Gene function rather than reproductive mode drives the evolution of RNA helicases in sexual and apomictic Boechera

Markus Kiefer¹§, Berit H. Nauerth¹, Christopher Volkert¹, David Ibberson², Anna Loreth¹, Anja Schmidt¹§*

¹ Centre for Organismal Studies (COS) Heidelberg, Heidelberg University, Im Neuenheimer Feld 345, D-69120, Heidelberg

² Deep Sequencing Core Facility, CellNetworks Excellence Cluster, University of Heidelberg, Im Neuenheimer Feld 267, D-69120, Heidelberg

§ contributed equally

*Author for correspondence: Anja Schmidt, Centre for Organismal Studies (COS) Heidelberg, Heidelberg University, Germany, telephone number: +49 6221 544683, fax: +49 6221 545508, email: anja.schmidt@cos.uni-heidelberg.de.

Data deposition: Raw data are deposited under SRA accession PRJNA563290. Sequence alignments will be made available on Dryad.
Abstract

In higher plants sexual and asexual reproduction through seeds (apomixis) have evolved as alternative strategies. Evolutionary advantages leading to coexistence of both reproductive modes are currently not well understood. It is expected that accumulation of deleterious mutations leads to a rapid elimination of apomictic lineages from populations. In this line, apomixis originated repeatedly, likely from deregulation of the sexual pathway, leading to alterations in the development of reproductive lineages (germlines) in apomicts as compared to sexual plants. This potentially involves mutations in genes controlling reproduction.

Increasing evidence suggests that RNA helicases are crucial regulators of germline development. To gain insights into the evolution of 58 members of this diverse gene family in sexual and apomictic plants, we applied target enrichment combined with Next Generation Sequencing to identify allelic variants from 24 accessions of the genus *Boechera*, comprising sexual, facultative and obligate apomicts. Interestingly, allelic variants from apomicts did not show consistently increased mutation frequency. Either sequences were highly conserved in any accession, or allelic variants preferentially harbored mutations in evolutionary less conserved C- and N-terminal domains, or presented high mutation load independent of the reproductive mode. Only for a few genes allelic variants harboring deleterious mutations were only identified in apomicts. To test if high sequence conservation correlates with roles in fundamental cellular or developmental processes, we analyzed *Arabidopsis thaliana* mutant lines in VASA-LIKE (VASL), and identified pleiotropic defects during ovule and reproductive development. This indicates that also in apomicts mechanisms of selection are in place based on gene function.

Keywords: apomixis, *Boechera*, evolution, reproduction, RNA helicases, sequence variation
Introduction

Apart from sexual reproduction, also asexual reproduction through seeds (apomixis) established in the course of evolution in more than 400 species (Carman 1997; Hojsgaard, et al. 2014). Unlike through sexual reproduction leading to genetic diversity, clonal offspring is formed by apomixis, making apomixis exceptionally attractive for potential applications in agriculture (Spillane, et al. 2004; Barcaccia and Albertini 2013; Conner and Ozias-Akins 2017). In terms of evolution however, asexual reproduction is perceived as a dead end, as deleterious mutations are predicted to agglomerate due to the absence of meiotic recombination (Muller 1964). While mathematical models support this theory of Muller’s ratchet under different assumptions (Audiffren and Pardoux 2013), empirical data on asexual taxa are inconsistent as exemplified by the unexpected age and fitness of the asexual Amazon molly contradicting the predictions (Warren, et al. 2018).

Nevertheless, from a study of apomixis in plants a slightly higher genome wide mutation rate has recently been described in sympatric apomictic as compared to sexual Boechera spatifolii (Lovell, et al. 2017).

As model system, the genus Boechera increasingly gains attention (Rushworth, et al. 2011; Lovell, et al. 2013; Mau, et al. 2015; Lee, et al. 2017; Li, et al. 2017; Lovell, et al. 2017; Kliver, et al. 2018; Schilling, et al. 2018). First, it is closely related to the sexual model species Arabidopsis thaliana, thus simplifying identification of homologs and comparative analyses. Also, in Boechera more than 100 taxa have been identified consistent of sexual, facultative apomictic, and obligate apomictic accessions at low ploidy levels including diploids (Aliyu, et al. 2010; Windham 2010; Mau, et al. 2015). Evolutionary, apomixis in Boechera likely recurrently derived by hybridizations, leading to heterozygocity and polyploidy typically associated with apomixis (Lovell, et al. 2013). Together with copy number variations, this might be instrumental to buffer the effects of deleterious mutations (Aliyu, et al. 2013; Lovell, et al. 2017).
According to common hypotheses, mutation load is not only expected to be high in apomicts, but also to differ based on ploidy (Hodač, et al. 2019). However, the question remains if mutation accumulation also relates to gene function. In contrast to genome wide studies, knowledge on the evolution of coding regions in apomicts is scarce to date. Specifically, deleterious mutations in genes essential for fundamental cellular processes such as regulation of translation initiation or for reproductive development might not be tolerated but lead to rapid extinction.

Developmental processes to form seeds are largely similar in apomicts and sexual plants, except for a few distinctive changes: While in sexual plants the first cell of the female germline, the megaspore mother cell (MMC), undergoes meiosis, meiotic recombination and reduction is circumvented in apomicts by the corresponding apomictic initial cell (AIC). Two major types of apomixis, diplospory and apospory, are thereby distinguished (Schmidt, et al. 2015). They differ by the origin and fate of the AIC: In diplospory, the AIC specifies instead of the MMC in the reproductive flower tissues. It typically undergoes a mitotic-like division instead of meiosis. In apospory, an additional sporophytic cell adjacent to the sexual MMC gets selected as AIC which completely omits meiosis. In contrast to the functional megaspore (FMS) in sexual plants, the founder cell of the gametophytic lineage remains unreduced in apomicts and typically genetically identical to the AIC. Through three rounds of mitotic divisions and cellularization female gametophytes are formed enclosing the female gametes (egg and central cell). To give rise to formation of the embryo and its nourishing tissue, the endosperm, during sexual reproduction both female gametes get fertilized. In contrast, in apomicts the egg cell forms the embryo parthenogenetically (without fertilization) and the central cell develops into endosperm either autonomously or dependent on fertilization.

Both in sexual plants and apomicts, germline development requires a tight regulatory system controlling cell fate specifications and other developmental decisions. An increasing number of genes involved have recently been identified (Schmidt, et al. 2015; Tekleyohans, et al. 2017; Zhou,
et al. 2017; Nakajima 2018; Nonomura 2018). The regulation of gene activity includes translational control, splicing, and degradation of previously transcribed RNAs (Hafidh, et al. 2011). This requires activity of RNA helicases that are often associated to ribonucleoprotein complexes mediating storage and subcellular localization of RNAs (Arkov and Ramos 2010; Hafidh, et al. 2011; Gao and Arkov 2013). Little is known about the roles of RNA helicases during plant reproduction despite their abundant expression upon germline formation, in contrast to the animal germline where the conserved roles to regulate different aspects of development are well described (Schmidt, et al. 2011; Friday and Keiper 2015). Here, especially RNA helicases from the VASA clade are crucial i.e. for stem cell maintenance, cell fate decisions, cell cycle progression, and for the protection of the germline from the activity of transposable elements (Yajima and Wessel 2011). While RNA helicases appear to be required for reproduction both in animals and plants, the VASA clade derived only during animal evolution (Gustafson and Wessel 2010).

Evolutionary, the gene family of RNA helicases is conserved throughout all kingdoms of life and has largely diversified in higher eukaryotes (Linder and Owttrim 2009). Of the six superfamilies (SF) described, SF2 represents the largest group (Linder and Owttrim 2009). Helicase genes share a conserved core consistent of two RecA-like domains (Linder and Jankowsky 2011). Generally, they are functional in NTP dependent unwinding of RNAs and contribute to basically any aspect of RNA metabolism (Linder and Owttrim 2009). Their target specificity largely resides in the N- and C-terminal domains flanking the conserved core domains that are frequently involved in mediating protein interactions (Jankowsky 2011). Based on at least 12 sequence motifs shared, the family is further subdivided in distinct clades, with SF2 containing the DEAD and DExD/H box helicase proteins named after the conserved amino acid motif, the Ski2-like proteins including SUV3 facilitating mitochondrial RNA degradation, RecQ-like proteins involved in regulation of meiotic recombination, and Swi1/SNF proteins typically acting in chromatin remodeling (Fairman-Williams, et al. 2010; Tang, et al. 2010; Higgins, et al. 2011; Jankowsky 2011; Khemici and Linder
Members of the *EUKARYOTIC TRANSLATION INITIATION FACTORS* (*EIF*) family also belong to the SF2 superfamily of RNA helicases (Marintchev 2013). The EIF proteins of the EIF4 family act in multimeric complexes to recruit mRNAs to the small subunits of the ribosomes upon translation initiation (Asano, et al. 2000; Hernández and Vazquez-Pianzola 2005). In *A. thaliana*, two isoforms exist, with EIF4A1 being involved in controlling cell cycle progression and ovule development (Bush, et al. 2015; Bush, et al. 2016).

From the 161 members of the gene family identified in *A. thaliana* (Xu, et al. 2013), requirement for germline development has been reported for a few genes only, i.e. *MAGATAMA3* (*MAA3*), *MATERNAL EFFECT EMBRYO ARREST29* (*MEE29*), *SLOW WALKER3* (*SWA3*), *CHROMATIN REMODELING 11* (*CHR11*) and *AtSUV3* (Huanca-Mamani, et al. 2005; Pagnussat, et al. 2005; Shimizu, et al. 2008; Liu, et al. 2010). In addition, MNEME (MEM) has previously been described, sharing 43% amino acid identity with SUV3 (Schmidt, et al. 2011). *MEM* is predominantly expressed in the MMC, and plants carrying a mutant *mem* allele likely show formation of unreduced gametophytes, representing an element of apomixis (Schmidt, et al. 2011). Furthermore, also the homolog of *Hypericum perforatum MEE29* is preferentially expressed in reproductive tissues from apomicts compared to sexual plants (Barcaccia and Albertini 2013). Members of the RNA helicase family might thus be involved in establishing or sustaining apomixis.

Based on investigations on *A. thaliana* and maize, also epigenetic regulatory pathways involving small RNA pathways and DNA methylation to control gene activity or to mediate silencing of transposable elements appear to be relevant for apomixis (Garcia-Aguilar, et al. 2010; Olmedo-Monfils, et al. 2010; Singh, et al. 2011). These pathways involve certain RNA helicases including *DECREASE IN DNA METHYLATION1* (*DDM1*), or genes encoding for DICER LIKE (*DCL*) proteins (Pikaard and Mittelsten Scheid 2014). However, to date, knowledge about the roles and evolution of RNA helicases in apomicts is largely lacking. Due to the diversity of this gene family it can be hypothesized that the functional impact of mutations predicted to accumulate in apomicts
differs for distinct members. While detrimental mutations in genes controlling basic cellular or essential developmental processes including reproduction are likely to be deleterious, mutations in others, e.g. MEM might even be instrumental for apomixis.

To study mutation accumulation of selected germline expressed RNA helicases and to gain insights into aspects of their evolution in sexual and apomictic Boechera, we applied in solution hybrid capture also referred to as target enrichment (Gnirke, et al. 2009). After enrichment we used Illumina HighSeq2500 to sequence genomic regions of 58 RNA helicases from 24 Boechera accessions representing sexual, facultative apomictic and obligate apomictic reproduction. We also included the gene encoding for the centromere specific histone variant CENH3 in this study for comparison, which is characterized by rapid evolution and higher expression levels in the germline of the triploid apomict B. gunnisoniana as compared to A. thaliana (Lysak 2014; Schmidt, et al. 2014).

Thereby, we identified different features of mutation accumulation and genetic variation presumably relating more to gene function than to the reproductive modes. While evidence for purifying selection was found for most of the genes, indications of near neutral evolution for allelic variants from apomictic as compared to sexual accessions were given only for a few helicases. As a prove of concept that high sequence conservation might be correlated with involvement of the proteins in regulating basic cellular or crucial developmental processes, we analyzed lines carrying mutations in a novel RNA helicase AT1G72730 in A. thaliana, now named VASA-LIKE (VASL). Consistent with our expectations these show defects during reproductive development including alterations in ovule development, gametogenesis, and embryogenesis. Taken together, our study provides new insights into the evolution of RNA helicases in sexual and apomictic Boechera, further supporting the relevance of RNA helicases to regulate germline and seed development.

Material and Methods
**Plant material**

Seeds of *Boechera* Å. Löve & D. Löve accessions included in this study were obtained from Timothy F. Sharbel (Global Institute for Food Security, University of Saskatchewan, Canada), Thomas Mitchell Olds (Duke University, USA), and John Carman (Utah State University; Table 1; Supporting material). B12-663 has been originally been collected from Big Horn Pass Montana, USA (Timothy Sharpel, Saskatoon, Canada, personal communication). Except for three accessions, *B. stricta* B12-663, *B. retrofracta x stricta* CO11010, and *B. stricta x retrofracta* B12-2582, the frequency of apomictic seed formation has previously been determined (Aliyu, et al. 2010; Mau, et al. 2015; Carman, et al. 2019). Therefore, these three accessions were excluded from all analyses summarizing groups by reproductive mode (also see “Supporting Information”). In this study, we classified accessions as obligate apomicts when no evidence for seed formation by sexual reproduction is given, and as facultative apomicts when sexual seed formation has been reported even at very low frequencies (Aliyu, et al. 2010; Mau, et al. 2015).

Seeds were stratified at 4°C for ~7 days before they were surface sterilized and grown on Murashige-Skoog (MS) plates for 2-3 weeks at 16 h light/darkness at 18°C night 21°C day cycle. One leaf per seedling was used for ploidy analysis by flow cytometry using a CyStainUV Precise P Kit (Sysmex Partec, Görlitz, Germany) on a Partec CyFlow Space instrument following manufacturer instructions.

*Arabidopsis thaliana* (L.) Heyn., ecotype Col-0 was used as wild-type and for generating CRISPR/CAS9 lines (see “Supporting Information”). Seeds were stratified over-night at 4°C before they were surface sterilized and germinated on MS plates similar to *Boechera*. Seedlings were transferred to soil (ED73, Einheitserde, Fröndenberg, Germany) and grown at 16 h light/darkness at 22°C. Lines SAIL_288_C08 and SALK_095627 with T-DNA insertions in the 3’ UTR were obtained from “The European Arabidopsis Stock Centre” (Arabidopsis.info) and referred to as *vasl_1* and *vasl_2*, respectively. Genotyping of plants carrying *vasl_1* or *vasl_2* alleles was done
with primers 5’-TGGTACTTGTGATGCTG-3’, 5’-CGATCTCTACGAGACACTGCG-3’, 5’-TGGATAAATAGCCTTGCTTCC-3’, or 5’-GGAAACTGAGACACTATGCG-3’, 5’-TTTGTCACTTGCGAGTTTTTG-3’, and 5’-ATTTTGCGATTTCCGAAC-3’, respectively. In addition, two heterozygous lines obtained by CRISPR/CAS9 harboring an insertion of a C after position 982 or T after position 980 are referred to as vasl_3 and vasl_4, respectively. Identification of homozygous lines in the next generation was done by CAPS genotyping (see “Supporting Information”).

**Phylogenetic analysis of RNA helicases with similarity to VASL**

For selection and alignment of protein sequences see “Supporting Information”. Phylogenetic analysis was done using RAxML v8.1.16 with model setting “PROTGAMMAAUTO” and including fast bootstrap analysis (100 replicates) (Stamatakis 2014).

**Target pulldown probe design**

To define genomic regions of interest, the closest homologs to the *A. thaliana* genes were identified in *B. stricta* LTM (Li, et al. 2017) by BLASTN with default settings on Phytozome v11 (https://phytozome.jgi.doe.gov). Homologs were determined based on highest sequence similarities, information on synteny and on annotations in Phytozome to belong to a gene family. The probe design for the pulldown was based on sequence information from *B. stricta* LTM genomic regions and up to 5kb upstream sequence downloaded from Phytozome v.11, and coding sequences (cds) from the annotated reference transcriptome of *B. gunnisoniana* (Schmidt et al., 2014), depending on availability of sequence data. 120 bp probes at 4x tiling density were designed and synthesized as myBaits 1-20K custom target capture kit by MYcroarray (Ann Arbor, MI, USA) (Table S2).

**DNA preparation, target pulldown, and library preparation**
DNA was isolated using Invisorb Plant Mini Kit (Stratec Molecular GmbH, Berlin, Germany) following manufacturer instructions from 2-3 seedlings per accession. Genomic DNA was quantified using QUBIT dsDNA HS Assay Kit (Thermo Fisher Scientific) following manufacturer instructions. Genomic DNA was sheared by Covaris before size selection of 300-400 bp fragments for library preparation. Libraries were prepared with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina and NEBNext® Ultra™ Multiplex Oligos for Illumina (New England Biolabs, Ipswich, USA) and amplified by 12 cycles of PCR prior to the pulldown. Subsequently, the mybaits 1-20K kit was used for target pulldown on the libraries following manufacturer instructions with hybridizations done at 60 °C. Prior to sequencing libraries were subjected to additional 10 cycles of PCR.

Next generation sequencing and variant detection

Libraries were sequenced on one lane using 125 bp Paired End (PE) sequencing on Illumina HighSeq2500 by the Deep Sequencing Core Facility (Heidelberg University). After quality control using FastQC v0.11.04 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) reads were trimmed with Trimmomatics 0.22 to remove adopters and low-quality bases (Bolger, et al. 2014). Reads shorter 40 bp after trimming were discarded. Subsequently, reads were aligned to the *B. stricta* genome using stampy v1.0.32 with default parameters (Lunter and Goodson 2011; Lee, et al. 2017). Samtools 1.6 and picardtools 1.95 (http://broadinstitute.github.io/picard/) were applied for sorting, removal of duplicates, and library indexing before realignment with GATK3.4.46 using RealignerTargetCreator and IndelRealigner (Li, et al. 2009; McKenna, et al. 2010). Regions with a base coverage >50 were included in further analysis. Variant calling was done with freebayes v.1.1.0-50 (Garrison 2012) using options for pooled libraries and setting ploidy to the sum of the ploidies of plants and the number of plants used for genomic DNA extraction (Table 1). Filtering of variants was applied with VCFtools 0.1.15 with the parameters DP > 10 and QUAL > 30 (Danecek, et al. 2011). Raw data are deposited under SRA accession PRJNA563290.
Alignment of sequence variants and identification of cds

Sequences of coding regions for *B. stricta* and *A. lyrata* were downloaded from Phytozome v11 (https://phytozome.jgi.doe.gov). Homologs in *A. lyrata* were identified using BLASTN with standard settings. Gene family annotations, highest sequence similarity and synteny were taken into account for determination of the closest homolog. Allelic variants (haplotypes) were sorted by genomic regions. Subsequently, *MAFFT* v7.271 was used to align cds of *B. stricta* and *A. lyrata* to the haplotypes from *Boechera* (Katoh and Standley 2013). Sequence alignments will be made available on Dryad (datadryad.org). To compare cds alignments, all sequences in the alignment were cropped to cds according to the sequence of the primary transcript annotated for *B. stricta* in Phytozome using Phyton. All alignments were checked and readjusted manually using PhyDE Phylogenetic Data Editor (http://www.phyde.de/).

Visualization of usage of nucleobases along genomic regions and mutation density along cds

To visualize the use of different nucleobases across the genomic regions of targets, Python was used to develop a script assessing and plotting the frequencies of distinct nucleobases used at a certain position. Another Phyton script was developed to plot mutation density and types of mutations along the cds. Python scripts developed for this study will be deposited on Github.

Assessment of mutation load

A Python script was developed to annotate changes in cds as compared to the *A. lyrata* reference sequence as silent mutations (synonymous SNPs), nonstop mutations (non-synonymous SNPs), insertions or deletions non causing any stop codons (In Frame Ins or In Frame Del, respectively), insertions or deletions leading to shifts of the open reading frame (Frame Shift Ins or Frame Shift Del, respectively), or nonsense mutations introducing premature stop codons. The information was transferred to a custom .maf file further used for visualization of mutation density and of the most deleterious mutation per allele using the Bioconductor R-package GenVisR and the waterfall
function (Skidmore, et al. 2016, Wagner, et al. 2016). The hierarchical categorization of the most
detrimental mutation per allele was applied in the order of mutations leading to premature stop
(nonsense mutations), INDELs causing frameshifts, in frame INDELS not affecting the reading
frame, non-stop mutations, and synonymous SNPs.

To compare mutation load per gene the numbers of non-synonymous SNPs (here including all
SNPs except the class of synonymous SNPs) per CDS was divided by protein length. Two way
anova and student’s t-test implemented in R were used to test for effects of gene identity and
reproductive mode (see “Supporting Information”).

Assessment of evolutionary rates

To estimate evolutionary rates of haplotypes from Boechera aligned to their A lyrata homologs we
used the yn00 program implemented in PAML (Yang 2007). To allow the analysis, stop codons
were replaced by gaps using PHYDE (http://www.phyde.de/). Variants with frame shifts resulting
in stop codons were excluded from further analysis. Ratios of nonsynonymous to synonymous
single nucleotide polymorphisms (SNP; \( \omega = dN/dS \)) were estimated using the LWL85m model
after (Li, et al. 1985; Yang 2007). Ratios from groups of sexual versus apomictic haplotypes of
one gene were compared by student’s t-test in R (https://www.r-project.org/). Fishers F-test
implemented in R was applied to test for higher variance of mean \( \omega \)-ratios in triploid as compared
to diploid accessions. Boxplot visualization was done using the boxplot function implemented in R
using default options.

Morphological investigation of reproductive defects

Morphological characterization of seed development was done by clearing in Hoyers solution as
described and DIC microscopy with a Zeiss Axio Imager M1 (Zeiss, Oberkochen, Germany)
(Schmidt, et al. 2011). Pictures captured were cropped and processed in Adobe Photoshop CS2
Version 9.0 (Adobe Systems Inc., San Jose, CA, USA)
Results

Genes with structural similarity to Drosophila VASA were identified amongst targets

RNA helicases are a large gene family involved in regulatory processes in any cell and tissue type. To our knowledge, the evolution of members of this family has so far not been studied in related sexual and apomictic plants. For this study, in addition to CENH3, we selected 58 RNA helicases that are active in cells of the female germline (Fig S1; Table S1) (Wuest, et al. 2010; Schmidt, et al. 2011; Schmidt, et al. 2014). They were chosen to be either enriched, or consistently active at key steps of reproduction in sexual or apomictic plants, or to show differences in activity between the reproductive modes. Hereafter, the A. thaliana gene identifier or name is used to refer to Boechera homologs.

These helicases comprised members of different clades, including i.e. genes with described roles for germline development or meiotic recombination, and also such encoding for uncharacterized proteins (Table S1). In order to identify similarities of the latter to described proteins, we modelled their B. stricta protein sequences to related protein structures using Phyre2 (Kelley, et al. 2015). Interestingly, similarities to the Drosophila VASA were predicted with high confidence for the homologs of AT1G72730, now named VASA-LIKE (VASL), and of AT1G51380 (Table S1). To establish their evolutionary relation to other genes in the study, we performed a phylogenetic analysis including 14 additional RNA helicases selected based on sequence similarity (Fig S2). Interestingly, VASL is grouped together with EIF4A1 (AT3G13920) and EIF4A2 (AT1G54270) and this group forms a cluster with AT1G51380.

Target enrichment and sequencing allowed identification of allelic variants in Boechera

To study allelic variation of candidate genes in apomictic versus sexual Boechera, target enrichment was performed (Gnirke, et al. 2009). 24 Boechera accessions representing 12 species or interspecific hybrids of parental species at three ploidy levels were selected to this aim including sexual, facultative and obligate apomictic accessions (Table 1) (Aliyu, et al. 2010; Beck, et al.
Target pulldown was performed on indexed libraries prepared from genomic DNA isolated from 2-3 pooled seedlings per accession (Table 1). Sequencing on the Illumina HighSeq2500 platform (125 bp, paired end (PE)) resulted in 3,763,336 to 13,825,485 paired reads per accession after quality control and trimming (Table 1). Variant calling resulted in identification of 147 alleles in total for all accessions after mapping of the reads to the B. stricta LTM genome, variant calling and filtering (Lee, et al. 2017). High quality data and annotations of allelic variants were achieved for all targeted regions.

**Sequence variation was generally higher in non-coding than in coding regions**

Evolutionary forces affect cds and non-coding regions differently. While changes causing an altered protein sequence can directly impact protein function, upstream regulatory regions are crucial for gene regulation. To visualize the overall sequence variation of target regions allelic variants (haplotypes) from all accessions were aligned. We used a sliding window approach to plot the number of distinct nucleobases used along the sequence with a binning of 12 positions separately for the groups of apomictic, facultative apomictic, or sexual accessions (Fig 1, Fig S3). Furthermore, we analyzed variation across all accessions to allow to identify positions conserved within the same reproductive mode but different between groups. Most differences in base usage and regions of overall high sequence variation in either sexual accessions or apomicts were identified in non-coding regions including upstream regulatory regions or introns, for example in AT2G40700 (*RNA HELICASE 17, RH17*) or AT3G13920 (*EIF4A1*) (Fig 1, Fig S3). However, also within these regions stretches of lower sequence variation can be identified, likely allowing to predict positions of conserved motifs. Interestingly, a region of high variability both within and between the groups was observed i.e. for *MEM* upstream of the start codon (Fig 1). Furthermore, for *VASL* sequence variation was observable mostly in the putative promoter region and in an intron within the 5’ untranslated region (Fig 1). While in case of *MAA3* differences in sequence variation between reproductive modes were mostly observable in intronic regions, also in the 17th exon high sequence variation was identified for facultative apomicts (Fig 1). Future investigations
are needed to elucidate if these regions are functionally important for the regulation of the genes and if differences are relevant for the reproductive modes, or if this high variability mainly results from the absence of selective pressure.

**Cds variants are mostly under purifying selection in diploid and triploid apomicts**

To describe mutation accumulation of targeted genes in more detail, we specifically investigated Cds. Therefore, haplotypes were aligned to Cds from *A. lyrate* as a suitable reference for variant calling in *Boechera* (Lovell, et al. 2017). Based on this, synonymous or non-synonymous Single Nucleotide Polymorphisms (SNPs), insertion and deletions (INDELs), and mutations leading to premature stop codons (nonsense mutations) were identified. Interestingly, the numbers of distinct alleles identified correlated with reproductive mode and ploidy level (Fig 2, Fig S4): In all diploid sexual accessions for at least half of all genes only one haplotype was identified, in contrast to two or more most frequently observed in apomicts (Fig 2). In the same line, the tendency was more pronounced in apomicts than in sexual accessions to show differences in the numbers of mutations per MB for different allelic variants of the same accession (Fig 3A). Nevertheless, also differences were observed depending on the gene. The maximum number of allelic variants identified per gene in any accession ranged between 18 and 117 (Table S1), with only few sequence variants identified e.g. for EIFs and higher numbers e.g. for members of chromatin remodelers and others (Fig S4, Table S1).

To obtain insights into the evolutionary mechanisms shaping diversification of the alleles in *Boechera*, we further estimated the ratios of nonsynonymous (dN) versus synonymous (dS) SNPs ($\omega = \text{dN/dS}$). This can be taken as an indication for the evolutionary rate, with $\omega < 1$ indicating purifying selection, $\omega = 1$ neutral evolution and $\omega > 1$ diversification (Yang and Bielawski 2000). For asexual reproduction an accumulation of slightly deleterious mutations under nearly neutral evolution has been hypothesized (Schiffels, et al. 2011; Hojsgaard and Hörandl 2015). To uncover if this holds true for the RNA helicases we estimated $\omega$-values using the yn00 program.
implemented in PAML (Yang 2007). While this test indicated purifying selection for cds both from sexual and apomictic accessions (Fig S5), differences in $\omega$ were still evident between genes, with the homologs of ESP3 (AT1G32490), EIF4A1 (AT3G13920), EIF4A2 (AT1G54270), AT1G51380, AT1G72730 (VASL), CHR12 (AT3G06010), CHR11 (AT3G06400), ABA OVERLY SENSITIVE 6 (ABO6; AT5G04895), and others being under strong purifying selection, while SNF2-RING-HELICASE LIKE5 (FRG5) (AT1G11100) and CHR40 (AT3G24340) showed relatively high $\omega$-values pointing towards near neutral evolution.

Furthermore, $\omega$-values averaged for alleles of obligate apomictic as compared to sexual accessions were only significantly different for 6 genes (student’s t-test; p-value < 0.05; Fig S5). While the homologs of RECQ4A (AT1G10930), RH5 (AT1G31970), AT2G35920, and AT3G02060 showed lower selection pressure on allelic variants of sexual accessions as compared to apomicts, only for ESP3 and AT1G58060 higher rates of mutation accumulation were identified for alleles from apomictic accessions.

As higher ploidy is commonly perceived as a strategy to buffer the effect of deleterious mutations, we tested in addition, if different allelic variants accumulate mutations at higher rates in triploid apomicts as compared to diploids. This would result in a larger variance of mean dN/dS ratios. We tested for this using the F-test implemented in R (p-value < 0.05; www.r-project.org/). However, this trend was not observed for any of the RNA helicases preselected based on broader distributions of $\omega$ values in all apomicts, except for AT5G61140 at near significant levels ($p = 0.632$) (Fig S6).

**Allelic variants with premature stop codons were identified in more than half of the genes**

To categorize mutations based on their effect on the protein encoded, a hierarchical categorization of the most detrimental mutation per allele was applied (Fig 3) (Skidmore, et al. 2016). As compared to *A. lyrata*, none of the genes showed only synonymous SNPs (Fig 3B). Nevertheless, EIF4A1 and EIF4A2 (AT3G13920 and AT1G54270, respectively), as well as the uncharacterized
proteins encoded by VASL (AT1G72730), and CHR17 (AT5G18620) showed only non-stop mutations as the most deleterious changes present. For additional 21 genes studied in frame INDELs were the most deleterious type of mutations represented in any accession (Fig 3B). In the cds of 11 helicases mutations leading to stop codons (nonsense mutations, INDELs) were exclusively found in allelic variants from apomicts in contrast to alleles from sexual accessions (Fig 3B). Interestingly, these comprise genes involved in epigenetic regulatory pathways, maintenance of genome stability, or regulation of reproduction, including DCL3 (AT3G43920), CHR40 (AT3G24340), FRG2 (AT1G50410), REGULATOR OF TELOMERE ELONGATION HELICASE1 (RTEL1; AT1G79950), and MAA3 (AT4G15570) (Shimizu, et al. 2008; Ahmad, et al. 2010; Groth, et al. 2014; Recker, et al. 2014; Zhou, et al. 2018). In contrast mutations leading to premature stop codons were only observed for alleles from sexual accessions for MEM (AT5G39840), in addition the INO80 orthologue AT3G57300, AT2G47250 and AT2G35920 (Fig 3B). For other genes, including DCL2 (AT3G03300), RH1 (AT4G15850) and RH5 (AT1G31970) mutations leading to premature stop were identified in accessions of both reproductive modes (Fig 3B, Table S1). Moreover, high abundance of alleles with deleterious mutations including absence of start codons and premature stop codons were observed for several genes, including CHR34 (AT2G21450), MEE29 (AT2G35340), and FASCIATED STEM4 (FAS4, AT1G33390) (Fig 3). The consistent identification of frame shift mutations is mis-leading for the homolog of PIGMENT DEFECTIVE (AT5G08610) and DECREASE IN DNA METHYLATION1 (DDM1, AT5G66750), as in these cases reference sequences from A. lyrata are much divergent from the sequences in Boechera. Taken together, the data indicate that the presence of deleterious mutations is more dependent on the gene and possibly functional implications rather than on the reproductive mode.

The load of deleterious mutations varies greatly between genes

Apart from the most deleterious mutations contained in a cds the mutation density and position of amino acid changes or INDELs typically impacts protein folding and function. While the core domains of RNA helicases are evolutionary highly conserved, aspects of gene evolution and
functionalization are driven by changes in flanking C- and N-terminal sequences (Sloan and Bohnsack 2018). To gain more detailed insights into sequence changes we plotted the mutation distribution along the cds. First, comparisons included all 147 allelic variants from any accession by showing the percentage of alleles holding a mutation when averaging 12 consecutive bp (Fig 4). Here we focused on MEM, VASL, and AT3G16840 all previously described as significantly enriched during germline formation in Arabidopsis (Schmidt, et al. 2011), in addition to FAS4, EIF4A1, and RECQ4A. Overall, mutation density varied greatly between these proteins with mainly synonymous and only few non-synonymous SNPs present in EIF4A and VASL, consistent with the strong purifying selection observed (Fig 4, Fig S5). While sharing the conserved PFAM domains “DEAD/DEAH box helicase” (PF00270) and “helicase conserved C-terminal domain” (PF00271) with the former, AT3G16840 is characterized by longer N- and C-termini showing a significant load of nonstop mutations and in frame INDELs. MEM showed mainly changes in the N-terminal domain, but for a few alleles also nonsense mutations close to the stop codon. A number of mutations are also represented in the conserved helicase domains, especially in the “mitochondrial degradasome RNA helicase C terminal” (PF12513) domain. Interestingly, both FAS4 and RECQ4A presented a high mutation load including deleterious frame shift mutations in different sequence contexts along the protein. Furthermore, mutation load and position might differ between reproductive modes. To investigate this, we compared the mutation plots of MEM and MAA3 between the groups of sexual or obligate apomictic accessions (Fig S7). Interestingly, most positions of mutations were shared between both groups with only few differences observed between variants from sexual or apomictic accessions (Fig S7). Nevertheless for MEM, a slightly higher amount of synonymous SNPs and nonstop mutations was identified in sequences from apomicts as compared to sexual Boechera accessions, in particular 5’ of the “helicase conserved C-terminal domain” (Fig S7). Similarly for MAA3, a tendency towards more nonstop mutations was apparent in apomicts. In addition, few deleterious mutations in the conserved “AAA-domain” (PF13086) leading to a premature stop
codon (frame shift INDELs and nonsense mutations) were identified only in allelic sequences of apomictic accessions.

Our analyses indicate greater differences in mutation load dependent on the gene identity than on the mode of reproduction. Consistently, gene identity was identified as significant factor using two way anova (p < 0.001). This difference was supported by pairwise comparisons of selected RNA helicases (Fig S8). While student’s t-test indicated differences in mutation accumulation between allelic variants of sexual accessions as compared to facultative or obligate apomictics for less than half of the genes (p < 0.05), highly significant differences were observed between the mutation accumulation of VASL alleles as compared to any other gene tested within sexual accessions (p < 0.001).

In summary, differences in mutation load largely differed between genes, with such RNA helicases encoding i.e. for VASL and the related evolutionary conserved EIF proteins remaining highly conserved both in sexual and apomictic species. In contrast, for the coding regions of other genes more divergence between allelic variants was observed, including such with known functions for germline development or candidate genes for elements of apomixis.

**High sequence conservation of VASL correlates with roles during reproductive development**

Our data indicated different evolutionary pressure to act on the RNA helicases both in apomicts and in sexuals. Interestingly, a small number of genes was identified to be highly conserved and under purifying selection independent of the reproductive mode, including the undescribed gene VASL.

To gain first insights into the potential roles of VASL for reproductive development we took advantage of the established model species A. thaliana. First, we analyzed two independent T-DNA insertion lines, vasl_1 and vasl_2. Furthermore, we generated vasl_3 and vasl_4 mutant
lines with single nucleotide insertions leading to premature stop codons using CRISPR/Cas9 (Wang, et al. 2015). Interestingly, in all lines similar defects of reproductive development and pleiotropic phenotypes were observed both in heterozygous and in homozygous plants, although at different frequencies (Fig 5, Fig S9, Table S3). In agreement with VASL expressed predominantly during germline formation (Schmidt, et al. 2011), in all lines sporophytic defects of ovule development and protruding gametophytes were observed (Fig 5, Table S3). Also, indications were given for the formation of ectopic gametophytes likely not derived from a FMS formed at the right position as in wild-type, in addition to occasional observations of odd numbers of nuclei in gametophytes (Fig 5, Table S3). Further developmental defects were observed during seed development (Fig S9, Table S3). While embryo development usually follows a series of well-defined cell divisions and developmental stages (Jenik, et al. 2007), in our lines over-proliferation of the embryo and abort of embryo development was observed at low frequencies (Fig S9, Table S2). Taken together, the analysis indicates the importance of VASL for plant reproduction and supports the hypothesis that genes involved in crucial cellular processes or important for developmental processes including reproduction are conserved also in apomicts.

**Discussion**

**Purifying selection observed for most RNA helicases challenges common hypothesis on mutation accumulation in apomicts**

The coexistence of sexual and asexual reproduction is a feature of reproduction across different kingdoms. From an evolutionary perspective, both sexual and asexual reproduction have their unique advantages and downsides. While sexual reproduction allows de novo combination of genetic information and genomic elements and thus holds great potential for adaptive evolution, also favorable allelic combinations can get lost by meiotic recombination. Furthermore, investment of energy and resources is required for the production of male gametes and for finding a suitable mating partner (Barton and Charlesworth 1998; Charlesworth 2006; Stelzer 2015). Asexual
reproduction on the other hand can lead to the rapid fixation of hybrid genotypes frequently outperforming their parents. Over time however, mutations are predicted to accumulate in genomes of apomictic lineages leading to their rapid extinction (Muller 1964; Stelzer 2015). Investigations on different asexual species have only partially proven this hypothesis. Confirmation of this prediction has previously been found for *Oenothera*, where higher rates of mutation accumulation have been identified in functionally asexual as compared to sexual taxa based on comparisons of transcriptomes (Hollister, et al. 2014). In the functional asexual taxa, recombination is nearly fully suppressed and combined with self-fertilization this results in offspring genetically identical to the mother plants by another system than apomixis (Hollister, et al. 2014). Moreover, in apomictic *Boechera*, genome wide higher frequencies of mutations have been described as compared to sexual lineages (Lovell, et al. 2017). While an elevated mutation load in apomicts is even presented at phylogenetically conserved sites, the effect is least observable for coding sites (Lovell, et al. 2017). This suggests that also in apomicts mutation accumulation is not stochastically affecting any position in the genome, but underlies certain evolutionary constraints. In the same line, from a study of *Ranunculus auricomus* indications are given that gene evolution in apomicts is not random but related to gene functions (Pellino, et al. 2013). Indeed, for cds, elevated dN/dS-ratios and thus indications either for accumulation of deleterious mutations or for diversifying selection are only given for certain genes, including such functionally important for meiosis and gametogenesis, while the majority of genes are under purifying selection (Pellino, et al. 2013). To explain this observation, a relatively high degree of persistent facultative sexuality has been postulated to counteract the genome wide effect of mutation accumulation (Hojsgaard and Hörandl 2015). Occasional processes of sexual reproduction would lead to an efficient unmasking of deleterious mutations and rapid purging from the population (Hojsgaard and Hörandl 2015; Hodač, et al. 2019). Interestingly, while obligate apomicts have been defined based on maturing seeds in *Boechera* (Aliyu, et al. 2010; Mau, et al. 2015), greater developmental plasticity during early stages of germline development have recently been described and also rare
reversions from apomictic to sexual reproductions are postulated (Carman, et al. 2019). In apomictic *Boechera*, both apospory and diplospory are common. Diplospory is most frequently of *Taraxacum*-type, where a first division restitution of meiosis takes place in the AIC followed by meiosis II, but also *Antennaria*-type of diplospory has been reported where the AIC omits meiosis to directly initiate gametogenesis (Carman, et al. 2019). Also, rare occurrence of second division restitution maintaining meiotic recombination cannot fully be excluded. With respect to the evolutionary young apomicts of *Ranunculus auricomus*, it has also been proposed that the time of divergence from sexual lineages was too short to allow substantial diversification (Hojsgaard and Hörandl 2015). In general, mutation accumulation in asexual species or lineages is largely influenced by the time of evolution, but also by population size (Brandt, et al. 2017). Large population sizes thereby efficiently counteract the tendency of mutation accumulation. This has recently been shown by comparing ancient asexual to sexual oribatid mites (Brandt, et al. 2017). Here, in contrast to common expectations, higher efficiency of purifying selection was observed in asexual rather than in sexual mites (Brandt, et al. 2017). Apart from the effect of large populations, it can also not fully be excluded that so far largely overlooked genome repair mechanisms contribute to a certain degree to the conservation of sequences in asexual species, similar to mechanisms acting on non-recombining genomes of mitochondria and chloroplasts (Khakhlova and Bock 2006; Brandt, et al. 2017). Indeed, unexpected mechanisms of evolution have previously been reported for the ameiotic bdelloid rotifer *Adenita vaga*, where gene conversion and horizontal gene transfer counteract the accumulation of deleterious mutations (Flot, et al. 2013). Taken together, the partially contradicting observations from different studies in general suggest versatile mechanisms shaping gene evolution in different asexual systems and populations.

More allelic variants per accession were identified from apomicts than from sexuals
In our study, rather than investigating a population, we selected accessions to represent part of the genetic diversity and distinct geographic origins in *Boechera* (Aliyu, et al. 2010; Mau, et al. 2015; Carman, et al. 2019). Likewise done previously (Lovell, et al. 2017), we defined mutations by comparison of sequences to the homologs of *A. lyrata*. Thereby, the finding of sequence changes largely at similar positions in sexual and apomictic *Boechera* reflects the general divergence of *Boechera* and *Arabidopsis* split about 10-14.5 Mya (timetree.org). Moreover, the number of haplotypes per gene in different accessions was correlated with the reproductive mode and ploidy levels, in agreement with previous studies indicating highest levels of heterozygocity for triploid apomicts, intermediate levels for diploid apomicts and lowest levels for diploid sexuals (Lovell, et al. 2013). In this line, especially in case of polyploid apomicts the frequencies of mutation accumulation are likely to differ between alleles (Hojsgaard and Hörandl 2015). In consequence, different evolutionary forces can shape mutation accumulation on different alleles to allow loss of function or even *de novo* functionalization of one allelic variant. While polyploidy is perceived as mechanism to buffer deleterious mutations the situation in diploid apomicts is less well understood as complex genetic interactions might play a role in addition to copy number variations previously reported (Aliyu, et al. 2013). Nevertheless, no consistent tendency of rapid diversification between alleles of the same accession was observed in triploid as compared to diploid apomicts. Also, disentangling the origins of the different allelic variants is not trivial, as many apomicts are derived by hybridization or intra-specific crosses several times (Lovell, et al. 2013).

**Cds regions are less variable than non-coding sequences**

To gain a better understanding of the diversifications of alleles in *Boechera*, we accessed sequence divergence along the genomic regions. Hereby, increased variability in the sequences was evident particularly for non-coding regions. The identification of conserved stretches in regulatory regions, i.e. when comparing different taxa, often indicates sequence motives important
for regulation, particularly transcription factor binding sites. Our comparison of variance allows to
distinguish regions of overall high variability and from such with differences between reproductive
modes. These might serve as an indication for changes in regulation, as previously been
described for the APOLLO-locus showing polymorphisms linked to apomixis in the 5’ UTR leading
to changes in transcription factor binding sites (Corral, et al. 2013). Therefore, identification of
differences in regulatory regions between sexual and apomictic accessions can serve as starting
points for investigations of potential functional roles of different alleles.

**Gene evolution relates to function in sexual and apomictic lineages**

Compared to non-coding regions, less variability was observed in cds. Based on their allelic
variants, Boechera RNA helicases could be classified into distinct groups: Genes with high
sequence conservation being under strong purifying selection independent of the reproductive
mode, genes showing more non-synonymous sequence changes in the evolutionary less
conserved N- and C-terminal domains, and genes harboring a number of allelic variants with
deleterious mutations. These differences seemed to be mainly dependent on gene function, with
e.g. members of the evolutionary conserved EIFs remaining conserved also in apomicts. In
contradiction to common predictions for a subset of helicases, even higher sequence conservation
of allelic variants was given in apomicts than in sexual accessions. However, it needs to be taken
into account that in our analysis stop codons were replaced by deletions and alleles with frame
shift mutations were not taken into account, as yn00 cannot account for stop codons. This likely
leads to an underestimation of sequence divergence for a subset of genes.

**Genes less conserved in apomicts involve members of epigenetic regulatory pathways**

While overall cds remained under purifying selection in any accession, near neutral evolution was
given for certain genes, including the homolog of FRG5. As FRG1 and FRG2 are involved in RNA-
directed DNA methylation (Groth, et al. 2014), also the Boechera homolog of FRG5 might serve
similar roles. Moreover, ESP3 as a player in epigenetic regulatory pathways, was identified to be
amongst the few genes with higher mutation load in apomicts or mutations causing stop codons only in alleles from apomicts. These findings are in agreement with the notion of apomixis being related to complex changes in the epigenetic landscape (León-Martínez and Vielle-Calzada 2019). However, as none of the deleterious mutations identified is consistently present in all apomicts investigated, such mutations are more likely to contribute to sustaining apomixis rather than being causal for the reproductive mode. Likewise, it is possible that different genes in similar epigenetic pathways are affected by mutations leading to similar deregulations of gene activity related to expression of components of apomixis. It is also important to note that genetically apomixis is typically linked to one or few recombination repressed loci (Barcaccia and Albertini 2013). In this respect, genes encoded on these loci should remain to be prone to mutation accumulation even at low rates of facultative sexuality. On the other hand, identification of elevated mutation frequency alone cannot be taken as sufficient indication for a gene to be linked to an apomixis locus.

Interestingly, also for MEM, evidence for few more potentially deleterious mutations in core regions of apomictic as compared to sexual alleles was given. As mutant mem alleles cause formation of an additional and unreduced gametophyte in the same ovule in A. thaliana (Schmidt, et al. 2011), such mutations might be relevant for aspects of apomixis in Boechera. Furthermore, high sequence variation was observed upstream of the MEM coding region, suggesting differences in the regulation of MEM activity. To regulate distinct developmental features, the observed diversity of MEM allelic variants might correlate to the developmental flexibility during germline development in Boechera (Carman, et al. 2019). Nevertheless, MEM alleles carrying deleterious mutations were identified only from sexual accessions. As these mutations were in the C-terminal domain close to the stop codon the mutation might not severely affect MEM functionality. Future functional investigations of allelic variants will be required to disentangle their potential relevance for aspects of apomixis.
Interestingly, for a subset of helicases, a high frequency of deleterious mutations was observed in both reproductive modes. Together with a high number of allelic variants observed for these genes this might indicate cases of redundancy or copy number variations, or exceptionally problems of accurate mapping due to close homologs present in the genome.

**High sequence conservation of certain RNA helicases indicates functions in development**

In contrast to the expected mutation accumulation in apomicts, we also identified a number of RNA helicases under strong purifying selection in all accessions. These include *CHR17* and members of the EIF family, involved in controlling fundamental cellular functions of translational control and regulation of nucleosome patterning, respectively, and in turn important to control gene or protein activity (Struhl and Segal 2013; Li, et al. 2014; Dutt, et al. 2015). From the animal germline it is well appreciated that EIF proteins have essential functions in the regulation of translation of stored mRNAs (Henderson, et al. 2009). Similar to the importance of such mechanisms in animals, evidence suggests that translational regulation of stored mRNAs is an essential mechanism also in the plant germline, in particular during meiosis (Schmidt, et al. 2011). Here, we identified the undescribed gene *VASL* not only to be highly conserved in *Boechera*, but also to have similarities to animal specific VASA proteins, while being related to *EIF4A1* and *EIF4A2* based on phylogenetic analysis. Nevertheless, low bootstrap support for certain nodes likely results from the overall high divergence of RNA helicases.

Our functional analysis identifies *VASL* as gene involved in regulating reproductive development. In agreement with a relation to *EIF4A1* defects in ovule development observed partially resemble phenotypes in *eif4A1* and *eif4A2* mutants (Bush, et al. 2015). EIF proteins typically act in multimeric complexes often involving different isoforms, some of which are not essential for translation initiation but rather modify the efficiency of the process (Hinnebusch and Lorsch 2012). Importantly, also VASA has previously been determined to play roles in translation initiation and cell cycle regulation, in particular with respect to genes required for germline development (Yajima
and Wessel 2011; Marintchev 2013). Therefore, VASL might serve similar roles in plant reproductive tissues. Strikingly, similar effects were observed in lines with insertions in the 3' UTR as for the lines with a premature stop codon, as well as in heterozygous and homozygous lines. As the 3' UTR is relevant for mRNA stability or translational regulation (Srivastava, et al. 2017), these lines likely result in altered levels of VASL. Similar to EIF4A1 acting in a level dependent manner (Vain, et al. 2011; Bush, et al. 2016), likely also VASL levels are relevant its mode of action.

Our observation of high conservation of genes important for reproductive development appears partially contradicting to the observations in Ranunculus auricomus where genes with roles for gametophyte development showed higher mutation accumulation in apomicts (Pellino, et al. 2013). For such genes higher rates of mutations in apomicts might be explained by dosage effects or recessive mutations being buffered by unreduced genomes in apomictic gametophytes. Nevertheless, even in sexual plants genes preferentially expressed in female gametes are prone to higher rates of protein evolution as compared to the genome wide average (Gossmann, et al. 2013). This is in agreement with the finding of accumulation of deleterious mutations in FAS4 and CHR34 in alleles of any accession, which are predominantly transcribed in the cells of the female gametophyte in A. thaliana (Wuest, et al. 2010). High evolutionary rates of genes in female gametogenesis might in general allow for a certain developmental flexibility during reproductive development, potentially also involved in sustaining the differences between sexual and apomictic reproduction.

In conclusion, our data indicate that versatile mechanisms are in place to shape the evolution of genes in apomictic Boechera. Thereby the gene function appears as a key determinant for the acquisition of mutations. Based on the hypothesis that a high degree of purifying selection can be taken as an indication for important functional roles, VASL has been identified as a novel gene playing a role to regulate reproductive development. This indicates that the data presented provide
an ideal starting basis for future functional investigations on the roles of the genes and allelic variants. Important new insights into aspects of the evolution of RNA helicases in sexual and apomictic Boechera are presented. A deeper understanding of aspects of the evolutionary forces acting on apomicts, in particular with respect to genes important for regulating reproduction is not only of great interest from a scientific point of view. In longer terms this understanding can also contribute to allow predictions on behavior of engineered apomicts released to fields for agricultural use.

Acknowledgement

We thank Timothy Sharbel (Global Institute for Food Security, Canada), Thomas Mitchell-Olds (Duke University, USA), and John Carman (Utah State University, USA) for providing Boechera seeds, Marcus A. Koch (Centre for Organismal Studies Heidelberg (COS), Germany) for providing instruments and facilities and support through his technical assistant. We also thank Jan U. Lohmann (COS Heidelberg) for contributing instruments, and Christiane Kiefer (COS Heidelberg) for critical reading of the manuscript and for helpful discussions. We are grateful to Anne Pfeiffer (COS Heidelberg) for advice concerning CRISPR/Cas9, and to Qi-Jun Chen (College of Biological Sciences, China Agricultural University, Beijing, China) for providing the pHEE401E plasmid. This work was supported by funding of Deutsche Forschungsgemeinschaft (DFG) SCHM2448/2-1 and SCHM2448/2-2 to AS.

References

Ahmad A, Zhang Y, Cao X-F. 2010. Decoding the Epigenetic Language of Plant Development. Mol Plant 3:719-728.
Aliyu OM, Schranz ME, Sharbel TF. 2010. Quantitative variation for apomictic reproduction in the genus Boechera (Brassicaceae). Am J Bot 97:1719-1731.
Aliyu OM, Seifert M, Corral JM, Fuchs J, Sharbel TF. 2013. Copy Number Variation in Transcriptionally Active Regions of Sexual and Apomictic Boechera Demonstrates Independently Derived Apomictic Lineages. Plant Cell 25:3808-3823.
Arkov AL, Ramos A. 2010. Building RNA-protein granules: insight from the germline. Trends Cell Biol 20:482-490.
Asano K, Clayton J, Shalev A, Hinnebusch AG. 2000. A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. Genes Dev 14:2534-2546.

Audiffren J, Pardoux E. 2013. Muller’s ratchet clicks in finite time. Stochastic Processes and their Applications 123:2370-2397.

Barcaccia G, Albertini E. 2013. Apomixis in plant reproduction: a novel perspective on an old dilemma. Plant Reprod 26:159-179.

Barton NH, Charlesworth B. 1998. Why Sex and Recombination? Science 281:1986-1990.

Beck JB, Alexander PJ, Allphin L, Al-Shebhab IA, Rushworth C, Bailey CD, Windham MD. 2012. DOES HYBRIDIZATION DRIVE THE TRANSITION TO ASEXUALITY IN DIPLOID BOECHERA? Evolution 66:985-995.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Brandt A, Schaefer I, Glanz J, Schwander T, Maraun M, Scheu S, Bast J. 2017. Effective purifying selection in ancient asexual oribatid mites. Nat Commun 8:873.

Bush MS, Crowe N, Zheng T, Doonan JH. 2015. The RNA helicase, eIF4A-1, is required for ovule development and cell size homeostasis in Arabidopsis. Plant J 84:989-1004.

Bush MS, Pierrat O, Nibau C, Mikitova V, Zheng T, Corke FMK, Vlachonasios K, Mayberry LK, Browning KS, Doonan JH. 2016. eIF4A RNA Helicase Associates with Cyclin-Dependent Protein Kinase A in Proliferating Cells and Is Modulated by Phosphorylation. Plant Phys 172:128-140.

Carman JG. 1997. Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bisporpy, tetrasporpy, and polyembryony. Biol J Linn Soc Lon 61:51-94.

Carman JG, Mateo de Arias M, Gao L, Zhao X, Kowallis BM, Sherwood DA, Srivastava MK, Dwivedi KK, Price BJ, Watts L, et al. 2019. Apospory and Diplospory in Diploid Boechera (Brassicaceae) May Facilitate Speciation by Recombination-Driven Apomixis-to-Sex Reversals. Front Plant Sci 10.

Charlesworth B. 2006. The Evolutionary Biology of Sex. Curr Biol 16:R693-R695.

Conner JA, Oziak-Akins P. 2017. Apomixis: Engineering the Ability to Harness Hybrid Vigor in Crop Plants. Methods Mol Biol 1669:17-34.

Corral JM, Vogel H, Aliyu OM, Hensel G, Thiel T, Kumlehn J, Sharbel TF. 2013. A Conserved Apomixis-Specific Polymorphism Is Correlated with Exclusive Exonuclease Expression in Premeiotic Ovules of Apomictic Boechera Species. Plant Physiol 163:1660-1672.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics (Oxford, England) 27:2156-2158.

Dutt S, Parkash J, Mebra R, Sharma N, Singh B, Raigon P, Joshi A, Chopra S, Singh BP. (Dutt2015 co-authors). 2015. Translation initiation in plants: roles and implications beyond protein synthesis. Biol Plantarum 59:401-412.

Fairman-Williams ME, Guenther U-P, Jankowsky E. 2010. SF1 and SF2 helicases: family matters. Curr Opin Struct Biol 20:313-324.

Flot J-F, Hesperks B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejno A, Henrissat B, Koszul R, Aury J-M, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature 500:453-457.

Friday AJ, Keiper BD. 2015. Positive mRNA Translational Control in Germ Cells by Initiation Factor Selectivity. BioMed Res 2015:327963.

Gao M, Arkov AL. 2013. Next Generation Organelles: Structure and Role of Germ Granules in Germline. Mol Reprod Devel 80:610-623.

Garcia-Aguilar M, Michaud C, Leblanc O, Grimanelli D. 2010. Inactivation of a DNA Methylation Pathway in Maize Reproductive Organs Results in Apomixis-Like Phenotypes. Plant Cell 22:3249-3267.

Garrison EM, Gabor. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T,
Giannoukos G, Fisher S, Russ C, et al. 2009. Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. Nat Biotechnol 27:182-189.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2013. Selection-Driven Evolution of Sex-Biased Genes Is Consistent with Sexual Selection in Arabidopsis thaliana. Mol Biol Evol 31:574-583.

Groth M, Stroud H, Feng S, Greenberg MVC, Vashisht AA, Wohlschlegel JA, Jacobsen SE, Ausin I. 2014. SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation. Proc Nat Acad Sci U S A 111:17666-17671.

Gustafson EA, Wessel GM. 2010. Vasa genes: emerging roles in the germ line and in multipotent cells. Bioessays 32:626-637.

Hafidh S, Čapková V, Honys D. 2011. Safe Keeping the Message: mRNP Complexes Tweaking after Transcription. In: Collins LJ, editor. RNA Infrastructure and Networks. New York, NY: Springer New York. p. 118-136.

Henderson MA, Cronland E, Dunkelbarger S, Contreras V, Strome S, Keiper BD. 2009. A germline-specific isoform of elf4E (IFE-1) is required for efficient translation of stored mRNAs and maturation of both oocytes and sperm. J Cell Sci 122:1529-1539.

Hernández G, Vazquez-Pianzola P. 2005. Functional diversity of the eukaryotic translation initiation factors belonging to elf4 families. Mech Develop 122:865-876.

Higgins JD, Ferdous M, Osman K, Franklin FCH. 2011. The RecQ helicase AtRECQ4A is required to remove inter-chromosomal telomeric connections that arise during meiotic recombination in Arabidopsis. Plant J 65:492-502.

Hinnebusch AG, Lorsch JR. 2012. The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb Perspect Biol 4.

Hodač L, Klatt S, Hojsgaard D, Sharbel TF, Hörandl E. 2019. A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. BMC Evol Biol 19:170.

Hojsgaard D, Hörandl E. 2015. A little bit of sex matters for genome evolution in asexual plants. Front in Plant Sci 6: 82.

Hollister J, Greiner S, Wang W, Wang J, Zhang Y, Ka-Shu Wong G, I Wright S, T J Johnson M. 2014. Recurrent Loss of Sex Is Associated with Accumulation of Deleterious Mutations in Oenothera. Mol Biol Evol 32:896-905.

Hojsgaard D, Klatt S, Baier R, Carman JG, Hörandl E. 2014. Taxonomy and Biogeography of Apomixis in Angiosperms and Associated Biodiversity Characteristics. Critical reviews in plant sciences 33:414-427.

Huanca-Mamani W, Garcia-Aguilar M, León-Martínez G, Grossniklaus U, Vielle-Calzada J-P. 2005. CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in Arabidopsis thaliana. Proc Nat Acad Sci U S A 102:17231-17236.

Jankowsky E. 2011. RNA Helicases at work: binding and rearranging. Trends Biochem Sci 36:19-29.

Jenik P, Stewart Gillmor C, Lukowitz W. 2007. Embryonic Patterning in Arabidopsis thaliana. Ann Rev Cell Dev Biol 23:207-236.

Katoh K, Stendle DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772-780.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845.

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

Khemic V, Linder P. 2018. RNA helicases in RNA decay. Biochemical Society Transactions 46:163-172.

Kliver S, Rayko M, Komissarov A, Bakin E, Zhernakova D, Prasad K, Rushworth C, Baskar R, Smetanin D, Schmutz J, et al. 2018. Assembly of the Boechera retrofracta Genome and Evolutionary Analysis of Apomixis-Associated Genes. Genes (Basel) 9.

Kotliński M, Knizewski L, Muszewska A, Rutowicz K, Lirski M, Schmidt A, Baroux C, Ginalska I, Jerzmanowski A. 2017. Phylogeny-Based Systematization of Arabidopsis Proteins with Histone H1 Globular Domain. Plant Physiol 174:27-34.
Lee C-R, Wang B, Mojica J, Mandáková T, Prasad KVS, Goicoechea JL, Perera N, Hellsten U, Hundley HN, Johnson J, et al. 2017. Young inversion with multiple linked QTLs under selection in a hybrid zone. Nat Ecol Evol 1:119-119.

León-Martínez G, Vielle-Calzada J-P. 2019. Chapter Twenty - Apomixis in flowering plants: Developmental and evolutionary considerations. In: Grossniklaus U, editor. Current Topics in Developmental Biology: Academic Press. p. 565-604.

Li F-W, Rushworth CA, Beck JB, Windham MD. 2017. Boechera microsatellite website: an online portal for species identification and determination of hybrid parentage. Database: The Journal of Biological Databases and Curation 2017:baw169.

Li G, Liu S, Wang J, He J, Huang H, Zhang Y, Xu L. 2014. ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis. Plant J 78:706-714.

Li H, Handsaker B, Wyssoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150-174.

Linder P, Jankowsky E. 2011. From unwinding to clamping — the DEAD box RNA helicase family. Nat Rev Mol Cell Biol 12:505.

Linder P, Owttrim GW. 2009. Plant RNA helicases: linking aberrant and silencing RNA. Trends Plant Sci 14:344-352.

Liu M, Shi D-Q, Yuan L, Liu J, Yang W-C. 2010. SLOW WALKER3, Encoding a Putative DEAD-box RNA Helicase, is Essential for Female Gametogenesis in Arabidopsis. J Integ Plant Biol 52:817-828.

Lovell JT, Aliyu OM, Mau M, Schranz ME, Koch M, Kiefer C, Song B-H, Mitchell-Olds T, Sharbel TF. 2013. On the origin and evolution of apomixis in Boechera. Plant Reprod 26:309-315.

Lovell JT, Williamson RJ, Wright SI, McKay JK, Sharbel TF. 2017. Mutation Accumulation in an Asexual Relative of Arabidopsis. PLoS Genet 13:e1006550.

Lunter G, Goodson M. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21:936-939.

Lysak MA. 2014. Live and let die: centromere loss during evolution of plant chromosomes. New Phytol 203:1082-1089.

Marintchev A. 2013. Roles of Helicases in Translation Initiation: A Mechanistic View. Biochim Biophys Acta 1829:799-809.

Mau M, Lovell JT, Corral JM, Kiefer C, Koch MA, Aliyu OM, Sharbel TF. 2015. Hybrid apomicts trapped in the ecological niches of their sexual ancestors. Proc Nat Acad Sci U S A 112:E2357-E2365.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303.

Muller HJ. 1964. THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. Mutat Res 106:2-9.

Nakajima K. 2018. Be my baby: patterning toward plant germ cells. Curr Opin Plant Biol 41:110-115.

Nonomura KI. 2018. Small RNA pathways responsible for non-cell-autonomous regulation of plant reproduction. Plant Reprod 31:21-29.

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP. 2010. Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature 464:628-632.

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V. 2005. Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development 132:603-614.

Pellino M, Hojsgaard D, Schmutzer T, Scholz U, Horandl E, Vogel H, Sharbel TF. 2013. Asexual genome evolution in the apomictic Ranunculus auricomus complex: examining the effects of hybridization and mutation accumulation. Mol Ecol 22:5908-5921.
Pikaard CS, Mittelsten Scheid O. 2014. Epigenetic regulation in plants. Cold Spring Harb Perspect Biol 6:a019315.

Recker J, Knoll A, Puchta H. 2014. The Arabidopsis thaliana homolog of the helicase RTEL1 plays multiple roles in preserving genome stability. Plant Cell 26:4889-4902.

Rushworth CA, Song BH, Lee CR, Mitchell-Olds T. 2011. Boechera, a model system for ecological genomics. Mol Ecol 20:4843-4857.

Schifflers S, Szöllösi GJ, Mustonen V, Lässig M. 2011. Emergent neutrality in adaptive asexual evolution. Genetics 189:1361-1375.

Schilling MP, Gompert Z, Li FW, Windham MD, Wolf PG. 2018. Admixture, evolution, and variation in reproductive isolation in the Boechera puberula clade. BMC Evol Biol 18:61.

Schmidt A, Schmid MW, Grossniklaus U. 2015. Plant germline formation: common concepts and developmental flexibility in sexual and asexual reproduction. Development 142:229-241.

Schmidt A, Schmid MW, Klostermeier UC, Qi W, Guthörl D, Sailer C, Waller M, Rosenstiel P, Grossniklaus U. 2014. Apomictic and Sexual Germline Development Differ with Respect to Cell Cycle, Transcriptional, Hormonal and Epigenetic Regulation. PLoS Genet 10:e1004476.

Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U. 2011. Transcriptome Analysis of the Arabidopsis Megaspore Mother Cell Uncovers the Importance of RNA Helicases for Plant Germline Development. PLoS Biol 9:e1001155.

Shimizu KK, Ito T, Ishiguro S, Okada K. 2008. MAA3 (MAGATAMA3) Helicase Gene is Required for Female Gametophyte Development and Pollen Tube Guidance in Arabidopsis thaliana. Plant Cell Physiol 49:1478-1483.

Singh M, Goel S, Meeley RB, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D. 2011. Production of Viable Gametes without Meiosis in Maize Deficient for an ARGONAUTE Protein. Plant Cell 23:443-458.

Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, Griffith M. 2016. GenVisR: Genomic Visualizations in R. Bioinformatics 32:3012-3014.

Sloan KE, Bohnsack MT. 2018. Unravelling the Mechanisms of RNA Helicase Regulation. Trends Biochem Sci 43:237-250.

Spillane C, Curtis MD, Grossniklaus U. 2004. Apomixis technology development-virgin births in farmers' fields? Nat Biotechnol 22:687-691.

Srivastava A, Lu Y, Zinta G, Lang Z, Zhu J-K. 2017. UTR-Dependent Control of Gene Expression in Plants. Bioinformatics 30:1312-1313.

Stelzer C-P. 2015. Does the avoidance of sexual costs increase fitness in asexual invaders? Proc Nat Acad Sci U S A 112:8851-8858.

Struhl K, Segal E. 2013. Determinants of nucleosome positioning. Nat Struct Mol Biol 20:267-273.

Tang L, Nogales E, Ciferri C. 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Prog Biophys Mol Biol 102:122-128.

Tekleyohans DG, Nakel T, Gross-Hardt R. 2017. Patterning the Female Gametophyte of Flowering Plants. Plant Physiol 173:122-129.

Vain P, Thole V, Worland B, Opanowicz M, Bush MS, Doonan JH. 2011. A T-DNA mutation in the RNA helicase eIF4A confers a dose-dependent dwarfing phenotype in Brachypodium distachyon. Plant J 66:929-940.

Wagner AH, Kunisaki J, Campbell KM, Lesurf R, Skidmore ZL, Griffith OL, Griffith M. 2016. GenVisR: Genomic Visualizations in R. Bioinformatics 32:3012-3014.

Wang ZP, Xing HL, Dong L, Zhang HY, Han CY, Wang XC, Chen QJ. 2015. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. Genome Biol 16:144.

Warren WC, García-Pérez R, Xu S, Lampert KP, Chalopin D, Stöck M, Loewe L, Lu Y, Kuderna L, Minx P, et al. 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. Nature ecology & evolution 2:669-679.

http://mc.manuscriptcentral.com/gbe
Windham A-Sa. 2010. Flora of North America Editorial Committee. Editorial Committee, eds. 1993+. New York and Oxford.
Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U. 2010. Arabidopsis Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes. Curr Biol 20:506-512.
Xu R, Zhang S, Huang J, Zheng C. 2013. Genome-Wide Comparative In Silico Analysis of the RNA Helicase Gene Family in Zea mays and Glycine max: A Comparison with Arabidopsis and Oryza sativa. PLoS ONE 8:e78982.
Yajima M, Wessel GM. 2011. The multiple hats of Vasa function and its regulation of cell cycle progression. Mol Reprod Dev 78:861-867.
Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24:1586-1591.
Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496-503.
Zhou LZ, Juranic M, Dresselhaus T. 2017. Germline Development and Fertilization Mechanisms in Maize. Mol Plant 10:389-401.
Zhou M, Palanca AMS, Law JA. 2018. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. Nature Genet 50:865-873.

Authors contributions

AS conceived the project. MK and AS planned and conducted the bioinformatics data analysis.
AS, DI, BN, CV, and AL planned and performed the experimental work. AS, MK, BN, and DI analyzed the data. AS wrote the manuscript. All authors approved the manuscript.

Figures and Tables

Figure 1: Sliding window approach to visualize the usage of nucleobases per position in genomic regions of selected RNA helicase genes. For VASL, MEM, and MAA3 the number of different nucleobases comprised in the allelic variants was plotted for all variants of apomictic (blue), facultative apomictic (green), sexual accessions (red) or all accessions (black) with a binning of 12 bases per datapoint. Orange boxes below symbolize exons, blue boxes UTRs and grey bars non-coding regions. For all genes, a higher sequence variability was observed in non-coding regions as compared to exonic regions or UTRs.
Figure 2: Box and whiskers plot depicting the number of allelic variants per accession for all genes. Boxes indicate the lower and upper quartile of numbers of unique cds variants. Horizontal lines in each box indicate the medium numbers of variants observed. Accession numbers are given below.

Figure 3: Visualization of mutation density per accession and most detrimental mutations per gene and accession as compared to A. lyrata as reference. (A) Plot of the abundance of synonymous and non-synonymous SNP per MB of cds sequences. (B) Visualization of the most deleterious mutation per haplotype and accession with a mutation hierarchy from deleterious to harmless from Nonesense_Mutation leading to a premature stop codon, Frame_Shift_Deletion, Frame_Shift_Insertion, In_Frame_Del and In_Frame_Ins leading to a deletion or insertion without causing a frame shift, and Nonstop_mutation causing a non-synonymous SNP. Rows in the plot represent different genes, while the 147 allelic variants identified in the study per gene are presented in columns. Boxes between (A) and (B) indicate reproductive mode and ploidy. The plot has been done using the waterfall function from GenVisR (Skidmore, et al. 2016).

Figure 4: Plot of mutation density along the cds of selected RNA helicase genes. For selected RNA helicases mutations as compared to A. lyrata used as reference were plotted along the cds in windows of 12 bp. Types of mutations depicted are silent mutations, nonstop mutations (non-synonymous SNPs), in frame INDELs, nonsense mutations introducing a stop codon, and frame shift INDELs. Depicted are the conserved proteins EIF4A1 and VASL harboring only silent or nonstop mutations in any accession, MEM and the undescribed helicase AT3G16840 accumulating mutations mainly in the N- and C-terminal regions, and RecQ4A and FAS4 accumulating all types of mutations along the cds including conserved domains.

Figure 5: Reproductive phenotypes in lines carrying mutant alleles of VASL in A. thaliana. Seed clearing and DIC microscopy was used to study reproductive development in wild-type and lines carrying mutant alleles. (A) Functional megaspore (FMS) in the wild-type. (B-D) Additional
or ectopic gametophytic cells in \textit{vasl\_1/VASL} (B, D) and \textit{vasl\_3/VASL} (C). (E) Wild-type like gametophyte after the first mitotic division harboring two nuclei in \textit{vasl\_3/VASL}. (F) Ectopic gametophyte in \textit{vasl\_1/VASL}. (G-I) Slightly protruding (I) or protruding (G, H) gametophytes in \textit{vasl\_2/vasl\_2} (G), \textit{vasl\_1/VASL} (H), and \textit{vasl\_3/VASL} (I), respectively. (ii) inner integuments; black stars denote FMS; dark grey stars ectopic cells, ectopic or protruding gametophytes. Scale bars are 20 µm.

\textbf{Table 1:} Summary of \textit{Boechera} accessions used, ploidy levels, number of seedlings pooled for DNA extraction, and read counts obtained.

\textbf{Supplemental Figures and Tables}

\textbf{Figure S1: Heatmaps of expression of candidate genes.} Heatmap of log scale expression values of the 59 candidate genes as analyzed by Arabidopsis GeneCHIPs as used in Kotliński, et al. 2017 (A). Heatmap of log2-scale expression values from normalized read counts as previously described (Schmidt, et al. 2014) (B). Hierachical clustering of genes and samples was based on agglomerative clustering and euclidean distance. Colors are scaled by row. Red denotes high and black low expression.

\textbf{Figure S2: Molecular phylogenetic analysis.} The evolutionary history of protein sequences of VASL homologs from \textit{B. stricta}, \textit{A. thaliana} and \textit{A. lyrata} and 15 related RNA helicases was inferred using the Maximum Likelihood method implemented in RAxML (Stamatakis 2014). The tree with the highest log likelihood is shown. Bootstrap values for clades are shown next to internal nodes. Closest relationships are supported for homologs from all three species.

\textbf{Figure S3: Plot of usage of nucleobases per position in genomic regions of target genes.} Allelic variants were plotted for all variants of apomictic (blue), facultative apomictic (green), sexual accessions (red) or all accessions (black) with a resolution of 12 bases averaged per position.
Coding and non-coding regions are visualized below with blue boxes representing cds or target genes. Coding regions of other than the target genes in the covered regions are not indicated.

**Figure S4: Number of allelic variants per gene and accession.** Genes were sorted by an increasing number of unique sequence variants per accession and the number of sequence variants observed is shown separately for sexual accessions and B12-663 (A), and facultative apomictic and apomictic accessions (B).

**Figure S5: dN/dS-ratios of allelic variants from sexual versus apomictic accessions.** Box and whiskers plot of ω-values that were estimated using the program yn00 implemented in PAML under the model LWL85m (Li, et al. 1985; Yang 2007). Boxes indicate the lower and upper quartile of values, horizontal lines in each box indicate the medium numbers. Pairwise student’s t-tests were performed in R to identify significant differences between sexual and apomictic accessions marked with * (p < 0.05).

**Figure S6: dN/dS-ratios of allelic variants from diploid and triploid apomicts for selected genes.** Box and whiskers plot summarizing ω-values that were estimated using the program yn00 implemented in PAML under the model LWL85m (Li, et al. 1985; Yang 2007). Differences in the variance of dN/dS-ratios between diploid and triploid accessions per gene were tested by Fishers F-test performed in R. Orange depicts diploid, red triploid accessions.

**Figure S7: Mutation density along the cds of selected genes in sexual or apomictic accessions.** Different types of mutations as compared to the A. lyrata reference were plotted along the cds of the homologues of MEM and MAA3 separately from all sequence variants of obligate apomictic accessions or sexual accessions at a resolution of 12 bp.

**Figure S8: Box and whiskers plot of mutation load of the homologs of selected A. thaliana RNA helicases.** Numbers of non-synonymous SNPs were normalized by protein length and plotted per gene and mode of reproduction (sexual, facultative apomictic, apomictic). Significant

http://mc.manuscriptcentral.com/gbe
differences indicated by * (p < 0.05) between mutation loads of alleles from sexual as compared to facultative apomictic or obligate apomictic accessions were identified by pairwise student’s t-test.

**Figure S9:** Seed and embryo development in lines carrying a mutant VASL allele. (A) Globular embryo in the wild-type. (B-D) Slight over-proliferation of the embryo at about globular stage of development in vasl_1/vasl_1 (B), vasl_2/vasl_2 (C), vasl_3/VASL (D). Heart stage embryo and over-proliferated embryo in vasl_4/VASL (E). Over-proliferated embryo in vasl_1/vasl_1, and embryo in vasl1/VASL probably developmentally delayed in contrast to the endosperm. emb; embryo, scale bars are 50 µm.

**Table S1: Gene loci included in this study.** Given are gene loci from *A. thaliana* including annotations according to TAIR (Araport11 genome release; www.arabidopsis.org) and homologues gene loci from *B. stricta* (Phytozome 12, https://phytozome.jgi.doe.gov). For selected genes lacking annotations homology modelling was performed based on *B. stricta* protein sequences as indicated using Phyre2 (Kelley, et al. 2015).

**Table S2: Sequences of probes used for target pulldown.** Probes were designed to either match die genomic regions of target genes in the *B. stricta* reference genome or the coding sequences from the reference transcriptome of *B. gunnisoniana* (Schmidt, et al. 2014; Lee, et al. 2017). After moderate filtering of non-unique probes, repeat masking, and collapsing of redundant sequences 18 941 probes were retained and synthesized. Locus identifiers and sequences are provided as fasta-file.

**Table S3: Frequencies in % of phenotypes observed in plants carrying mutant alleles of VASL.** (A) During early gametophyte development. (B) During seed development ranging from globular to torpedo stage.
Figure 1: Sliding window approach to visualize the usage of nucleobases per position in genomic regions of selected RNA helicase genes. For VASL, MEM, and MAA3 the number of different nucleobases comprised in the allelic variants was plotted for all variants of apomictic (blue), facultative apomictic (green), sexual accessions (red) or all accessions (black) with a binning of 12 bases per datapoint. Orange boxes below symbolize exons, blue boxes UTRs and grey bars non-coding regions. For all genes, a higher sequence variability was observed in non-coding regions as compared to exonic regions or UTRs.
Figure 2: Box and whiskers plot depicting the number of allelic variants per accession for all genes. Boxes indicate the lower and upper quartile of numbers of unique cds variants. Horizontal lines in each box indicate the medium numbers of variants observed. Accession numbers are given below.
Figure 3: Visualization of mutation density per accession and most detrimental mutations per gene and accession as compared to A. lyrata as reference. (A) Plot of the abundance of synonymous and non-synonymous SNP per MB of cds sequences. (B) Visualization of the most deleterious mutation per haplotype and accession with a mutation hierarchy from deleterious to harmless from Nonesence_Mutation leading to a premature stop codon, Frame_Shift_Deletion, Frame_Shift_Insertion, In_Frame_Del and In_Frame_Ins leading to a deletion or insertion without causing a frame shift, and Nonstop_mutation causing a non-synonymous SNP. Rows in the plot represent different genes, while the 147 allelic variants identified in the study per gene are presented in columns. Boxes between (A) and (B) indicate reproductive mode and ploidy and accession numbers are indicated accordingly. The plot has been done using the waterfall function from GenVisR (Skidmore, et al. 2016).
Figure 4: Plot of mutation density along the cds of selected RNA helicase genes. For selected RNA helicases mutations as compared to A. lyrata used as reference were plotted along the cds in windows of 12 bp. Types of mutations depicted are silent mutations, nonstop mutations (non-synonymous SNPs), in frame INDELs, nonsense mutations introducing a stop codon, and frame shift INDELs. Depicted are the conserved proteins EIF4A1 and VASL harboring only silent or nonstop mutations in any accession, MEM (AT5G39840) and the undescribed helicase AT3G16840 accumulating mutations mainly in the N- and C-terminal regions, and RecQ4A and FAS4 accumulating all types of mutations along the cds including conserved domains.

214x295mm (300 x 300 DPI)
Figure 5: Reproductive phenotypes caused of lines carrying mutant alleles of VASL in A. thaliana.

Seed clearing and DIC microscopy was used to study reproductive development in wild-type and lines carrying mutant alleles. (A) Functional megaspore (FMS) in the wild-type. (B-D) Additional or ectopic gametophytic cells in vasl_1/VASL (B, D) and vasl_3/VASL (C). (E) Wild-type like gametophyte after the first mitotic division harboring two nuclei in vasl_3/VASL. (F) Ectopic gametophyte in vasl_1/VASL. (G-I) Slightly protruding (I) or protruding (G, H) gametophytes in vasl_2/vasl_2 (G), vasl_1/VASL (H), and vasl_3/VASL (I), respectively. (ii) inner integuments; black stars denote FMS; dark grey stars ectopic cells, ectopic or protruding gametophytes. Scale bars are 20 µm.
Table 1: Summary of Boechera accessions used, ploidy levels, number of seedlings used, and read counts obtained.

| genus          | species                  | ploidy | accession | reproductive mode | no of seedlings used | raw reads paired | paired readcounts after trimming | publication of accession               |
|----------------|--------------------------|--------|-----------|-------------------|---------------------|------------------|-------------------------------|----------------------------------------|
| Boechera       | divaricarpa              | 2x     | ES517     | facultative       | 3                   | 9135540          | 7280704                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | retrofracta x stricta    | 2x     | CO11010   | apomictic         | 2                   | 5600566          | 3763336                       | Carman et al., 2019                   |
| Boechera       | retrofracta              | 2x     | B12-1131  | apomictic         | 3                   | 8803860          | 5936003                       | Mau et al., 2015                     |
| Boechera       | williamsii               | 2x     | B12-1524  | apomictic         | 3                   | 7660767          | 4708395                       | Mau et al., 2015                     |
| Boechera       | pallidifolia             | 2x     | B12-1578  | apomictic         | 3                   | 10065290         | 8650998                       | Mau et al., 2015                     |
| Boechera       | pallidifolia             | 3x     | B12-1599  | apomictic         | 3                   | 9739570          | 7210650                       | Mau et al., 2015                     |
| Boechera       | spatifolia               | 2x     | B12-956   | apomictic         | 3                   | 9460515          | 7965700                       | Mau et al., 2015                     |
| Boechera       | stricta x retrofracta    | 4x     | B12-2582  | apomictic         | 2                   | 9979526          | 8602569                       | Mau et al., 2015                     |
| Boechera       | divaricarpa              | 2x     | ES524     | facultative       | 3                   | 8394949          | 6744721                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | divaricarpa              | 3x     | ES598     | apomictic         | 3                   | 9442364          | 6404632                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | divaricarpa              | 3x     | ES704     | apomictic         | 3                   | 9411828          | 6550887                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | holboellii               | 2x     | ES911     | facultative       | 2                   | 9840426          | 6007448                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | holboellii               | 2x     | ES805     | facultative       | 3                   | 9686119          | 8430991                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | stricta                  | 2x     | B12-693   | facultative       | 3                   | 10100186         | 7702794                       | Mau et al., 2015                     |
| Boechera       | pallidifolia             | 2x     | B12-1591  | facultative       | 3                   | 10446466         | 8772196                       | Mau et al., 2015                     |
| Boechera       | stricta                  | 2x     | B12-663   | sexual           | 2                   | 10041301         | 8626705                       | Mau et al., 2015                     |
| Boechera       | stricta                  | 2x     | B12-1268  | sexual           | 3                   | 8809041          | 7094779                       | Mau et al., 2015                     |
| Boechera       | crandalli                | 2x     | B12-1397  | sexual           | 3                   | 10464956         | 7190352                       | Mau et al., 2015                     |
| Boechera       | crandalli                | 2x     | ES727     | sexual           | 3                   | 11340638         | 7882158                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | leisiocarpa              | 2x     | B12-357   | sexual           | 3                   | 9516505          | 5840745                       | Mau et al., 2015                     |
| Boechera       | williamsii               | 2x     | B12-558   | sexual           | 3                   | 9441753          | 8116152                       | Mau et al., 2015                     |
| Boechera       | holboellii               | 2x     | ES786     | sexual           | 2                   | 5259327          | 3966639                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | holboellii               | 2x     | ES820     | sexual           | 3                   | 9699227          | 8126082                       | Aliyu et al., 2010; Mau et al., 2015   |
| Boechera       | stricta                  | 2x     | LTM       | sexual           | 3                   | 16073343         | 13825485                      | Lee et al., 2017                     |

http://mc.manuscriptcentral.com/gbe