SUPPLEMENTAL FIGURES & TABLES

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

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 (*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 classify non-protein-coding RNAs (npcRNAs) from candidate genes (*e.g.* into microRNA (miRNA), transfer RNA (tRNA), random messenger RNA (mRNA) or ribosomal RNA (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. (Bonnet et al., 2004).

The MFE also depends on the length and the base composition of a folded sequence (Washietl et al., 2005). Therefore, the exceptional stability of a candidate npcRNA structure from Bsp*UPG* its complementary strand was tested using sliding windows of sizes (step size = 10 nt, impact of different step sizes was examined in Kavanaugh and Dietrich (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). Any windows producing a significant Z-score during the scanning process were considered candidate regions for a structural npcRNA. 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 (*EA*) 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 *EA* with the individual MFEs (*E*) results in the structural conservation

index (SCI=*EA/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).

Protocol S2. Small RNA Northern blot. Total RNA and small RNA were extracted from 0.150 – 0.250g flowers at differentstages (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 NaPO4 pH 7.2; 7 % SDS), and membrane washing in 2x SSC / 2% SDS at room temperature.

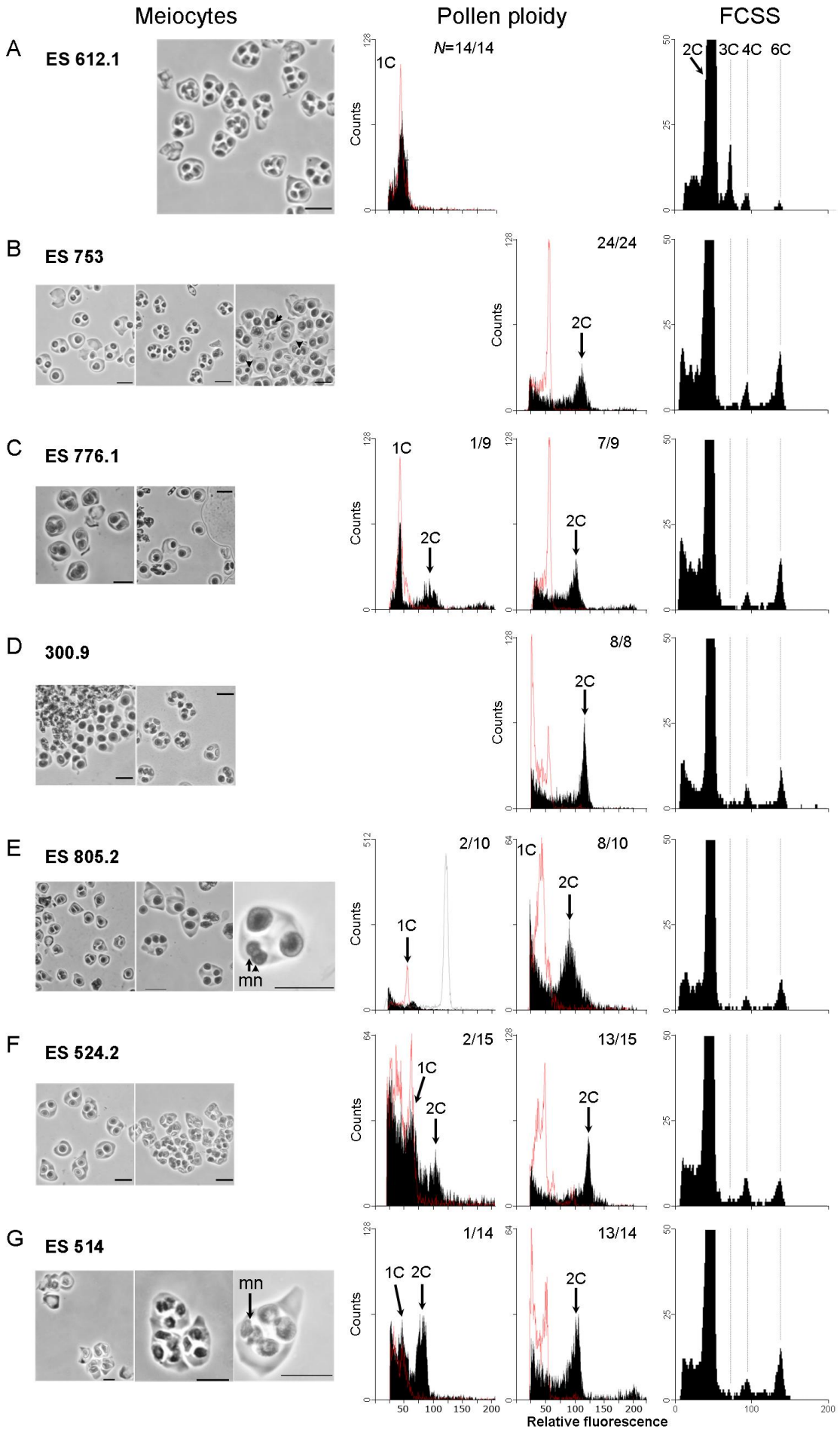
Meiocytes
Pollen ploidy
FCSS
A
ES 612.1
N=14/14
1C
2C
3C
4C
6C
B
ES 753
24/24
2C
C
ES 776.1
1/9
7/9
1C
2C
D
300.9
8/8
2C
E
ES 805.2
2/10
8/10
1C
2C
mn
F
ES 524.2
2/15
13/15
1C
2C
G
ES 514
1/14
13/14
1C
2C
mn
Counts
Relative fluorescence

Figure S1. 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 = [1Cmaternal + 1Cpaternal] embryo and a 3C = [1Cm + 1Cm] + [1Cp] 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=[2Cm] embryo and 6C=[2Cm + 2Cm] + [2Cp] 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 G2 of the embryo. *N*=individuals per genotype; mn/black arrows=micronuclei; the grey peak in subfigure E marks the 2C leaf reference.

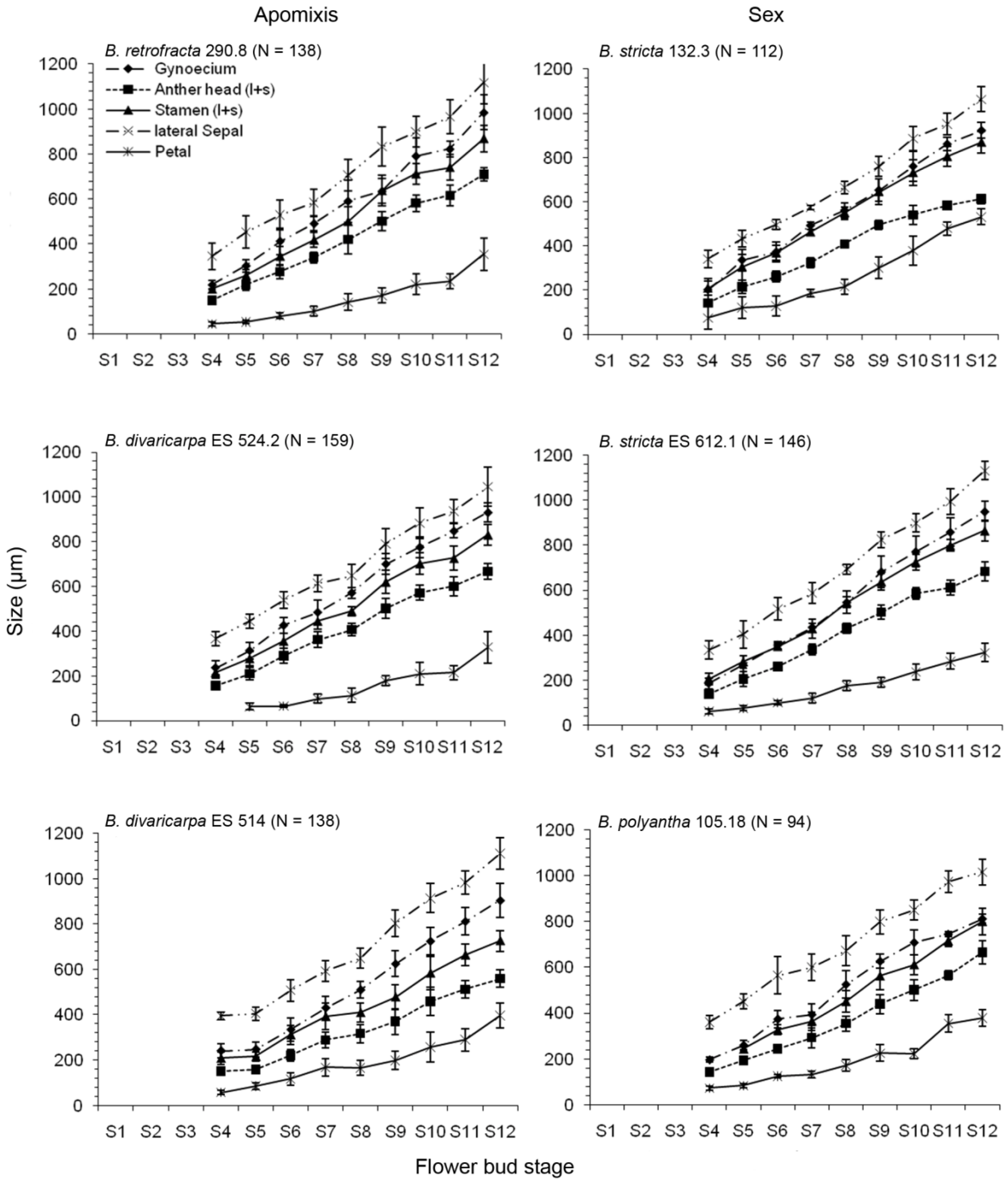

Figure S2. Correlation of flower organs and flower bud length for sexual and apomictic *Boechera* genotypes (Figure 1, Table S1). Error bars show the standard deviation (SD) of single flower organ sizes for each flower bud stage. Total number of examined flower buds per genotypes given in parentheses.

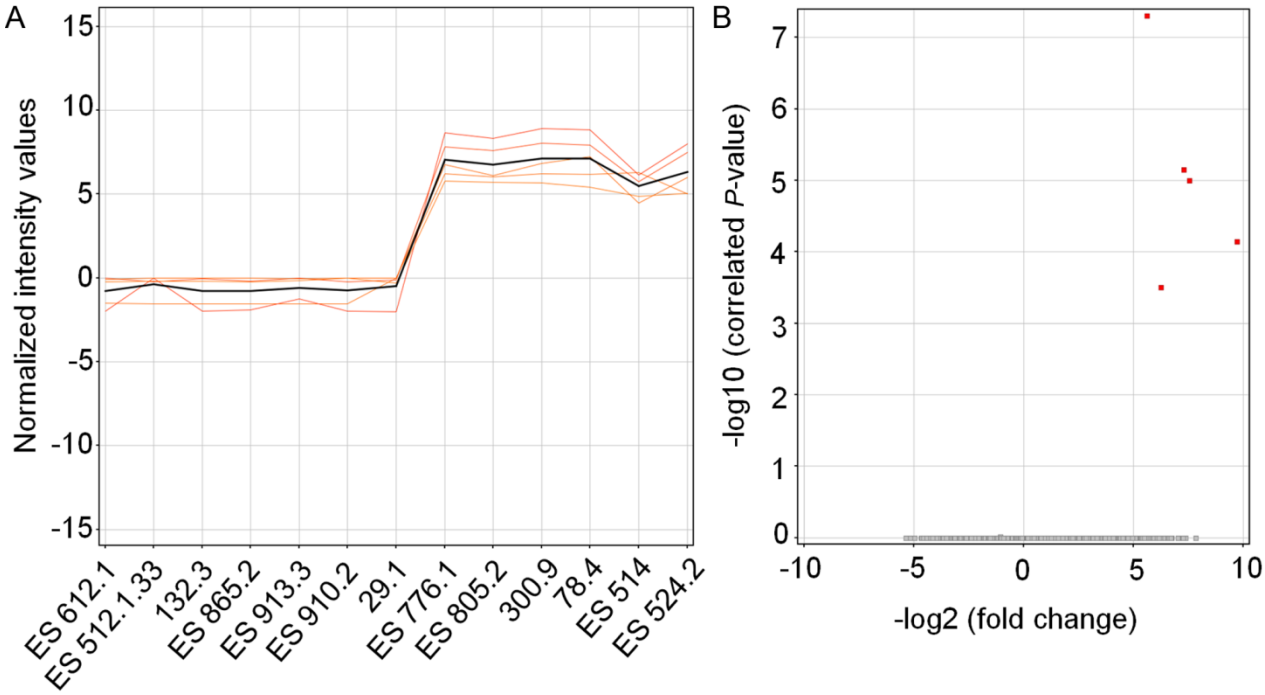

Figure S3. Constantly differentially regulated microarray probes in apomictic compared to sexual *Boechera* genotypes. A line chart (A) and a Volcano plot (B) display the relative expression ratio of five microarray probes which are constantly upregulated in apomictic compared to sexual *Boechera* genotypes. The Y-axes of the line chart (A) displays the averaged Log2 normalized fluorescence intensity, while the X-axes show single *Boechera* genotypes and their specific mode of reproduction. The black line represents the median expression level of the five candidate probes. The volcano plot (B) displays the negative log (base 10) of *P* values from unpaired *t*-tests on the Y-axes, while the X-axes 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.

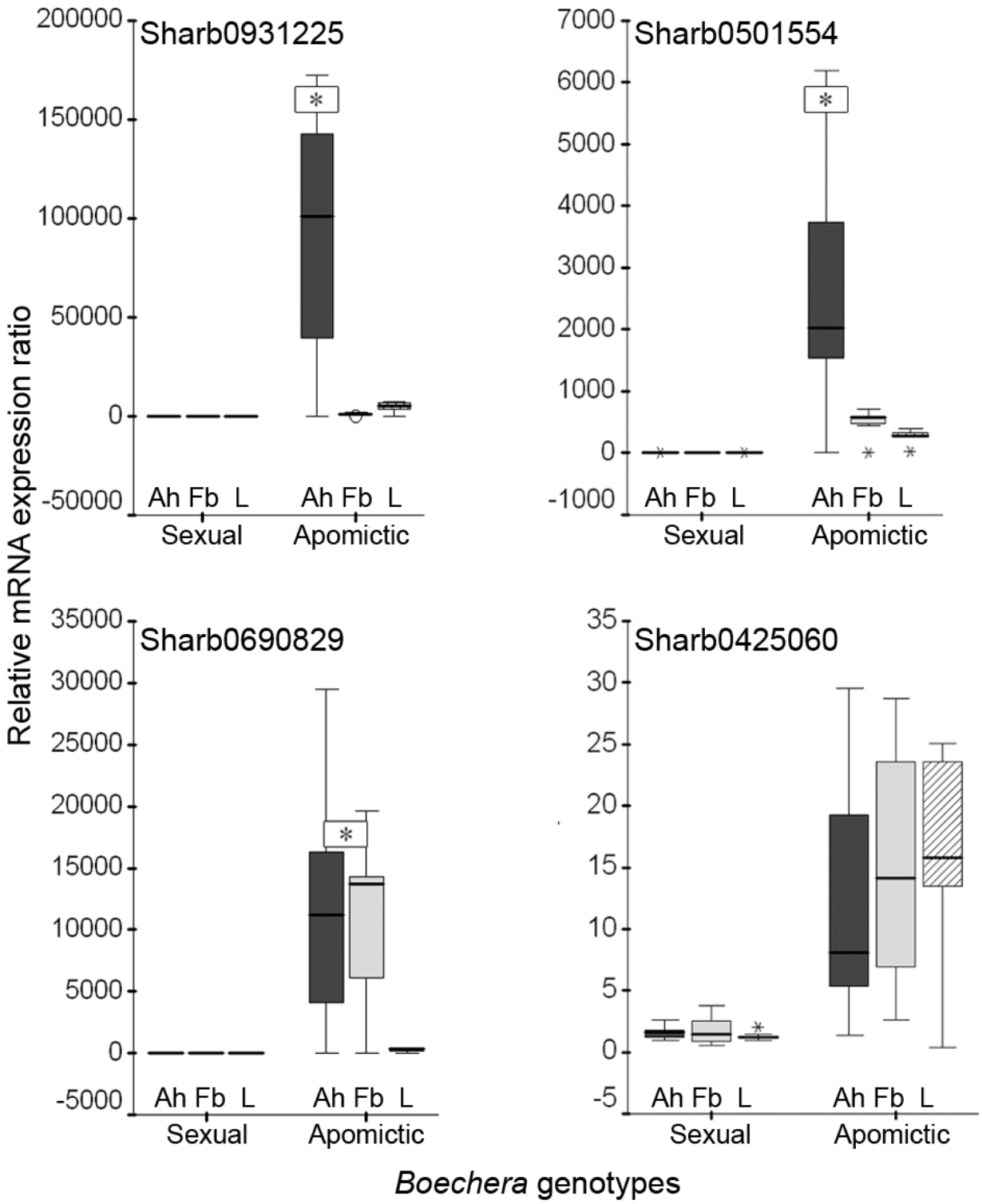

Figure S4. Validation of differentially-expressed microarray probes by qRT-PCR.

Expression of candidate microarray probes was tested in antherhead (Ah), flower bud withoutantherhead (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 (*$p<0.05$).

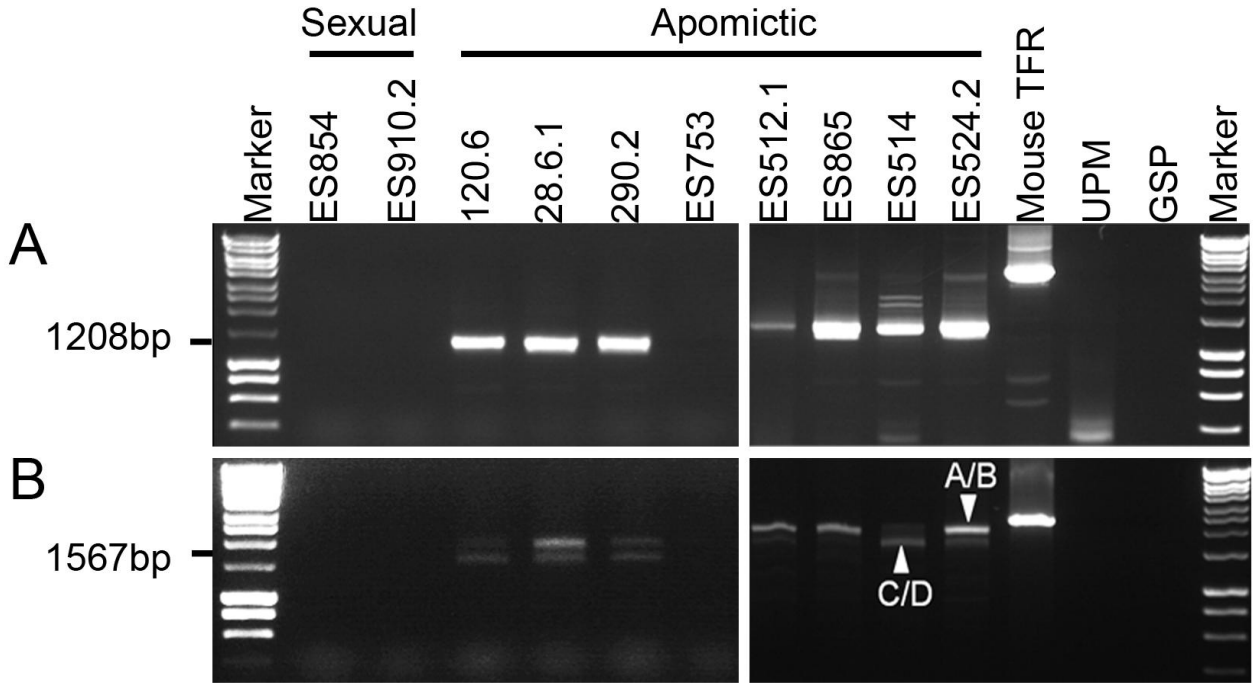

Figure S5. Rapid amplification of cDNA 3`- and 5`-ends of the candidate gene *UPGRADE*. 3`-end (A) and 5`-end (B) RACE cDNA fragments were obtained from sexual and apomictic *Boechera* whole flower cDNA. Transferrin receptor (TFR) primers on cDNA from mouse heart total RNA with optimal cycling parameters were used as positive control. A single universal primer (UPM) from the SMARTer RACE kit and a single gene-specific primer (GSP4) were used as negative control. A/B and C/D marks indicate the different putative splicing forms of the candidate transcript.

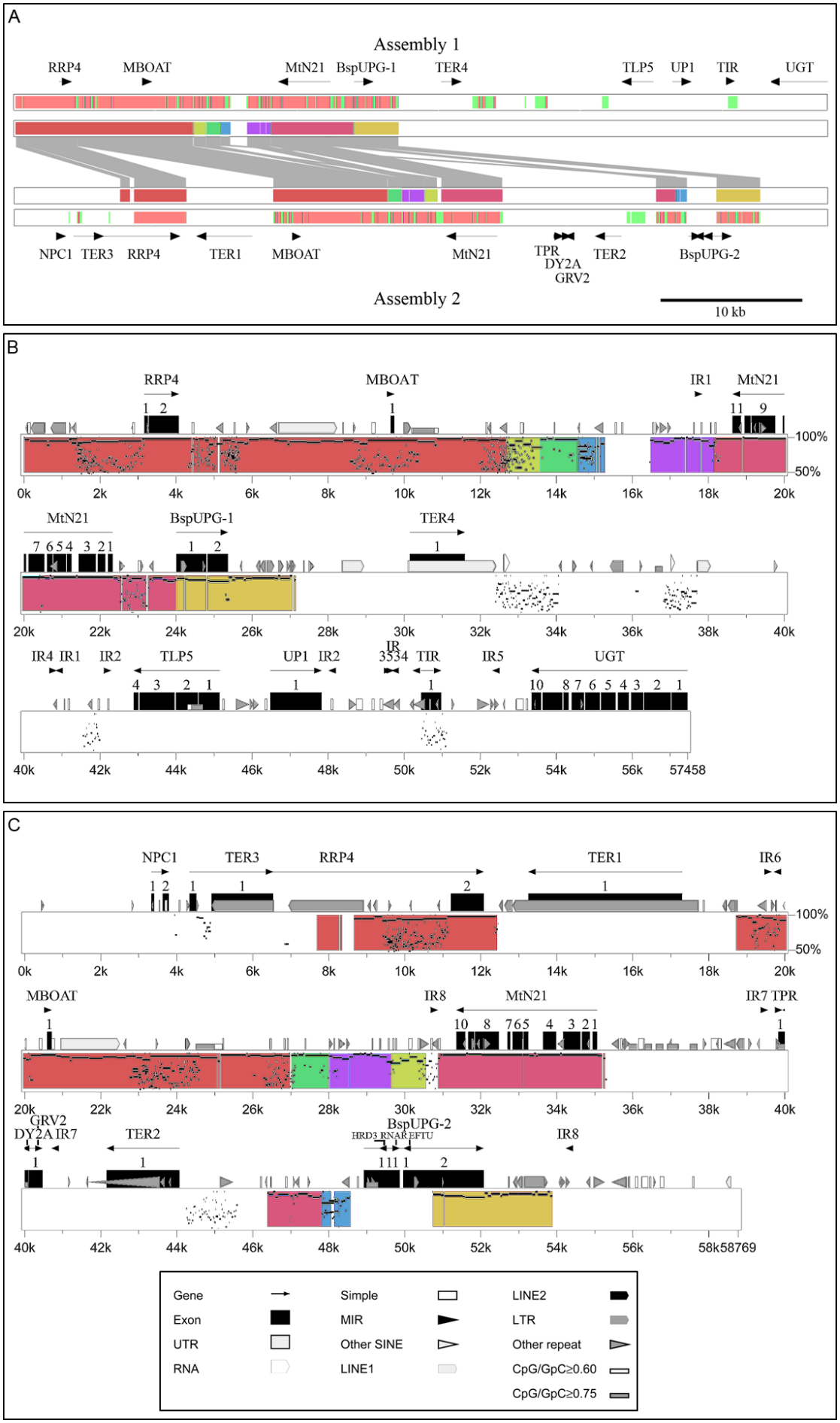
A
Assembly 1
RRP4
MBOAT
MtN21
BspUPG-1
TER4
TLP5
UP1
TIR
UGT
NPC1
TER3
RRP4
TER1
MBOAT
MtN21
TPR
DY2A
GRV2
TER2
BspUPG-2
Assembly 2
10 kb
B
RRP4
MBOAT
IR1
MtN21
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
MtN21
BspUPG-1
TER4
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
IR4
IR1
IR2
TLP5
UP1
IR2
IR
3534
TIR
IR5
UGT
40k
42k
44k
46k
48k
50k
52k
54k
56k
57458
C
NPC1
TER3
RRP4
TER1
IR6
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
MBOAT
IR8
MtN21
IR7
TPR
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
GRV2
DY2A
IR7
TER2
BspUPG-2
HRD3
RNAR
EFTU
IR8
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
58769
Gene
Exon
UTR
RNA
Simple
MIR
Other SINE
LINE1
LINE2
LTR
Other repeat
CpG/GpC≥0.60
CpG/GpC≥0.75

Figure S6. 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 C1, 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).

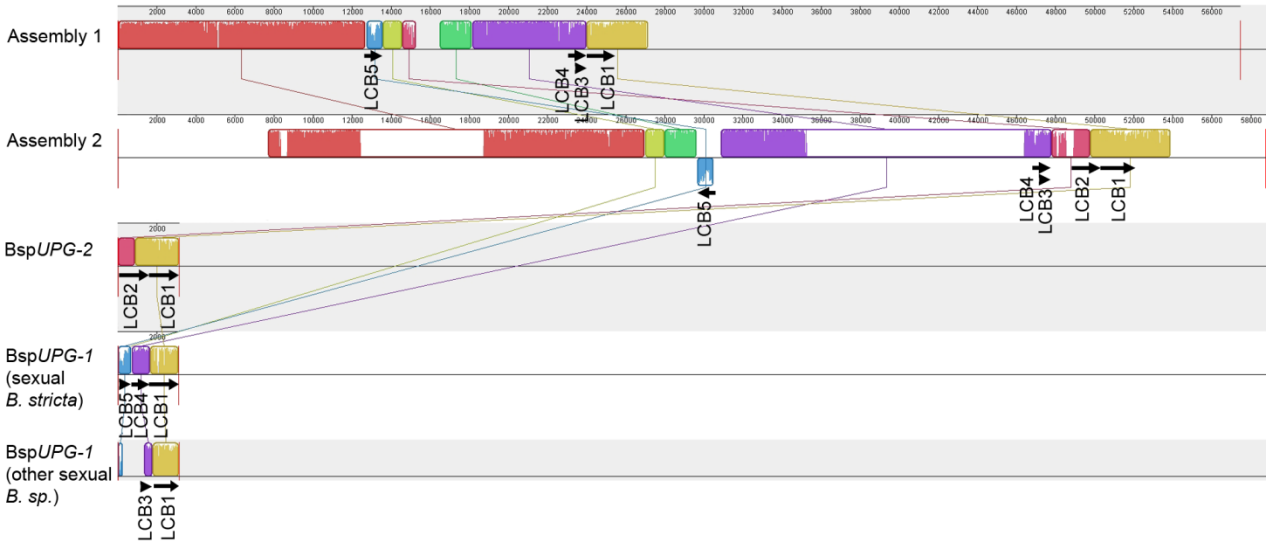

Figure S7. Distribution of locally collinear sequence blocks along the BAC clone Assembly 1 and Assembly 2. Locally collinear sequence (LCB) between Bsp*UPG-1*, Bsp*UPG-2*, the original locus (Assembly 1) and the duplicated locus (Assembly 2) derived from MAUVE (Darling et al., 2004).

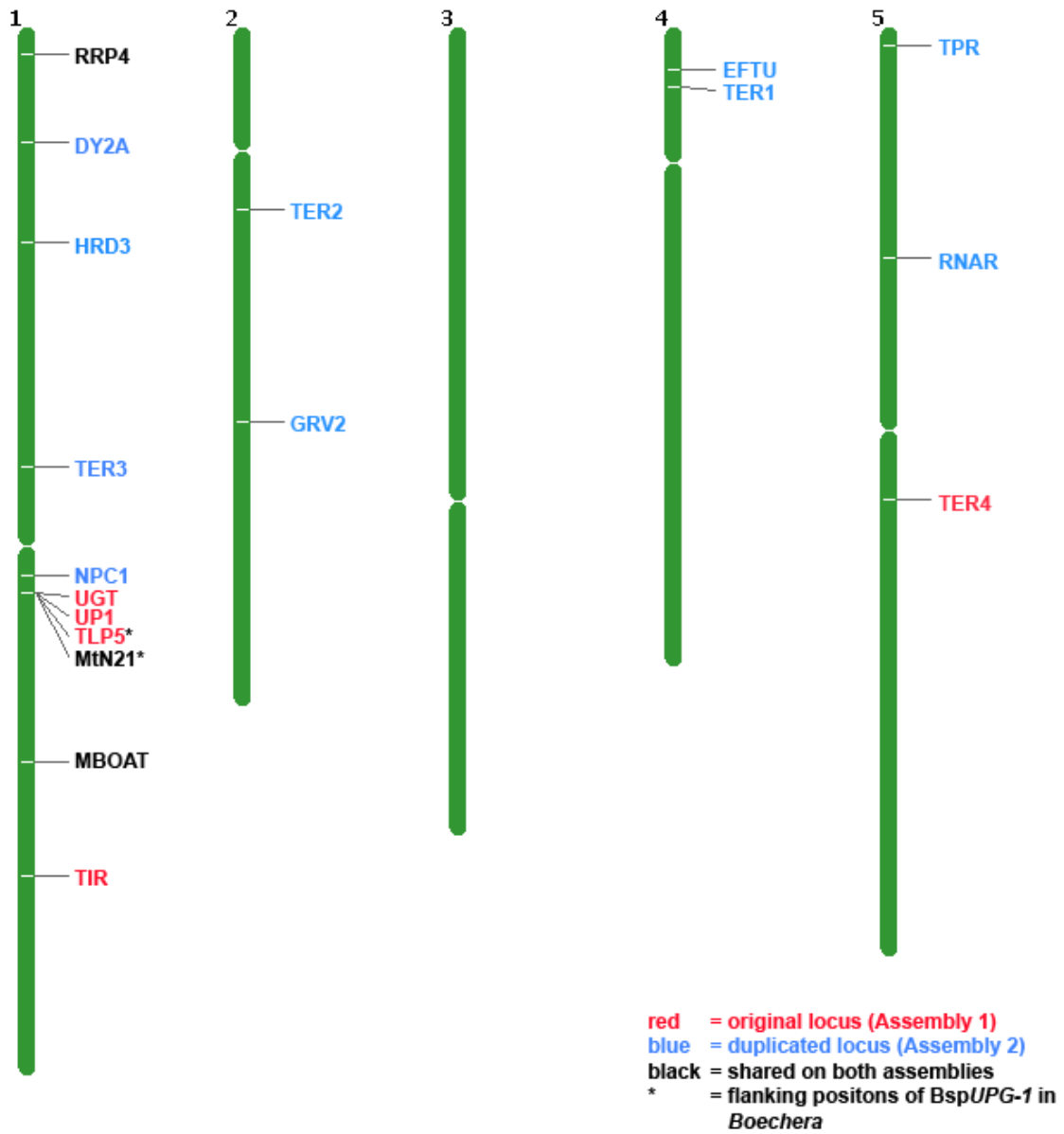

Figure S8. Distribution of *Arabidopsis* homologous loci corresponding to Assembly 1 and Assembly 2 in *Boechera* along *Arabidopsis* chromosomes. The Arabidopsis homologs from (red) the original locus (Assembly 1), (blue) the duplicated locus (Assembly 2) and (black) from both loci in *Boechera*, were mapped along *Arabidopsis* chromosomes using 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).

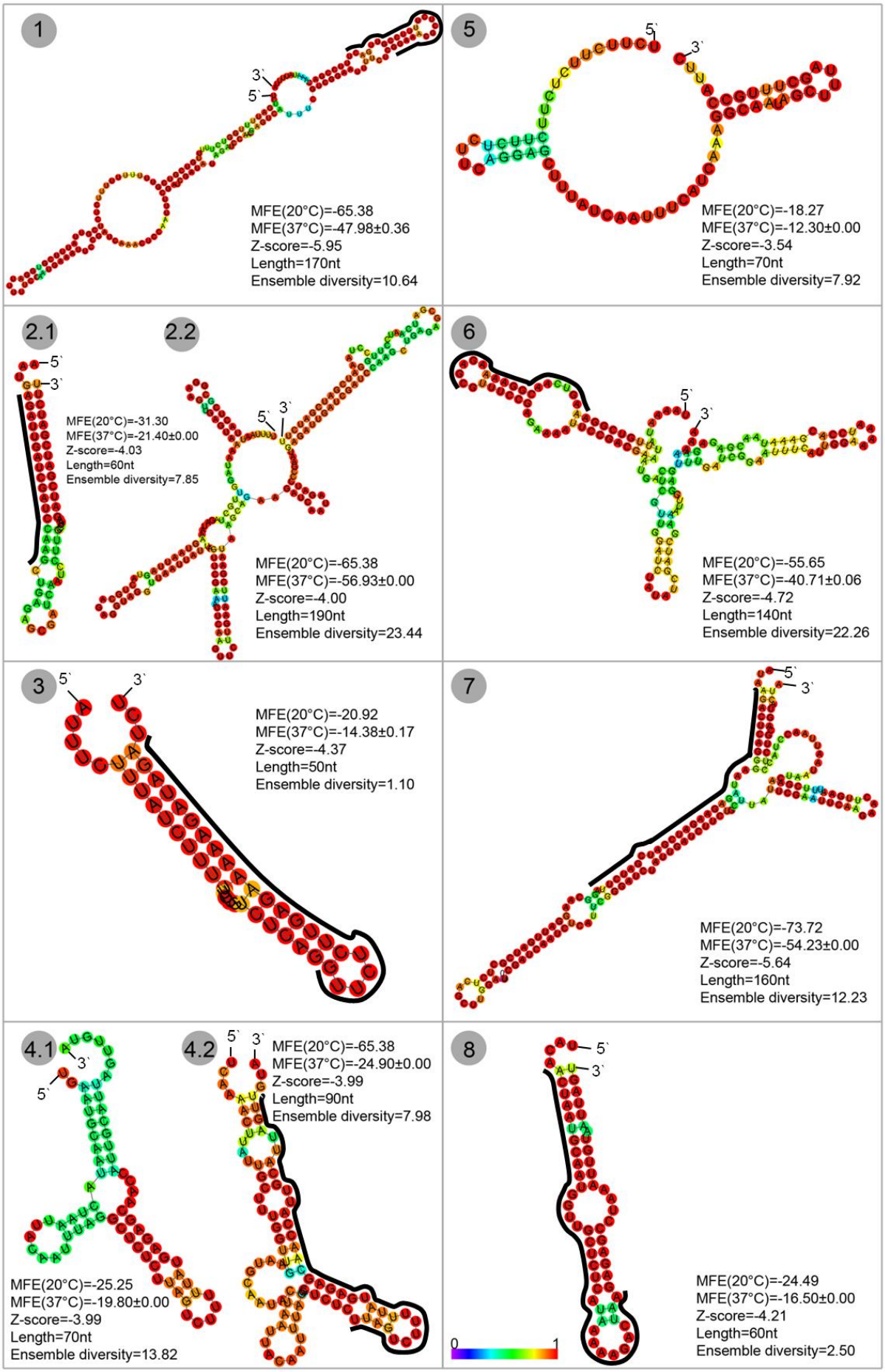
1
3`
5`
MFE(20°C)=-65.38
MFE(37°C)=-47.98±0.36
Z-score=-5.95
Length=170nt
Ensemble diversity=10.64
5
5`
3`
MFE(20°C)=-18.27
MFE(37°C)=-12.30±0.00
Z-score=-3.54
Length=70nt
Ensemble diversity=7.92
2.1
5`
3`
MFE(20°C)=-31.30
MFE(37°C)=-21.40±0.00
Z-score=-4.03
Length=60nt
Ensemble diversity=7.85
2.2
5`
3`
MFE(20°C)=-65.38
MFE(37°C)=-56.93±0.00
Z-score=-4.00
Length=190nt
Ensemble diversity=23.44
6
5`
3`
MFE(20°C)=-55.65
MFE(37°C)=-40.71±0.06
Z-score=-4.72
Length=140nt
Ensemble diversity=22.26
3
5`
3`
MFE(20°C)=-20.92
MFE(37°C)=-14.38±0.17
Z-score=-4.37
Length=50nt
Ensemble diversity=1.10
7
5`
3`
MFE(20°C)=-73.72
MFE(37°C)=-54.23±0.00
Z-score=-5.64
Length=160nt
Ensemble diversity=12.23
4.1
3`
5`
MFE(20°C)=-25.25
MFE(37°C)=-19.80±0.00
Z-score=-3.99
Length=70nt
Ensemble diversity=13.82
4.2
5`
3`
MFE(20°C)=-65.38
MFE(37°C)=-24.90±0.00
Z-score=-3.99
Length=90nt
Ensemble diversity=7.98
8
5`
3`
0
1
MFE(20°C)=-24.49
MFE(37°C)=-16.50±0.00
Z-score=-4.21
Length=60nt
Ensemble diversity=2.50

Figure S9. Thermodynamically stable non-random secondary structures of Bsp*UPG-2*. A CLUSTALW alignment of Bsp*UPG-2* from six different *Boechera* genotypes was analyzed for non-protein-coding RNA (npcRNA) discovery using the RNAz program (Washietl et al., 2005). Minimum folding energies (MFE) measure the thermodynamic stability of a candidate secondary structure and Z-scores define random from non-random secondary structures and were computed from a sliding window approach (step size=10nt, window delta=10nt) using window sizes between 50 and 300nt. Any windows producing a significant Z-score during the scanning process were considered candidate regions for structural npcRNAs. The thermodynamically stable secondary structure with the most negative (*i.e.* most stable and non-random) Z-score for each candidate region is shown. The structures were drawn using the RNAfold software (Hofacker et al., 1994). Consensus MFEs (*i.e.* consensus from single MFEs of aligned sequences) were computed at 37°C and are shown with standard deviations. All MFEs were expressed as negative kcal/mol. The ensemble diversity is a measure to indicate how much time the secondary strucure stays in the actual "target" shape. Black lines denote regions of npcRNAs homologous to potential target sites (Table S10).

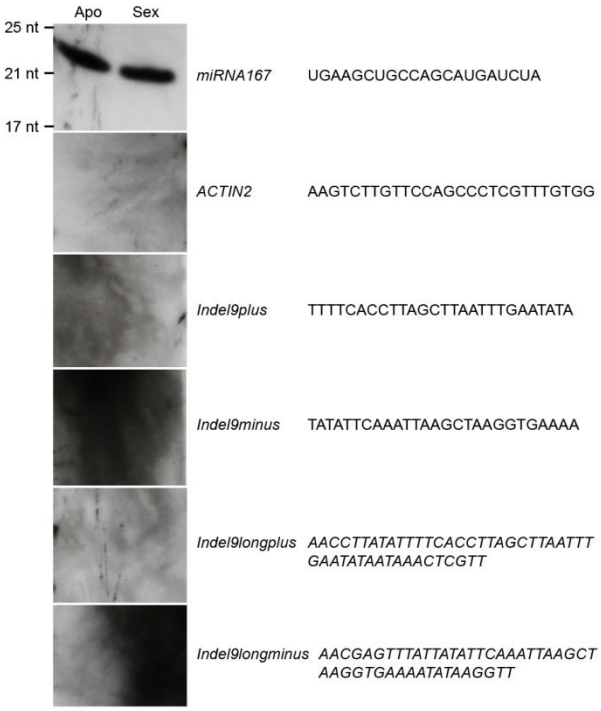

Figure S10. Northern blot analysis for putative small RNAs homologous to indel 9 at Bsp*UPG-2*. Pooled flower tissues from different developmental stages for each four pooled sexual and four pooled apomictic *Boechera* genotypes were used.

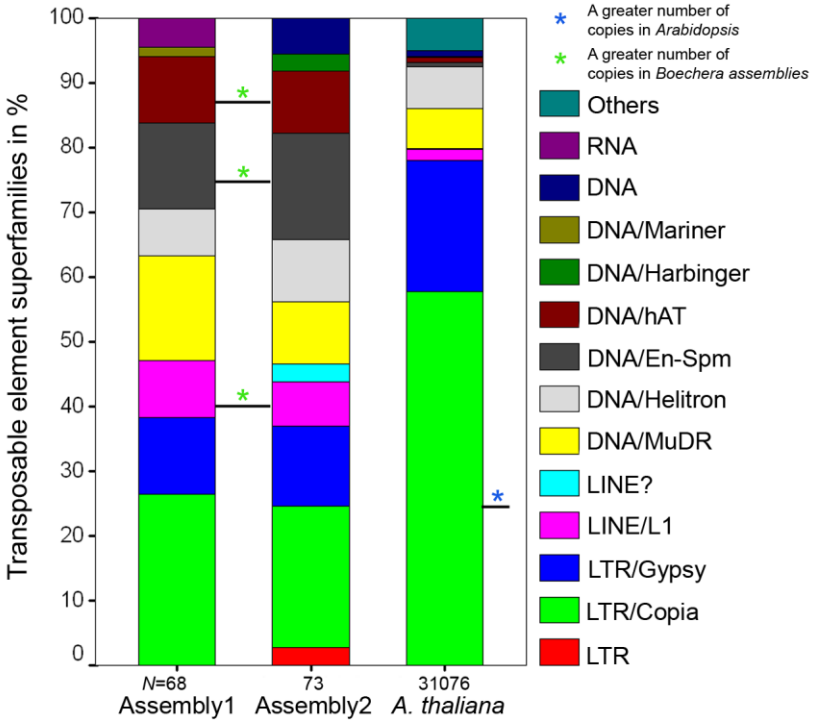

Figure S11. Distribution of the TE superfamilies along the original and duplicated *UPGRADE* locus. TE superfamily distributions along the original and duplicated *UPGRADE* locus in *Boechera* were examined in comparison with the distribution of TE superfamilies along the complete *A. thaliana* genome.

Table S1. List of *Boechera* genotypes with population information used for histological examination of microsporogenesis and candidate gene analyses.

Genotype No. ES/TS ID[b]	Species[a]	Map ID[b]	SPF[b]	cpDNA Haplotype[a]	Location name[b]	Flow cytometry[c] Pollen	Flow cytometry[c] Seed (Cm:Cp)	MOR[c*]	Experiment[d]
ES 612.1	*B. stricta*	20	UT-02	ah	Duck Creek, UT	1C	2C:3C	obligate sexual	1, 3, 5, 6, 7
ES 512.1.33	*B. stricta*	0	SAD12	dg	Taylor River, CO	1C	2C:3C	obligate sexual	1, 3, 4,5
132.3*	*B. stricta*	10*	CACR5	AV*	Canyon Creek/Vipond Park, Beaverhead County, MO	1C	2C:3C	obligate sexual	1, 3, 4,5, 7
ES 865	*B. stricta*	43	MT-56	ah	Joseph Creek, MO	1C	2C:3C	obligate sexual	1, 3, 5
ES 854	*B. stricta*	37	ID-73	ah	Sleeping Deer, Custer County, ID	1C	(Aliyu et al., 2010)*	obligate sexual	3
ES 685	*B. stricta*	7	CO-25	as	Sourdough-north, CO	1C	(Aliyu et al., 2010)*	low-level facultative apomict	3
ES 558.2	*B. stricta*	17	CO-28	bh	Buffalo Pass, CO	1C	2C:3C	low-level facultative apomict	5
ES 913.3	*B. polyantha**	66	ID-90	ci	Panther Creek, ID	1C	2C:3C	obligate sexual	1, 3, 5
ES 910.2	*B. polyantha**	32	ID-75	n/a	Parker Meadow, Custer County, ID	1C	2C:3C	obligate sexual	1, 3, 5
ES 903.1	*B. polyantha**	n/a	ID-022	n/a	Haystack Saddle, ID	1C	n/a	obligate sexual	3
29.1*	*B. polyantha**	8*	BHP2	CI*	Big Hole Pass, Beaverhead County, MO	1C	2C:3C	obligate sexual	1, 3, 5
105.18*	*B. polyantha**	7*	MT-45 Banra	n/a	Bandy Ranch, Missoula County, MO	1C	(Aliyu et al., 2010)*	obligate sexual	3, 5, 7
ES 776.1	*B. polyantha**	63	ID-29	dn	Morgan Switchback, ID	2C	2C:6C + 2C:8C	obligate apomict	1, 3, 5, 6
ES 805.2	*B. polyantha**	68	MT-33	ah	Ranch Creek Rd, Sliderock, MO	2C	2C:6C	high-level facultative apomict	1, 3, 4, 5, 6
300.9*	*B. retrofracta**	5*	BC-12	B*	Birch Creek, Ravalli County, MO	2C	2C:6C + 2C:8C	obligate apomict	1, 3, 4, 5, 6
78.4*	*B. retrofracta**	21*	MT-14	n/a	Ranch Creek, Granite County, MO	2C	2C:6C / 4C:8C	high-level facultative apomict	1, 2, 3, 5
28.6*	*B. polyantha X retrofracta**	5*	BC-13	CT*	Birch Creek, Ravalli County, MO	2C	(Aliyu et al., 2010)*	high-level facultative apomict	3, 4, 5
120.6	*B. stricta X retrofracta**	19*	MURA2	n/a	Mule Ranch, Beaverhead County, MO	2C	(Aliyu et al., 2010)*	obligate apomict	3, 4, 5
ES 753	*B. lignifera**	74	UT-15	u	Beckwith Spring, Summit County, UT	2C	2C:6C	obligate apomict	1, 3, 5, 6
ES 514	*B. divaricarpa*	55	MT-15	bs	Vipond Park, Beaverhead County, MO	2C	2C:6C / 2C:8C / 4C:6C	high-level facultative apomict	1, 3, 4, 5, 6, 7
ES 524.2	*B. divaricarpa*	60	MT-12	n/a	Mule Ranch, Beaverhead County, MO	2C	2C:6C / 2C:8C / 4C:6C	high-level facultative apomict	1, 3, 4, 5, 6, 7
290.8*	*B. retrofracta**	31*	BS-30	CI*	Sonora Pass, Toiyabe NF, CA	2C	-*	high-level facultative apomict	3, 4, 7

[a] Species marked with asterisk have been identified based upon silique orientation, trichome morphology, cpDNA marker (M. Koch, personnel comm., Koch et al., (2003); Kiefer et al., (2009)) and were formerly classified as *B. holboellii*. Species without asterisks according to the nomenclature formerly established in Windham and Al-Shebaz, 2007; cpDNA haplotypes from Schranz et al., (2005). n/a, data not applicable. Numbers behind dot mark if seeds were taken from F1 population.

[b] Map ID and population data as described in Schranz et al., (2005) (ES ID) and Sharbel et al., (2005) (TS ID, marked with asterisks) respectively.

[c] Mode of reproduction was confirmed using the flow cytometric analysis of seeds (FCSS, Matzk et al., (2000); * low level (1-3%) and high level (87-99%) apomict classification *sensu* Aliyu et al., (2010)) and pollen nuclei.

[d] Sample grouped for experimental parts presented in this publication: 1, Gene expression; 2, Chromosome walking and BAC library screen; 3, qRT-PCR-based validation; 4, RACE; 5, SNP Analysis; 6, Cytology; 7, Developmental staging

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

Genotype	Reproductive mode (FCSS)	Plant ID	Flower ID	Anther ID	Meiocyte count	(%)			
						Monad[b]	Dyad	Triad[c]	Tetrad
ES 612.1	**obligate sexual**	**B10-204**	1	2	323	0.62	8.36	17.34	73.68
		B10-203	1	2	242	8.68	0.41	2.48	88.43
		B11-403	1	5	32	0.00	0.00	9.38	90.63
			2	1+2	60	0.00	0.00	0.00	100.00
				3+4	49	0.00	0.00	4.08	95.92
		B11-397	1	1	41	0.00	0.00	0.00	100.00
				2	10	0.00	0.00	0.00	100.00
				3	86	0.00	0.00	0.00	100.00
				4	80	0.00	0.00	2.50	97.50
				5	152	0.00	0.00	12.50	87.50
				6	69	0.00	0.00	1.45	98.55
			2	1	77	0.00	0.00	12.99	87.01
				2+3	95	0.00	0.00	3.16	96.84
				4	36	0.00	0.00	13.89	86.11
		Group I[a]			**1352**	**1.70**	**2.07**	**7.91**	**88.31**
		Group II[a]			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
ES 805.2	**high facultative apomictic**	**B10-220**	1	1	131	12.98	64.89	13.74	8.40
				2	140	10.00	86.43	0.71	2.86
				3	47	17.02	70.21	4.26	8.51
			2	1	33	21.21	30.30	42.42	6.06
		B10-430	1	1	1010	2.67	97.13	0.00	0.20
				2	315	5.71	94.29	0.00	0.00
		B11-1199	1	1	703	15.50	84.35	0.14	0.00
			2	1	635	49.13	50.87	0.00	0.00
		B11-1205	1	1	480	47.08	52.92	0.00	0.00
			2	1	164	17.68	82.32	0.00	0.00
				2	1024	59.96	40.04	0.00	0.00
		B11-1042	1	1	63	42.86	57.14	0.00	0.00
			2	1	296	11.15	88.85	0.00	0.00
		B11-1043	1	1	610	7.38	92.62	0.00	0.00
		B11-1197	1	1	886	55.98	44.02	0.00	0.00
		Group I[a]			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		Group II[a]			**6537**	**30.32**	**68.78**	**0.55**	**0.35**
ES 753	**obligate apomictic**	**B11-1185**	1	1	17		5.88	58.82	35.29
				2	415	0.00	0.48	17.59	81.93
				3	158	0.00	0.00	24.68	75.32
			3	1	297	0.34	2.36	40.74	56.57
			2	1	293	0.34	3.75	32.42	63.48
				2	73	0.00	4.11	41.10	54.79
				3	464	0.22	0.65	20.69	78.45
		B11-1183	1	1	93	38.71	61.29	0.00	0.00
				2	352	65.91	34.09	0.00	0.00
			2	1	479	50.73	49.27	0.00	0.00
				2	812	60.10	39.90	0.00	0.00
				3	221	38.01	61.99	0.00	0.00
			3	1	457	32.60	67.18	0.22	0.00
		B11-1178	1	1	81	69.14	2.47	18.52	9.88
				2	23	26.09	0.00	21.74	52.17
				3	266	0.00	0.00	40.60	59.40
				4	39	100.00	0.00	0.00	0.00
				5	109	0.00	0.92	44.95	54.13
			2	2	254	0.00	0.39	31.10	68.50
				3	128	0.00	3.13	57.03	39.84

				4	351	0.28	2.85	51.28	45.58
		B11-412	1	1	61	4.92	91.80	3.28	0.00
				2	125	3.20	94.40	1.60	0.80
				3	12	0.00	83.33	16.67	0.00
				4	231	4.33	93.07	2.60	0.00
				5	81	8.64	88.89	2.47	0.00
		Group I[a]			**2968**	**3.54**	**1.52**	**32.78**	**62.16**
		Group II[a]			**2924**	**42.95**	**56.50**	**0.51**	**0.03**
ES 776.2	**obligate apomictic**	**B11-415**	1	1	203	34.98	65.02	0.00	0.00
				2	74	58.11	41.89	0.00	0.00
				3	223	49.33	50.67	0.00	0.00
		B11-418	1	1	112	72.32	27.68	0.00	0.00
			2	1	54	59.26	40.74	0.00	0.00
		B11-420	1	1	276	23.91	76.09	0.00	0.00
			2	1	546	6.78	93.22	0.00	0.00
		Group I[a]			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		Group II[a]			**1488**	**29.57**	**70.43**	**0.00**	**0.00**
300.9	**obligate apomictic**	**B11-344**	1	1	179	0.00	0.00	30.73	69.27
				2	233	7.30	1.72	33.05	57.94
				3	108	0.93	0.93	36.11	62.04
		B11-525	1	1	147	93.20	6.80	0.00	0.00
				2	39	100.00	0.00	0.00	0.00
			2	1	156	0.00	5.13	27.56	67.31
				2	363	0.00	0.83	17.08	82.09
				3	347	7.78	0.86	26.80	64.55
			3	2+3	96	19.79	65.63	9.38	5.21
				4	104	9.62	59.62	7.69	23.08
		B11-345	1	1	14	0.00	0.00	71.43	28.57
				2	339	0.88	0.59	40.41	58.11
			2	1	331	0.30	3.63	26.59	69.49
				2	174	14.37	0.57	20.69	64.37
		B11-346	1	1	189	75.13	24.34	0.00	0.53
				2	22	72.73	27.27	0.00	0.00
				3	37	94.59	2.70	2.70	0.00
			2	1+2	94	95.74	3.19	1.06	0.00
			3	1	253	79.84	20.16	0.00	0.00
		B11-527	1	1	585	34.53	65.47	0.00	0.00
			2	1+2	251	91.24	8.76	0.00	0.00
		Group I[a]			**2630**	**10.61**	**6.43**	**24.98**	**57.98**
		Group II[a]			**1431**	**64.01**	**35.78**	**0.14**	**0.07**
ES 524.2	**high facultative apomictic**	**B11-652**	1	1+2	660	22.73	77.27	0.00	0.00
				3+4	666	13.96	86.04	0.00	0.00
			2	1+2	378	32.54	67.20	0.26	0.00
			3	1+2	336	53.27	46.73	0.00	0.00
				3+4	589	11.71	88.29	0.00	0.00
				5+6	839	19.79	79.98	0.12	0.12
		B11-924	1	1+2	50	0.00	0.00	12.00	88.00
				3	347	0.00	0.29	23.34	76.37
				4	112	0.00	2.68	26.79	70.54
		B11-926	1	1+2	179	24.58	0.56	21.23	53.63
				3	16	0.00	0.00	12.50	87.50
				4	52	44.23	0.00	15.38	40.38
			2	1	243	0.00	0.41	25.51	74.07
				2+3	282	0.00	0.00	12.77	87.23
		B11-932	1	1	64	4.69	95.31	0.00	0.00
				2+3	79	24.05	75.95	0.00	0.00
				4+5	186	26.88	73.12	0.00	0.00
			2	1	198	0.00	0.00	7.07	92.93

				2+3	133	3.01	0.00	15.04	81.95
				4+5	25	4.00	0.00	40.00	56.00
		B11-927	1	1+2	52	0.00	0.00	15.38	84.62
				3+4	37	0.00	0.00	18.92	81.08
		Group I[a]			**2055**	**7.01**	**12.80**	**15.67**	**64.53**
		Group II[a]			**3468**	**22.49**	**77.42**	**0.06**	**0.03**
ES 514	**high facultative apomictic**	**B11-912**	1	1	77	6.49	93.51	0.00	0.00
				2	52	9.62	90.38	0.00	0.00
				3	90	4.44	95.56	0.00	0.00
				4	124	6.45	93.55	0.00	0.00
				5	214	3.74	95.79	0.47	0.00
			2	1	65	20.00	80.00	0.00	0.00
				2+3	99	31.31	68.69	0.00	0.00
			3	1	47	21.28	76.60	2.13	0.00
				2+3	80	31.25	68.75	0.00	0.00
		B11-371	1	1+2	73	0.00	0.00	31.51	68.49
				3+4	223	0.00	0.00	23.32	76.68
				5	222	0.00	0.00	8.11	91.89
				6	23	0.00	0.00	13.04	86.96
			2	1+2	359	0.00	0.00	12.81	87.19
				3+4	60	0.00	0.00	10.00	90.00
				5+6	128	0.00	0.00	15.63	84.38
		B11-366	1	1+2	56	0.00	0.00	21.43	78.57
				3+4	59	0.00	0.00	8.47	91.53
			2	1+2	89	21.35	0.00	15.73	62.92
				3	48	41.67	0.00	6.25	52.08
				4+5	19	0.00	0.00	10.53	89.47
		B11-364	1	1+2	138	0.00	0.00	4.35	95.65
				3+4	186	0.00	0.00	6.45	93.55
				5	69	0.00	0.00	1.45	98.55
			2	1+2	42	0.00	0.00	11.90	88.10
			3	1+2	18	0.00	0.00	11.11	88.89
			4	1+2	242	0.00	0.00	4.55	95.45
				3	26	0.00	0.00	3.85	96.15
				4+5	62	0.00	0.00	1.61	98.39
		B11-1344	1	1+2	47	25.53	74.47	0.00	0.00
		B11-367	1	2+3	112	0.89	0.00	8.93	90.18
		Group I[a]			**2254**	**1.77**	**0.00**	**11.22**	**87.00**
		Group II[a]			**895**	**13.52**	**86.26**	**0.22**	**0.00**

[a]Group I - plants producing primarily (>50 %) reduced gametes; Group II - plants producing primarily (>50%) unreduced gametes (chimeric plants were assigned to one group based on the prevalent meiocyte status).

[b]Subpopulation of monads might be pseudomonads which develop later into dyads.

[c]Triad count might be overestimated due to subpopulation of pseudotriads, representing tetrads with a loss of one nucleus.

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

Individual plant ID[a]	Genotype	Embryo: endosperm ratio	Seed frequency (*N*)	Seed frequency (%)	Pollen type expectation from embryo:endosperm ratio	Pollen type frequency (%)
B11-1178	ES 753	2C:5C	3	6.38	0+C	49.72
		2C:6C	40	85.11	0+2C	1.44
		other	4	8.51	other	48.84
B11-525	300.9	2C:5C	0	0.00	0+C	52.40
		2C:6C	49	92.45	0+2C	11.90
		other	4	7.55	other	35.70
B11-345	300.9	2C:5C	0	0.00	0+C	63.29
		2C:6C	53	100.00	0+2C	1.75
		other	0	0.00	other	34.96
B11-926	ES 524.2	2C:5C	5	11.11	0+C	72.15
		2C:6C	38	84.44	0+2C	0.26
		other	2	4.44	other	27.59
B11-932	ES 524.2	2C:5C	6	13.04	0+C	44.82
		2C:6C	40	86.96	0+2C	37.52
		other	0	0.00	other	17.66
B11-366	ES 514	2C:5C	5	16.13	0+C	73.32
		2C:6C	20	64.52	0+2C	0.00
		other	6	19.35	other	26.68

[a]Individual plant IDs are identical to IDs in Table S2.

Table S4. Frequencies of microspores, meiotic and pollen mother cells (PMC) relative to antherhead length. Statistical significance of an antherhead size-specific PMC enrichment was calculated with a two-tailed Fisher's exact test. Antherhead sizes correlate to bud stages S8 to S12 only.

Antherhead size (µm)	**290-430**		**430-490**		**490-550**		**550-610**		**610-670**		**670-758**	
Reproduction	Sex	Apo	Sex	Apo	Sex	Apo	Sex	Apo	Sex	Apo	Sex	Apo
PMC	16	20	15	13	21	12	6	6	1	0	0	0
Meiosis	0	2	1	13	7	20	35	14	12	6	5	1
Microspores	0	0	1	6	11	15	11	15	15	17	7	13
Total count	**16**	**22**	**17**	**32**	**39**	**47**	**52**	**35**	**28**	**23**	**12**	**14**
***P*-value**	1.79E-08	1.54E-10	4.29E-06	9.57E-02	7.28E-03	8.11E-01	1.00E+00	9.81E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00
Significance	***	***	***	n. s.	**	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.	n. s.

Table S5. Microarray probes demonstrating significant absolute fold change (AFC) upregulation in apomictic genotypes. Significance values of differential expression of microarray probes between all apomicts *versus* all sexuals were confirmed by unpaired Student's *t*-tests with asymptotic *P*-value computation and Bonferroni FWER multiple testing correction. *E*-values correspond to the best hit of the *Boechera* cDNA in a BLASTN analysis to GenBank nucleotide collection.

Microarray probe No.	Corrected *P*-value	AFC	EMBL No. of cDNA read homologous[a]	length (bp)	TAIR 10 peptide AGI code	Predicted function	*E*-value
Sharb1199059	7.08E-06	158.54	ERS317552, ERS317553, ERS317554	255 260 269	AT1G11410	*S-locus* lectin protein kinase	2.00E-30
Sharb0931225	7.13E-05	849.31	ERS317555	217	no hit	n/a	n/a
Sharb0501554	1.00E-05	188.3	ERS317556	277	AB180901	S-12 SRK gene for *S-locus* receptor kinase	6.00E-05
Sharb0690829	3.12E-04	76.51	ERS317557	269	no hit	n/a	n/a
Sharb0425060	4.90E-08	48.74	ERS317558	186	no hit	n/a	n/a

Table S6. QRT-PCR primers used for validation of microarray candidate probes. Housekeeping genes *ACTIN2* and *EFα1* were used as internal control. *GAPDH* was used as positive control. Efficiencies were calculated with Real Time PCR Miner software v2.0 (Zhao and Fernald 2005).

Microarray probe no.	**Primer name**	**Primer sequence 5`→ 3`**	**Amplicon size (bp)**	**Amplification efficiency ± SD**
Sharb0931225	CON234B4_L CON234B4_R	TTGCTTTGGTTGAATGCAATAC AATTACTAAATTTGCACACCACCTG	177	0.91 ± 0.019
Sharb0501554	CON234B3_L CON234B3_R	TGTGTTTGCTGTGCACTTACAG TCTCAAGAGAACCTGAGACACAAA	134	0.76 ± 0.011
Sharb0690829	CON234B2_L CON234B2_R	TCTTCTTCGCCATCGTTCAT TTCACAAATCTAGATGAAGAACCCT	182	0.90 ± 0.009
Sharb0425060	CON5B9_L CON5B9_R	TGGATGAGAAATACAAACTTGG AGGAACACGCCCTCAAATTC	100	0.95 ± 0.01
ACTIN2	RT_Act2_T7_L RT_Act2_T7_R	GTTCCACCACTGAGCACAATGTTACC AGTCTTGTTCCAGCCCTCTTTTGTG	132	0.91 ± 0.006
EFα1	RT_EFα1_M13_L RT_EFα1_M13_R	CCAAGGGTGAAAGCAAGGAGAGC CACTGGTGGTTTTGAGGCTGGTATCT	75	0.95 ± 0.01
GAPDH	GAPDH_For GAPDH_Rev	CAAGGTCATCCATGACAACTTTG GTCCACCACCCTGTTGCTGTAG	496	-

Table S7. Relative mRNA expression values for four microarray probes in somatic and reproductive tissues. The values are means calculated from Ct values of four technical replicates per sample. Relative mRNA expression was normalized independently against tissue-specific tested *Boechera* housekeeping genes *ACTIN2* and *EF1-α (ACTIN2* normalized results shown). Results for a one-way ANOVA with Tukey-HSD *post-hoc* test ($P \leq 0.05$) for each tissue comparison are listed at the bottom of the table.

Genotype	Reprod.	Sharb0931225			Sharb0501554			Sharb0690829			Sharb0425060		
		Anther	Flower	Leaf	Anther	Flower	Leaf	Anther	Flower	Leaf	Anther	Flower	Leaf
ES_612.1	Sex	2.32	0.25	2.46	1.04	0.52	0.66	1.36	0.37	3.52	1.75	0.53	1.18
ES_512.1.33	Sex	1.01	0.02	1.53	1.34	0.34	0.50	2.74	0.91	2.17	1.00	1.49	1.20
132.3	Sex	1.17	0.91	1.65	1.17	1.11	0.47	0.85	0.68	5.13	1.61	2.41	1.43
ES865	Sex	3.11	0.07	1.39	3.93	0.36	0.61	1.86	0.86	1.73	2.61	2.58	1.09
ES_913.3	Sex	1.33	0.17	1.98	1.13	0.38	5.03	1.53	1.79	4.54	1.85	3.77	2.00
ES_910.2	Sex	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
29.1	Sex	4.08	0.07	1.31	1.52	0.51	1.50	1.81	1.45	1.34	1.38	0.78	1.22
ES_776.1	Apo	112938.72	1171.28	7145.08	2083.91	491.33	279.92	14558.31	12204.14	332.15	8.06	9.82	14.15
ES_805.2	Apo	63913.70	1647.27	7461.92	2020.97	712.26	302.17	8150.26	19661.02	382.98	9.81	28.71	23.48
300.9	Apo	370851.67	1102.93	5136.35	6188.98	450.94	362.99	29487.02	13733.44	330.13	29.54	19.40	15.80
78.4	Apo	101234.30	1388.11	3689.34	1594.73	580.25	271.53	11241.40	14066.27	228.78	6.33	27.68	25.07
ES 753	Apo	40.25	0.06	57.10	1.61	6.11	21.38	0.94	0.48	70.45	1.34	2.65	0.35
ES_514	Apo	15657.15	958.52	3827.51	1481.44	600.78	279.45	2.52	0.57	0.23	4.42	3.98	12.82
ES_524.2	Apo	172436.97	2103.58	6033.22	5384.77	601.53	391.62	18058.93	14538.90	419.71	28.69	14.09	23.60
Mean ratio	Apo vs. Sex	59705.62	3362.15	2946.16	1685.13	815.92	195.40	7308.30	10510.60	90.81	7.88	8.47	12.64
Tukey-HSD	Anther	-	0.018	0.022	-	0.015	0.008	-	0.963	0.027	-	0.887	0.767
	Flower	0.018	-	0.995	0.015	-	0.947	0.963	-	0.045	0.887	-	0.971
	Leaf	0.022	0.995	-	0.008	0.947	-	0.027	0.045	-	0.767	0.971	-

Table S8. 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 S4). 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 sequences BAC clone.

[d]100% identical.

Table S9. Transcription factor binding sites (TFBS) upstream of Bsp*UPG-2* (+47809 to +48921 nt of Assembly 2).

Factor	Strand	Seq	Species	Source
AGL3	+	acaaCCATAattcgacga	*Arabidopsis*	TRANSFAC
AGL3	+	aattCCATAataaataac	*Arabidopsis*	TRANSFAC
AG	-	gcgatggatTTTGGtagt	*Arabidopsis*	TRANSFAC
AG	-	tggtcggaaTTTGGtcgg	*Arabidopsis*	TRANSFAC
AG	-	tcgttggatTTTGGtagt	*Arabidopsis*	TRANSFAC
AG	-	tggtcggatTTTGGtcgg	*Arabidopsis*	TRANSFAC
AG	+	aaatCCAAAatccgacca	*Arabidopsis*	TRANSFAC
AG	+	ccgaCCAAAattatgctt	*Arabidopsis*	TRANSFAC
AG	-	tcgatggttTTTGGtaga	*Arabidopsis*	TRANSFAC
AG	-	tagtcggatTTTGGttga	*Arabidopsis*	TRANSFAC
AG	+	ctgaCCAAActgaaattt	*Arabidopsis*	TRANSFAC
AG	+	cctaCCAAActgactaat	*Arabidopsis*	TRANSFAC
Athb-1	+	ccaaaATTATgctt	*Arabidopsis*	TRANSFAC
Athb-1	-	gaaaATAATtacaa	*Arabidopsis*	TRANSFAC
Athb-1	-	aaccATAATtcgac	*Arabidopsis*	TRANSFAC
Athb-1	+	aaaatATTATatta	*Arabidopsis*	TRANSFAC
Athb-1	+	aacatATTATtaaa	*Arabidopsis*	TRANSFAC
Athb-1	+	gcgagATTATttgc	*Arabidopsis*	TRANSFAC
Athb-1	-	atctATAATtccat	*Arabidopsis*	TRANSFAC
Athb-1	-	ttccATAATaaata	*Arabidopsis*	TRANSFAC
Athb-1	+	aaaacATTATatag	*Arabidopsis*	TRANSFAC
ATHB-5	-	aAATAAtta	*Arabidopsis*	TRANSFAC
ATHB-5	-	tAATAAcaa	*Arabidopsis*	TRANSFAC
ATHB-5	+	ataTTATTa	*Arabidopsis*	TRANSFAC
ATHB-5	-	aAATAAaaa	*Arabidopsis*	TRANSFAC
ATHB-5	-	aAATAAaac	*Arabidopsis*	TRANSFAC
ATHB-5	-	gAATAAatg	*Arabidopsis*	TRANSFAC
ATHB-5	+	aagTTATTt	*Arabidopsis*	TRANSFAC
ATHB-5	+	agaTTATTt	*Arabidopsis*	TRANSFAC
ATHB-5	-	tAATAAata	*Arabidopsis*	TRANSFAC
ATHB-5	-	aAATAAcac	*Arabidopsis*	TRANSFAC
ATHB-9	-	tacagaaATCATaaataaa	*Arabidopsis*	TRANSFAC
ATHB-9	-	aatttcaATCATtttaaat	*Arabidopsis*	TRANSFAC
ATHB-9	+	aatgtttATGATtgtgata	*Arabidopsis*	TRANSFAC
ANAERO1CONSENSUS	-	TTTGTTT	*Maize/Arabidopsis/pea/barley/rice*	PLACE
AP1	+	TTTTTGG	*Arabidopsis*	AGRIS
ARR10	+	AGATTATT	*Arabidopsis*	JASPER
ARR1AT	-	AATCC	*Arabidopsis*	PLACE
ARR1AT	-	AATCC	*Arabidopsis*	PLACE
ARR1AT	+	GGATT	*Arabidopsis*	PLACE
ARR1AT	-	AATCC	*Arabidopsis*	PLACE
ARR1AT	+	GGATT	*Arabidopsis*	PLACE
ARR1AT	+	GGATT	*Arabidopsis*	PLACE
ARR1AT	-	AATCC	*Arabidopsis*	PLACE
ARR1AT	-	AATCC	*Arabidopsis*	PLACE
ARR1AT	+	GGATT	*Arabidopsis*	PLACE
ARR1AT	-	AATCG	*Arabidopsis*	PLACE
ARR1AT	+	CGATT	*Arabidopsis*	PLACE
ARR1AT	-	AATCA	*Arabidopsis*	PLACE
ARR1AT	-	AATCA	*Arabidopsis*	PLACE
ARR1AT	-	AATCG	*Arabidopsis*	PLACE
ARR1AT	+	TGATT	*Arabidopsis*	PLACE
ARR1AT	+	AGATT	*Arabidopsis*	PLACE
Athb-1	+	CAATCATT	*Arabidopsis*	JASPER
Bellringer	-	ACTAATTT	*Arabidopsis*	AGRIS
Bellringer	+	AAATTAAT	*Arabidopsis*	AGRIS
CARGCW8GAT	+/-	CTAATTTTTG	*Arabidopsis*	PLACE
CARGCW8GAT	+/-	CAAAATTATG	*Arabidopsis*	PLACE

CARGCW8GAT	+/-	CATTATATAG	*Arabidopsis*	PLACE
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
Core	-	TAAT	*Arabidopsis*	AGRIS
Core	+	ATTA	*Arabidopsis*	AGRIS
DPBFCOREDCDC3	+	ACACTCG	*carrot/Arabidopsis*	PLACE
DPBFCOREDCDC3	+	ACACTTG	*carrot/Arabidopsis*	PLACE
E2FCONSENSUS	-	GCGGCAAA	*Arabidopsis/tobacco/rice*	PLACE
GATABOX	+	GATA	*petunia/Arabidopsis/rice*	PLACE
GATABOX	+	GATA	*petunia/Arabidopsis/rice*	PLACE
GATABOX	+	GATA	*petunia/Arabidopsis/rice*	PLACE
GATABOX	+	GATA	*petunia/Arabidopsis/rice*	PLACE
GATABOX	-	TATC	*petunia/Arabidopsis/rice*	PLACE
GT1CONSENSUS	+	GGAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAT	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	TTTTTC	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	TTTACC	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	TTTACC	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GGAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAT	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GGAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAT	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	ATTTCC	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GGAAAT	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAT	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	ATTTTC	*pea/oat/rice/tobacco/Arabid*	PLACE

			opsis	
GT1CONSENSUS	+	GGAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	+	GAAAAA	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
GT1CONSENSUS	-	TTTTTC	*pea/oat/rice/tobacco/Arabidopsis*	PLACE
LTRECOREATCOR15	+	CCGAC	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	+	CCGAC	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	+	CCGAC	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	+	CCGAC	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
LTRECOREATCOR15	-	GTCGG	*Arabidopsis/rape*	PLACE
MYB1AT	-	TGGTTT	*Arabidopsis*	PLACE
MYB1AT	-	TGGTTT	*Arabidopsis*	PLACE
MYB1	+	ATCCTACC	*Arabidopsis*	AGRIS
MYB2CONSENSUSAT	+	TAACGG	*Arabidopsis*	PLACE
MYB4	+	ACCAAAC	*Arabidopsis*	AGRIS
MYB4	-	GGTAGTT	*Arabidopsis*	AGRIS
MYB4	-	GGTAGTT	*Arabidopsis*	AGRIS
MYB4	-	GGTTGGT	*Arabidopsis*	AGRIS
MYB4	+	ACCAAAC	*Arabidopsis*	AGRIS
MYBCOREATCYCB1	+	AACGG	*Arabidopsis*	PLACE
MYBCORE	-	TAACGG	*Arabidopsis/petunia*	PLACE
MYBPLANT	-	GGTTGGTT	*snapdragon/bean/petunia/Arabidopsis*	PLACE
MYCCONSENSUSAT	+/-	CAAATG	*Arabidopsis*	PLACE
MYCCONSENSUSAT	+/-	CACTTG	*Arabidopsis*	PLACE
POLASIG1	+	AATAAA	*pea/rice/Arabidopsis*	PLACE
POLASIG1	+	AATAAA	*pea/rice/Arabidopsis*	PLACE
POLASIG1	+	AATAAA	*pea/rice/Arabidopsis*	PLACE
POLASIG1	+	AATAAA	*pea/rice/Arabidopsis*	PLACE
RAV1AAT	+	CAACA	*Arabidopsis*	PLACE/AGRIS/TRANSFAC
RAV1AAT	-	TGTTG	*Arabidopsis*	PLACE/AGRIS/TRANSFAC
TBOXATGAPB	+	ACTTTG	*Arabidopsis*	PLACE
WBOXATNPR1	-	GTCAA	*Arabidopsis*	PLACE

Table S10. Primers used for chromosome walking on microarray candidate probes and the source gene Bsp*HRD3*.

Target sequence	Primer name	Primer sequence 5`→ 3`
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
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
Sharb0425060	TSP5_1L	AATTTACAGACCCTGCGATCT
	TSP5_1R	CTTTCTCCTCGATTTCTGATTG
	TSP5_2L	CCATAAGCACAACCAATCGAAA
	TSP5_2R	TTTCGATTGGTTGTGCTTATGG
	TSP5_3L	CAGAAATCGAGGAGAAAGAGACA
	TSP5_3R	TCGCAGGGTCTGTAAATTGG
Bsp*HRD3*	TSP9_1L	TTCTCAGGCTTGCTTGTTGA
	TSP9_1R	CCAAAGCAAGCCAAAACATT
	TSP9_2L	GCCTTCAAATGCAGGCAAGAG
	TSP9_2R	TCCAAGGTTAAAATGCCCACT
	TSP9_3L	GGTGGGCGTGGCAGAGATTA
	TSP9_3R	AAATCAACAAGCAAGCCTGAGA

Table S11. 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	Elongation factor Tu	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.

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

Symbol	3`-end position	5`-end position	%matches	Score	Gaps	Identities	Assembly
IR 1	17799-17838	40869-40830	92	99	0	37/40	1
IR 2	42257-42290	48044-48011	79	53	0	27/34	1
IR 3	49604-49664	49724-49667	91	106	3	53/58	1
IR 4	40820-40859	50264-50224	87	73	1	35/40	1
IR 5	49697-49730	52335-52321	88	62	1	30/34	1
IR 6	19583-19649	19771-19706	81	102	1	54/66	2
IR 7	39432-39552	40824-40704	92	300	0	112/121	2
IR 8	30741-30868	54357-54231	83	222	1	106/127	2

Table S13. Detection of candidate regions for stable non-protein-coding RNA secondary structures of Bsp*UPG-2*.

		Single Window Size (step size = 10, window delta = 10, Z-score threshold ≤ -3.5)[c]																									
Forward strand																											
Name[a]	**Position**[b]	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**	**210**	**220**	**230**	**240**	**250**	**260**	**270**	**280**	**290**	**300**
npcRNA 1	1478-1936	-4.82	-5.37	-4.24	-4.46	-3.92	-3.59	-3.55	-4.17	-4.66	-4.97	-4.75	-5.68	**-5.95**	-5.58	-5.25	-5.20	-4.78	-4.90	-4.77	-4.61	-4.34	-4.25	-4.26	-3.82	-4.01	-4.23
npcRNA 2	2256-2455		**-4.03**	-3.74			-3.52		-3.56	-3.86		-3.50	-3.63	-3.62		**-4.00**	-3.71										
npcRNA 3	2795-2854	**-4.37**	-4.20																								
npcRNA 4	2955-3044			**-3.99**	-3.58	**-3.99**																					
Reverse strand																											
npcRNA 5	1091-1160			**-3.54**																							
npcRNA 6	1487-1836	-3.84	-4.65	-3.79	-4.26	-4.31	-3.79	-3.87	-3.82	-4.26	**-4.72**	-4.29	-4.59	-4.53	-4.46	-4.34	-4.35	-4.04	-4.40	-3.99	-4.05	-4.13	-3.83	-4.24	-3.89	-3.53	-3.52
npcRNA 7	2207-2505	-3.98	-4.05	-3.55	-4.95	-4.97	-5.31	-4.99	-5.28	-5.29	-5.24	-4.98	**-5.64**	-5.43	-4.84	-5.11	-5.00	-4.84	-4.80	-4.74	-4.62	-4.59	-4.14	-3.95	-4.01	-3.96	-3.86
npcRNA 8	2955-3054		**-4.21**	-3.89	-3.92	-3.67	-3.62																				

[a] The npcRNA name was chosen to follow the naming convention established by Hirsch et al. (2006).

[b] The position marks the total length of a candidate region for structural npcRNAs produced by multiple overlapping windows. Windows with most negative Z-scores for each npcRNA are displayed in bold letters.

[c] A blank space means that the npcRNA was undetected using the window size specified at the top of the column.

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

Name	Region[a]	Length	Position[b]	Length	Dir[c]	Z[d]	%A	%U	%G	%C	%A+U	MFE[e]	AMFE[f]	MFEI[g]	Best hit[h]	*E*-value	Best hit[i]	*E*-value
npcRNA 1	1478-1936	459	1617-1786	170	+	-5.95	20.70	43.20	15.40	20.70	63.90	-65.38	-38.46	-1.07	AT3G07340	1.5	AT5G50690	0.73
																	AT5G50690	0.73
																	AT4G32300	0.73
																	AT3G50530	0.73
																	AT1G48110	0.73
npcRNA 2	2256-2455	200	2386-2445	60	+	-4.03	30.00	26.70	21.70	21.70	56.70	-31.30	-52.17	-1.20	AT1G64020	1.2	AT1G64020	0.053
																	AT1G43725	0.053
npcRNA 3	2795-2854	60	2805-2854	50	+	-4.37	24.00	46.00	16.00	14.00	70.00	-20.92	-41.84	-1.39	AT1G04160	0.056	AT1G04160	0.010
npcRNA 4	2955-3044	90	2955-3044	90	+	-3.99	28.90	40.00	15.60	15.60	68.90	-65.38	-72.64	-2.34	AT3G06530	0.015	AT1G11450	1.4
npcRNA 5	1091-1160	70	1091-1160	70	-	-3.54	20.00	42.90	11.40	25.70	62.90	-18.27	-26.10	-0.70	AT4G02930	7.00E-24	AT4G02930	5.00E-18
npcRNA 6	1487-1836	350	1647-1786	140	-	-4.72	39.30	24.30	22.10	14.30	63.60	-55.65	-39.75	-1.09	AT3G07340	1.2	AT5G50690	0.73
																	AT5G50690	0.73
																	AT4G32300	0.73
																	AT3G50530	0.73
																	AT1G48110	0.73
npcRNA 7	2207-2505	299	2306-2465	160	-	-5.64	30.60	30.60	18.80	20.00	61.20	-73.72	-46.08	-1.19	AT1G64020	4.7	AT1G64020	0.053
npcRNA 8	2955-3054	100	2985-3044	60	-	-4.21	40.00	28.30	16.70	15.00	68.30	-24.49	-40.82	-1.29	AT3G06530	0.008	AT1G11450	0.83

[a] Position of the candidate region for npcRNAs from Bsp*UPG-2*.

[b] Position of the specific candidate npcRNA demonstrating the highest (most negative) Z-score within a candidate region.

[c] Orientation of the candidate npcRNAs on Bsp*UPG-2*.

[d] *Z*-score

[e] MFEs were conducted 20°C temperature as standard environmental condition for *Boechera* considering that the mean annual temperature for >90% of all tested *Boechera* accessions ranges between -5 and 15°C (M. Mau, unpublished data).

[f] Adjusted MFE is calculated by (MFE/lenght of RNA sequence)x100 according to Zhang et al. (2006).

[g] Minimal folding free energy index (MFEI) was calculated by AMFE/(G+C)% according to Zhang et al. (2006).

[h] BLASTN results of a GenBank nucleotide collection search with candidate npcRNAs.

[i] BLASTN results of a TAIR10 Genes database (including introns and untranslated regions) search with candidate npcRNAs.

Table S15. Sequence divergence of Bsp*UPG-1* and Bsp*UPG-2*. Comparing end-to-end sequenced gene copies by CLUSTALW analysis (Slow/Accurate, IUB). 5`-end: from +1 to +1820 nt, 3`-end: from +1821 to +3153 nt.

Comparison	**avg %similarity**					
	Complete	**SE**	**3`-end**	**SE**	**5`-end**	**SE**
Bsp*UPG-2* in apomicts vs. Bsp*UPG-1* in sexual *B. stricta*	64.62	0.02	75.92	0.04	39.95	0.02
Bsp*UPG-2* in apomicts vs. Bsp*UPG-1* in other sexual genotypes	61.21	0.40	91.29	0.22	38.64	0.03
Bsp*UPG-2* in apomictic genotypes	99.34	0.13	99.33	0.13	99.34	0.13
Bsp*UPG-1* in sexual *B. stricta*	99.69	0.05	99.85	0.02	99.34	0.05
Bsp*UPG-1* in other sexual genotypes	98.87	0.34	99.22	0.16	98.57	0.23
Bsp*UPG-1* all sexual genotypes	74.02	0.53	81.73	1.66	58.05	0.52
Bsp*UPG-2* in all apomictic vs. Bsp*UPG-1* in all sexual genotypes	63.10	0.26	82.75	0.86	39.37	0.17

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

ID	Position[a]	Length (nt)	Sequence 5` → 3`	Specificity[b]
Indel 1	2931	1	C/T	Non-stricta sexuals
Indel 2	2833	3	TTG	Non-stricta sexuals
Indel 3	2796	1	T	Sexual *B. stricta*
Indel 4	2733	1	G/T	Non-stricta sexuals
Indel 5	2693	5	TTAAA	ES 913.3
Indel 6	2661	9	TTTGCTGTG	Sexual *B. stricta*
Indel 7	2645	1	C/T	Apomicts (except ES 753)
Indel 8	2641	4	GAAC	Non-stricta sexuals
Indel 9	2603	27	TTTTCACCTTAGCTTAATTTGAATATA	All sexuals
Indel 10	2475	2	GA	Apomicts (except ES 753 and ES 514)
Indel 11	2453	1	C	Non-stricta sexuals
Indel 12	2441	20	TTCTCTAAGATCGATCAATC	Sexual *B. stricta* (except 132.3 and ES 558.2)
Indel 13	2212	1	T	ES 753 and ES 910.1
Indel 14	2190	2	GT	ES 910.1
Indel 15	2188	1	G	ES 910.1
Indel 16	2110	58	TGTAGATTTACTTTCCGAAAATTGATAAATCACCTTAAC TAAATTATATAAGTAAATT	Sexual *B. stricta*
Indel 17	2031	12	TTGGTTGATCAT	Sexual *B. stricta*
Indel 18	2009	16	TCGTATGTTTACTTCC	ES 865
Indel 19	1884	1	A	Non-stricta sexuals
Indel 20	1826	3	GGA	ES 514
Indel 21	1652	2	TT	ES 753
Indel 22	1653	1	T	ES 514
Indel 23	1479	1	G	ES 753 and ES 514
Indel 24	1367	3	TTT	ES 753 and ES 514
Indel 25	564	1	T	ES 753
Indel 26	104	1	A	ES 514
Indel 27	103	2	CA	ES 753
Indel 28[†]	791	1	T	132.3
Indel 29[†]	833	3	GAA	132.3 and ES 558.2
Indel 30[‡]	1339	1	A	ES 910.2
Indel 31[‡]	1337	2	AA	29.1
Indel 32[‡]	1186	30	ACTTTTTTCCCATTGTTGGTTTTTTCCACA	105.18
Indel 33[‡]	470	2	AA	ES 910.2 and 29.1
Indel 34[‡]	467	3	AAA	105.18
Indel 35[‡]	292	5	TTTGT	105.18

[a] Position on genomic sequence of Bsp*UPG-2*, and Bsp*UPG-1* in sexual *B. stricta*[†] and other sexual genotypes[‡].

[b] Occurrence of the sequence variant

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

Genus	Name	Query[a]		Hit[b]		Repeat class	Dir[c]	%Similarity	Score
		From	To	From	To				
Brachypodium	Copia-4_BD-I	47	91	1577	1622	LTR/Copia	c	0.7826	207
Arabidopsis	ATCOPIA20I	181	551	1097	1455	LTR/Copia	c	0.6885	908
Arabidopsis	ATCOPIA20I	734	1099	152	498	LTR/Copia	c	0.6667	641
Arabidopsis	AT9TSD1	1197	1241	1885	1927	DNA/MuDR	d	0.7955	211
Sorghum	EnSpm11_SB	1262	1362	6319	6420	DNA/EnSpm	c	0.7200	231
Aegilops	EnSpm-N1_AT	2828	2912	482	579	DNA/EnSpm	c	0.7528	243
Arabidopsis	ATHPOGON3	3228	3277	229	276	DNA/Mariner	d	0.8163	208
Solanum	hAT-2_SD	5049	5240	3700	3871	DNA/hAT	c	0.7232	285
Glycine	MuDR-2_GM	5853	5927	4122	4192	DNA/MuDR	d	0.7887	238
Glycine	MuDR-11_GM	6466	6650	7767	7940	DNA/MuDR	c	0.7486	213
Arabidopsis	ATLINE1A	6690	8232	630	2145	NonLTR	d	0.6521	1909
Zea	Gypsy66-ZM_LTR	8666	8715	5523	5580	LTR/Gypsy	d	0.8654	213
Sorghum	REP2_SB	9147	9240	172	262	Simple/Sat	d	0.7889	256
Glycine	Gypsy-66_GM-LTR	9985	10172	28	195	LTR/Gypsy	d	0.7414	240
Vitis	Copia16-VV_I	12200	12289	762	847	LTR/Copia	c	0.8140	237
Glycine	MuDR-14_GM	12516	12705	322	513	DNA/MuDR	c	0.7568	281
Arabidopsis	ATENSPM4	14567	14613	4811	4855	DNA/EnSpm	c	0.8043	223
Glycine	MuDR-12_GM	15007	15233	7912	8114	DNA/MuDR	c	0.7413	215
Glycine	TGM1_GM	16615	16689	762	838	DNA/EnSpm	d	0.7532	202
Arabidopsis	ATHATN3	16727	16902	1	162	DNA/hAT	d	0.7423	441
Arabidopsis	ATHATN3	16929	17010	45	129	DNA/hAT	d	0.7262	216
Carica	Copia-11_CP-I	18799	18868	2704	2764	LTR/Copia	c	0.7656	223
Zea	Copia-47_ZM-I	19194	19249	1842	1897	LTR/Copia	c	0.7500	225
Arabidopsis	HELITRONY1D	19384	19528	2203	2361	DNA/Helitron	d	0.7432	244
Glycine	MuDR-7_GM	20678	20878	6823	7035	DNA/MuDR	c	0.7010	242
Vitis	MUDRAVI2	22515	22647	10459	10578	DNA/MuDR	d	0.7680	240
Sorghum[d]	Helitron-N9_SBi	23078	23133	2238	2293	DNA/Helitron	d	0.7895	234
Arabidopsis[d]	AT9TSD1	23294	23404	1864	1976	DNA/MuDR	c	0.7670	240
Arabidopsis[d]	ATHILA4D_LTR	24123	24279	440	625	LTR/Gypsy	d	0.7212	276
Glycine[d]	Copia-33_GM-I	24698	24776	1237	1315	LTR/Copia	d	0.6875	214
Populus	Copia-1_PTri-I	25682	25729	292	341	LTR/Copia	c	0.8163	252
Arabidopsis	TA1-2_I	26106	26211	2093	2194	LTR/Copia	c	0.7019	283
Arabidopsis	ATCOPIA66_I	26219	26357	1828	1966	LTR/Copia	c	0.7338	496
Brassica	Copia-46_BRa-I	26389	26581	3380	3570	LTR/Copia	d	0.6984	374
Glycine	Copia-82_GM-I	26696	26821	2676	2806	LTR/Copia	d	0.7000	310
Arabidopsis	ATCOPIA75_I	26913	26989	3919	3998	LTR/Copia	d	0.6538	209
Populus	Copia-77_PT-I	26990	27142	3864	4024	LTR/Copia	d	0.6772	379
Solanum	hAT-6_STu	27496	27634	63	183	DNA/hAT	d	0.7661	278
Arabidopsis	ATLINE1_6	28360	28941	544	1125	NonLTR/L1	d	0.6500	862
Arabidopsis[e1]	ATLINE1_6	30103	32420	1459	3746	NonLTR/L1	d	0.6578	3065
Lycopersicon	LycEPRV_I	32619	32775	1505	1675	Caulimovirus	d	0.7205	233
Arabidopsis	ATHILA8B_I	34068	34140	4950	5021	LTR/Gypsy	c	0.8493	365
Hordeum	CEREBA_I	34425	34528	8985	9077	LTR/Gypsy	d	0.7582	228
Sorghum	Copia-78_SB-I	34889	34953	3657	3726	LTR/Copia	c	0.7424	210
Arabidopsis	ATHILA8A_I	35414	35765	2765	3118	LTR/Gypsy	c	0.6844	481
Brassica	Copia-34_BRa-I	36201	36258	5676	5727	LTR/Copia	d	0.8148	260
Lycopersicon	LycEPRV_I	37017	37115	772	859	Caulimovirus	d	0.7447	210
Arabidopsis	ATLINE1_6	37706	38063	3745	4103	NonLTR/L1	d	0.6676	722
Malus	L1-11_Mad	39731	39816	5318	5405	NonLTR/L1	d	0.6552	246
Vitis	ENSPM2_VV	40774	40864	2665	2737	DNA/EnSpm	c	0.8421	248
Arabidopsis	ATREP15	41679	41792	4	121	DNA/Helitron	c	0.7193	279
Arabidopsis	ATHATN5	45579	45918	3	346	DNA/hAT	d	0.7368	884
Arabidopsis	ATREP10	45943	46019	489	566	DNA/Helitron	c	0.7089	273
Oryza	ENSPM5_OS	46026	46167	9179	9323	DNA/EnSpm	d	0.7092	283
Glycine	Copia-35_GM-LTR	48073	48129	36	90	LTR/Copia	c	0.8036	206
Arabidopsis	HELITRONY1B	48522	48867	366	745	DNA/Helitron	d	0.7126	503
Sorghum	EnSpm2_SB	49450	49731	6669	6913	DNA/EnSpm	c	0.7673	426
Sorghum	EnSpm6_SB	49735	49881	6096	6238	DNA/EnSpm	c	0.7426	258
Vitis	EnSpm-3_VV	50145	50246	3088	3180	DNA/EnSpm	c	0.7684	213
Glycine	MuDR-12_GM	50250	50547	7626	7924	DNA/MuDR	c	0.7352	233
Carica	Gypsy-5_CP-LTR	50826	50907	128	196	LTR/Gypsy	c	0.7746	244
Vitis	MuDR-6_VV	51251	51312	1749	1816	DNA/MuDR	d	0.7778	201
Arabidopsis	ATHATN3	51929	52071	401	542	DNA/hAT	d	0.7518	437
Vitis	MuDR-8_VV	52269	52353	9804	9892	DNA/MuDR	d	0.7727	226
Arabidopsis	ATREP10C	52372	52480	1	89	DNA/Helitron	d	0.8043	283
Glycine	Gypsy-32_GM-I	52603	52646	2126	2169	LTR/Gypsy	c	0.7727	229
Arabidopsis	TA1-2_I	53421	53453	2375	2407	LTR/Copia	c	0.8788	243
Oryza	LINE-1	54647	54707	2472	2528	NonLTR	d	0.7458	209

[a] Positions of corresponding fragment on assembled *Boechera* BAC DNA.

[b] Start and end positions of fragment on corresponding sequence.

[c] Values indicate orientation of repeat fragment ('d' for direct, 'c' for complementary).

[d] Repeat fragment matching Bsp*UPG-1*.

[e] Repetitive element confirmed with GenBank BLASTX: 1 – TER4.

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

Genus	Name	Query[a]		Hit[b]		Repeat class	Dir[c]	%Similarity	Score
		From	To	From	To				
Malus	Copia-90_Mad-I	440	510	5706	5778	LTR/Copia	d	0.7671	363
Medicago	LINE1D_MT	2819	2862	4147	4190	NonLTR/L1	d	0.7727	224
Malus	hAT-N3_Mad	3341	3379	621	659	DNA/hAT	c	0.8205	251
Arabidopsis	ATHILA2_I	4263	4331	2942	3007	LTR/Gypsy	d	0.7164	212
Arabidopsis	AT9TSD1	4529	4573	1885	1927	DNA/MuDR	d	0.8182	226
Arabidopsis[e1]	ATCOPIA20I	4919	6555	2735	4398	LTR/Copia	c	0.6885	3877
Arabidopsis	ATCOPIA93_I	6940	8920	341	2338	LTR/Copia	c	0.6682	2533
Arabidopsis	ATCOPIA20I	9019	9105	152	239	LTR/Copia	c	0.6818	281
Medicago	GYPSY3-I_MT	9167	9263	102	200	LTR/Gypsy	c	0.6869	240
Zea	Copia-37-LTR_ZM	9573	9642	1084	1152	LTR/Copia	d	0.7714	266
Sorghum	MuDR-6_SBi	10125	10173	8328	8377	DNA/MuDR	c	0.82	250
Aegilops	EnSpm-N1_AT	10854	10921	482	552	DNA/EnSpm	c	0.7826	239
Carica	Copia-5_CP-LTR	12479	12583	271	381	LTR/Copia	c	0.7308	241
Vitis	Copia-88_VV-LTR	12620	12826	31	254	LTR/Copia	c	0.7583	516
Vitis[e2]	Copia-89_VV-I	12837	17729	195	5139	LTR/Copia	c	0.7705	20053
Carica	Copia-5_CP-LTR	18360	18464	271	381	LTR/Copia	c	0.7308	241
Vitis	Copia-88_VV-LTR	18501	18707	31	254	LTR/Copia	c	0.7583	518
Solanum	hAT-2_SD	19275	19510	3611	3853	DNA/hAT	c	0.7082	270
Arabidopsis	ATCOPIA53_I	19635	19724	2156	2259	LTR/Copia	d	0.7957	246
Arabidopsis	ATLINE1A	20945	22496	630	2148	NonLTR	d	0.6567	1840
Oryza	COWARD-2	22603	22669	75	137	DNA	c	0.7576	204
Solanum	Gypsy-47_STu-I	23452	23505	5798	5854	LTR/Gypsy	d	0.8214	223
Vitis	Gypsy-23_VV-LTR	24227	24350	445	582	LTR/Gypsy	d	0.7559	236
Solanum	DNA3-5_STu	28007	28099	811	905	DNA	d	0.7766	200
Arabidopsis	ATHATN3	28261	28437	3	164	DNA/hAT	d	0.7195	523
Arabidopsis	ATHAT3	28466	28550	4051	4139	DNA/hAT	d	0.7442	295
Vitis	MUDRAVI1	30401	30627	5081	5305	DNA/MuDR	d	0.7604	257
Arabidopsis	ATREP3	30713	30870	2	159	DNA/Helitron	c	0.7662	309
Carica	Copia-11_CP-I	31531	31600	2704	2764	LTR/Copia	c	0.7656	223
Cucumis	GYCUME1_LTR	31762	31865	172	270	LTR/Gypsy	d	0.76	230
Malus	DNA3-11_Mad	31887	31965	478	551	DNA	c	0.72	240
Arabidopsis	HELITRONY1D	32100	32244	2203	2361	DNA/Helitron	d	0.7315	206
Vitis	TE-7-1_VV	34026	34192	767	919	DNA/hAT	c	0.7383	208
Solanum	MuDR-5_STu	34792	34851	12604	12666	DNA/MuDR	c	0.7541	226
Vitis	MuDR-8_VV	35442	35634	9402	9583	DNA/MuDR	c	0.7219	248
Arabidopsis	ATENSPM4	35714	35768	4800	4854	DNA/EnSpm	c	0.8	260
Populus	EnSpm3_PT	35782	35914	27	162	DNA/EnSpm	d	0.7239	309
Brassica	RT_BC	36951	37012	122	183	NonLTR	c	0.7581	257
Arabidopsis	HELITRONY1B	37852	37930	632	715	DNA/Helitron	d	0.7949	224
Vitis	Harbinger-1_VV	38191	38252	4184	4248	DNA/Harbinger	c	0.8095	223
Arabidopsis	BRODYAGA1A	38391	38703	459	749	DNA	c	0.7597	298
Physcomitrella	Gypsy-7N_PPa-LTR	38728	38921	1178	1372	LTR/Gypsy	c	0.7354	280
Malus	DEM1_LTR	39048	39121	220	294	LTR	c	0.76	289
Solanum	L1-1_STu	39276	39330	8200	8257	NonLTR/L1	c	0.8393	291
Arabidopsis	ATENSPM5	39766	39888	4988	5092	DNA/EnSpm	d	0.7615	225
Medicago	COPMET_LTR	41149	41189	920	959	LTR	d	0.8537	220
Brassica	BRENSPM1	41617	41660	3038	3081	DNA/EnSpm	c	0.7727	209
Arabidopsis[e3]	ATENSPM2	41680	43557	1252	3185	DNA/EnSpm	c	0.6759	4005
Arabidopsis	ATENSPM2	43562	43679	973	1081	DNA/EnSpm	c	0.7387	265
Arabidopsis	ATENSPM6	43786	43843	764	825	DNA/EnSpm	c	0.7667	211
Vitis	ENSPM2_VV	45140	45481	2386	2708	DNA/EnSpm	d	0.7556	288
Solanum	NonLTR-1_STu	46324	46372	9	59	NonLTR	c	0.82	237
Vitis	Harbinger-3_VV	47067	47108	1718	1757	DNA/Harbinger	d	0.8537	211
Arabidopsis	AT9TSD1	47115	47223	1849	1951	DNA/MuDR	c	0.7551	259
Arabidopsis	ATHILA8BLTR	47452	47504	1	53	LTR/Gypsy	d	0.7736	266
Sorghum	EnSpm4_SB	48344	48541	6604	6798	DNA/EnSpm	c	0.7326	294
Populus[d]	Gypsy-26_PTr-LTR	49019	49076	984	1042	LTR/Gypsy	d	0.7458	263
Zea[d]	MuDR-7N1_ZM	49121	49216	1097	1205	DNA/MuDR	d	0.7755	234
Medicago[d]	GYPSY2-LTR_MT	50212	50369	170	331	LTR/Gypsy	d	0.7438	253
Medicago[d]	HAT1_MT	50985	51020	2107	2142	DNA/hAT	d	0.8333	219
Arabidopsis	TA1-2_I	52851	52954	2092	2191	LTR/Copia	c	0.6961	277
Arabidopsis	ATCOPIA66_I	52961	53100	1828	1966	LTR/Copia	c	0.7214	454
Brassica	Copia-46_BRa-I	53132	53644	3380	3882	LTR/Copia	d	0.6667	568
Arabidopsis	ATCOPIA75_I	53656	53751	3901	3999	LTR/Copia	d	0.6531	213
Sorghum	EnSpm1_SB	54052	54101	5922	5968	DNA/EnSpm	c	0.7959	233
Arabidopsis	ATREP3	54110	54216	2	118	DNA/Helitron	d	0.7727	273
Arabidopsis	ATREP3	54234	54336	7	118	DNA/Helitron	d	0.7905	311
Oryza	TWIFBIG	54777	54877	3132	3241	DNA/hAT	c	0.7477	254
Arabidopsis	HELITRONY3A	54972	55255	350	634	DNA/Helitron	d	0.7185	390
Sorghum	EnSpm3_SB	55440	55830	504	894	DNA/EnSpm	c	0.7268	579
Arabidopsis	ATREP3	56543	56598	19	78	DNA/Helitron	d	0.7966	218
Arabidopsis	ATLINE1_5	58452	58573	4133	4258	NonLTR/L1	c	0.6855	383

[a] Positions of corresponding fragment on assembled *Boechera* BAC DNA.

[b] Start and end positions of fragment on corresponding sequence.

[c] Values indicate orientation of repeat fragment ('d' for direct, 'c' for complementary).

[d] Repeat fragment matching Bsp*UPG-2*.

[e] Repetitive element confirmed with LTR FINDER and GenBank BLASTX: 1 – TER3; 2 – TER1; 3 – TER2.