence analysis and assembly was
carried out with Lasergene 8 (DNASTar, Madison, WI, USA). The *E. coli* TOP10 sub-cloned,
PCR amplified and finally gel purified DNA walking products (NucleoSpin® Extract II kit,
Macherey-Nagel, Düren, Germany) were hybridized against a gridded 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 *Hind*III, *Bgl*II and *Bam*HI (Amersham Pharmacia Biotech) at 37°C (*Hind*III,
*Bgl*II) and 30°C (*Bam*HI) for 8 h and examined on a 1% agarose gel. Based on their partial
overlapping restriction patterns DNA was isolated from four 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 using 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 using BLASTN and BLASTX searches against the non-redundant GenBank nucleotide and protein databases, respectively (http://www.ncbi.nlm.nih.gov/genbank/).

**Processing of DNA sequences and sequence analysis**

The allelic constitution of *BspUPG-2* in nine each sexual and apomictic genotypes was determined by 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’-TCCGACCTAATCTACAAACTGA-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’). Lasergene 8 (DNASTar) was used for assembly and similarity analysis of Phred 20 quality-trimmed sequences. Coding potential was calculated from all six frames of all *BspUPG-2* splicing forms using the Coding Potential Calculator software (Kong et al., 2007).

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 *BspUPG-1* and *BspUPG-2* between different genotypes.

Local BLASTN search of the complete genomic SRA (SRP007750, http://www.ncbi.nlm.nih.gov/sra/, 454 GS FLX) and sequence extraction were conducted with CLC Genomics Workbench (costs = match 1, mismatch 3, existence 5, extension 2; E-value = 10; word size 11, filter complexity = yes). For mapping of both, 454 WGS and CNV sequence tags (Aliyu et al., unpublished results) the genomic sequence of *BspUPG-2* was divided into eighteen 200 bp bins with 20 bp overlap. All existing CNVs for each single sequence tag are
represented and relative CNV frequencies are considered independently for sexual and apomorphic genotypes.

BAC sequence assemblies were annotated for transposable elements 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). Programs, einverted and EMBOSS (Rice et al., 2000) were employed to identify inverted repeats with $\geq 80\%$ 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).

SUPPLEMENTAL MATERIALS

**Protocol S1.** Computational analysis of RNA folding probabilities of BspUPG-2.

**Protocol S2.** Small RNA Northern blot.

**Figure S1.** Representative meiocytes at tetrad stage and flow cytometric ploidy confirmation.

**Figure S2.** Correlation of flower organs and flower bud length for sexual and apomorphic Boechera genotypes.

**Figure S3.** Constantly differentially regulated microarray probes in apomorphic compared to sexual Boechera genotypes.

**Figure S4.** Validation of differentially-expressed microarray probes by qRT-PCR.

**Figure S5.** Rapid amplification of cDNA 3'- and 5'-ends of the candidate gene UPGRADE.

**Figure S6.** Percent identity plots showing the original UPGRADE locus and its duplicated variant.

**Figure S7.** Distribution of locally collinear sequence blocks along the BAC clone Assembly 1 and Assembly 2.

**Figure S8.** Distribution of Arabidopsis homologous loci corresponding to Assembly 1 and Assembly 2 in Boechera along Arabidopsis chromosomes.

**Figure S9.** Thermodynamically stable non-random secondary structures of BspUPG-2.

**Figure S10.** Northern blot analysis for putative small RNAs homologous to indel 9 at BspUPG-2.

**Figure S11.** Distribution of the TE superfamilies along the original and duplicated UPGRADE locus.
Table S1. List of Boechera genotypes with population information used for histological examination of microsporogenesis and candidate gene analyses.

Table S2. Analysis of meiocyte constitution at the tetrad stage in diploid sexual and high facultative and obligate apomictic Boechera genotypes.

Table S3. High-level 1C-pollen individuals with corresponding flow cytometric seed screen data.

Table S4. Frequencies of microspores, meiotic and pollen mother cells relative to antherhead length.

Table S5. Microarray probes demonstrating significant absolute fold change (AFC) upregulation in apomictic genotypes.

Table S6. QRT-PCR primers used for validation of microarray candidate probes.

Table S7. Relative mRNA expression values for four microarray probes in somatic and reproductive tissues.

Table S8. Overview of BAC clones carrying one to three candidate microarray probes.

Table S9. Transcription factor binding sites upstream of BspUPG-2.

Table S10. Primers used for chromosome walking on microarray candidate probes and the source gene BspHRD3.

Table S11. Gene annotation of Boechera BAC clone Assembly 1 and Assembly 2.

Table S12. Distribution of inverted repeats (IR) on Assembly 1 and Assembly 2.


Table S14. Characteristics of thermodynamically stable secondary non-protein-coding RNA structures with highest Z-score from BspUPG-2 in apomictic Boechera genotypes.

Table S15. Sequence divergence of BspUPG-1 and BspUPG-2.

Table S16. Indels on BspUPG-2 isolates of sexual and apomictic genotypes using a CLUSTALW multiple sequence alignment.

Table S17. Green plant (Viridiplantae) homologous repetitive DNA element sequences mapping on Assembly 1.

Table S18. Green plant (Viridiplantae) homologous repetitive DNA element sequences mapping on Assembly 2.
ACKNOWLEDGEMENTS

The authors thank H. Block, V. Katzorke, M. Wiesner and I. Walde for technical support. We thank Prof. I. Schubert for the constructive comments on the manuscript, and to Thomas Mitchell-Olds and Eric Schranz for seed material. This work was also supported in part by the COST action HAPRECI “Harnessing Plant Reproduction for Crop Improvement”.

REFERENCES


Bancroft I (1997) A strategy involving the use of high redundancy YAC subclone libraries facilitates the contiguous representation in cosmid and BAC clones of 1.7 Mb of the genome of the plant Arabidopsis thaliana Weeds World 4: 1-9


Hunt AG (1994) Messenger RNA 3' end formation in plants. Annual Reviews in Plant Physiology and Plant Molecular Biology 45: 47–60


Ma J, Yan B, Qu Y, Qin F, Yang Y, Hao X, Yu J, Zhao Q, Zhu D, 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: 136-146


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

**Ozias-Akins P, Roche D, Hanna WW** (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


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


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


Zhang BH, Pan XP, Cox SB, Cobb GP, And TA (2006) Evidence that miRNAs are different from other RNAs. Cellular and Molecular Life Science 63: 246-254


FIGURE LEGENDS

Figure 1. 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. Arrowheads show synapsed homologues in sexual (A) and non-synapsed homologues in apomictic Boechera (H and O). In addition, homologues with close juxtaposition were occasionally observed in apomictic Boechera (asterisk). Scale bars, 5 μm.

Figure 2. Meiocyte constitution at tetrad stage and correlation of flower bud size and antherhead length with different gametophyte stages.

(A) Squashes of single antherheads from several individuals per genotype demonstrate meiocyte constituiion at tetrad stage. Roman numbers denote (I) groups of individuals per
genotype primarily (>50%) producing reduced gametes versus (II) groups of individuals per genotype primarily (> 50%) producing unreduced gametes. (B) Relationship between flower bud stages S3 to S12 and proportion of anthers at a particular developmental stage of gametophytes in Boechera. The abbreviations used for each developmental stage are shown in Table 1. (C) Horizontal bars above boxplots demonstrate significant comparisons between gametophytic stages per antherhead length 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. Despite observations of antherhead related gametophyte stages for flower bud stages S3 to S12, the correlation analysis focuses only on S8 to S12.

Figure 3. Light microscopic 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.

Figure 4. Schematic representation of putative splicing forms of BspUPG-2. All four splice forms could be extracted by end-to-end sequencing in five apomorphic Boechera. Exons are represented by grey boxes and numbers indicate intron-exon boundary positions. BLAST search results from GenBank nucleotide collection and TAIR10 gene database are given below.

Figure 5. Original and duplicated UPGRADE locus.

(A) Schematic representation of the gene annotation of Assembly 1 containing the original BspUPG-1 locus and Assembly 2 containing the duplicated apo-specific BspUPG-2 locus. Syntenic regions of the two assemblies are shown below. Inverted repeats are denoted as grey numbered arrowheads. Insertions of genic fragments marked in red letters. (B) Mapping of genomic sequences of BspUPG-1 in apomorphic (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. LCB5 position on Assembly 1 and Assembly 2 is displayed in Fig. S5. Black arrows denote primers: 1 - CON234X2L, 2 - CON234X14L, 3 - CON234X10R, 4 - PC1pol1L, 5 - PC1pol1R, 6 - GSP4, 7 - TSP33R, 8 - CON234X5R, 9 - Indel9minus. (C) CLC Genomics Workbench (v4.5.1) output file showing different distribution of separate mapped cDNA reads from sexual and apomictic genotypes onto BspUPG-2. Both sense-oriented (green) and antisense-oriented (red) cDNA reads are displayed below grey regions which demonstrate cDNA coverage (in pink). Coloured vertical bars on cDNA reads show SNPs in comparison with genomic DNA of BspUPG-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 in a different position in the genome. Primer locations are shown as numbered black arrows in part (B). White asterisk denotes missing band for genotype 105.18, which lacks the priming site for primer 6. (E) PCR with primers (4 and 5) combining LCB1 and LCB2 shows apo-specific presence of BspUPG-2, whereas BspUPG-1 is present in both sexual and apomictic genomes (primer 6 and 9). In contrast to BspUPG-1 which is not transcribed in sexuals and apomicts (primer 6 and 9), BspUPG-2 is solely transcribed from apomicts (primer CON234B2L and CON234B2R, Table S6). White arrow marks faint band for genotype ES 753 and white asterisk marks missing band for genotype 105.18, which lacks the priming site for primer 6.

RRP4, Exosome complex component RRP4; MBOAT, membrane bound O-acyl transferase-like protein; MtN21, nodulin MtN21/EamA-like transporter protein; BspUPG-1, original locus of candidate gene UPGRADE; TER1-4, transposable element related protein; TLP5, Tubby-like F-box protein 5; UP1, uncharacterized protein; TIR, TIR-NBS class of disease resistance protein; UGT, sterol 3beta-glucosyltransferase; NPC1, Niemann-Pick C1 protein; TPR, tetratricopeptide repeat domain-containing protein; DY2A, Dynamin-2A; GRV2, DNAJ heat shock N-terminal domain-containing protein; HRD3, HRD3-like protein; RNAR, RNA recognition motif-containing protein; EFTU, Elongation factor Tu; BspUPG-2; duplicated locus of UPGRADE.

Figure 6. Secondary structure prediction for BspUPG-2.

(A) The Z-score for the sliding window with length between 50nt and 300nt (step size=10) is plotted vs. position on BspUPG-2. 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 (Fig. S8; Tables S13 and S14). (B) Proposed minimum free energy structures (MFE) of putative npcRNAs with significant BLASTN hits in GenBank and TAIR10 Genes databases. 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. (C) Overlap of npcRNA 5 with exon 3 of a GTP binding Elongation factor Tu family protein (EFTU/EF-1A; AT4G02930). Red letters indicate dissimilarities between sequences.

Figure 7. Genesis of *BspUPG-2* via duplication from the original locus *BspUPG-1* and sequential insertion of genic fragments.

(A) Structural overview of full-length *BspUPG-2* in apomictic and sexual *Boechera* (see Table 1) adapted and modified from the Progressive Mauve algorithm, exhibits locally collinear blocks (LCB) between the candidate genes in various genotypes and (B) between *BspUPG-2* and the source genes from which *BspUPG-2* hosts fragments (identical colour display highly similar sequences; black zigzag line denotes truncated sequences due technical reasons). (C) Distribution of 454 whole genome sequencing (WGS) reads mapping across sliding windows (bins) of *BspUPG-2* versus copy number variations (CNV) in 10 sexual versus 10 apomictic genotypes (Aliyu et al., unpublished results, details in “Materials and Methods”; note that “genic inserts” refers to position +199 nt to +1191 nt in (B) and that WGS reads did not overlap between LCB1 and LCB2). (D) Schematic representation of parental and chimeric *BspUPG* genes. Boxed numbers illustrate different LCBs. (+) Transcription activity and (–) no transcription activity.

Figure 8. Correlation between *BspUPG-2* and *APOLLO* with apomictic *Boechera*.

The high correlation of *BspUPG-2* (primers 4 and 5 as denoted in Fig. 5B) with diploid and triploid high-level and obligate apomicts (see Appendix in Aliyu et al. (2010)) is also mirrored by similar results for the female meiosis marker *APOLLO* (see Corral et al., submitted to Plant Physiology in parallel with this manuscript). Numbers above the columns denote count of tested genotypes per mode of reproduction as determined by Aliyu et al. (2010).
Table 1. Developmental markers of pollen formation corresponding to gametophyte and flower stages.

<table>
<thead>
<tr>
<th>Flower stage</th>
<th>Size range (mm)</th>
<th>Arabidopsis flower stage*</th>
<th>Pollen stageb</th>
<th>Developmental marker ID</th>
<th>Developmental marker and histological events</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0 – 0.1</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>S2</td>
<td>0.1 – 0.2</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>S3</td>
<td>0.2 – 0.3</td>
<td>7-8</td>
<td>Sp</td>
<td>Sporocytes, premeiotic, anther in differentiation stage, occurrence of 1° and 2° parietal cells, vascular region initiated.</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>0.3 – 0.4</td>
<td>9</td>
<td>3</td>
<td>Pattern of anther defined, all four locules present.</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>0.4 – 0.5</td>
<td></td>
<td></td>
<td>Rapid lengthening of all flower organs, especially antherheads and filament.</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>0.5 – 0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>0.6 – 0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>0.7 – 0.8</td>
<td></td>
<td>PMC</td>
<td>Pollen mother cell enlarged and clear separated from tapetum, prior to meiosis, no callose deposition.</td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>0.8 – 0.9</td>
<td>4</td>
<td></td>
<td>PMC enters pollen meiosis, crushed middle layer, vacuolated tapetum, initiaal callose deposition, binucleate tapetum.</td>
<td></td>
</tr>
<tr>
<td>S10</td>
<td>0.9 – 1.0</td>
<td>4 – 6</td>
<td>Me</td>
<td>Meiosis, tetrads of microspores, accumulated callose deposition.</td>
<td></td>
</tr>
<tr>
<td>S11</td>
<td>1.0 – 1.1</td>
<td>5 – 6</td>
<td>Msp</td>
<td>Microspores in interphase, early-vacuolated, minorities of tetrads or bi-nucleate microspores.</td>
<td></td>
</tr>
<tr>
<td>S12</td>
<td>1.1 – 1.2</td>
<td>6 – 8</td>
<td></td>
<td>Microspores vacuolated and undergo first mitotic division, callose degradation, bi-nucleate microspores.</td>
<td></td>
</tr>
</tbody>
</table>

*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.