New Phytologist

Research

Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size

Julia M. Kreiner[1,2], Paul Kron[1] and Brian C. Husband[1]

[1]Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; [2]Present address: Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St, Toronto, ON, Canada M5S 3B2

Author for correspondence:
Julia M. Kreiner
Tel: +1 519 830 9741
Email: julia.kreiner@mail.utoronto.ca

Received: *24 October 2016*
Accepted: *3 December 2016*

New Phytologist (2017) **214:** 879–889
doi: 10.1111/nph.14423

Key words: asexual, Brassicaceae, deleterious trait, flow cytometry, natural populations, polyploidy, reproductive mode, unreduced gametes.

Summary

- Fertilization involving unreduced (2*n*) gametes is considered the dominant mechanism of polyploid formation in angiosperms; however, our knowledge of the prevalence of and evolutionary mechanisms maintaining 2*n* gametes in natural populations is limited. We hypothesize that 2*n* gametes are deleterious consequences of meiotic errors maintained by mutation–selection balance and should increase in species with relaxed opportunities for selection on sexual processes (asexuality), reduced efficacy of selection (asexuality, selfing) and increased genome instability (high chromosome number).
- We used flow cytometry to estimate male 2*n* gamete production in 60 populations from 24 species of Brassicaceae. We quantified variation in 2*n* gamete production within and among species, and examined associations with life history, reproductive mode, genome size and chromosomal number while accounting for phylogeny.
- Most individuals produced < 2% 2*n* male gametes, whereas a small number had > 5% (up to 85%) production. Variation in 2*n* gamete production was significant among species and related to reproductive system; asexual species produced significantly more 2*n* gametes than mixed-mating and outcrossing species.
- Our results, unique in their multi-species perspective, are consistent with 2*n* gametes being deleterious but maintained when opportunities for selection are limited. Rare individuals with elevated 2*n* gamete production may be key contributors to polyploid formation.

Introduction

Polyploidy, the possession of greater than two genomes (complete chromosome sets) per nucleus, has been the focus of much plant research since its discovery a century ago (Lutz, 1907; Müntzing, 1936; Ramsey & Ramsey, 2014; Soltis *et al.*, 2016). Genome multiplication occurred early in the diversification of flowering plants and has been occurring repeatedly since then (Cui *et al.*, 2006; Wood *et al.*, 2009; Jiao *et al.*, 2011). The population-level process by which it evolves is typically broken into two phases: formation and establishment (Thompson & Lumaret, 1992). Establishment refers to the mechanisms by which neopolyploids spread in a population or colonize new populations. This stage has received considerable research attention (Fowler & Levin, 1984; Rodriguez, 1996; Husband & Sabara, 2004; Rausch & Morgan, 2005), particularly on the factors that may overcome the frequency-dependent disadvantage experienced by new polyploids (Levin, 1975a). In comparison, polyploid formation, the rate and underlying mechanisms by which new polyploids arise, is less well studied (Soltis *et al.*, 2010) despite being necessary to predict which species are predisposed to genome multiplication and under what conditions.

There is strong support for the hypothesis that polyploidy is derived primarily from unreduced gametes, which have the somatic (2*n*), rather than haploid (*n*) chromosome number (Harlan & De Wet, 1975). Unreduced gametes typically arise due to environmental or genetic factors leading to meiotic (or mitotic) failure during micro- and megasporogenesis (Bretagnolle & Thompson, 1995; Brownfield & Köhler, 2011). Polyploids form when two unreduced (2*n*) gametes unite to produce, for example, a tetraploid (2*n* + 2*n*, bilateral pathway), or when 2*n* and reduced gametes combine to form an odd ploidy (e.g. triploid from 2*n* + *n*, unilateral pathway), which can then backcross to a diploid plant (Den Nijs & Peloquin, 1977; Husband, 2004) to produce higher polyploids. All else being equal, the rate of polyploid formation is expected to be positively correlated with the frequency of 2*n* gametes in a population. The higher the rate of polyploid formation, the more likely polyploids will overcome the strong fitness disadvantage experienced by new polyploids due to the lack of compatible mates (minority cytotype exclusion principle: Levin, 1975a; Felber, 1991). The importance

of 2*n* gametes to polyploid evolution raises two fundamental questions: how frequent and variable are 2*n* gametes in natural populations and, given that these gametes are a product of meiotic dysfunction, what maintains them in natural populations?

Two possible evolutionary mechanisms may favour the persistence of 2*n* gametes in natural populations when they have some genetic basis. First, 2*n* gametes may be adaptive, in that they confer a direct fitness advantage that is positively selected under certain conditions. For example, the production of 2*n* gametes might be favoured along with parthenogenesis during the evolution of gametophytic apomixis (Bicknell & Koltunow, 2004; Lovell *et al.*, 2013). Such fitness advantages should give rise to high frequencies of 2*n* gametes in some species, yielding strong differentiation among species or populations. Alternatively, 2*n* gamete production may be maladaptive but persist in populations when the rate of mutations affecting meiosis exceeds the rate at which selection can purge them. A maladaptive role may be expected because dysfunctional meiotic processes are the basis of most 2*n* gamete production (Ramsey, 2007; Mason & Pires, 2015), and individuals with elevated 2*n* pollen production would likely have reduced fitness due to lower pollen viability and male siring success in diploid populations. Assuming reproductive isolation between cytotypes, a diploid individual that produces a high proportion of 2*n* gametes is effectively reducing its contribution to the diploid population gamete pool, wasting the resources that went into forming each 2*n* gamete. In this scenario, 2*n* gamete production should be relatively rare in populations, and should be elevated only when opportunities for meiotic mistakes are enhanced or selection to remove them is weakened.

Estimates of the frequency and variation of 2*n* gametes in natural populations are limited. Most published values come from ornamental or agricultural cultivars and from hybrid crosses grown in controlled environments (Veilleux *et al.*,1985; Dewitte *et al.*, 2011; Younis *et al.*, 2014), which are not necessarily representative of 2*n* gamete formation in natural populations. Nevertheless, these studies have found variation in 2*n* gamete production among individuals (Eenink, 1975), flowers (Matsuda, 1936; Lebedeff, 1940; Heyn, 1977), and anthers (De Mol, 1929), suggesting that it can be a highly labile trait. There are few quantitative studies of 2*n* gamete production in natural populations (e.g. De Haan *et al.*, 1992; Maceira *et al.*, 1992; Bretagnolle, 2001; Ramsey, 2007; Sheidai *et al.*, 2009; Kovalsky & Solís-Neffa, 2012). Based on a limited number of populations in a small number of species, these studies suggest that the frequency of 2*n* gametes is low. However, to date there have been no attempts to compare this variation within and among taxa or to relate that variation to relevant ecological and genetic attributes under natural conditions.

Here, we examine the frequency and correlates of variation in 2*n* male gamete production among several ecologically and chromosomally diverse species in the family Brassicaceae using flow cytometry. Flow cytometry, which assesses ploidy directly by evaluating the DNA content of pollen nuclei, is a high-throughput alternative to traditional morphological methods (Costich *et al.*, 1993; Ramsey & Schemske,1998; Bretagnolle, 2001; Ramsey, 2007), which limit sample size and are subject to inaccuracies due to natural variation in, for example, pollen size (Jansen & Den Nijs, 1993; Van Laere *et al.*, 2009; Dewitte *et al.*, 2011). Historically, quantifying 2*n* gamete formation using flow cytometry has been limited by technical difficulties, but recent developments in the technique enable population sampling of 2*n* male gametes on a scale rarely achieved before (Van Tuyl *et al.*, 1989; Bino *et al.*, 1990; Kron & Husband, 2012, 2015).

We sampled multiple species and populations to address the following questions: What is the overall rate of 2*n* male gamete production in natural plant populations?; Does 2*n* gamete production differ among populations within a species and among species?; To what extent is 2*n* gamete production related to the genomic and ecological characteristics of species such as chromosome number, ploidy, genome size, life history and reproductive mode?

As a first step in investigating the evolution of 2*n* gametes, we tested the hypothesis that 2*n* gametes are maladaptive in natural populations and thus are maintained at low frequencies through mutation–selection balance. If this is true, we expect population frequencies to be consistently low, overall, and to increase in populations where selection on sexual function is either rare (due to infrequent sex) or inefficient (due to low effective population size, N_e, and low recombination). Specifically, we predict that 2*n* gamete production will be higher in perennials compared to annuals. In any given year, perennial plants can forego sexual reproduction in favour of vegetative perennation or clonal reproduction, which may reduce the opportunity of selection on sexual function (Weismann, 1889; Levin, 1975b; Barton & Charlesworth, 1998) compared to annuals, which are obligately sexual. We also predict that 2*n* gametes will vary as a function of reproductive system (asexual vs selfing vs outcrossing) because selection against meiotic mutations will be weaker in predominantly asexual species due to infrequent sexual reproduction, and less efficient in asexual and selfing species because of lower N_e and effective recombination than in sexual, outcrossing species (Crow & Kimura, 1965; Orive, 1993; Charlesworth & Wright, 2001; Glémin, 2007). Alternatively, aspects of genome size, such as higher chromosome number (Snoad, 1955), polyploidy (Gottschalk, 1959; cited by Thompson, 1962; Britton & Hull, 1957; Selmecki *et al.*, 2015) and hybridization (Ramsey & Schemske, 1998), may increase the likelihood of 2*n* gamete formation by increasing chromosomal instability, chromosomal misalignments and unbalanced chromosomal segregation in premeiotic or meiotic cells.

Materials and Methods

Study system

The family Brassicaceae (formerly Cruciferae) comprises *c.* 325 genera and 3740 species (Stevens, 2016) that vary widely in life history, reproductive system, ploidy, genome size and chromosome number. Warwick & Al-Shehbaz (2006) have shown haploid chromosome numbers to vary among species of Brassicaceae, and even within genera (ranging from $n = 7$ to $n = 19$ in *Brassica*). Ploidy in Brassicaceae ranges from 2× to

>30× and genome size (1C) ranges from 0.15 and 4.73 picograms (pg) (Bennett & Leitch, 2012), with the family having undergone at least three whole genome duplication events in its history (Bowers *et al.*, 2003). Mating system ranges from obligately outcrossing, due to a widespread sporophytic self-incompatibility system, to predominantly selfing (Dickinson *et al.*, 1998). Asexual reproduction dominates in some species due to gamete sterility and vegetative reproduction; although apomictic seed production has been reported in three genera (Boechera, Draba and Arabis; Böcher, 1951; Mulligan & Findlay, 1970; Naumova *et al.*, 2001; Schranz *et al.*, 2005), it has not been substantiated in any of the genera we studied. The ecological and genetic variation within Brassicaceae make it an ideal family for a comparative study, and pollen nuclei characteristics in this family make it a good candidate for flow cytometric estimation of $2n$ male gametes (Kron & Husband, 2015).

Experimental design

During the summers of 2014 and 2015, we sampled 1647 individuals from 60 populations of 24 species (Table 1). These collections represented 15 genera, occupying environments ranging from open and disturbed to forested and relatively undisturbed, predominantly in Wellington County, Ontario, Canada (Supporting Information Table S1). To assess the variation in $2n$ gamete production within species, we sampled up to seven populations per species, ranging from 4 to 97 individuals per population (mean individuals per population = 28). Species were identified using Alex & Switzer (1988), Gleason & Cronquist (1991) and Reznicek & Voss (2012).

Within each population, individuals were sampled along a transect. For vegetatively reproducing species, individuals were sampled *c.* 2 m apart to maximize the number of genets sampled. We collected whole individuals or flowering stems from the field, placed them in water, and kept them in the lab overnight. This method allowed for the anthers of each individual to dehisce further overnight, allowing for cleaner collections of more pollen. We harvested pollen from each individual the day after it was collected by removing anthers (if large enough) or whole flowers and placing them into 1.5-ml plastic tubes to be dehydrated with silica gel for storage. A leaf sample was also collected, dehydrated and stored to estimate plant 2C DNA content (pg). For each population, a voucher individual was collected and deposited in the BIO Herbarium, University of Guelph, Canada (accession nos. 98259–98298).

Cytometry

We prepared pollen samples following the filter bursting method outlined by Kron & Husband (2012). Pollen samples were suspended and vortexed in LBO1 buffer (Doležel *et al.*, 2007),

Table 1 Summary statistics of the proportion of unreduced gamete production for 24 species in Brassicaceae sampled in southern Ontario, Canada

Species	Reproductive mode	*n* (Pop)	*n* (Ind)	Mean	Median	Min	Max	s^2 within species	Min s^2 within pops	Max s^2 within pops
Alliaria petiolata	M	7	253	0.016	0.014	0	0.128	1.20E-04	1.86E-05	6.69E-04
Barbarea vulgaris	O	3	81	0.008	0.007	0.001	0.031	3.40E-05	1.37E-05	1.37E-05
Brassica napus	S	3	91	0.01	0.009	0.003	0.041	3.34E-05	7.84E-06	8.90E-05
Capsella bursa-pastoris	S	2	42	0.019	0.017	0.003	0.048	1.38E-04	6.70E-05	2.14E-04
Cardamine bulbosa	O	2	62	0.015	0.013	0	0.035	5.52E-05	3.62E-05	5.55E-05
Cardamine concatenata (L)	A	3	44	0.102	0.035	0	0.71	9.60E-03	1.52E-02	5.37E-02
Cardamine concatenata (M)	A	1	25	0.054	0.014	0	0.398	3.46E-02	~	~
Cardamine concatenata (H)	A	2	23	0.088	0.024	0.005	0.856	2.28E-02	3.40E-02	6.22E-02
Cardamine diphylla	A	3	55	0.062	0.057	0.022	0.141	7.32E-04	4.10E-04	7.35E-04
Cardamine pensylvanica	S	2	62	0.017	0.013	0	0.073	2.19E-04	1.93E-04	2.07E-04
Descurainia sophia	S	2	50	0.09	0.09	0.01	0.19	2.46E-03	1.97E-03	2.87E-03
Diplotaxis tenuifolia	O	3	88	0.006	0.004	0.001	0.028	2.83E-05	1.75E-05	3.59E-05
Erucastrum gallicum	S	3	80	0.004	0.003	0.001	0.018	8.18E-06	1.68E-06	1.45E-05
Erysimum cheiranthoides	S	3	81	0.01	0.007	0	0.044	6.17E-05	7.29E-06	9.82E-05
Erysimum hieraciifolium	S	3	66	0.007	0.006	0	0.031	5.79E-05	1.32E-05	1.05E-04
Hesperis matronalis	M	3	91	0.012	0.011	0.007	0.083	6.71E-05	1.68E-05	1.18E-04
Lepidium campestre	S	2	85	0.009	0.007	0	0.039	3.71E-05	1.87E-05	1.11E-04
Nasturtium microphyllum	M	3	79	0.016	0.013	0	0.06	1.28E-04	8.39E-05	1.76E-04
Nasturtium officinale	M	1	18	0.01	0.009	0.004	0.02	2.06E-05	~	~
Sinapis arvensis	O	3	101	0.006	0.004	0.001	0.026	2.24E-05	1.26E-05	3.36E-05
Sisymbrium altissimum	S	1	31	0.017	0.014	0.003	0.069	1.69E-04	~	~
Sisymbrium loeselii	M	3	71	0.01	0.008	0	0.153	3.24E-04	2.89E-06	8.71E-04
Sisymbrium officinale	S	2	55	0.01	0.007	0.004	0.06	9.20E-05	9.21E-05	9.65E-05
Thlaspi arvense	S	3	62	0.007	0.006	0	0.041	5.46E-05	3.91E-06	6.86E-05

Number of populations and individuals, mean, median, minimum, and maximum columns represent values taken among individuals within species. s^2 within species represents variance among individuals within species, which incorporates within and between population variation. s^2 within populations represents variance within populations among individuals. Reproductive mode abbreviations: O, predominantly outcrossing; M, mixed-mating; S, predominantly selfing; A, predominantly asexual.

filtered through a coarse filter to exclude any large debris, and then filtered through a fine filter that effectively captured all pollen grains. To determine the appropriate size for the coarse and fine filters, we measured pollen diameter range from at least 100 pollen grains stained with DAPI (4′,6-diamidino-2-phenylindole) for each species using OpenLab v.4.0.3 (Improvision Inc., Waltham, MA, USA) and a Leica DMR microscope (Leica Microsystems, Wetzlar, Germany). We crushed the pollen against the fine filter using a plastic rod to release the pollen nuclei and then rinsed the free nuclei through the filter using 0.5 ml LBO1 buffer with 100 μg ml^{-1} of a DNA selective dye, propidium iodide, and 50 μg ml^{-1} of RNAse. Samples were allowed to stain for 20 min to 2 h and were then tested for DNA content and nuclei number using a FACSCalibur flow cytometer (BD Biosciences, San Jose, CA, USA) and CellQuest Pro v.6 (BD Biosciences, San Jose, CA, USA).

In order to estimate the 2C pg DNA content of each species, leaf samples of at least two individuals per population were tested along with an internal DNA content standard. Because ploidy varied within populations of *Cardamine concatenata*, 2C DNA content was measured for every individual. We chopped leaf samples in *c.* 700 μl of LB01 buffer with the same concentrations of propidium iodide and RNAse as above, and passed the suspended nuclei solution through a 30 μm filter. The DNA content standard was *Verbena officinalis* L. (2C pg DNA content = 0.678, Martin *et al.*, 2014) except for species with higher DNA contents: *Sisymbrium loeselii*, *Brassica napus* and *Descurainia sophia* were tested with *Raphanus sativus* L. 'Saxa' (2C pg DNA content = 1.11), *Hesperis matronalis* was tested with *Pisum sativum* L. 'Ctirad' (2C pg DNA content = 9.09), *Cardamine concatenata* was tested with *Secale cereale* L. 'Dankovské' (2C pg DNA content = 16.19) and *Cardamine diphylla* was tested with *Zea mays* L. 'CE-777′ (2C pg DNA content = 5.43) (Doležel *et al.*, 2007). The calculation of 2C pg DNA content for each species was based on the fluorescence of the sample species' 2C peak mean relative to that of the standard, where fluorescence area (FL-A, 485 nm) peak means were measured using ModFit LT v.3.3.11 (Verity Software House, Topsham, ME, USA).

2*n* gamete calculations and quality assessment

For each sample we calculated the proportion of 2*n* gamete nuclei (*u*) as:

$$u = \frac{2C_{nuclei}}{1C_{nuclei} + 2C_{nuclei}}$$

Methods for gating debris and doublets, as well as for estimating numbers of 1C (reduced gamete) and 2C (2*n* gamete) nuclei, followed Kron & Husband (2015). Debris was excluded by gating on a fluorescence area (FL-A, 485 nm) by fluorescence height (FL-H, 670 nm) scatterplot to eliminate particles with distinctly different fluorescence properties than nuclei. In some cases, we gated additional debris using a FL-A (485 nm) by side scatterplot. We applied pulse analysis (Sharpless & Melamed, 1976; Kron & Husband, 2015) to differentiate between 2*n* gametes and aggregates (two or more adhering nuclei). Specifically, we eliminated doublets by removing 2C events that had a distinctly lower FL-H per FL-A relative to that of single 2C nuclei. We manually gated the 1C and 2C nuclei clusters after excluding debris and doublets and calculated *u*. To minimize the error for estimates of rare (2C) events, we excluded samples with fewer than 3000 total nuclei. We corrected for possible somatic tissue contamination using the method described in Sora *et al.* (2016), and we performed our analyses (see Life history, reproductive and genomic associations with 2*n* gamete production) using both uncorrected and corrected estimates.

Species-level traits

We characterized each species by five characteristics: reproductive mode, life history, genome size (2C DNA content measured in pg, as mentioned above), somatic chromosome number (2*n*) and ploidy level (Table S2). The sources used differed among variables from plant databases and floras to primary literature. We classified mode of reproduction as predominantly selfing, mixed mating, predominantly outcrossing and predominantly asexual (pollen sterility, vegetative reproduction) for each species. Species were classified using a range of information, listed in order of priority: genetic estimates of selfing, flower morphology and pollinator interactions, and incidence of self-compatibility (for details see annotations to Table S2). We used flow cytometry to assess the 2C pg DNA content of each species as described above. Ploidy and chromosome number were inferred from a combination of the 2C DNA content and published species descriptions as described in the annotations to Table S2 (Methods S1).

Quantifying variation in 2*n* production

We quantified variation in 2*n* production among species, among populations within species, and among individuals within populations using a nested ANOVA implemented in the stats package in R (R Core Team, 2014). The population term was nested within species and the residual represented variation among individuals. Genus was not included in the analysis because in our sample only four genera comprised more than one species. To improve the fit of the data, we log-transformed the proportion of 2*n* gametes produced.

Life history, reproductive and genomic associations with 2*n* gamete production

We tested for an association between 2*n* gamete proportion and the four species-level traits (life history, reproductive mode, chromosome number, genome size (2C), and no interaction terms) while accounting for nonindependence due to phylogeny with phylogenetic generalized least squares (PGLS) using a single model that included all predictors. Ploidy was dropped from the analysis as it was highly correlated with chromosome number (and because our estimate of ploidy was often calculated from chromosome number; see Methods S1). We used the corPagel option in the R package nlme (v.3.1-128) to estimate the PGLS

correlation structure, which evaluates the phylogenetic signal in the character and 2*n* gamete dataset by estimating Pagel's λ (0 = no signal, 1 = complete phylogenetic dependence). Proportion of 2*n* gametes and chromosome number were log-transformed to improve normality of residuals (based on Q-Q plots) and ensure homoscedasticity of the data (Levene's test for reproductive mode: $F = 0.8649$, $P = 0.4755$). We calculated the likelihood ratio chi-square values (LR χ^2) of each predictor variable, and for any significant predictor variable examined all PGLS pairwise *t*-tests between treatment levels.

We used a modified version of Huang *et al.*'s (2016) Brassicaceae chronogram to control for phylogeny, which originally estimated divergence from the concatenation of 113 nuclear markers. Huang *et al.*'s (2016) phylogeny comprises 55 species, 29 out of 51 tribes, and 11 of 15 genera for which we had data. We added four missing genera to the tree using the bind.tip function in the R package PHYTOOLS v.0.5-39 (Revell, 2012). This function adds a new tip to the tree by positioning the new branch at a randomly generated position below the specified node. However, we constrained this position according to known taxonomic affiliations (Carlsen *et al.*, 2009; Couvreur *et al.*, 2010). We also used the add.species.to.genus function in PHYTOOLS v.0.5-39 (Revell, 2012) to add congeners at a random position along the genus branch. We then pruned the phylogeny to include only our 24 species of interest. We accounted for uncertainty from the addition of missing species by repeating our analysis using a set of 1000 random-addition trees (for example tree see Fig. S1). This approach was modified from Algar & Mahler (2015).

Results

The 24 species sampled differed greatly in reproductive mode, life history and genomic characteristics (Table S2). With respect to life history, 11 species were annual, five biennial, and eight perennial. It terms of reproductive mode, 11 species were predominantly selfing, five were mixed-mating, four were predominantly outcrossing, and four were predominantly asexual (Table S2). 2C pg DNA content ranged from 0.41 to 17.03, exceeding the previously reported range for this family (0.3–9.47; Bennett & Leitch, 2012). Chromosome number and ploidy ranged from 2*n* = 2x = 14 to 2*n* = 30× = 240. There was discrete variation in 2C DNA content within populations of the asexual *Cardamine concatenata*, leading us to tentatively identify three chromosomal species by clustering DNA contents of low (7.13–8.37 pg/2C, putative 20x), medium (8.42–10.22 pg/2C, putative 24x), and high (11.41–14.96 pg/2C, putative 30x) DNA content (see Table S2 annotations; Methods S1, for details).

Frequency and variation of 2*n* gamete production

The distribution of 2*n* male gamete frequencies among all individuals sampled was positively skewed (skewness = 9.8) (Fig. 1). The proportion of 2*n* gametes per plant ranged from 0 to 85.6%, where the average was 1.93% (SE = 0.001; median = 0.97%) among all plants and 2.52% across species means (SE = 0.006; median = 1.10%; Table 1). Of the 1696 individuals sampled, 51.4% of individuals produced < 1% 2*n* gametes, whereas 6.7% produced > 5% 2*n* gametes. Individuals with > 5% 2*n* gametes were primarily from the species *Descurainia sophia, Cardamine diphylla* and *Cardamine concatenata*.

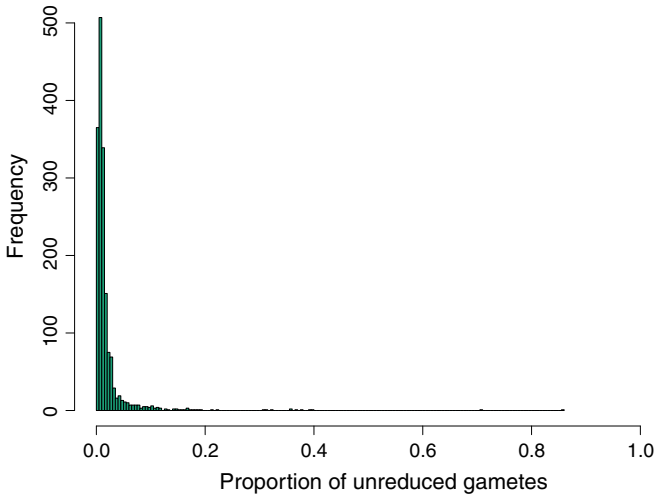

Fig. 1 Distribution of the proportion of unreduced gametes produced by all 1696 sampled individuals. Individuals are representative of multiple populations in 24 species of the family Brassicaceae (see Table 1 for list of species).

2*n* gamete production varied greatly among and within species (Table 1). Based on the nested ANOVA analysis, species explained almost double the variation in 2*n* gamete production compared to populations within species (30.5% compared to 17.3%) and both were significant (species: $F = 2.9$, $P = 0.0013$; Population: $F = 13.85$, $P < 2 \times 10^{-6}$). The majority of variation in 2*n* gamete production was among individuals within populations (residual) (52.2%). Variance in 2*n* gamete production among individuals within populations ranged from 2.89×10^{-6} in *Sisymbrium loeselii* to 6.22×10^{-2} in *Cardamine concatenata* (H) (Table 1). The highest three variances within populations occurred in *Cardamine concatenata* (20x, 30x) and *Descurainia sophia*, corresponding to their high mean production of 2*n* gametes.

In the most intensively sampled species, *Alliaria petiolata* ($n = 253$ individuals, $n = 7$ populations), the distribution of 2*n* gametes among individuals was less positively skewed (skewness = 4.7) than the distribution for all species pooled, but had a similar mean value of 1.59% and a median of 1.39% (Fig. 2). The distribution of 2*n* gamete production within populations of *A. petiolata* varied widely, from a somewhat normal distribution in the highly sampled Preservation Park population to positively skewed in the Speed River population (Fig. 2). Additionally, the minimum value of 2*n* gamete production in the Water Rd population (1.74%) was elevated relative to other populations and species, which typically show lower minimum values (mean minimum across *A. petiolata* populations = 0.46%; mean across species = 0.27%).

The predominantly asexual species, *Cardamine concatenata*, composed of three chromosomal species (20x, 24x, 30x), had the highest 2*n* gamete production (Table 1). The distribution of all

individuals combined across *C. concatenata* cytotypes was the least positively skewed out of all the species (skewness = 2.9). Only 27.2% of individuals produced < 1% 2*n* gametes, whereas 35.9% of individuals produced > 5% 2*n* gametes with levels of production up to 82.6% (Fig. S2). The average frequency of 2*n* gamete production across cytotypes was 8.57% ($n = 92$), more than four times greater than the mean frequency of 2*n* production across all samples and species (means for each chromosomal species were 10.18%, 5.38%, and 8.80%, for 20x, 24x and 30x taxa, respectively). Despite the high means, the shape of the distribution was similarly leptokurtic to other species, with a strong peak at the low end of production, < 5% (Fig. 3).

Life history, reproductive and genomic associations with 2*n* gamete production

We found a significant effect of reproductive mode on 2*n* gamete production (PGLS: $\lambda = 0.8716$, LR $\chi^2 = 4.67$, LR χ^2 SE = 0.028, $P = 0.019$; for distribution of χ^2 and P values see Fig. S3). PGLS pairwise comparisons across reproductive categories showed that predominantly asexual species had significantly more 2*n* gametes compared to mixed-mating species ($t = -2.246$, $P = 0.0292$) and predominantly outcrossing species ($t = -3.3175$, $P = 0.007$), and was nearly significantly different from predominantly selfing species ($t = -2.104$, $P = 0.0542$) (Fig. 4). The mean PGLS predicted value of the back-transformed proportion of 2*n* gamete production was lowest in outcrossing species (0.0096) out of all reproductive modes (0.0129, 0.0144 and 0.05193 for mixed mating, predominantly selfing and predominantly asexual species, respectively; Fig. 3). We found no significant effect of life history, chromosome number or DNA content on 2*n* gamete production. Annual species had the smallest predicted value (0.0158, compared to 0.0191 and 0.0178 for biennial and perennial species, respectively). Chromosome number had a slightly positive association (slope = 0.103) with log 2*n* gamete production; DNA content had a weak linear dependence (slope = 0.0017). The full analysis (four factors) conducted on data before correcting for somatic nuclei gave quantitatively the same results, in that mating system remained the only significant predictor of 2*n* gamete production. Because pollen was sampled over 2 yr, we tested for a difference among years using four species, sampled in both years, and found no difference (2-way mixed-model ANOVA; year effect, $F = 1.49$, $P = 0.31$).

Discussion

The union of 2*n* gametes is considered to be the primary mechanism of polyploid formation, yet we understand little about how common such gametes are and how they persist in natural populations. These questions are of particular interest given that 2*n* gametes likely arise due to meiotic dysfunction and are generally expected to have negative effects on male fitness in diploid populations. Here, we estimated male 2*n* gamete production in a large number of individuals and populations from 24 species of Brassicaceae using flow cytometry. The multispecies approach and intensive sampling is unique among existing studies of 2*n* gametes in natural populations. We found marked variation in 2*n* gamete production among individuals, populations and species. Reproductive mode was a significant predictor of the proportion of 2*n* gametes per species. Specifically, asexual species produced more 2*n* gametes than self-fertilizing, mixed-mating and outcrossing species (the latter two, statistically more). This pattern can be explained by weaker selection against deleterious mutations expressed during gametogenesis in asexual species.

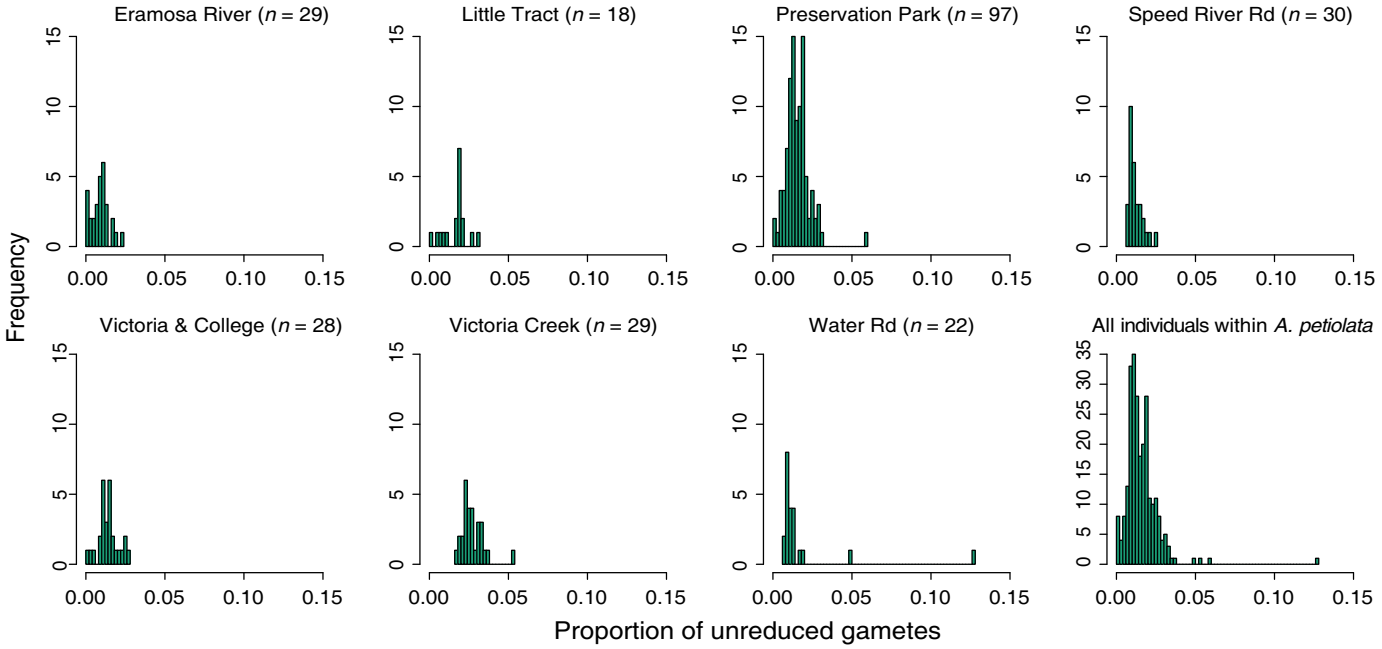

Fig. 2 Distribution of the proportion of unreduced gametes produced for seven populations of *Alliaria petiolata,* and for all individuals within *A. petiolata*.

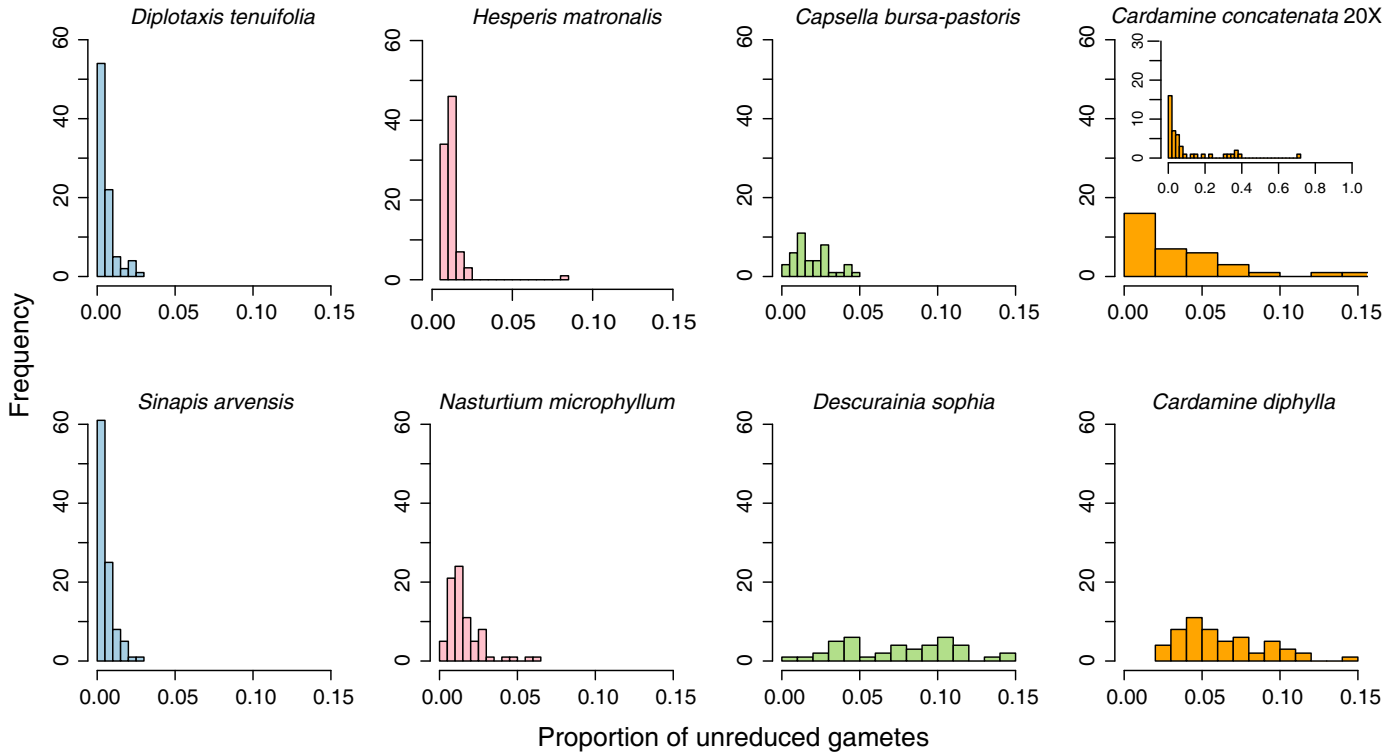

Fig. 3 Examples of the distribution of the proportion of unreduced gametes for species with different mating systems (blue, predominantly outcrossing; pink, mixed-mating; green, predominantly selfing; orange, predominantly asexual). The *Cardamine concatenata* 20× inset shows the full range of the distribution, with individuals producing unreduced gametes up to *c.* 75%.

Distribution of 2*n* gametes

The distribution of 2*n* gamete production among individuals was similar to that reported in other natural plant populations (Bretagnolle, 2001; Ramsey, 2007), in that the majority of individuals produce low frequencies of 2*n* gametes, whereas a minority produce higher frequencies. For example, in our study, 75.1% of individuals produced between 0.1 and 2.0% 2*n* gametes, the typical range described by Ramsey (2007) for population mean production in nonhybrid species. Further, we found 6.8% of all individuals produced > 5% 2*n* gametes (range among species, 0–70%), which was similar to 6.3% in *Anthoxanthum alpinum* ($n = 174$; Bretagnolle, 2001) but greater than the 1.6% in *Achillea borealis* ($n = 250$; Ramsey, 2007). Along with the skewed distributions, we also observed significant variance attributed to populations within species (17.3%) and among species (30.5%). We could not, however, compare these variances for the Brassicaceae to other plant families because of the lack of among-population and among-species data in the literature.

Despite the similarities we found to previous reports, our species mean (2.5% 2*n* gametes) is high compared to other reports. Even if we exclude *Cardamine concatenata* as an outlier, bringing our mean down to 1.7%, our estimate is almost four times higher than the mean estimate of 0.465% found by Ramsey & Schemske (1998) based on published reports for nine nonhybrid species. Several reasons may explain this discrepancy, including the fact that the Ramsey and Schemske mean included no Brassicaceae, a majority of diploid species (with no ploidies higher than 4x), and no predominantly asexual species. Also, more than a third of the estimates were based on realized 2*n* gamete production rates (i.e. post seed production rather than direct measures of 2*n* gamete production), as well as a small number of female 2*n* gamete estimates, which averaged lower than males. These differences highlight the need for more empirical data from natural populations using comparable methodologies while suggesting that 2*n* gamete production in natural populations may be higher than previously assumed.

The literature often describes individuals as '2*n* gamete producers' and 'non-producers' (e.g. Parrott & Smith, 1986; De Haan *et al.*, 1992; Maceira *et al.*, 1992; Kovalsky & Solís-Neffa, 2012). Typically, 2*n* pollen are identified using pollen size or other traits as indicators; 'producers' are identified by the 'presence' of large pollen, based on sampling some limited number of pollen grains, whereas 'absence' indicates nonproducers. This categorization implies a simple bimodality that may reflect the limitations of sample size in early studies using pollen size (and other traits) as indicators of 2*n* gametes (Bretagnolle & Thompson, 1995). For example, using pollen size, Bretagnolle (2001) observed 10% and Ramsey (2007) found that 44% of individuals have 2*n* gametes. By contrast, using flow cytometry and somatic correction, we found that 95.9% of individuals had some detectable 2*n* gametes. This suggests that as nuclei sample size increases (ranges up to 10 000 nuclei with flow cytometry) we can better detect rare events and quantify their frequency. Although we cannot exclude the possibility that this is a methodological artifact (background noise inflating very low estimates;

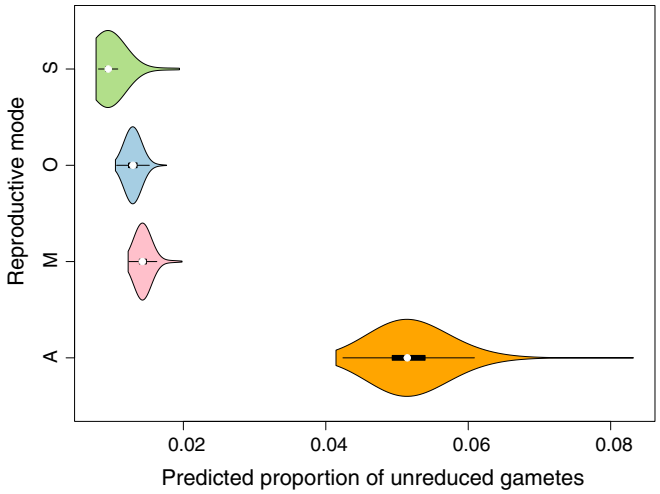

Fig. 4 Violin plots of 1000 iterations of phylogenetic generalized least squares (PGLS) model predicted values for the back-transformed proportion of unreduced gamete production by four reproductive modes (O, predominantly outcrossing; M, mixed-mating; S, predominantly selfing; A, predominantly asexual; see Table 1 for reproductive mode annotations for species). Asexual species have significantly higher predicted unreduced gamete production compared to outcrossing and mixed mating species, and marginally significantly higher production compared to selfing species. For each reproductive mode, colored distributions represent the kernel density distribution of predicted values, white circles represent the median predicted values, black boxes represent the interquartile range, and the tails represent the minimum and maximum predicted values.

Kron & Husband, 2015), it raises the possibility that 2*n* gamete frequencies in many natural populations may best be described as continuous, at least in Brassicaceae. Although previous studies have shown a relationship between flow cytometric 2*n* gamete estimates and measures of large pollen frequency (Van Tuyl *et al.*, 1989) or average pollen size (Sora *et al.*, 2016), future studies comparing these methods more directly could shed light on how methodology may have influenced our understanding of 2*n* gamete frequency and variation in natural populations.

Our evolutionary hypotheses of how 2*n* gametes are maintained in natural populations are based on the premise that 2*n* gamete production has a strong genetic component. However, it is possible that low rates of production in the majority of individuals may be the result of environmental effects. Studies have suggested that 2*n* gamete production is usually controlled by a monogenic recessive allele whose expression and frequency is modified by other genes (Bretagnolle & Thompson, 1995). In *Arabidopsis*, 2*n* gamete formation has been observed in several mutants acting in different pathways, including AtPS1 (d'Erfurth *et al.*, 2008), JASON (Erilova *et al.*, 2009) and OSD1 (d'Erfurth *et al.*, 2009). Interacting genes are especially likely, considering Bretagnolle & Thompson (1995) showed that there are at least 11 cytological abnormalities that lead to 2*n* gamete production. Nevertheless, extreme environments such as fluctuating temperature have been shown to interact with genotype to elevate the frequency of production (Belling, 1925; De Mol, 1929; Sax, 1936; Lewis, 1943; McHale, 1983; Ramsey & Schemske, 1998; Mason *et al.*, 2011). Thus, despite the underlying genetic basis of 2*n* gamete production playing a key role in its evolution, environmental factors are sure to explain a portion of the variation seen within and between populations and species, as are other factors, such as the ages or developmental stages of plants. Unfortunately, little is known about the relative contributions of genetics and environment to 2*n* gamete production in natural populations (but see Ramsey, 2007), and future studies incorporating both spatial and temporal variation in environmental conditions will hopefully fill this gap in our understanding. This could help determine whether polyploids tend to form more often in lineages genetically predisposed to produce 2*n* gametes, or whether environmental conditions are better predictors of polyploid formation.

Reproductive mode & the evolution of 2*n* gametes

Reproductive mode was the only factor we examined that was associated with 2*n* male gamete production. Species that are predominantly asexual (in the genus *Cardamine*) produced significantly higher frequencies of 2*n* gametes than mixed-mating or outcrossing species and weakly higher frequencies than self-fertilizing species. The effect of reproductive mode remained after phylogenetic correction, despite predominantly asexual species being confined to one genus. The significant effect of reproductive mode can be explained even if the fitness effects of 2*n* gamete production are deleterious due to variation in the opportunity and efficiency of selection. Selection against deleterious mutations is expected to be less efficient in predominantly asexual species and selfing species compared to outcrossing species owing to their reduced effective population size and reduced effective recombination (Crow & Kimura, 1965; Orive, 1993; Nordborg, 2000; Charlesworth & Wright, 2001; Hartfield *et al.*, 2016). As a result, neutral or mildly deleterious mutations will have an increased probability of spreading or fixation due to genetic drift. However, this mechanism doesn't explain the similarity in 2*n* gamete production between selfing and outcrossing species. This result may primarily reflect the difference in opportunity for selection among reproductive modes. Predominantly asexual species are distinct from selfing and outcrossing modes of reproduction, because sexual reproduction occurs less often or not at all, thus reducing the exposure of sexual function to selection.

Although our results can be explained without invoking an adaptive effect of 2*n* gamete production, we cannot exclude the possibility that 2*n* gamete production is positively selected in some asexuals. For example, 2*n* gametes may result from selection to reduce investment in meiosis and gametogenesis in organisms that rely on vegetative reproduction. However, this circumstance could only operate in obligately asexual species. In those with occasional sexual reproduction, selection would quickly restore meiotic function. In addition, the reduced male fitness due to 2*n* gametes would further outweigh the benefits of reduced investment in meiosis. Selection for 2*n* production in asexuals might also be expected in gametophytic apomict species, which require the joint occurrence of 2*n* gametes and parthenogenesis (Bicknell & Koltunow, 2004). Our study organisms, however, do not include any apomicts of which we are aware. Rather, asexual

reproduction results strictly from gamete sterility and vegetative reproduction. Therefore, it is unlikely that $2n$ gamete production in our study reflects selection for $2n$ gamete production in apomictic species. Given the evidence, we argue that it is more parsimonious to conclude that $2n$ production is elevated due to relaxed selection, as opposed to positive selection.

We observed no differences in $2n$ gamete production among species with different life histories, genome size or chromosome number. With respect to life history, perhaps any differences in the opportunity for selection between annuals and perennials is small compared to differences between asexuals and sexuals, which was readily detectable. Furthermore, our annual species are most often selfing (only 1 was biennial), which may have the opposite effect on $2n$ gamete production due to reduced efficiency of selection, and likely counter any effect of being annual. The association between life history and mating system may thus explain the lack of difference in $2n$ gamete production among life histories. The expected effect of chromosome number, ploidy and genome size was based largely on the premise that chromosome instability will increase with bigger genomes. We have not observed this directly in our species; however, it is likely to be most important in odd-numbered ploidy levels, which are not represented in our study. Moreover, where instability exists, unreduced gametes are only one of an array of possible meiotic outcomes; thus their effect on rates of $2n$ gametes may be undetectable with our sampling. Clearly additional study, using comparative and experimental approaches are required to fully understand the genetic and ecological circumstances favouring the origin and persistence of $2n$ gametes in natural populations. In particular, this should involve replication in other families that include predominantly asexual (agamospermous and vegetatively clonal) species.

Conclusions

$2n$ gametes are largely regarded as the primary mechanism of polyploid formation, despite the lack of investigations into frequency and occurrence of $2n$ gametes across a large sample size, multiple populations and species. The results of this study underscore that $2n$ gametes are remarkably widespread among individuals at low frequencies and can be particularly high in select individuals of a population, despite $2n$ gametes being previously thought of as relatively rare. If autopolyploids form through bidirectional sexual polyploidization via the union of two $2n$ gametes, and assuming completely random mating with equal $2n$ pollen and ovule population mean rates, then the likelihood that an autopolyploid forms is equivalent to the squared rate of $2n$ gamete production (Husband, 2004). Based on the mean proportion of $2n$ gamete production across species in this study (0.0252) the rate of autotetraploid formation should be 6.35×10^{-4}. This is one order of magnitude higher than predicted by Ramsey & Schemske (1998). Even when asexual *C. concatenata,* which produced an exceptionally high proportion of $2n$ gametes, is excluded from this calculation, the rate of autotetraploid formation remains high (2.96×10^{-4}). Thus, $2n$ gametes may be more common than previously thought, accounting for the discrepancy between what is considered to be their foremost role in polyploid formation and perceived rarity. Moreover, the mean rate of $2n$ gamete production may underestimate the rate of polyploid formation; individuals with high rates of $2n$ gamete production (i.e. the 6.7% of individuals producing $>5\%$) can enhance the probability of polyploid formation under certain conditions (e.g. positively correlated male and female rates and some selfing). The relatively high rates of polyploid formation predicted by the rates of $2n$ gamete production found in this study may help explain the frequent recurrence of polyploidy in plant lineages. Future studies directly comparing $2n$ gamete production to rates of neopolyploid formation in natural populations would be a useful next step in expanding our understanding of polyploid evolution.

Acknowledgements

We would like to thank Dylan Sora, James Seery, Tyler Kent, Corlett Wood and Luke Mahler for helping with collections and providing useful advice, as well as Carole Ann Lacroix for help with identification and herbarium deposition, and anonymous reviewers. We thank the rare Charitable Research Reserve (Cambridge, Ontario) for permission to collect on their property. This study was supported by grants from Natural Science and Engineering Research Council of Canada, Canada Research Chair Program and Canadian Foundation for Innovation to BCH, and NSERC USRA and Stobie Summer Research Assistantship to JMK.

Author contributions

J.M.K., P.K. and B.C.H. planned and designed the research; J.M.K. collected the plants and analysed the data; J.M.K. and P.K. performed the flow cytometry; and J.M.K., P.K. and B.C.H. wrote the manuscript.

References

Alex J, Switzer C. 1988. *Ontario weeds: descriptions, illustrations and keys to their identification.* Toronto, ON, Canada: Ontario Ministry of Agriculture and Food.

Algar AC, Mahler DL. 2015. Area, climate heterogeneity, and the response of climate niches to ecological opportunity in island radiations of *Anolis* lizards. *Global Ecology and Biogeography* **25**: 781–791.

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

Belling J. 1925. The origin of chromosomal mutations in *Uvularia*. *Journal of Genetics* **15**: 245–266.

Bennett M, Leitch I. 2012. *Plant DNA C-values database* (release 6.0, Dec. 2012). [WWW document] URL http://www.kew.org/cvalues/ [accessed 14 July 2015].

Bicknell RA, Koltunow AM. 2004. Understanding apomixis: recent advances and remaining conundrums. *Plant Cell* **16**: S228–S245.

Bino RJ, Van Tuyl JM, De Vries JN. 1990. Flow cytometric determination of relative nuclear DNA contents in bicellulate and tricellulate pollen. *Annals of Botany* **65**: 3–8.

Böcher TW. 1951. Cytological and embryological studies in the amphi-apomictic Arabis holboellii complex. *Kong Dansk Vidensk Selsk Biol Skr VI* **7**: 1–57.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.

Bretagnolle F. 2001. Pollen production and spontaneous polyploidization in diploid populations of *Anthoxanthum alpinum*. *Biological Journal of the Linnean Society* **72**: 241–247.

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

Britton DM, Hull JW. 1957. Mitotic instability in *Rubus*. *Journal of Heredity* **48**: 11–20.

Brownfield L, Köhler C. 2011. Unreduced gamete formation in plants: mechanisms and prospects. *Journal of Experimental Botany* **62**: 1659–1668.

Carlsen T, Bleeker W, Hurka H, Elven R, Brochmann C. 2009. Biogeography and phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden* **96**: 215–236.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics and Development* **11**: 685–690.

Costich DE, Ortiz R, Meagher TR, Bruederle LP, Vorsa N. 1993. Determination of ploidy level and nuclear DNA content in blueberry by flow cytometry. *Theoretical and Applied Genetics* **86**: 1001–1006.

Couvreur TL, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution* **27**: 55–71.

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

Cui L, Wall P, Leebens-Mack L, Lindsay B, Soltis D, Doyle J, Soltis P, Carlson J, Arumuganathan A, Barakat A *et al.* 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* **16**: 738–749.

De Haan A, Maceira NO, Lumaret R, Delay J. 1992. Production of 2*n* gametes in diploid subspecies of *Dactylis glomerata* L. 2. Occurrence and frequency of 2*n* eggs. *Annals of Botany* **69**: 345–350.

De Mol W. 1929. The originating of diploid and tetraploid pollen-grains in Duc van Thol-Tulips (*Tulipa suaveolens*) dependent on the method of culture applied. *Genetica* **11**: 119–212.

Den Nijs T, Peloquin S. 1977. 2*n* gametes in potato species and their function in sexual polyploidization. *Euphytica* **26**: 585–600.

Dewitte A, Laere K, Huylenbroeck J. 2011. Use of 2*n* gametes in plant breeding. In: Abdurakhmonov IY, ed. *Plant breeding*. Rijeka, Croatia: InTech Open Access, 59–86.

Dickinson H, Doughty J, Hiscock S, Elleman C, Stephenson A. 1998. Pollen-stigma interactions in *Brassica*. *Symposia of the Society for Experimental Biology* **51**: 51–57.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2**: 2233–2244.

Eenink AH. 1975. Matromorphy in *Brassica oleracea* L. VII. Research on products of microsporogenesis and gametogenesis from prickle pollinated plants. *Euphytica* **24**: 45–52.

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R. 2009. Turning meiosis into mitosis. *PLoS Biology* **7**: e1000124.

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Simon M, Jenczewski E, Mercier R. 2008. Mutations in AtPS1 (*Arabidopsis thaliana* parallel spindle 1) lead to the production of diploid pollen grains. *PLoS Genetics* **4**: e1000274.

Erilova A, Brownfield L, Exner V, Rosa M, Twell D, Scheid OM, Hennig L, Köhler C. 2009. Imprinting of the polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. *PLoS Genetics* **5**: e1000663.

Felber F. 1991. Establishment of a tetraploid cytotype in a diploid population – effect of relative fitness of the cytotypes. *Journal of Evolutionary Biology* **4**: 195–207.

Fowler NL, Levin DA. 1984. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *American Naturalist* **5**: 703–711.

Gleason H, Cronquist A. 1991. *Manual of the vascular plants of Northeastern United States and adjacent Canada*, 2nd edn. New York, NY, USA: New York Botanical Garden.

Glémin S. 2007. Mating systems and the efficacy of selection at the molecular level. *Genetics* **177**: 905–916.

Gottschalk W. 1959. Die Bildung von Keimzellen mit euploiden Chromosomenzahlen aus triploiden und aneuploiden Pollenmutterzellen abregulierender poly-ploider Pflanzen. *Molecular and General Genetics* **90**: 198–214.

Harlan J, De Wet J. 1975. On *Ö winge* and a prayer: the origins of polyploidy. *Botanical Review* **41**: 361–390.

Hartfield M, Wright SI, Agrawal AF. 2016. Coalescent times and patterns of genetic diversity in species with facultative sex: effects of gene conversion, population structure, and heterogeneity. *Genetics* **202**: 297–312.

Heyn F. 1977. Analysis of unreduced gametes in the Brassicaceae by crosses between species and ploidy levels. *Journal of Plant Breeding* **78**: 13–30.

Huang CH, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP *et al.* 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution* **33**: 394–412.

Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* **82**: 537–546.

Husband BC, Sabara HA. 2004. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* **161**: 703–713.

Jansen RC, Den Nijs APM. 1993. A statistical mixture model for estimating the proportion of unreduced pollen grains in perennial ryegrass (*Lolium perenne* L.) via the size of pollen grains. *Euphytica* **70**: 205–215.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Kovalsky IE, Solís-Neffa VG. 2012. Evidence of 2*n* microspore production in a natural diploid population of *Turnera sidoides* subsp. *carnea* and its relevance in the evolution of the *T. sidoides* (Turneraceae) autopolyploid complex. *Journal of Plant Research* **125**: 725–734.

Kron P, Husband BC. 2012. Using flow cytometry to estimate pollen DNA content: improved methodology and applications. *Annals of Botany* **110**: 1067–1078.

Kron P, Husband BC. 2015. Distinguishing 2*N* gamete nuclei from doublets in pollen using flow cytometry and pulse analysis. *Cytometry Part A* **87**: 943–957.

Lebedeff GA. 1940. Failure of cytokinesis during microsporogenesis in *Zea mays* following heat treatment. *Cytologia* **10**: 434–442.

Levin D. 1975a. Minority cytotype exclusion in local plant populations. *Taxon* **24**: 35–43.

Levin D. 1975b. Pest pressure and recombination systems in plants. *American Society of Naturalists* **109**: 437–451.

Lewis D. 1943. The incompatibility sieve for producing polyploids. *Journal of Genetics* **45**: 261–264.

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

Lutz AM. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science* **26**: 151–152.

Maceira N, De Haan A, Lumaret R, Billon M, Delay J. 1992. Production of 2*n* gametes in diploid subspecies of *Dactylis glomerata* L. 1. occurrence and frequency of 2*n* pollen. *Annals of Botany* **69**: 335–343.

Martin S, Sauder C, James T, Cheung K, Razeq F, Kron P, Hall L. 2014. Sexual hybridization between *Capsella bursa-pastoris* (L.) Medik (♀) and *Camelina sativa* (L.) Crantz (♂) (Brassicaceae). *Plant Breeding* **134**: 212–220.

Mason AS, Nelson MN, Yan G, Cowling WA. 2011. Production of viable male unreduced gametes in *Brassica* interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biology* **11**: 103.

Mason AS, Pires JC. 2015. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in Genetics* **31**: 5–10.

Matsuda H. 1936. The effect of abnormal temperature upon the pollen formation of *Petunia*. *Journal of the College of Agriculture* **14**: 71–92.

McHale NA. 1983. Environmental induction of high frequency 2*n* pollen formation in diploid *Solanum*. *Canadian Journal of Genetics and Cytology* **25**: 609–615.

Mulligan GA, Findlay JN. 1970. Sexual reproduction and agamospermy in the genus Draba. *Canadian Journal of Botany* **48**: 269–270.

Müntzing A. 1936. The evolutionary significance of autopolyploidy. *Hereditas* **21**: 363–378.

Naumova TN, van der Laak J, Osadtchiy J, Matzk F, Kravtchenko A, Bergervoet J, Ramulu KS, Boutilier K. 2001. Reproductive development in apomictic populations of *Arabis holboellii* (Brassicaceae). *Sexual Plant Reproduction* **14**: 195–200.

Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* **154**: 923–929.

Orive ME. 1993. Effective population size in organisms with complex life-histories. *Theoretical Population Biology* **44**: 316–340.

Parrott W, Smith R. 1986. Recurrent selection for 2*n* pollen formation in red clover. *Crop Science* **26**: 1132–1135.

R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. . [WWW document] URL http://www.R-project.org/. [accessed 15 January 2015]

Ramsey J. 2007. Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). *Heredity* **98**: 143–150.

Ramsey J, Ramsey TS. 2014. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B* **369**: 20130352.

Ramsey J, Schemske D. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Rausch JH, Morgan MT. 2005. The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishment. *Evolution* **59**: 1867–1875.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Reznicek AA, Voss EG. 2012. *Field manual of Michigan flora*. Michigan, MI, USA: University of Michigan Press.

Rodriguez DJ. 1996. A model for the establishment of polyploidy in plants. *American Naturalist* **147**: 33–46.

Sax K. 1936. The experimental production of polyploidy. *Journal of the Arnold Arbor* **17**: 153–159.

Schranz ME, Dobeš C, Koch MA, Mitchell-Olds T. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). *American Journal of Botany* **92**: 1797–1810.

Selmecki A, Maruvka Y, Richmond P, Guillet M, Shoresh N, Sorenson A, De S, Michor F, Dowell R, Pellman D. 2015. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**: 349–352.

Sharpless TK, Melamed MR. 1976. Estimation of cell size from pulse shape in flow cytofluorometry. *Journal of Histochemistry & Cytochemistry* **24**: 257–264.

Sheidai M, Jafari S, Taleban P, Keshavarzi M. 2009. Cytomixis and unreduced pollen grain formation in *Alopecurus* L. and *Catbrosa* Beauv. (Poaceae). *Cytologia* **74**: 31–41.

Snoad B. 1955. Somatic instability of chromosome number in *Hymenocallis calathium*. *Heredity* **9**: 129–133.

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS. 2010. What we still don't know about polyploidy. *Taxon* **59**: 1387–1403.

Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: pitfalls and paths to a paradigm. *American Journal of Botany* **103**: 1146–1166.

Sora D, Kron P, Husband BC. 2016. Genetic and environmental determinants of unreduced gamete production in *Brassica napus, Sinapis arvensis* and their hybrids. *Heredity* **117**: 440–448.

Stevens PF. 2016. *Angiosperm phylogeny website, v.12, July 2012* [and more or less continuously updated since]. [WWW document] URL http://www.mobot.org/MOBOT/research/APweb/ [accessed 20 September 2016].

Thompson MM. 1962. Cytogenetics of *Rubus*. III. Meiotic instability in some higher polyploids. *American Journal of Botany* **49**: 575–582.

Thompson J, Lumaret R. 1992. The evolutionary dynamics of polyploid plants: origin, establishment and persistence. *Trends in Ecology and Evolution* **7**: 302–307.

Van Laere K, Dewitte A, Van Huylenbroeck J, Van Bockstaele E. 2009. Evidence for the occurrence of unreduced gametes in interspecific hybrids of *Hibiscus*. *Journal of Horticultural Science and Biotechnology* **84**: 240–247.

Van Tuyl JM, de Vries JN, Bino RJ, Kwakkenbos AM. 1989. Identification of 2*n* pollen producing interspecific hybrids of *Lilium* using flow cytometry. *Cytologia* **54**: 737–745.

Veilleux RE, Booze-Daniels J, Pehu E. 1985. Anther culture of a 2*n* pollen producing clone of Solanum phureja Juz. & Buk. *Canadian Journal of Genetics and Cytology* **27**: 559–564.

Warwick S, Al-Shehbaz A. 2006. Brassicaceae: chromosome number index and database on CD-Rom. *Plant Systematics and Evolution* **259**: 237–248.

Weismann A. 1889. The significance of sexual reproduction in the theory of natural selection. *Essays Upon Heredity and Kindred Biological Problems* **1**: 255–332.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* **106**: 13875–13879.

Younis A, Hwang YJ, Lim KB. 2014. Exploitation of induced 2*n*-gametes for plant breeding. *Plant Cell Reports* **33**: 215–223.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Example phylogenetic tree.

Fig. S2 Distribution of 2*n* gamete production for the three chromosomal races of *Cardamine concatenata*.

Fig. S3 Distribution of χ^2 and *P*-value for reproductive mode across 1000 iterations of the phylogenetic generalized least squares.

Table S1 GPS coordinates of sampling locations across Ontario

Table S2 Classification table of collected Brassicaceae species

Methods S1 Annotations and methods for assigning values in classification table.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.