MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 5908–5921 doi: 10.1111/mec.12533

Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation

MARCO PELLINO,* DIEGO HOJSGAARD,† THOMAS SCHMUTZER,* UWE SCHOLZ,* ELVIRA HÖRANDL,† HEIKO VOGEL‡ and TIMOTHY F. SHARBEL*

*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben, Germany, †Department of Systematic Botany, Albrecht-von-Haller Institute for Plant Sciences, Georg-August-University of Goettingen, Untere Karspuele 2, D-37073 Goettingen, Germany, ‡Department of Entomology, Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany

Abstract

Asexual lineages are thought to be prone to extinction because of deleterious mutation accumulation (Muller's ratchet). Here, we analyse genomic effects of hybridity, polyploidy and allelic divergence in apomictic plants, and identify loci under divergent selection among sexual/apomictic lineages. RNAseq was used to sequence the flower-specific transcriptomes of five genotypes of the *Ranunculus auricomus* complex, representing three sexual and two apomictic reproductive biotypes. The five sequence libraries were pooled and *de novo* assembly performed, and the resultant assembly was used as a backbone for a subsequent alignment of each separate library. High-quality single-nucleotide (SNP) and insertion–deletion (indel) polymorphisms were mined from each library. Annotated genes for which open reading frames (ORF) could be determined were analysed for signatures of divergent versus stabilizing selection. A comparison between all genotypes supports the hypothesis of Pleistocene hybrid origin of both apomictic genotypes from *R. carpaticola* and *R. cassubicifolius*, with subsequent allelic divergence of apomictic lineages (Meselson effect). Pairwise comparisons of nonsynonymous (dN) to synonymous (dS) substitution rate ratios between apomictic and sexual genotypes for 1231 genes demonstrated similar distributions for all comparisons, although 324 genes demonstrated outlier (i.e. elevated) dN/dS ratios. Gene ontology analyses of these outliers revealed significant enrichment of genes associated with reproduction including meiosis and gametogenesis, following predictions of divergent selection between sexual and apomictic reproduction, although no significant signal of genome-wide mutation accumulation could be identified. The results suggest that gene function should be considered in order to understand effects of mutation accumulation in asexual lineages.

Keywords: apomixis, hybridization, Meselson effect, Muller's ratchet, polyploid, *Ranunculus*, single-nucleotide polymorphism

Received 22 May 2012; revision received 21 August 2013; accepted 26 August 2013

Introduction

Asexual reproduction has evolved repeatedly and independently from sexual ancestors in many species of animals and plants (Suomalainen 1950; Mittwoch 1978; Barton & Charlesworth 1998). A number of factors are hypothesized to make asexuality advantageous, including avoidance of the twofold cost of sex (Maynard Smith 1978), and adaptation to environmental stability (Bell 1982). On the other hand, sex can be advantageous in changing environments (Van Valen 1973), for generating genetic variance (Fisher 1930), and for purging deleterious mutations (Muller 1964; Kondrashov 1982). The lack of meiotic recombination and syngamy defines the asexual genome with a number of effects, including

Correspondence: Timothy F. Sharbel, Fax: +049 (0)39482 5137; E-mail: sharbel@ipk-gatersleben.de

decreased probability of fixing advantageous mutations (Crow & Kimura 1965), the accumulation of deleterious mutations (i.e. Muller's ratchet; Felsenstein 1974; Muller 1964) and the loss of meiotic recombinational DNA repair mechanisms (Bernstein *et al.* 1988; Hörandl 2009), all of which play a central role in the evolutionary theory of sex (Maynard Smith 1978; Kondrashov 1993).

Many plants reproduce through apomixis, a reproductive phenomenon that allows the mother to produce clonal progeny via seeds (Nogler 1984). Apomixis is found naturally in more than 400 plant species and comprises various forms (http://www.apomixis.uni-goettingen.de; Hojsgaard *et al.*, submitted). Gametophytic development is characterized by three steps: the production of a meiotically unreduced egg cell (apomeiosis), parthenogenetic development of this egg cell without fertilization and production of a functional endosperm with (pseudogamy) or without (autonomous) fertilization of the binucleate central cell of the ovule (Koltunow & Grossniklaus 2003). While the evolutionary origin and molecular mechanisms behind apomixis remain enigmatic, it is generally accepted that deregulation in the timing of developmental steps characteristic of the sexual reproductive pathway (Koltunow 1993; Grossniklaus *et al.* 2001), the result of global gene regulatory effects associated with polyploidy and/or hybridity (Carman 1997; Grossniklaus 2001), leads to the induction of apomixis. In fact, almost all apomictic plants are polyploid (Koltunow & Grossniklaus 2003) with some exceptions of diploid apomictic hybrids (Kantama *et al.* 2007).

Polyploidy increases the load of deleterious mutations as more potential mutation sites are available; all else being equal, the mean fitness of a population is approximately reduced by deleterious mutations by cU (where c is the ploidy level and U the mutation rate of the haploid genome; Gerstein & Otto 2009). In the short term, polyploidy can buffer negative effects of the mutational load by masking recessive deleterious mutations, with a temporary rise in the fitness of a newly formed polyploid (Otto & Whitton 2000). In the long run, masked mutations persist longer and with higher frequencies in the population before being eliminated by selection (Otto & Whitton 2000). In sexual lineages, the meiosis-mixis cycle further exposes haploid gametes or gametophytes with deleterious mutations and low fitness to purifying selection (Hörandl 2009). In polyploid apomicts, this possibility of purging deleterious recessive mutations is often decreased considering that female gametes are by definition meiotically unreduced, while both meiotically reduced and – unreduced male gametes characterize some populations. Consequently, asexual polyploid lineages are expected to have short-term benefits, but might in the long run be more prone to decreasing fitness and eventual extinction.

When mutations are beneficial and partially to fully dominant, their effects are expected to have more significant relative positive effects in polyploid populations (Otto & Whitton 2000). Asexuality is beneficial in newly formed polyploids and hybrids as it helps to bypass the effects of meiotic irregularities during gamete formation (Comai 2005). In the majority of apomictic plants, the Mendelian factors regulating apomixis are indeed dominant (Ozias-Akins & Van Dijk 2007). On the other hand, highly heterozygous, vigorous hybrid genotypes could have a selective advantage because of heterosis effects (Comai 2005); in this case, the heterozygous hybrid genotype is expected to be favoured by stabilizing selection (Otto 2009). Heterozygosity could even be fixed in sexual allopolyploids by preferential homolog chromosome pairing at meiosis (Comai 2005) rather than being correlated with the shift to asexuality. The expected genomic signature of such stabilizing selection in allopolyploids would be a genome-wide increase in heterozygosity, but not necessarily a rapid evolution in genes associated with apomixis.

All apomictic plants appear scattered on the tips of phylogenies and are thought to be evolutionarily young (Van Dijk & Vijverberg 2005). However, estimating ages of lineages from phylogenies is biased by potential reversals from apomixis to sexuality (Schwander & Crespi 2009; Hörandl & Hojsgaard 2012). For asexual plants, hardly any empirical age estimate is available as reproductive mode cannot be assessed from the fossil record. Ancient asexuality can be inferred from intraindividual genomic divergence of alleles. In the absence of recombination, allele pairs are expected to independently accumulate mutations and diverge from one another; this so-called Meselson effect was described from ancient asexual animals (Mark Welch & Meselson 2000) and has been tested in plants (Corral *et al.* 2009). Alternatively, similar high levels of heterozygosity may rather reflect the hybrid nature of apomicts (Beck *et al.* 2012).

The *Ranunculus auricomus* complex is becoming a model system for a comparative genome evolution between sexual and asexual taxa. It comprises hundreds of apomictic species that are divided into two main subcomplexes that constitute morphologically distinct groups: *R. auricomus* and *R. cassubicus* (Hörandl *et al.* 2009). The 'cassubicus' subcomplex is a morphologically well-defined subgroup (Hörandl 1998; Hörandl *et al.* 2009) of a few sexual diploid and apomictic polyploid cytotypes, from which two sexual species and one apomictic taxon have been particularly studied. *Ranunculus carpaticola* is a diploid sexual species widespread in the Carpathians and central Slovakia, and the closely related sexual, diploid and autotetraploid *R. cassubicifolius* is distributed from eastern Austria to Switzerland (Hörandl *et al.* 2009). Hexaploid apomicts, which likely

originated from ancient interspecific hybridization between diploid *R. carpaticola* and the autotetraploid cytotype of *R. cassubicifolius* from Lower Austria, are present in central Slovakia (Paun *et al.* 2006a). Population genetic marker analyses (AFLPs, SSRs), DNA sequence analysis and morphometric data support the hybrid origin hypothesis (Paun *et al.* 2006a; Hörandl *et al.* 2009). As with many taxa (Richards 2003), apomictic *Ranunculus* are thought to have arisen during the last glacial period (115 000–15 000 years before present; Paun *et al.* 2006a). A more distantly and geographically separated taxon, which is also part of the '*auricomus*' subcomplex, *Ranunculus notabilis*, is a sexual diploid species found in southeastern Austria (Hörandl *et al.* 2009). Genetic distances between populations suggest that *R. notabilis* separated from the *R. cassubicifolius-R. carpaticola* ancestor c. 900 000 years ago, while the latter speciated c. 300 000 years ago (Hörandl, 2004).

Rapid advances in DNA sequencing technology are, by nature of increasing genome coverage per individual, leading to more confident use of single-nucleotide polymorphisms (SNPs) in studies of ecology, conservation and evolution. SNPs have been successfully used to uncover individual (Cork & Purugganan 2005), population (Song *et al.* 2009), and species-specific (Trick *et al.* 2009; Bajgain *et al.* 2011) differences, and have furthermore been employed to identify disease-linked markers (Hampe *et al.* 2007). Next-generation sequencing (NGS) technology has been successfully used for detecting SNPs in maize (Barbazuk *et al.* 2007), *Eucalyptus (*Novaes *et al. 2008), Brassica napus* (Trick *et al.* 2009) and rye (Haseneyer *et al.* 2011), to name but a few.

SNP analysis of annotated genes provides insights into functional sequence diversity and putative divergent selection (Bajgain *et al.* 2011). For example, a high ratio of nonsynonymous vs. synonymous substitutions could suggest that specific genes are under divergent selection. While sexual species in allopatry are expected to diversify slowly, independently evolving asexual lineages are expected to rapidly diverge, the result of little to no recombination and syngamy, followed by the accumulation of divergent alleles (Lynch & Conery 2000; Mark Welch & Meselson 2000; Van Dijk 2003; Corral *et al.* 2009). Polyploidy and hybridity would further leave genomic signatures in asexual organisms as a higher number of possible mutational sites are available, while the shift to apomictic reproduction could have pleiotropic effects on functional traits and genes connected to apomixis.

Here, we describe high-quality SNP discovery based upon RNAseq data collected from 5 sexual and apomictic individuals of four taxa of the *Ranunculus auricomus* complex. The goals of this project were to: (i) identify high-quality SNPs using strict quality and coverage parameters, to (ii) compare mutation accumulation (i.e. Muller's ratchet), divergence and the genomic signature of hybridization between sexual and apomictic forms, and (iii) to identify gene classes under divergent selection between sexual species, apomictic lineages and apomictic versus sexual taxa.

Materials and methods

RNA collection and extraction

Flowers were collected from 5 different genotypes of the *Ranunculus auricomus* complex (Table 1). For each plant, total RNA was isolated from 5 different harvested flower sizes using the Qiagen RNeasy Plant Mini Kit (QIAGENE, www.qiagene.com) following the manufacturer's instructions. The isolated RNA was treated with Qiagen RNase-Free DNase according to the producer's protocol in order to eliminate any contaminating trace of DNA. A second purification step was performed using a Qiagen RNeasy Mini Kit to eliminate contaminating polysaccharides, proteins and the DNase enzyme. The final concentration and quality were checked using an Agilent Technologies 2100 Bioanalyzer NanoChip (Agilent Technologies, - Santa Clara, CA, United States).

RNA normalization

In order to avoid over-representation of the most commonly transcribed genes, full-length enriched, normalized cDNA libraries were generated using a combination of a mint-universal cDNA synthesis kit (Evrogen, Moscow, Russia) and the Trimmer Direct cDNA normalization kit (Evrogen, http://www.evrogen.com), which utilizes the Kamchatka crab duplex-specific nuclease method (Zhulidov *et al.* 2004; Shcheglov *et al.* 2007). The procedure generally followed the manufacturer's protocol, but included several important modifications, as described by Vogel *et al.* (2010).

Each step of the normalization procedure was carefully monitored to avoid the generation of artefacts and overcycling. The optimal condition for ds-cDNA synthesis was empirically determined by subjecting the cDNA to a range of thermocycle numbers and their products checked by electrophoresis. The optimal cycle number was defined as the maximum number of PCR cycles without any signs of overcycling. Optimization of the complete cDNA normalization procedure was essentially performed as described in Vogel & Wheat (2011).

Sequencing, assembly and annotation

RNAseq, using 15 μl of normalized cDNA (200 ng/μl), was outsourced to Fasteris (www.fasteris.com). As the

Table 1 *Ranunculus* samples used in SNP analysis

Taxon	Sample	Ploidy*	Reprod	Locality	Collectors and vouchers
R. carpaticola	REV1	2x	Sex	Slovakia, Slovenské rudohorie,Revúca, hill Skalka (forest)	Hörandl, 8483, 01.05. 1998 (WU)
R. carpaticola x cassubicifolius	TRE	6x	Apo	Slovakia, Strázovské vrchy (near Trençín), between Kubra and Kubrica, close to the bus-stop Kyselka (margin of Carpinus forest and meadow)	Hörandl, Paun, Mládenková, s.n., 30.04. 2004 (SAV)
R. carpaticola x cassubicifolius	VRU 2	6x	Apo	Slovakia, Turçianska kotlina,Vrútky-Piatrová, behind cottage (meadow)	Hörandl, Paun, Mládenková, s.n.,01.05. 2004 (SAV)
R. cassubicifolius	YBB 1	4x	Sex	Austria, Lower Austria, Wülfachgraben, SE Ybbsitz (forest)	Hörandl, 8472 12.04. 1998 (WU)
R. notabilis	NOT1	2x	Sex	Austria, Burgenland, Strem valley on 1,5 km ENE Strem, 220 m, wet meadow and forest margin	Hörandl 5612 + 7220, 0.04.1996(WU); AY680033; AY954115d; FJ619873

*see Hörandl *et al*., 2000; Hörandl & Greilhuber, 2002; Paun *et al*. 2006a,b.

genome of *Ranunculus* is uncharacterized, and furthermore, as no genomic information was available from any other closely related species, a dual-sequencing approach was taken in order to balance costs and the ability to *de novo* assemble the sequencing data. Thus, a single *R. carpaticola* was sequenced using a 108-mer paired-end (PE) approach, while the remaining four samples (*R. cassubicifolius*, *R. notabilis* and the 2 hybrid apomicts) were analysed by 54-mer single-end (SE) sequencing, both using the HiSeq™ 2000 Sequencing System from Illumina® (http://www.illumina.com/).

CLC genomic workbench (CLC bio version. 4.9, www.clcbio.com) was used for sequence assembly. At first, the libraries were trimmed for vector contamination, length and quality score, and two methods were attempted to create the best reference assembly from which SNPs would be inferred. In a first assembly (pooled *de novo*) approach, all libraries were pooled and assembled using the following CLC parameters: nucleotide mismatch cost = 2, insertion-deletion costs = 3; length fraction = 0.5, similarity = 0.9 and any conflicts among the individual bases were resolved by voting for the base with highest frequency, while contigs shorter than 300 bp were removed from the final analysis. A second (iterated *de novo*) approach performed a *de novo* assembly on each species (i.e. library) separately using the same parameters, followed by a *de novo* assembly of all contigs from each individual libraries together, including the reads that were initially discarded in each library. Once again, contigs smaller than 300 bp were removed from the data set. The different assemblies were evaluated using the number of matched reads and N50 values. The final contig set was annotated by searching for similarities against the complete protein database of UniProt (UniProt Release 2012_08), using the 2.2.18 version of Blastx (ftp://ftp.ncbi.nih.gov/blast/executables/). The resultant homology data were filtered based upon E-value, and only hits having $E \leq 1^{-10}$ were kept.

SNP calling

The transcriptome reads were processed in a long-read alignment to the chosen reference assembly (see above) using BWA (version: 0.5.9-r16, Li & Durbin 2010). For *R. carpaticola*, reads were mapped in paired-end mode (PE), while the remaining genotypes were mapped in single-end mode (SE). BWA mapping was performed using the default settings for Illumina sequence reads, except setting the minimal base quality for read trimming to 20 (see Tables S1 and S2, Supporting information). Read trimming was performed down to a minimal required read length of 35 bp if necessary. These conservative default settings consider only two nucleotide differences in the seeding process of the initial read placement. Gaps were penalized more strictly in comparison with mismatches. Within the mapping process, insertions or deletions (indels) with less than 5-bp distance to the contig end were discarded (see BWA documentation).

Putative single-nucleotide polymorphisms (SNPs) and indels were called using Samtools (version 0.1.18 r982; Li *et al*. 2009). The raw set of potential variations was generated using VCFtools (version 0.1.17; Danecek *et al*. 2011). Artificial variations caused by unambiguous bases in the reference (inserted Ns) were removed, and remaining variations were filtered using the vcfutils module of VCFtools.

SNP mining was performed for each of the genotypes individually. To access relevant information on SNP calling, we developed a Perl script to combine filtered SNPs from different genotypes to identify shared versus

genotype-specific SNPs. Variation calling was run under default settings and considered the following criteria: high-quality bases at SNP position, minimal read mapping quality (10), the probabilities for end distance bias (0.000) and biases in base quality (1e-100; Li *et al.* 2012). To exclude SNP calls located in collapsed regions, we applied adjustable read depth settings and discarded positions with extreme read depth. The tablet alignment viewer (Milne *et al.* 2010) was used for graphical visualization and inspection of putative SNP calls. Developed Perl scripts are freely available by the authors, but should be considered as in developmental stage.

Open reading frame and synonymous/nonsynonymous mutation analysis

In order to identify synonymous and nonsynonymous mutations, SNP variation was assigned to codon position based upon open reading frame (ORF) analysis using contigs in which at least one position had a high-quality polymorphism across all the lineages. The sequences were analysed to find the starting codon and the ORF using the module GetOrf from the EMBOSS package (Version 1.5 Rice *et al.* 2000). When two ORFs were found overlapping, a portion of the same contig, or in the case of an ORF located inside a longer one, and then the longest ORF were selected. In case of two similarly long (within 90% of each other) ORFs located in opposite strands of the contigs, the polarity of the ORFs was then compared with the strand orientation from the Blastx analysis and the appropriate one selected.

The proportion of dN/dS for all genotypes was pairwise-calculated for all resultant annotated contigs where ORFs could be assigned using the Bio::Align::DNAStatistic BioPerl module (http://www.bioperl.org). In order to correct for cases with ratios having 0 in the denominator or in the numerator, a value of 1 was added to both synonymous and nonsynonymous data (dN and dS) before calculating the ratios, as already described by Bajgain *et al.* 2011 and Novaes *et al.* 2008. Boxplots of the pairwise distributions of dN/dS between sexual, sexual and apomictic, and apomictic genotypes were generated using SPSS (Version 21.0: http://www-01.ibm.com/software/analytics/spss/products/statistics/), the program marks as outliers values that are equal or higher than one and a half box length, defined as the distance between the first and the third quartile (IQR). Lower outliers are those values that are equal or lower than one and a half IQR.The genes characterized by outlier values were selected for gene ontology (GO) analysis. Gene ontology analyses were performed using Blast2GO (Conesa *et al.* 2005), using the default annotation parameters of the program. Enrichment analyses were performed using a Fisher's exact test with FDR (false discovery rate) correction, and the complete set of annotated genes from *Ranunculus* (see above) was used as a reference data set. A second GO investigation was performed with the web-based program AgriGO (Du *et al.* 2010) using *Arabidopsis* homologous genes and the default annotation parameters of the program. ORFs were blasted against the *Arabidopsis* database (TAIR: http://www.arabidopsis.org) and filtered with a BlastX (E-value cut-off of $E \leq 1^{-5}$). A Fisher's exact test with FDR correction was used for defining enriched GO terms, and the AgriGo results were filtered for *P* value ($E \leq 1^{-5)}$ and FDR ($E \leq 1^{-5)}$.

Hybrid origin

Previous hypothesis that the apomictic accessions would be hybrid derivatives from 2× *R. carpaticola* and 4× *R. cassubicifolius* were based upon extensive AFLP analyses of 450 individuals from the geographical distribution area (Paun *et al.* 2006a). However, as AFLPs are dominant markers with unknown genetic background, this data set could not provide information regarding allelic composition and its divergence in apomictic lineages, and hence neither on the relative evolutionary ages of these apomicts. Here, we wanted to (i) date the origin of hybrids and (ii) test for single vs. multiple origins and posthybrid divergence of apomictic lineages. Based on this phylogenetic framework, we were able to test whether the SNP data set can discriminate between heterozygosity derived from hybrid origin and that resulting from putative mutations accumulated after the hybrid origin of each apomictic lineage (see under *g* below). A phylogenomic analysis was performed on the SNP data set whereby the presence/absence of SNPs in parents and the 6× hybrids bears information on evolutionary relationships and genomic evolution (Bajgain *et al.* 2011). The original SNP data set was converted into a presence/absence matrix of SNPs, considering shared presence only (i.e. SNPs appearing in at least two taxa and differing in at least two taxa). Because our data set did not provide detailed information on allelic ratios in the different cytotypes, a direct analysis of heterozygosity in hybrids was not feasible. To compare normal SNPs (e.g. A) called from our assembled reference transcriptome with those heterozygous ones (e.g. A/T) at the same site - which likely reflect different alleles in polyploids; Buggs *et al.* 2010; Trick *et al.* 2012), the latter ones were coded with the standard IUPAC code for ambiguous sites (Bajgain *et al.* 2011). Indel polymorphisms greater than 2 bp were coded as a single phylogenetic character (i.e. equivalent to one site), based on the assumption that these polymorphisms

would represent single evolutionary events (Simmons & Ochoterena 2000). All other SNPs with low coverage were coded as gaps in the data matrix and treated as missing data in the phylogenetic analysis.

The expected conflict of phylogenetic signals in hybrids was resolved by using split network methods (Huson *et al.* 2010) as implemented in Splitstree 4.0 (Huson & Bryant 2006). By using uncorrected P-distances, we applied a distance-based method of analysis to test for the expected reticulate data structure. A neighbour net analysis was done, shown as a hybridization network and rooted with the more distantly related *R. notabilis* as out-group taxon (see phylogeny by Hörandl *et al.* 2009). Neighbour net calculates the support for 'splits' (relationships) from genetic distances and displays these splits in a graph (i.e. a 'splits graph' or 'split network'). The algorithm determines a circular ordering of taxa (i.e. based on the extent of differences between the SNP data, the taxa are ordered around a circle). The layout on the circle determines what splits occur in the data and is displayed in a planar graph. The support for each of these splits is then measured using a least squares method that adjusts the lengths of the splits in the splits graph so as to minimize the difference with the pairwise distances in the original data matrix (Huson & Bryant 2006).

Allelic sequence divergence in apomicts

To test the hypothesis that alleles at a single locus in parthenogenetic lineages are expected to accumulate mutations and diverge from each other in the absence of meiosis (Mark Welch & Meselson 2000), intraindividual allelic sequence divergence was measured in apomictic lineages after their hybrid origin. We selected a sample including 10000 annotated SNPs for *R. carpaticola*, *R. cassubicifolius* and the hybrid *R. carpaticola* × *R. cassubicifolius*. In order to check for the presence of intraindividual allelic sequence divergence, we looked first for heterozygous SNPs (e.g. A/T at one site) per sequence/contig that were exclusive to either one of the two hexaploid apomictic lineages (see also Trick *et al.* 2012). Such SNPs could reflect mutation accumulation between homologous chromosomes that were subsequently maintained through a lack of homogenizing mechanisms (i.e. homologous pairing and recombinational DNA repair during meiosis). In contrast, heterozygous sites derived from hybrid origin would be located on homologous chromosomes, which usually do not pair and recombine during meiosis (Comai 2005), and would be shared both between allopolyploid apomicts and in either of their parents (see also Trick *et al.* 2012). These latter heterozygous sites were not considered for calculating the age of apomictic genotypes.

Finally, as the number of sequence changes should be proportional to the time back to the hybridization event, we estimated the age of the apomictic genotypes by the number of generations needed to obtain the observed values. A theoretical number of expected mutations/neutral substitutions per generation was calculated using: (i) a standard substitution rate for plant nuclear genes (u = 5e-9; Wolfe *et al.* 1989), (ii) the total number of possible mutational sites in our sample (n = 10000) and (ii) a multiplying factor of three (hexaploids had three times the number of possible sites). The generation time in *Ranunculus auricomus* was considered to be three years (from seedling to flowering and reproduction).

Results

Sequencing, assembly and high-quality SNP calling

Five normalized cDNA libraries were synthesized from RNA extracted from pooled flower buds at five different stages. The one paired-end (PE) and four single-end (SE) RNAseq runs generated a total of 221 744 460 reads, corresponding to a total of 19.5 Gbp of sequence. In detail, the 54-bp SE sequencing run of two apomictic genotypes (TRE and VRU2) and two sexual genotypes (YBB and NOT) generated from 19 to 22 million reads per library, while the 108-mer PE sequencing run of one sexual genotype (REV) generated ca. 138 million reads (Table 2).

The five libraries were screened for adaptor contamination, sequence quality and minimum length, resulting in 30% of the PE library and an average of 60% for the four SE libraries being trimmed. The pooled *de novo* assembly yielded 62 102 contigs greater than 300 bp in length and incorporated 91% of all initial reads. The final library had N50 = 1064 bp, with the largest contig = 11 942 bp (Fig. 1), while the iterated *de novo* assembly produced 72 915 contigs with N50 = 1039. While the PE carpaticola de novo assembling incorporated 89% of the PE reads, the 4 SE libraries result in 50% of unassembled reads. Hence, the pooled *de novo* assembly was kept for subsequent analyses.

In total 19 977 contigs could be annotated, with the maximum number of homologies found with *Arabidopsis thaliana* (56%,), *Oryza* (7.6%) and *Nicotiana* (2%). The percentage identity ranged from 20% to 100% (see Fig S1, Supporting information). A total of 447 428 high-quality SNPs and 18 732 indels were mined from the libraries (Table 3). Of these, 73% were detected in the PE *R. carpaticola* (REV1) library, due to the significantly higher number of reads for this sample. The greater number of polymorphisms in the *R. carpaticola* library did not represent a bias as only shared polymorphisms

Table 2 Raw sequencing results from RNAseq runs on five *Ranunculus* genotypes

	Genotype				
	REV1	TRE(Apo)	VRU 2(Apo)	YBB1	NOT
Sequencing technology	Illumina PE	Illumina SE	Illumina SE	Illumina SE	Illumina SE
Number of reads	137 304 156	22 203 280	21 330 140	20 934 596	19 972 288
Number of reads after quality filtering	96 855 485	8 799 138	8 958 973	8 218 588	8 056 033
Average read length	108	54	54	54	54
Number of matched reads	120 154 734 (91.79%)				

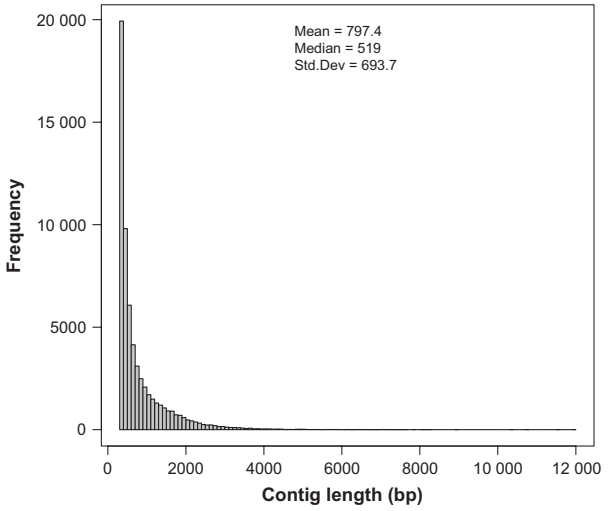

Fig. 1 Contig length distribution from *de novo* assembly of RNAseq reads from five *Ranunculus* cDNA libraries.

Table 3 Numbers of high-quality SNPs and indels called from *Ranunculus* RNAseq data

Genotype	SNPs	Indel
REV1	323 622	12 937
TRE(Apo)	34 361	1 564
VRU2(Apo)	30 785	1 360
YBB1	25 976	1 267
NOT	32 684	1 604
Total	447 428	18 732

between libraries were considered in downstream analyses.

SNP variation supports Pleistocene hybrid origin of apomictic Ranunculus

A total of 1677 characters with SNPs shared by at least two taxa (1408 single-base or ambiguous sites, the rest short-sequence polymorphisms) were identified and used to generate a well-supported hybridization network having a Fit value of 100% and 100% bootstrap support for all nodes (Fig. 2). Using *R. notabilis* as the out-group, the analysis shows that the two apomictic genotypes are closely related to one another, and their intermediate relationship between *R. cassubicifolius* and *R. carpaticola* supports the hypothesis of hybrid origin from these two taxa (Fig. 2). The hybridization network suggests a single origin of the hybrid with subsequent divergence of the two apomictic lineages. The apomict VRU1 shares more SNPs with *R. cassubicifolius* while the apomictic lineage TRE was closer to *R. carpaticola* (Fig. 2). Branch lengths of the hybridization network of the sexual parental taxa slightly exceed that of apomictic derivatives, which is consistent with an earlier evolutionary origin and divergence in allopatry. Divergence of the two apomictic accessions is higher than that between origin and split of the two accessions.

A relatively small number of mutations (and higher than those of the sexual reproducing parents) were accumulated in the asexual hexaploids after hybridization (Table 4). The estimated number of generations needed to reach the observed accumulation of multiple SNPs per locus suggests an approximate evolutionary origin of the hybrids to be maximally around 80 000 years before present.

Nonsynonymous/synonymous mutation rates in sexual and apomictic genotypes

The analysis of nonsynonymous to synonymous mutation rates was performed on 1231 genes that fulfilled the following criteria: the genes could be annotated, the ORFs could be assigned, and high-quality SNPs could be measured and compared for each genotype across the length of the complete gene (data S1). The overall distributions of pairwise comparisons of the dN/dS ratios for apomictic, apomictic and sexual, and sexual genotypes were strikingly similar, with the exception of outlier values (Fig. 3). The number of genes showing outlier dN/dS values increased from apomict (n = 163), to sexual (n = 373) to apomictic-sexual (n = 386) comparisons.

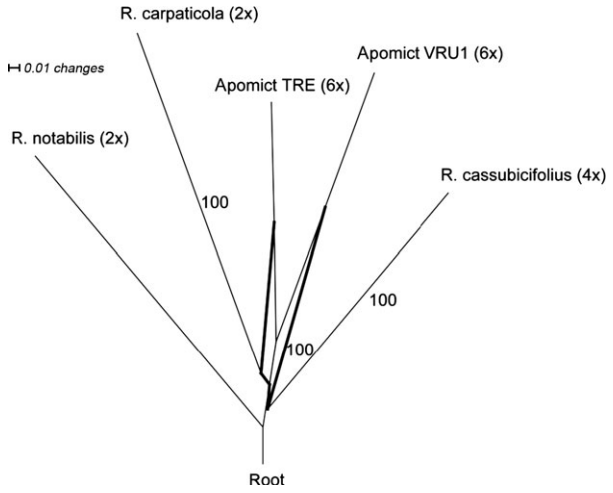

Fig. 2 Hybridization network (Splitsgraph) of five *Ranunculus* based on 1477 phylogenetically informed single-nucleotide polymorphisms(SNPs), with *R. notabilis* defined as out-group (Fit = 100%). Bootstrap percentages are indicated beside branches. Lines in bold indicate the hybrid relationship of the apomictic lineages reconstructed from SNPs shared by apomicts and parents. The numbers in parenthesis indicate the ploidy levels.

Table 4 Evolutionary divergence time as calculated from differential mutation accumulation observed between clonal hexaploids *R. carpaticola* × *cassubicifolius* genotypes and their putative parental sexual species

Genotype	Mutation rate (m)*	Loci with heterozygous SNPs†	Number of generations‡	Time (years)
TRE	9,1485e-5	3	20000	60000
VRU2	9,1485e-5	4	26667	80000

*expected frequency of mutation/substitution per generation.
†accumulated after the hybridization event.
‡number of generations needed to reach the observed number of loci with heterozygous SNPs when lack of meiosis is assumed.

A gene ontology (GO) analysis of gene enrichment for outlier values (Fig. 3) in all three comparisons, using all *Ranunculus* genes that could be annotated (n = 1846) as custom reference set, showed no statistically significant results. Interestingly, genes with elevated outlier dN/dS ratios exclusively found in the apomictic–sexual comparison were associated with processes involved in meiosis and gametogenesis (Fig. 4). Moreover, the proportion of GO terms associated with reproduction was significantly higher in the apomictic–sexual comparison (Table 5). The outliers in the apomict–apomict comparison were associated with 'metabolic' and 'cellular' GO terms. Finally the comparison between sexuals yielded

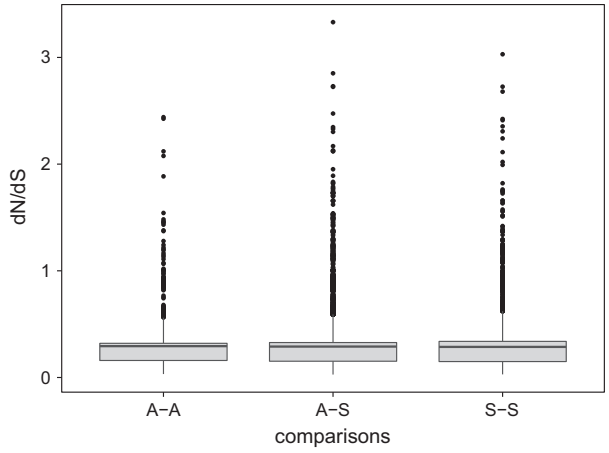

Fig. 3 Boxplot distributions of pairwise nonsynonymous (dN) to synonymous (dS) nucleotide substitution ratios for 574 open reading frames that shared at least a single single-nucleotide polymorphism at one common position across all *Ranunculus*. The three distributions represent apomict–apomict (A–A), apomict–sexual (A–S) and sexual–sexual (S–S) pairwise comparisons (circles are outliers values).

'metabolic process', 'cellular processes' and 'response to stimulus' as the 3 most representative terms (Figs S2 and S3, Supporting information).

Discussion

Whole transcriptome EST sequencing and gene annotation in nonmodel organisms, once the extremely expensive products of coordinated efforts between multiple laboratories, are becoming standardized analytical pipelines heralded by rapid advances in NGS and bioinformatics technology (Novaes *et al.* 2008; Bajgain *et al.* 2011). Here, we have taken advantage of the high coverage offered by one NGS platform (Illumina HiSeq™ 2000 Sequencing System) to generate and assemble the first transcriptome in the *Ranunculus auricomus* complex, a naturally occurring biological system for studying apomixis, polyploidy and hybrid genome evolution. Using a number of methods for qualification and quantification of DNA sequence polymorphisms (see methods), we have furthermore generated a robust data set of high-quality SNP and indel polymorphisms in order to compare five *Ranunculus* genotypes characterized by sexual and apomictic reproduction, ploidy variation and hybridization (Table 1).

Sequencing strategy, a balance between costs, goals and natural history

The decision to generate both SE and PE libraries from the different *Ranunculus* taxa being analysed here was based upon consideration of costs (*i.e.* SE libraries are

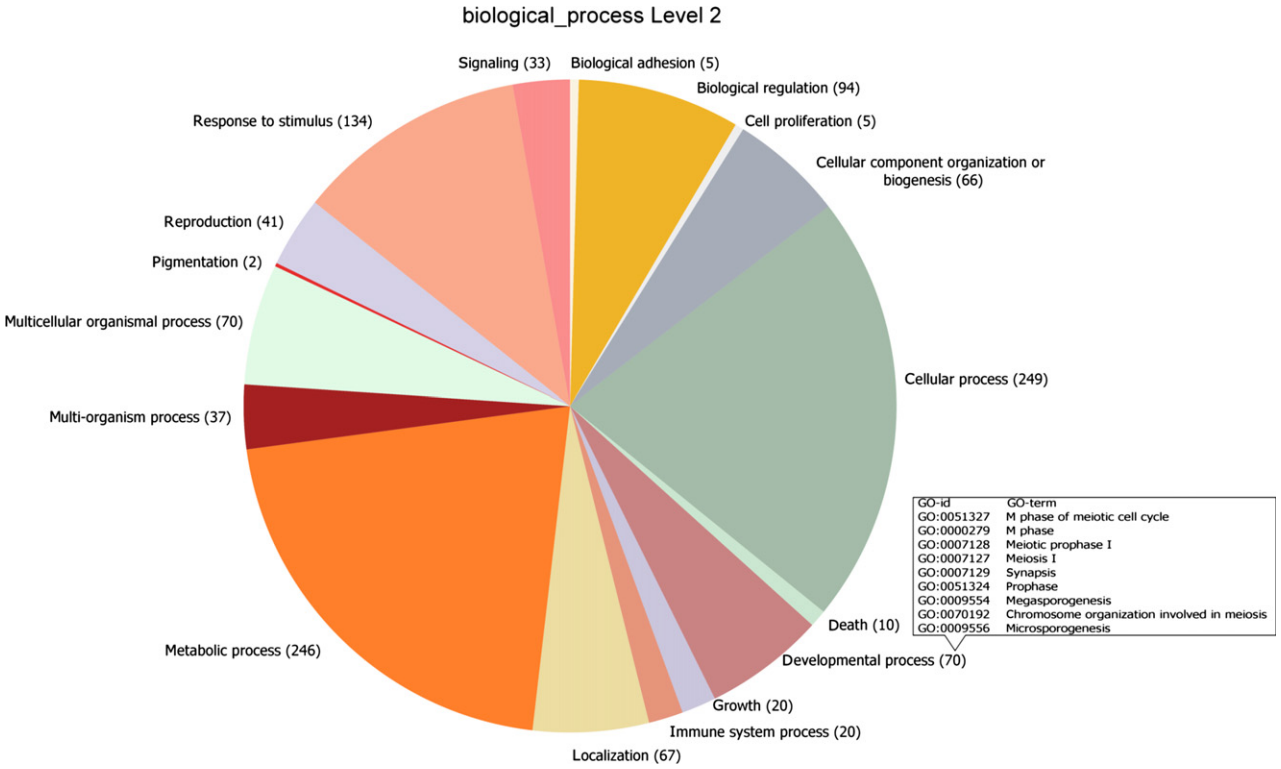

Fig. 4 Pie chart of second-level gene ontology (GO) terms (biological process) of genes showing outlier dN/dS ratios in the apomict–sexual comparison. Numbers in brackets refer to the number of GO terms in each category. Terms listed in the developmental process bubble cloud represent level 5–7 GO terms.

Table 5 Number of GO terms associated with reproduction in comparisons of genes showing outlier values of dN/dS

	Comparison	Reproductive GO	Other GO	Total
Apo–Apo	1	6	157	163
Apo–Sex	2	41	345	386
Sex–Sex	3	15	358	363

Fisher's exact test of comparisons: 1–2 ($P = 0.007$), 2–3 ($P = 0.0004$), 1–3 ($P = 1$).

cheaper than PE) and the fact that this genome is uncharacterized (*i.e.* PE libraries can be more successfully *de novo* assembled). We analysed RNA rather than DNA as we would expect the highest amount of information on effects of mutational load in the coding regions. Furthermore, we used an RNA normalization step, as our ultimate goal was to maximize genic (and allelic) sequence representation for their use in designing gene expression and CGH (comparative genome hybridization; see Aliyu *et al.*, in review) microarrays for future experiments. Working within a finite project budget, this approach has had both advantages and disadvantages.

As *R. carpaticola* is the only diploid progenitor in the hybridization event being studied here (Paun *et al.* 2006a; Table 1), it was chosen for the generation of the PE library as we assumed that (i) it would be characterized by ancestral polymorphisms of the whole 'cassubicus' group, (ii) that the complexity of sequence information resulting from ploidy variation would be minimized and (iii) having a high coverage from one of the parents would increase the confidence level when searching new single-nucleotide variations after hybrid origin. Indeed, the use of a *de novo* assembly strategy whereby trimmed reads from all libraries were pooled to generate a backbone sequence to which each individual library was reference assembled enabled the assembly of 91% of the original quality-trimmed sequencing reads (Table 2). The high proportion of removed reads included reads filtered by size (short reads), quality threshold and vector contamination (e.g. illumina adaptors and those from the cDNA normalization procedure). This has naturally led to a reduction in the gene coverage and SNP calling, but the point of this strict set of criteria was to identify high-quality polymorphisms for subsequent analyses. While most (73%) of the high-quality polymorphisms were identified in the *R. carpaticola* (PE) library, only shared polymorphisms between libraries were considered in our

subsequent analyses, and hence, this did not represent a bias in our interpretations.

Homology to known sequences was found for about one-third of the total contig number (19 977), which is likely a reflection of genomic divergence of *Ranunculus* over millions of years to the quite distantly related reference species (*Arabidopsis*) for which significant sequence information exists. *Ranunculus* belongs to basal eudicots that are relatively under-represented in sequence databases, and thus, our data may fill an important gap for understanding of transcriptome evolution in angiosperms. Alternatively, the presence of orphan genes in the genus, as have been found in other plant species with relatively little transcriptome information, for example, *Artemisia tridentata* (Bajgain *et al.* 2011) and *Epimedium sagittatum* (Zeng *et al.* 2010), could additionally contribute to the low genomic similarity of *Ranunculus* to other databases.

The resolving power of SNPs

A high number of high-quality SNPs were detected among all libraries, based upon their high coverage and quality control filtering applied by the vcfutils module of bcftools. Because of the low number of individuals selected for the analysis, we cannot exclude that a fraction of the SNPs represent individual rather than population-specific polymorphisms. Nevertheless, considering the high number of SNPs, a sample of just five individuals representing four taxa comprised sufficient information for the reconstruction of their evolutionary history. In comparison, previous population genetic approaches using AFLPs required 450 individuals out of the complete range of the species to reconstruct the ancient hybridization event between *R. carpaticola* and *R. cassubicifolius* (Paun *et al.* 2006a), but could neither resolve timing of events nor divergence between apomictic accessions. Our study shows that large amounts of SNP data allow not only for a reliable reconstruction of phylogenetic relationships with a few individuals, but also provide more information on evolutionary history (e.g. mutation accumulation). Moreover, AFLP markers could not provide information on genetic background, and DNA sequence analyses of standard plastid and nuclear markers from a few individuals could not resolve taxonomic relationships (Hörandl *et al.* 2009). SNP analysis is thus powerful for studying evolutionary processes in groups with young evolutionary histories and shallow phylogenies, even on a minimal number of samples.

Divergence and evolutionary origin

After a single hybrid origin, the apomictic lineages diversified rapidly from each other and in comparison with their sexual progenitors. The branch lengths of apomicts and sexuals are almost the same despite a much younger evolutionary age (max. 80 000 years, while diploid sexuals diverged c. 900 000 years ago; Hörandl 2004); hence, divergence of apomicts must have happened at a faster rate. This effect is even underestimated in the hybridization network, as only shared SNPs (i.e. without autapomorphic SNPs) were considered for this analysis (Fig. 3). Analysis of SNPs exclusive to apomicts provided information regarding divergence times and places the hybridization event reliably into the last glacial period. The method is thus powerful for dating of shallow phylogenies that is otherwise difficult without a fossil as calibration point (e.g. Forest 2009). Overall, our age estimate is concordant with the general divergence time estimates of the *Ranunculus* phylogeny (Emadzade *et al.* 2011), where the *R. auricomus* complex is placed on terminal nodes (Emadzade *et al.* 2011). The age estimate for origin of hybrids is younger than the divergence time of sexual parents of the hybrids (c. 0.3 Mill. years; see Hörandl 2004). Our dating confirms the general assumption that apomictic plant lineages are evolutionarily young and associated with the Pleistocene (Van Dijk & Vijverberg 2005).

In sexually reproducing, outcrossing diploids, allelic variation is reshuffled within a population to maintain a balance of homo- and heterozygous individuals. Divergence between alleles in the same individual cannot accumulate, being inhibited by the homogenizing effect of homolog pairing and recombination at meiosis. In contrast, diploid apomictic lineages characterized by decreased frequencies of meiotic crossing-over and chromatid interchange are expected to gradually accumulate inter- and intraindividual allelic sequence variation as a result of independent mutation accumulation. The resultant interallelic divergence is known as the Meselson effect and has been observed in Bdelloid rotifers, with some evidence demonstrated in apomictic *Boechera* (Mark Welch & Meselson 2000; Corral *et al.* 2009). Interestingly, in parthenogenetic Bdelloid rotifers, divergent alleles at the same locus can acquire distinct but related functions that produce proteins that act synergistically (Pouchkina-Stantcheva *et al.* 2007). Whether functional divergence between alleles similarly occurs in young asexual *Ranunculus* lineages remains an open question.

Our approach to identify intraindividual allelic divergence at single loci, and its correlation with apomixis in *R. auricomus*, provides evidence that apomictic plants can accumulate Meselson-effect-like changes. In the absence of gene conversion, highly efficient DNA repair or other homogenizing mechanisms, in addition to disturbed DNA repair mechanisms associated with

homologous recombination during meiosis, mutations should accumulate faster in apomicts despite low levels of sexual reproduction (Ceplitis 2003). We suggest that the observed number of divergent gene copies exclusively found in allohexaploid apomicts can be attributed to posthybridization mutational events associated with the young evolutionary history of this taxon.

The rapid accumulation of changes within a short time period is most likely enhanced by polyploidy, as on the one hand more possible mutational sites are available, while on the other hand, the availability of multiple gene copies would mask the effect of deleterious recessive mutations, and consequently, masked mutations could reach higher frequencies before elimination by selection. As neopolyploids, apomictic lineages would benefit from masking effects of deleterious mutations while they do not yet suffer from a highly accumulated mutational load (Gerstein & Otto 2009). However, detailed models of such effects on hexaploids are still not available. Polyploidy might explain that, in spite of their evolutionarily young age, genomic divergence was observed in the hexaploids analysed here, in contrast to ancient diploid asexual animals (e.g. mites) for which no signs of the Meselson effect could be found (Schaefer *et al.* 2006). However, considering that Oribatid mites reproduce via automixis and inverted meiosis (Heethoff *et al.* 2009), genomic evolution in these ancient asexuals may not be directly comparable to plants.

Whether the accumulation of mutational changes measured here reflects positive or negative fitness effects, however, cannot be directly inferred from our data set. Nonetheless, divergent selection interestingly occurs in genes associated with the cell cycle (see below), an observation that suggests a selective process associated with the shift from sexual to apomictic reproduction.

dN/dS ratios as signals for divergent selection

Considering the dearth of genomic information related to the genus, we approached the estimation of the differences in number of synonymous and nonsynonymous mutations between sexual and asexual *Ranunculus* similarly to what has been previously done in other nonmodel species (i.e. *Eucalyptus grandis*; Novaes *et al. 2008*). We focused on those ORFs that shared at least one SNP across all the lineages. While this approach has the drawback of loosing information from rare alleles and hotspots of genetic diversity, it minimized false positives due to alignment errors.

The analysis of dN/dS frequencies in pairwise comparisons between sexual and asexual genotypes revealed similar distributions, in addition to 345 outliers within and between reproductive groups (Fig. 3). While no significant enrichment was discovered in the GO analysis of outlier genes within each comparison, outliers between different comparisons of apomicts and sexual showed a significant enrichment of genes associated with reproduction (Table 5). Considering misregulation of sexual genes as the hypothesized induction mechanism for apomictic seed formation, a hybridization event that brings together relatively divergent alleles (and their associated regulatory factors) of genes associated with reproduction is consistent with the enrichment of such genes with outlier dN/dS ratios in the sex–apomict comparison.

Interestingly, outlier genes associated with 6 meiosis and gametogenesis terms were identified (Fig. 4), one of which was homologous to the ASY1 protein from *Arabidopsis* (E-value: 8e-24 with 45% identity). ASY1 is homologous to the *Saccharomyces cerevisiae* HOP1gene (Caryl *et al.* 2000), which is associated with meiosis, and promotes synapses formation preceding meiotic recombination. HOP1 mutants in yeast show reduced levels of meiotic recombination (Hollingsworth & Byers 1989) while the mutant for ASY1 in *A. thaliana* showed an almost complete absence of homologous chromosome synapsis during male meiosis (Ross *et al.* 1996). Indeed, lack of chromosome pairing characterizes sterility and low pollen viability in aposporous apomictic *Cassava* (Nassar 2001) as well as other apomictic taxa, included *Ranunculus* and *Paspalum* (Hörandl *et al.* 1997; Hörandl 2008; Izmailow 1996; Podio *et al.* 2012). These meiotic disturbances are thought to be a consequence of hybridization and/or polyploidy (Izmailow 1967; Paun *et al.* 2006a), while the functional connection to apomixis remains to be studied.

The finding of divergent selection (i.e. genes with high dN/dS ratios), albeit in a small number of samples, supports the hypothesis that in the short term, apomixis might be induced or enhanced by selection on particular loci. For example, the hexaploid apomictic lineages are geographically more widespread, occupy more different niches and are more abundant than the sexual progenitors (Paun *et al.* 2006a; Hörandl *et al.* 2009), and hence divergent selection could reflect a heterosis-like effect in the hybrid apomicts that is correlated with their evolutionary success. On the other hand, the sexual–apomictic comparison additionally revealed a number of genes with low dN/dS ratios, a result that points to stabilizing selection acting on otherwise highly heterozygous genotypes. Taken together, the mutations observed in these young polyploids may have not necessarily negative effects on fitness, as would be predicted by Muller's ratchet. Divergence among apomictic lineages, as observed in our SNP data set, might reflect the ecological potential of lineages (i.e. niche specialization) as observed in natural populations (Paun *et al.* 2006b). Therefore, mutation accumulation does not necessarily

confirm the doomed view of Stebbins (1950) that asexual plant lineages would rapidly go to extinction and would represent dead ends of evolution. An in-depth analysis of gene families, their evolution and their functions is required to understand the actual effects of mutation accumulation on evolutionary histories.

Acknowledgements

This project was funded under the DFG Grant number SH 337/5-1 granted to T.F.S., the FWF Grant I 310 B16 granted to E.H. and the COST action HAPRECI 'Harnessing Plant Reproduction for Crop Improvement.

References

Bajgain P, Richardson BA, Price JC, Cronn RC, Udall JA (2011) Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *BMC Genomics*, **12**, 370.

Barbazuk WB, Emrich S, Schnable PS (2007) SNP mining from maize 454 EST sequences. *CSH Protoc*, **2007**. doi: 10.1101/pdb.prot4786.

Barton NH, Charlesworth B (1998) Why sex and recombination? *Science*, **281**, 1986–1990.

Beck JB, Alexander PJ, Allphin L *et al.* (2012) Does hybridization drive the transition to asexuality in Boechera? *Evolution*, **66**, 985–995.

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

Bernstein H, Byerly H, Hopf F, Re M (1988) Is meiotic recombination an adaptation for repairing DNA, producing genetic variation, or both?. In: *The Evolution of Sex* (ed Michod RE LB), pp. 139–160. Sinauer Associates Inc Publishers, Sunderland, Massachusetts.

Buggs RJA, Chamala S, Wu W *et al.* (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid Tragopogon miscellus by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular ecology*, **19**, 132–146.

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

Caryl AP, Armstrong SJ, Jones GH, Franklin FCH (2000) A homologue of the yeast HOP1 gene is inactivated in the *Arabidopsis* meiotic mutant asy1. *Chromosoma*, **109**, 62–71.

Ceplitis A (2003) Coalescence times and the Meselson effect in asexual eukaryotes. *Genetical Research*, **82**, 183–190.

Comai L (2005) The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, **6**, 836–846.

Conesa A, Gotz S, Garcia-Gomez J, M.Terol J, Talon M, Robles M (2005) *Bioinformatics*, **21**, 3674–6.

Cork JM, Purugganan MD (2005) High-diversity genes in the *Arabidopsis* genome. *Genetics*, **170**, 1897–1911.

Corral J, Piwczynski M, Sharbel TF (2009) Allelic Sequence Divergence in the Apomictic *Boechera holboellii* Complex. In: *Lost Sex*(eds Schön I., Martens K. & Van Dijk P.), pp. 495–516. The evolutionary Biology of Parthenogenesis , Springer.

Crow JF, Kimura M (1965) Evolution in sexual and asexual populations. *American Naturalist*, **99**, 439–450.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acid Research*, **38**, 64–70.

Emadzade K, Gehrke B, Linder HP, Hörandl E (2011) The biogeographical history of the cosmopolitan genus *Ranunculus* L. (*Ranunculaceae*) in the temperate to meridional zones. *Molecular Phylogenetics and Evolution*, **58**, 4–21.

Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics*, **78**, 737–756.

Fisher RA (1930) *The Genetical Theory of Natural Selection* Clarendon Press, Oxford.

Forest F (2009) Calibration of the tree of life: fossils, molecules and evolutionary timescales. *Annals of Botany*, **104**, 789–794.

Gerstein AC, Otto SP (2009) Ploidy and the causes of genomic evolution. *Journal of Heredity*, **100**, 571–581.

Grossniklaus U (2001) From sexuality to apomixis: molecular and genetic approaches. In: *The Flowering of Apomixis: From Mechanisms to Genetic Engineering*(eds Savidan Y, Carman JG & Dresselhaus T), pp. 168–211. IRD CIMMYT, Mexico.

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

Hampe J, Franke A, Rosenstiel P *et al.* (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics*, **39**, 207–211.

Haseneyer G, Schmutzer T, Seidel M *et al.* (2011) From RNA-seq to large-scale genotyping genomics resources for rye (*Secale cereale* L.). *BMC Plant Biology*, **11**.

Heethoff M, Norton RA, Scheu S *et al.* (2009) Parthenogenesis in Oribatid mites (Acari, Oribatida): evolution without sex. In: *Lost Sex*(eds Schön I, Martens K & Van Dijk P), pp. 241–257. The evolutionary Biology of Parthenogenesis, Springer.

Hollingsworth NM, Byers B (1989) Hop1 – a yeast meiotic pairing gene. *Genetics*, **121**, 445–462.

Hörandl E (1998) Species concepts in agamic complexes: applications in the *Ranunculus auricomus* complex and general perspectives. *Folia Geobotanica*, **33**, 335–348.

Hörandl E (2008) Evolutionary implications of self-compatibility and reproductive fitness in the apomictic *Ranunculus auricomus* polyploid complex (*Ranunculaceae*). *International Journal of Plant Sciences*, **169**, 1219–1228.

Hörandl E (2009) A combinational theory for maintenance of sex. *Heredity (Edinb)*, **103**, 445–457.

Hörandl E, Greilhuber J (2002) Diploid and autotetraploid sexuals and their relationships to apomicts in the *Ranunculus cassubicus* group: insights from DNA content and isozyme variation. *Plant Systematics and Evolution*, **234**, 85–100.

Hörandl E, Hojsgaard D (2012) The evolution of apomixis in angiosperms: a reappraisal. *Plant Biosystems*, **146**, 681–693.

Hörandl E, Dobes C, Lambrou M (1997) Chromosomen- und Pollenuntersuchungen an österreichischen Arten des apomiktischen *Ranunculus auricomus*-Komplexes. *Botanica Helvetica*, **107**, 195–209.

Hörandl E, Greilhuber J, Dobes C (2000) Isozyme variation and ploidy levels within the apomictic *Ranunculus auricomus*

complex: evidence for a sexual progenitor species in south-eastern Austria. *Plant Biology*, **2**, 53–62.
Hörandl E, Greilhuber J, Klimova K *et al.* (2009) Reticulate evolution and taxonomic concepts in the *Ranunculus auricomus* complex (*Ranunculaceae*): insights from analysis of morphological, karyological and molecular data. *Taxon*, **58**, 1194–1215.
Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
Huson DH, Rupp R, Scornavacca C (2010) *Phylogenetic Networks. Concepts, Algorithms and Applications*. Cambridge Univ. Press, New York.
Izmailow R (1967) Observations in embryo and endosperm development in various chromosomic types of apomictic species *Ranunculus Cassubicus* L. *Acta Biologica Cracoviensia Series Botanica*, **10**, 99–111.
Izmailow R (1996) Reproductive strategy in the *Ranunculus auricomus* complex (*Ranunculaceae*). *Acta Societatis Botanicorum Poloniae*, **65**, 167–170.
Kantama L, Sharbel TF, Schranz ME *et al.* (2007) Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 14026–14031.
Koltunow AM (1993) Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. *Plant Cell*, **5**, 1425–1437.
Koltunow AM, Grossniklaus U (2003) Apomixis: a developmental perspective. *Annual Review of Plant Biology*, **54**, 547–574.
Kondrashov AS (1982) Selection against harmful mutations in large sexual and asexual populations. *Genetical Research*, **40**, 325–332.
Kondrashov AS (1993) Classification of hypotheses on the advantage of amphimixis. *Journal of Heredity*, **84**, 372–386.
Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, **26**, 589–595.
Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
Li JW, Schmieder R, Ward RM *et al.* (2012) SEQanswers: an open access community for collaboratively decoding genomes. *Bioinformatics*, **28**, 1272–1273.
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155.
Mark Welch D, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211–1215.
Maynard Smith J (1978) *The Evolution of Sex* Cambridge University Press, Cambridge.
Milne I, Bayer M, Cardle L *et al.* (2010) Tablet—next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.
Mittwoch U (1978) Parthenogenesis. *Journal of Medical Genetics*, **15**, 165–181.
Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research*, **106**, 2–9.
Nassar NM (2001) The nature of apomixis in cassava (*Manihot* esculentum, Crantz). *Hereditas*, **134**, 185–187.
Nogler GA (1984) Gametophytic apomixis. In: *Embryology of Angiosperms*(ed Johri BM), pp. 475–518. Springer Verlag, New York.
Novaes E, Drost DR, Farmerie WG *et al.* (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. *BMC Genomics*, **9**, 312.
Otto SP (2009) The evolutionary enigma of sex. *Am Nat*, **174**, S1–S14.
Otto SP, Whitton J (2000) Polyploidy incidence and evolution. *Annual Review of Genetics*, **34**, 401–437.
Ozias-Akins P, Van Dijk PJ (2007) Mendelian genetics of apomixis in plants. *Annual Review of Genetics*, **41**, 509–537.
Paun O, Stuessy TF, Hörandl E (2006a) The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex. *New Phytologist*, **171**, 223–236.
Paun O, Greilhuber J, Temsch E, Hörandl E (2006b) Patterns, sources and ecological implications of clonal diversity in apomictic *Ranunculus carpaticola* (*Ranunculus auricomus* complex, *Ranunculaceae*). *Molec Ecol*, **15**, 897–910.
Podio M, Siena LA, Hojsgaard D *et al.* (2012) Evaluation of meiotic abnormalities and pollen viability in aposporous and sexual tetraploid *Paspalum notatum* (*Poaceae*). *Pl Syst Evol*, **298**, 1625–1633.
Pouchkina-Stantcheva NN, McGee BM, Boschetti C *et al.* (2007) Functional divergence of former alleles in an ancient asexual invertebrate. *Science*, **318**, 268–271.
Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, **16**, 276–277.
Richards AJ (2003) Apomixis in flowering plants: an overview. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **358**, 1085–1093.
Ross KJ, Fransz P, Jones GH (1996) A light microscopic atlas of meiosis in *Arabidopsis thaliana*. *Chromosome Research*, **4**, 507–516.
Schaefer I, Domes K, Heethoff M *et al.* (2006) No evidence for the "Meselson effect" in parthenogenetic oribatid mites (*Oribatida, Acari*). *Journal of Evolutionary Biology*, **19**, 184–193.
Schwander T, Crespi BJ (2009) Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Molec Ecol*, **18**, 28–42.
Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369–381.
Shcheglov A, Zhulidov P, Bogdanova E, Shagin D (2007) Normalization of cDNA libraries. In: *Nucleic Acids Hybridization: Modern Applications* (eds Buzdin A. & Lukyanov S.), pp. 97–124. Springer, Germany. ISBN: 978-1-4020-6039-7.
Song Y, Leonard SW, Traber MG, Ho E (2009) Zinc deficiency affects DNA damage, oxidative stress, antioxidant defenses, and DNA repair in rats. *Journal of Nutrition*, **139**, 1626–1631.
Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia Univ Press, New York.
Suomalainen E (1950) *Parthenogenesis in animals. Advan. Genet.*, **3**, 193–253.
Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. *Plant Biotechnology Journal*, **7**, 334–346.
Trick M, Adamski NM, Mugford SG, Jiang CC, Febrer M, Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biology*, **12**, 14.
Van Dijk P (2003) Ecological and evolutionary opportunities of apomixis: insights from *Taraxacum* and *Chondrilla*. *Phil Trans R Soc Lond B*, **358**, 1113–1121.

Van Dijk P, Vijverberg K (2005). The significance of apomixis in the evolution of the angiosperms: a reappraisal. In: Bakker F, Chatrou L, Gravendeel B, Pelser PB (eds). *Plant Species-level Systematics: New Perspectives on Pattern and Process*. Gantner Verlag: Ruggell, Liechtenstein. pp 101–116.

Van Valen L (1973) *A new evolutionary law*. *Evolutionary Theory*, **1**, 1–30.

Vogel H, Wheat CW (2011) Accessing the transcriptome: how to normalize mRNA Pools. In: *Molecular Methods for Evolutionary Genetics*, 772. pp 105–128

Vogel H, Heidel AJ, Heckel DG, Groot AT (2010) Transcriptome analysis of the sex pheromone gland of the noctuid moth *Heliothis virescens*. *BMC Genomics*, **11**, 29.

Wolfe KH, Gouy ML, Yang YW, Sharp PM, Li WH (1989) Date of the monocot dicot divergence estimated from chloroplast dna-sequence data. *Proc Natl Acad Sci U S A*, **86**, 6201–6205.

Zeng S, Xiao G, Guo J, *et al*. (2010) Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics*, **11**, 94.

Zhulidov PA, Bogdanova EA, Shcheglov AS, *et al*. (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res*, **32**, e37–e37.

M.P., T.F.S and E.H. conceived and designed the study. M.P, T.F.S, E.H, and D.H. performed experiments and data analyses, and T.S. and U.S. performed the SNP analysis. M.P. and H.V performed the cDNA normalization. M.P. wrote the article with the support of all authors.

Data accessibility

Annotated contigs: Dryad entry doi:10.5061/dryad.nk151; NCBI SRA: SRX333612, SRX340844, SRX340856, SRX340870, SRX340890.

Presence/absence SNPs matrix: Dryad entry doi:10.5061/dryad.nk151.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Default parameters used for vcfutils module of VCFtools.

Table S2 Default and applied parameters used for samtool.

Fig. S1 Frequency distribution of blasted sequences and their similarity percentage with known species.

Fig. S2 Pie chart of second level GO terms (Biological Process) of gene showing outliers dN/dS ratio in the sexual-sexual comparison. Number in the bracket refers to the number of terms in each category.

Fig. S3 Pie chart of second level GO terms (Biological Process) of gene showing outliers dN/dS ratio in the apomictic-apomictic comparison. Number in the bracket refers to the number of terms in each category.

Data S1 List of the polymorphisms for each species for dN/dS analysis and substitution type in the 3 group comparisons.