JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12734

Microsatellite evidence for obligate autogamy, but abundant genetic variation in the herbaceous monocarp *Lobelia inflata* (Campanulaceae)

P. W. HUGHES* & A. M. SIMONS†

*Max Planck Institute for Plant Breeding Research, Cologne, Germany
†Department of Biology, Carleton University, Ottawa, ON, Canada

Keywords:

autogamy;
lineage selection;
Lobelia inflata;
self-fertilization;
semelparity.

Abstract

Although high levels of self-fertilization (>85%) are not uncommon in nature, organisms reproducing entirely through selfing are extremely rare. Predominant selfers are expected to have low genetic diversity because genetic variation is distributed among rather than within lineages and is readily lost through genetic drift. We examined genetic diversity at 22 microsatellite loci in 105 individuals from a population of the semelparous herb *Lobelia inflata* L. and found (i) no evidence of heterozygosity through outcrossing, yet (ii) high rates of genetic polymorphism (2–4 alleles per locus). Furthermore, this genetic variation among lineages was associated with phenotypic traits (e.g. flower colour, size at first flower). Coupled with previous work characterizing the fitness consequences of reproductive timing, our results suggest that temporal genotype-by-environment interaction may maintain genetic variation and, because genetic variation occurs only among lineages, this simple system offers a unique opportunity for future tests of this mechanism.

Introduction

The fitness consequences of self-fertilization (selfing) have been the subject of considerable attention in evolutionary ecology (Lande & Schemske, 1985; Schemske & Lande, 1985; Barrett & Eckert, 1990; Hereford, 2010). Self-fertilization is favoured over outcrossing when fitness gains due to reproductive assurance, transmission advantage, isolation from maladaptive genes, or density-dependent interactions outweigh negative effects associated with inbreeding depression or gamete discounting (Jain, 1976; Lande & Schemske, 1985; Schemske & Lande, 1985; Barrett & Eckert, 1990; Goodwillie *et al.*, 2005). Inbreeding depression may have slight-to-severe fitness consequences and may vary in intensity (Eckert & Barrett, 1994; Husband & Schemske, 1996; Willis, 1999; Armbruster & Reed, 2005). The strength of inbreeding depression is negatively associated with the proportion of self-fertilization present in a species with a mixed mating system (Johnston & Schoen, 1996; Dudash *et al.*, 1997; Crnokrak & Barrett, 2002)

Although 10–20% of species are predominantly (>50%) selfing (Barrett, 2002), only a small minority are highly (>95%) selfing (Wright *et al.*, 2013). In these species, reproductive assurance is cited as the main evolutionary mechanism favouring extreme selfing over a mixed mating system (Takebayashi & Morrell, 2001; Wright *et al.*, 2013; Zhang *et al.*, 2014). However, Wright *et al.* (2013) also note that 'marker-based estimates of self-fertilization rates suggest that very few if any plant species are completely selfing' – that is, obligate selfing is extremely rare or unknown in nature.

Extant mating-system models generally assume that – in the absence of gene flow among lineages within a population – genetic diversity will gradually erode for several reasons. As the effective population size of selfers is half that of outcrossers, the effects of genetic drift may be strong (Lande & Schemske, 1985; Schemske & Lande, 1985; Jarne & Charlesworth, 1993; Charlesworth & Charlesworth, 1995). Alternative alleles are 'captured' within lineages and cannot be recombined; thus, both drift and selective purging of deleterious

Correspondence: Patrick William Hughes, Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany.
Tel.:+49 1762 7663 089; fax: +49 221 5062 413; e-mail: whughes@mpipz.mpg.de

alleles will result in low genetic diversity (Husband & Schemske, 1996; Dudash *et al.*, 1997; Crnokrak & Barrett, 2002). This leads to the speculation that selfing may be irreversible and is therefore an 'evolutionary dead end' (Darwin, 1877; Schoen & Brown, 1991; Barrett & Harder, 1996; Takebayashi & Morrell, 2001; Igic & Busch, 2013) and may explain the rarity of very high selfing rates in nature.

In a comparison of predominantly inbreeding and outbreeding plant species, allelic diversity of inbreeders was found to be less than half that of outcrossers, although the variance in allelic diversity among selfing populations was higher (Schoen & Brown, 1991). There are also numerous examples of organisms showing lower allelic diversity in selfing than in outcrossing populations or congeners (e.g. Jarne & Charlesworth, 1993; Dudash *et al.*, 1997). For example, lower genetic diversity has been observed in highly selfing populations of *Arabidopsis thaliana* than in populations with a higher degree of outcrossing (Abbot & Gomes, 1989). However, high levels of within-population allelic diversity have also been found within predominantly selfing populations of groups including fungi (Winton *et al.*, 2006), legumes (Siol *et al.*, 2007), proteaceous trees (Ayre *et al.*, 1994), ginger (Zhang & Li, 2008), brittle stars (Boissin *et al.*, 2008), snails (Viard *et al.*, 1996; Trouvé *et al.*, 2003) and killifish (Tatarenkov *et al.*, 2007).

There is thus no simple relationship between mating system and intrapopulation genetic diversity, and little is known about how genetic diversity is maintained in highly selfing species, despite the fact that the partitioning of genetic variation in species with mixed mating systems is an important and well-studied question in evolutionary ecology (Stebbins, 1974; Barrett & Eckert, 1990; Holsinger, 1991; Kalisz *et al.*, 2004; Goodwillie *et al.*, 2005).

In this study, we determined the selfing rate and assessed the degree of genetic diversity in the monocarpic herb *Lobelia inflata* L. (Campanulaceae). *Lobelia inflata* has been assumed to be obligately (autogamously) self-fertilizing in previous studies; because a closed corolla tube prevents pollen release (and instead deposits it on the stigma of the same flower), it is not clear how outcrossing may occur in this species (Ames, 1901; Simons & Johnston, 2000; Hughes & Simons, 2014b,c). However, the distribution of phenotypic variation among lineages in *L. inflata* is suggestive of genetic variation (Hughes *et al.*, 2014), despite the fact that studies of outcrossing congeners show limited allelic diversity at genotyped loci (Antonelli, 2008; Geleta & Bryngelsson, 2012). Self-fertilization may result in multigenerational genetic lineages, as each parent produces homozygous and genotypically identical offspring. The null expectation in self-perpetuating genetic lineages is the rapid erosion of genetic variation through drift. However, genetic variation could be maintained under these circumstances by fluctuating selection on lineages with distinct phenotypes (Winton *et al.*, 2006; Tatarenkov *et al.*, 2007; Leys *et al.*, 2014). Moreover, models of genetic variation in predominantly selfing species have predicted that relatively high levels of genetic variation can be maintained by fluctuating selection (Ellner & Sasaki, 1996; Brys *et al.*, 2011). Therefore, if – through autogamous self-fertilization – *L. inflata* forms genetically identical lineages, temporal fluctuations in fitness among reproductive phenotypes may preserve the diversity of genetic lineages, and therefore of alleles, through time. For *L. inflata*, fitness variance within populations may be related to flowering phenology and reproductive effort, as abiotic conditions vary over the course of the season (Hughes & Simons, 2014a,c).

To date, obligate autogamy in *L. inflata* has been an assumption, based only on circumstantial evidence: anthesis is extremely brief, and fruit development occurs almost immediately after the initiation of flower formation (Simons & Johnston, 2000; Hughes & Simons, 2014a,c). Moreover, in each flower the short stigma is completely enclosed by a tube of anthers, and pollen does not appear to disperse outside flowers. Given the rarity of complete selfing in nature, in this study we collected further evidence concerning the true selfing rate of *L. inflata*. We also assessed the phenotypic consequences of a high degree of selfing, as this should be associated with the formation of genetic lineages, and that there is substantial variation in key fitness traits – especially flowering phenology – in field populations of *L. inflata* (Simons & Johnston, 2003; Hughes & Simons, 2014c). Accordingly, we performed two main analyses. First, we assessed the degree of autogamous selfing and genetic structure of a wild population of *L. inflata* using microsatellite marker diversity found in specimens sampled from a field site. Second, to determine whether genetic lineages were associated with variation in flowering phenology, we evaluated the degree of variation in phenotype that was associated with variation among genetic lineages. Low levels of heterozygosity in a field population despite high allelic diversity would support the hypothesis that selfing is predominant in *L. inflata*; the detection of high levels of heterozygosity would instead be consistent with outcrossing. Results revealing abundant genetic variation among completely selfing lineages would require an explanation. We propose that consistent differences in reproductive trait values among lineages could be explained by temporally variable selection in which the rank order of genotypic fitness is environment-dependent.

Materials and methods

Source population

Lobelia inflata L. (Campanulaceae) is a rosette-forming monocarpic herb distributed throughout eastern North

America. The genus *Lobelia* contains more than 400 species, distributed globally, and is structured by sections and subgenera confined to particular geographic regions; *L. inflata* has been placed in *Lobelia* sect. *Lobelia*, along with most other eastern North American *Lobelia* species – for example *L. cardinalis*, *L. siphilitica* and *L. dortmanna* (Lammers, 2011). *Lobelia inflata* is the only species in this section that is semelparous (Simons & Johnston, 2000; Hughes & Simons, 2014a). All *L. inflata* plants used in this study were descended from mothers sourced from a wild population in the Petawawa Research Forest (Petawawa, Canada 45°99′N, 77°30′W). Although the degree of gene flow between this population and others is unknown, plants from the Petawawa population do not have markedly different phenotypes from plants from populations found elsewhere; common characteristics include a monocarpic life history (as a summer annual or biennial), similar flowering phenology and three flower colour phenotypes (white, pink, and purple) (Anderson, 2003). Wild-grown plants ($n = 21$) represented the parental generation (which we term 'P_1'; here, we use P_1/S_1 parental/selfed labels to indicate generations, not hybrids); seed samples were collected in situ in October 2007.

We constructed progeny arrays representing the selfed offspring (S_1) of the P_1 plants; these arrays were grown under laboratory and field conditions from 2008 to 2010, upon which they were phenotyped and genotyped. Our 21 progeny arrays contained 811 S_1 offspring in total, with 23–42 replicates for each P_1 mother. The S_1 generation was grown under laboratory conditions before bolting to adequately replicate all lineages and to control for maternal effects that might differ among field-harvested plants; however, after bolting, S_1 plants were split between indoor and outdoor growth locations to allow for a test of whether ambient growth conditions affected phenology, either independently or interactively with plant genotype.

Experimental growth conditions

Progeny arrays were created from seed collected from P_1 parents. Batches of 100–200 seed were placed on 70-mm circles of moistened Whatman #5 filter paper (Thermo Fisher Scientific, Waltham MA, USA) inside 4″ Petri plates. Seed plates were germinated in a Biochambers (Winnipeg, MB, Canada) SG-30 germination chamber for 10–14 days under a day/night light cycle of 12 h/12 h, an ambient temperature of 20 °C and an ambient humidity level of 85%.

Upon germination, seedlings were planted in 4 × 4 cm cellpacks of autoclaved topsoil, in trays of 32, and moved to a Biochambers AC-40 growth chamber where they were left to grow under a 16 h at 24 °C/8 h at 18 °C day/night light and temperature cycle, with an ambient humidity of approximately 30%. Each tray was watered with 1 L of water twice per week, and 15 mL of a solution of 5% by weight liquid fertilizer (15-5-15) was added once every 2 weeks. Seedlings were left to grow for 55–65 days until they had formed small rosettes, and were undisturbed until they initiated reproduction (i.e. 'bolted'), which we assessed as the formation of an inflorescence taller than 40 mm. Once they had begun bolting, plants were randomly allocated for translocation to one of two sites: (i) a field site at Carleton University (Ottawa, Canada 45°23′N, 75°41′W), or (ii) another AC-40 growth chamber, designed to emulate the prevailing photoperiod, light intensity and day/night temperature fluctuation at the field site; these laboratory conditions allowed genotypes to respond to photoperiod cues without the confounding effect of stochastic variation in temperature or precipitation (see Hughes & Simons, 2014c). Field plants were planted in the ground and laboratory plants were transferred to the new chamber in their rosette planting trays, and both groups of plants were left to grow until their semelparous reproductive episode was completed and all individuals had senesced. This experiment was replicated for three seasons from 2008 to 2010, in order to expose genotypes assigned to the field group to different environmental conditions. Demographic data for the plants included in the study are shown in Tables S1–S2 in the supplementary material.

Genotyping

We genotyped P_1 and S_1 individuals at 22 microsatellite loci found to be polymorphic in the source population (Hughes *et al.*, 2014), to (i) evaluate the distribution of alleles at microsatellite loci in (S_1) offspring for evidence of outcrossing in the field-raised (P_1) generation, and (ii) assess whether P_1 parents were genetically identical to S_1 offspring, and thus were members of identifiable genetic lineages. Microsatellite genotyping was performed, followed by allele identification and lineage scoring.

Amplification of all microsatellite alleles was performed using a Phire II Direct PCR Kit (Thermo Fisher Scientific), according to the method described in Hughes *et al.* (2014). We used a direct PCR protocol in which DNA extraction and PCR are combined into a single step. A microsample of unprocessed plant tissue was used as the genomic DNA template for amplification, acquired using a 0.35-mm Harris Unicore Micro-Punch (Thermo Fisher Scientific) to obtain tissue from a dried leaf or fruit. We performed PCR according to the kit's recommended protocol for 20 μL reaction volumes: each tube contained a fresh punch of plant tissue, 0.5 μM of each primer, 1.5 mM $MgCl_2$, 1 X Phire Plant PCR Buffer, and 1 μL of Phire Hot Start II DNA Polymerase per reaction. The PCR protocol, which was performed in a T-3000 thermocycler (Biometra, Goettingen, Germany) included the following: (i) an initial

denaturation step of 5 min at 98 °C; (ii) 30 cycles of PCR with a denaturation step of 5 s at 98 °C, an annealing step of 5 s at 50–58 °C (T_m values and other details concerning SSR loci can be found in Hughes *et al.*, 2014) and an extension step of 20 s at 72 °C; and (iii) a final extension step of 10 min at 72 °C, after which the sample was held at 4 °C.

Immediately following amplification, tubes were removed from the thermocycler and PCR products were prepared for high-resolution melt (HRM) analysis. A total of 2.0 μL of a 1/4000 dilution of SYBR Green I (Life Technologies, Carlsbad, CA, USA) was added to each tube; SYBR Green was not included in the original PCR mixture because it may affect the efficiency of the PCR. Tubes were then placed in a Rotor-Gene 6000 thermocycler (QIAGEN Inc., Valencia, CA, USA) for HRM analysis as per Arthofer *et al.* (2011). The HRM protocol included the following: (i) an initial denaturation step of 5 min at 95 °C; (ii) a cooling period of 5 min at 72 °C; and (iii) a melting period, where the temperature ranged from 75 °C to 95 °C, rising by 0.1 °C every 5 s. The HRM curve analysis was performed using Rotor-Gene ScreenClust HRM Software (QIAGEN Inc.), which provided peak melting temperatures and normalized fluorescence difference plots for all samples, which were used to identify which allele was present at a given locus for a given individual. We confirmed differences in allele size using agarose gel electrophoresis to separate amplicons; 6% agarose gels were run at 60V for 75 min and scored relative to a 100-bp GeneRuler DNA Ladder (Life Technologies).

Estimation of selfing and outcrossing rates

We assessed genetic diversity by calculating estimates of F_{IS} (inbreeding coefficient), H_O/H_E using version 4.2 of the GENEPOP software package (Rousset, 2008) to analyse allele frequencies and to calculate linkage disequilibrium between pairwise loci. F_{IS} estimates were computed according to the estimation procedure found in Weir & Cockerham (1984). Multilocus index of association and r_d values were computed using Multilocus 1.3b (Agapow & Burt, 2001). As the molecular markers we used (SSR loci) are codominant, we obtained estimates of the multilocus outcrossing rate using MLTR (Ritland, 2002) and RMES (Gaiotto *et al.*, 1997; Enjalbert & David, 2000; David *et al.*, 2007; Koelling *et al.*, 2012).

We examined multilocus patterns of allelic variation to determine whether *L. inflata* individuals are members of one of a number of nonoutcrossing, genetically identical lineages (see Saltonstall, 2003 and Ivors *et al.*, 2006), as well as how many genetic lineages were represented by our sample of 21 P_1 individuals from the Petawawa Research Forest. Lineage identification was straightforward: across all 22 loci, eight unique patterns of microsatellite alleles were found; hence, all plants were assigned a lineage based on which unique pattern of microsatellite loci they displayed. Across all SSR loci, all plants were found to be 100% homozygous.

High-throughput phenotyping

To assess whether phenotypic variation exists among microsatellite lineages, we examined the association between genetic lineage and four reproductive traits observed under growth-chamber conditions: (i) size at formation of the first flower; (ii) size at formation of the 50th percentile flower; (iii) size at formation of the last flower; and (iv) flower colour. Plants were monitored every 2–4 days from bolting until natural senescence to track size (stalk height) at flower formation. Upon the emergence of each flower, plant stalk height (±0.1 mm) was recorded; only the stalk heights at the formation of the first flower (i.e. Flower 1), the last flower (i.e. Flower x) and the 50th percentile flower (i.e. Flower x/2) are reported here. The total number of flowers produced per plant was determined after the plant had senesced. These phenotypic traits (time of formation of first, last and 50th percentile flower) were chosen because of the strong evidence for phenotypic plasticity in timing of initiation and cessation of semelparous reproduction in *L. inflata*. In particular, the timing of the formation of the 50% percentile flower is a reliable indicator of the 'front-' or 'back-loadedness' of a semelparous reproductive episode (Hughes & Simons, 2014c). Plants expressing a higher proportion of reproductive effort in the days leading up to senescence – and therefore a later date for the formation of the 50th percentile flower – realize a more 'semelparous' pattern of reproduction than those that produce flowers more regularly throughout the season. Throughout an inflorescence, plants produce flowers of only a single, fixed colour; flower colour was determined by visual inspection and was classified as pink, purple or white. Although, because *L. inflata* has no known pollinators, there is no known functional difference between flower colours, flower colour can be used to determine whether microsatellite genotypes are associated with intraspecies variation in phenotypic traits. Randomized positions of genotypes ensured that measurements were not susceptible to subjective bias.

Statistical analysis

To ask whether genetic lineage was a significant predictor of the three phenological traits, we used a hierarchical analysis that included models both with and without genetic lineage (as a random effect), and used likelihood ratio tests to assess the effect of the inclusion of lineage. The first model in each pair was a generalized linear model (GLM) that included only fixed effects, including year (2008, 2009 or 2010), environment (laboratory or field) and year × environment as

fixed effects, as well as prebolting rosette size (in mm) as a covariate. The second model included these predictors, but with added lineage (A-H), lineage × year, lineage × environment and lineage × year × environment as random effects. We used restricted maximum likelihood (REML)-based estimation of variance components for each of the random effects, as it results in a more accurate estimator of variance and covariance than maximum likelihood estimation (Bolker *et al.* 2009). The GLM/GLMM used a Poisson distribution and a logarithmic canonical link function. A likelihood ratio test was used to determine whether the GLMM explained a significantly greater proportion of variation in the value of the reproductive trait than did the GLM (i.e. whether the inclusion of genetic lineage and its associated interaction effects resulted in a model with greater explanatory power). Post hoc tests were performed on genetic lineage and related interaction effects when the inclusion of genetic lineage was deemed significant.

We used chi-square goodness-of-fit tests to determine whether the observed proportions of flower colour within lineages are significantly different from the proportions of flower colour that would be expected by chance, given the proportion in the population as a whole (i.e. white = 0.348, pink = 0.508, purple = 0.144). One goodness-of-fit test was performed for each genetic lineage.

Results

Genotyping

An analysis of microsatellite marker data of 105 plants at 22 marker loci revealed significant polymorphism, but no heterozygous loci, nor any parent–offspring (i.e. P_1-S_1) pairs with different alleles at any SSR locus. Population genetics parameters calculated by GENEPOP (v4.2) are shown in Table S3, and pairwise estimates of linkage disequilibrium are shown in Table S4. The tests of multilocus linkage disequilibrium showed a significant deviation from the null expectation of independence between loci ($I_A = 2.56$, $r_d = 0.125$, $P = 0.001$). Notably, observed heterozygosity was much lower (i.e. zero) than expected heterozygosity given observed allele frequencies in a panmictic population. Our analysis found no evidence of outcrossing in the Petawawa population of *L. inflata*. Both estimates of multilocus selfing rate were higher than 99.9% (mean ± SE: MTLR = 0.9999 ± 2.45E-5; RMES: 0.9998 ± 1.76E-4).

A small sample ($n = 16$) of S_2 individuals – offspring of the S_1 generation reared in 2008 – including plants descended from all P_1 individuals, was also genotyped (data not shown). Allelic genotypes of S_2 individuals were identical to those of their respective S_1 parents and P_1 grandparents. All S_2 plants genotyped were obtained from S_1 plants raised in the field sample, as – given that there is no putative pollination mechanism for *L. inflata* – it was desirable to avoid the confounding fact that S_1 laboratory plants were raised in an environment without access to insect pollinators.

After genotyping P_1 and S_1 plants, we identified eight distinct *L. inflata* genetic lineages from our original sample of 21 individuals from the source population, each of which had a unique pattern of alleles at our 22 SSR loci (shown in Table S5).

Phenotyping

Using a hierarchical modelling approach to assess the significance of genetic lineage as a predictor of three reproductive traits, we found that GLMM predictive models explained significantly more variation for size at first flower and size at 50th percentile flower than did the GLM. However, the GLMM did not explain significantly more variance for size at last flower than the GLM (Table 1). The individual F-tests for all random effects in all GLMMs – including those that did not show better fit than the GLM alternative – are shown in Table 2.

Genetic lineage, genetic lineage × environment and genetic lineage × year were not significant predictors of any phenological traits, but the interaction between genetic lineage, environment and year was a highly significant predictor of variation in size at first flower ($F_{12,\ 766} = 2.13$, $P = 0.004$), size at 50th percentile flower ($F_{12,\ 766} = 2.04$, $P = 0.02$) and size at last flower ($F_{12,\ 766} = 1.97$, $P = 0.02$).

Chi-square goodness-of-fit tests revealed that lineages were strongly associated with flower colour (Table 3) and that all lineages were invariant for colour. Lineages with the same flower colour had a higher proportion of SSR alleles in common (Fig. 1).

Discussion

Lobelia inflata is highly self-fertilizing, yet has high allelic diversity

Complete selfing is extremely rare in nature. Despite the high degree of polymorphism present in the Petawawa population of *L. inflata*, we found no heterozygotes – and hence no evidence of outcrossing – at any loci in any plants we genotyped. The F_{IS} for this sample was 1.00, indicating purely inbred lineages. Estimates of multilocus selfing rates show that the maximum multilocus outcrossing rate consistent with our data is 0.012%, which is much lower than outcrossing rates typically found in mixed mating systems (Wright *et al.* 2013). An outcrossing rate this low supports the hypothesis that *L. inflata* is at the extreme high selfing end of the mating-system continuum and is likely obligately autogamous in the studied population (Lande & Schemske, 1985; Schemske & Lande, 1985). Furthermore – again despite the ubiq-

Table 1 Likelihood ratio test results comparing GLM and GLMM predictive validity for three reproductive traits in *Lobelia inflata*.

Dependent Variable	Model	-2RLL	AIC	Parameters	Chi-square	d.f.	*P*	GLMM>GLM
Size at first flower	GLM	8625.1	8627.1	13	20.65	8	8.01 E-03	Y
	GLMM	8604.4	8622.4	21				
Size at 50th percentile flower	GLM	8768.4	8770.4	13	46.83	8	1.65 E-07	Y
	GLMM	8721.6	8739.6	21				
Size at last flower	GLM	8974.4	8976.4	13	12.46	8	0.13	N
	GLMM	8961.9	8979.9	21				

Table 2 Random effects included in GLMMs predicting three phenological reproductive traits. F-tests were conducted using REML-based estimation of variance parameters.

	Random Effect			
Dependent variable	Genetic lineage	Genetic lineage × Environment	Genetic lineage × Year	Genetic lineage × Environment × Year
Size at first flower				
F	3.00	1.03	1.01	2.13
d.f.	7, 3.52	7, 13.15	12, 12.01	12, 766
P	0.17	0.46	0.49	<0.01**
Size at 50th percentile flower				
F	4.25	0.88	1.31	2.04
d.f.	7, 4.33	7, 13.21	12, 12.01	12, 766
P	0.08	0.55	0.33	0.02*
Size at last flower				
F	2.65	0.79	1.17	1.97
d.f.	7, 3.41	7, 13.25	12, 12.01	12, 766
P	0.21	0.61	0.39	0.02*

*$P < 0.05$
**$P < 0.01$

Table 3 Chi-square goodness-of-test results for flower colour by lineage. Expected frequency of each colour was calculated based on overall population frequencies: white = 0.347, pink = 0.508 and purple = 0.144.

		Observed frequency of flower colour						
Genetic lineage	n	White	Pink	Purple	Expected frequency of observed colour	Chi-square	d.f.	*P*
A	219	219	0	0	73.15	410.82	2	4.94 E-6
B	62	0	0	62	8.94	367.98	2	1.59 E-6
C	190	0	190	0	96.52	184.02	2	3.98 E-6
D	123	0	123	0	62.49	119.10	2	5.51 E-6
E	55	0	0	55	27.94	326.46	2	1.95 E-6
F	24	0	24	0	12.19	23.25	2	3.13 E-4
G	75	0	75	0	38.10	76.64	2	2.17 E-5
H	63	63	0	0	32.00	118.15	2	5.54 E-6

uity of polymorphism – we found that all P_1 and S_1 plants had identical allelic genotypes at all microsatellite loci, and this finding supports the hypothesis that, via autogamy, *L. inflata* forms genetically distinct lineages.

Why *L. inflata* is obligately self-fertilizing – despite the fact that many highly selfing plant species retain the ability to outcross – is unknown. We suggest that the importance of reproductive assurance cannot be considered independently of life history; thus, in a semelparous species such as *L. inflata*, extreme or obligate selfing may be favoured because reproductive assurance becomes relatively more important when the possibility of reproductive failure due to low pollen availability, or mistimed coordination represents a substantial risk (Lloyd & Schoen, 1992; Agren & Schemske, 1993).

The degree of allelic polymorphism present at the 22 microsatellite loci – with a mean of 2.50 alleles per microsatellite locus – is also high given that *L. inflata*

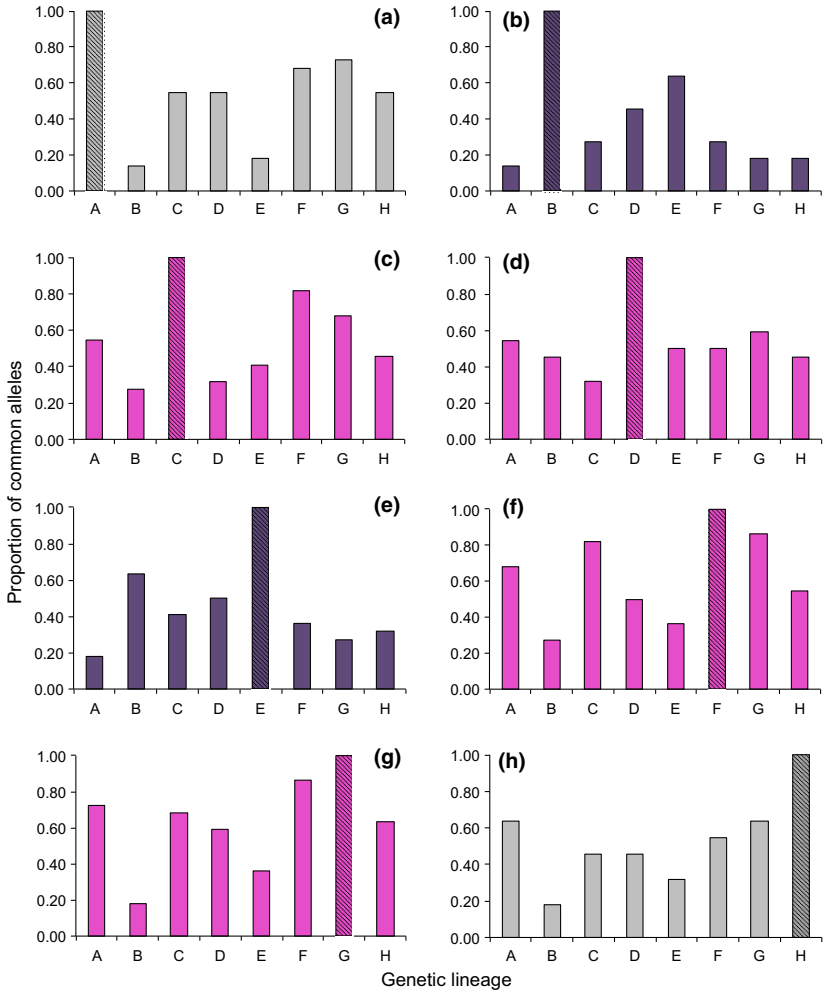

Fig. 1 Proportion of common alleles at 22 polymorphic SSR loci between eight genetic lineages identified in this study. Panels a-h represent genetic lineages A-H, and bar colours indicate flower colour of that lineage (white, purple or pink). Hatched bars show proportions of shared alleles between a genetic lineage and itself (always 22/22 or 1.0). Data on alleles found by lineage are shown in Table S5.

is obligately selfing. Because new haplotypes cannot be generated through recombination in populations of highly inbred selfers, the erosion of allelic diversity is expected over time (Nordborg *et al.*, 2014). Many species with mixed mating systems show lower allelic diversity than *L. inflata*: an analysis of microsatellite loci in the angiosperm *Leavenworthia uniflora* (Brassicaceae), with a selfing rate of approximately 90%, showed a mean allelic diversity of 1.42 alleles per locus (Busch & Werner, 2012), and a similar analysis in the freshwater snail *Lymnaea truncatula*, with a selfing rate of ~80%, showed a mean allelic diversity of 2.36 alleles per locus (Trouvé *et al.*, 2003). Thus, the level of allelic diversity found in the Petawawa population of *L. inflata* was notably higher than has been found in other species with a predominantly selfing mating system. The causes of genetic diversification at these loci are likely worth studying further, as expectations of allelic diversity are relative to effective population size and the strength of background selection, and we cannot exclude the possibility that, at least in part, observed allelic diversity in this population is a result of neutral mutation.

It should also be noted that, because we did not initially know whether *L. inflata* was entirely autogamous, whether genetic variation existed among lineages or how many lineages were present among source plants, the sample of genetic lineages included here is limited, and the proportion of field plants that each lineage accounts for is unknown. Second, although we did not find any heterozygotes at any of the SSR loci we tested, we cannot conclude that absence of proof is proof of absence. Thus, we cannot rule out that outcrossing occurs, if extremely infrequently, in the Petawawa population of *L. inflata* or that outcrossing is more common among individuals from elsewhere in its native range.

Genetic lineages are associated with variation in phenology

Variable selection on genetic lineages may be the mechanism that is responsible for preserving genetic diver-

sity. Our analyses revealed eight distinct genetic lineages in our sample of *L. inflata*, all of which showed allelic differences at three loci or more. We also found significant variation associated with genetic lineage in four traits: size at first flower, size at 50th percentile flower, size at last flower and flower colour. Moreover, we found extensive linkage disequilibrium in this sample. Multilocus and pairwise between-locus measures of linkage disequilibrium suggest nonrandom patterns in inheritance among these alleles, supporting that lineages result from obligate selfing over long timescales. Thus, we conclude that the microsatellite loci we used in this study not only function as reliable genetic labels, but also identify genetic lineages that differ in observable phenotypes such as flower colour.

Notably, rather than genetic lineage alone, it was the interaction between genetic lineage, environment and year that was the most important predictor of the three phenological traits. This makes sense given that substantial variation in key environmental parameters – that is temperature, rainfall and wind – existed between experimental environments as well as between years (see Figs S1–S3). This, coupled with the fact that *L. inflata*'s reproductive phenotype is highly sensitive to environmental variation (see Simons & Johnston, 2000; Hughes & Simons, 2014b,c), makes it unsurprising that interactions including genetic lineage, rather than genetic lineage alone, predict variation in reproductive traits.

We speculate that genetic variation among lineages is maintained by genotype-by-environment interaction through time: coarse-grained fluctuating selection favours different phenotypes (associated with the genetic lineages) at different times. This hypothesis has also been proposed as the cause of relatively high allelic diversity in other highly selfing species, although in these cases the association between microsatellite lineage and phenotype has not been established (e.g. Winton *et al.*, 2006). We acknowledge that a rigorous test of this hypothesis is beyond the scope of this study; our data come from a single source population in a single year, whereas providing proof of fluctuating selection in the field would involve measuring relative lineage fitness in the field over many seasons. However, we suggest that our data are more consistent with temporal fluctuating selection than with alternative explanations including selection under spatial variation. Further research on genetic lineage fitness *in situ* should be performed to better substantiate this hypothesis. Studying the relationship between populations of *L. inflata* at Petawawa and elsewhere in its native range may further clarify the degree to which gene flow shapes environmental adaptation; for example, it is possible that the Petawawa population of *L. inflata* may be a refugial sink, and the high degree of genetic variation reflects a history of expansion and contraction over long timescales.

In conclusion, we have shown that the population of *Lobelia inflata* studied is obligately or near-obligately self-fertilizing and – although there is no genetic variation among offspring – substantial genetic variation occurs among lineages. Furthermore, genetic lineages accounted for significant differences in reproductive traits. We hypothesize that genetic variation among lineages may be maintained by fluctuating selection, specifically by temporal genotype-by-environment interaction in which different lineages are favoured at different times.

Acknowledgments

The authors thank M. Compton, J. Graham, K. Côté, H. Rundle and T. Sherratt for their helpful contributions to study design, and S. Aitken, C. Davis and A. Jaworski for their assistance in designing the genotyping protocol. Finally, the authors also thank P. Arbour, the Petawawa Research Forest, and the Canadian Wood Fibre Centre for access to long-term field sites. This work was supported through an NSERC CGS to PWH and an NSERC Discovery Grant to AMS.

References

Abbot, R. & Gomes, M. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**: 411–418.

Agapow, P.-M. & Burt, A. 2001. Indices of multilocus linkage disequilibrium. *Mol. Ecol. Notes* **1**: 101–102.

Agren, J. & Schemske, D. 1993. Outcrossing and inbreeding depression in two annual monoecious herbs *Begonia hirsuta* and *Begonia semiovata*. *Evolution* **47**: 125–135.

Ames, O. 1901. *Lobelia inflata* x *cardinalis*. *Rhodora* **3**: 296–298.

Anderson, M.K. 2003. *Indian Tobacco—USDA Plant Guide*. USDA NRCS National Plant Data Center, Davis, CA, USA.

Antonelli, A. 2008. Higher level phylogeny and evolutionary trends in Campanulaceae subfam. Lobelioideae: molecular signal overshadows morphology. *Mol. Phylogenet. Evol.* **46**: 1–18.

Armbruster, P. & Reed, D. 2005. Inbreeding depression in benign and stressful environments. *Heredity* **95**: 235–242.

Arthofer, W., Steiner, F. & Schlick-Steiner, B. 2011. Rapid and cost-effective screening of newly identified microsatellite loci by high-resolution melting analysis. *Mol. Genet. Genomics* **286**: 225–235.

Ayre, D., Whelan, R. & Reid, A. 1994. Unexpectedly high levels of selfing in the Australian shrub *Grevillea barklyana* (Proteaceae). *Heredity* **72**: 168–174.

Barrett, S. 2002. The evolution of plant sexual diversity. *Nat. Rev. Genet.* **3**: 274–284.

Barrett, S. & Eckert, C. 1990. Variation and evolution of mating systems in seed plants. In: *Biological Approaches and Evolutionary Trends in Plants* (S. Kawano, ed.), pp. 229–254. Academic Press, Tokyo, Japan.

Barrett, S. & Harder, L. 1996. Ecology and Evolution of Plant Mating. *Trends Ecol. Evol.* **11**: 73–79.

Boissin, E., Hoareau, T., Féral, J. & Chenuil, A. 2008. Extreme selfing rates in the cosmopolitan brittle star species complex *Amphipholis squamata*: data from progeny-array and heterozygote deficiency. *Mar. Ecol. Prog. Ser.* **361**: 151–159.

Bolker, B., Brooks, M., Clark, C., Geange, S., Poulsen, J., Stevens, M., *et al.* 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**: 127–135.

Brys, R., de Crop, E., Hoffmann, M. & Jacquemyn, H. 2011. Importance of autonomous selfing is inversely related to population size and pollinator availability in a monocarpic plant. *Am. J. Bot.* **98**: 1834–1840.

Busch, J.W. & Werner, W.J. 2012. Population structure has limited fitness consequences in the highly selfing plant *Leavenworthia uniflora* (Brassicaceae). *Int. J. Plant Sci.* **173**: 495–506.

Charlesworth, D. & Charlesworth, B. 1995. Quantitative genetics in plants: the effect of the breeding system on genetic variability. *Evolution* **49**: 911–920.

Crnokrak, P. & Barrett, S. 2002. Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56**: 2347–2358.

Darwin, C. 1877. *Effects of Cross and Self-Fertilization in the Vegetable Kingdom*. John Murray, London, UK.

David, P., Pujol, B., Viard, F., Castella, V. & Goudet, J. 2007. Reliable selfing rate estimates from imperfect population genetic data. *Mol. Ecol.* **16**: 2474–2487.

Dudash, M., Carr, D. & Fenster, C. 1997. Five generations of enforced selfing and outcrossing in *Mimulus guttatus*: inbreeding depression variation at the population and family level. *Evolution* **51**: 54–65.

Eckert, C. & Barrett, S. 1994. Inbreeding depression in partially self-fertilizing *Decodon verticillatus* (Lythraceae): population-genetic and experimental analyses. *Evolution* **48**: 952–964.

Ellner, S. & Sasaki, A. 1996. Patterns of genetic polymorphism maintained by fluctuating selection with overlapping generations. *Theor. Popul. Biol.* **50**: 31–65.

Enjalbert, J. & David, J. 2000. Inferring recent outcrossing rates using multilocus individual heterozygosity: application to evolving wheat populations. *Genetics* **156**: 1973–1982.

Gaiotto, F., Bramucci, M. & Grattapaglia, D. 1997. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. *Theor. Appl. Genet.* **95**: 842–849.

Geleta, M. & Bryngelsson, T. 2012. Population genetic analysis of Lobelia rhynchopetalum Hemsl. (Campanulaceae) using DNA sequences from ITS and eight chloroplast DNA regions. *Sci. World J.* **26471**: 1–10.

Goodwillie, C., Kalisz, S. & Eckert, C. 2005. The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **36**: 47–79.

Hereford, J. 2010. Does selfing or outcrossing promote local adaptation? *Am. J. Bot.* **97**: 298–302.

Holsinger, K. 1991. Mass-action models of plant mating systems: the evolutionary stability of mixed mating systems. *Am. Nat.* **138**: 606–622.

Hughes, P. & Simons, A. 2014a. Changing reproductive effort within a semelparous reproductive episode. *Am. J. Bot.* **101**: 1323–1331.

Hughes, P. & Simons, A. 2014b. Secondary reproduction in the herbaceous monocarp *Lobelia inflata*: time-constrained primary reproduction does not result in increased deferral of reproductive effort. *BMC Ecol.* **14**: 1–10.

Hughes, P. & Simons, A. 2014c. The continuum between semelparity and iteroparity: plastic expression of parity in response to season length manipulation in *Lobelia inflata*. *BMC Evol. Biol.* **14**: 90.

Hughes, P., Jaworski, A., Davis, C., Aitken, S. & Simons, A. 2014. Development of polymorphic microsatellite markers for Indian Tobacco, *Lobelia inflata* (Campanulaceae). *Appl. Plant Sci.* **2**: 1300096.

Husband, B. & Schemske, D. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**: 54–70.

Igic, B. & Busch, J.W. 2013. Is self-fertilization an evolutionary dead end? *New Phytol.* **198**: 386–397.

Ivors, K., Garbelotto, M., Vries, I.D.E., Ruyter-Spira, C., Hekkert, B.T.E., Rosenweig, N. *et al.* 2006. Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Mol. Ecol.* **15**: 1493–1505.

Jain, S. 1976. The evolution of inbreeding in plants. *Annu. Rev. Ecol. Syst.* **7**: 469–495.

Jarne, P. & Charlesworth, D. 1993. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annu. Rev. Ecol. Syst.* **24**: 441–466.

Johnston, M. & Schoen, D. 1996. Correlated evolution of self-fertilization and inbreeding depression: An experimental study of nine populations of *Amsinckia* (Boraginaceae). *Evolution* **50**: 1478–1491.

Kalisz, S., Vogler, D. & Hanley, K. 2004. Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature* **430**: 884–887.

Koelling, V., Monnahan, P. & Kelly, J. 2012. A Bayesian method for the joint estimation of outcrossing rate and inbreeding depression. *Heredity* **109**: 393–400.

Lammers, T.G. 2011. Revision of the infrageneric classification of *Lobelia* L. (Campanulaceae, Lobelioideae). *Ann. Mo. Bot. Gard.* **98**: 37–62.

Lande, R. & Schemske, D. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**: 24–40.

Leys, M., Petit, E., El-Bahloul, Y., Liso, C., Fournet, S. & Arnaud, J.-F. 2014. Spatial genetic structure in *Beta vulgaris* subsp. maritima and *Beta macrocarpa* reveals the effect of contrasting mating system, influence of marine currents, and footprints of postglacial recolonization routes. *Ecol. Evol.* **4**: 1828–1852.

Lloyd, D. & Schoen, D. 1992. Self- and cross-fertilization in plants. I. Functional dimensions. *Int. J. Plant Sci.* **153**: 358–369.

Nordborg, M., Charlesworth, B. & Charlesworth, D. 2014. Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc. R. Soc. B* **263**: 1033–1039.

Ritland, K. 2002. Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**: 221–228.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**: 103–106.

Saltonstall, K. 2003. Microsatellite variation within and among North American lineages of *Phragmites australis*. *Mol. Ecol.* **12**: 1689–1702.

Schemske, D. & Lande, R. 1985. The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. *Evolution* **39**: 41–52.

Schoen, D. & Brown, A. 1991. Intraspecific variation in population gene diversity and effective population size correlates with mating system in plants. *Proc. Natl Acad. Sci. USA* **88**: 4494–4497.

Simons, A. & Johnston, M. 2000. Plasticity and the genetics of reproductive behaviour in the monocarpic perennial, *Lobelia inflata* (Indian tobacco). *Heredity* **85**: 356–365.

Simons, A. & Johnston, M. 2003. Suboptimal timing of reproduction in *Lobelia inflata* may be a conservative bet-hedging strategy. *J. Evol. Biol.* **16**: 233–243.

Siol, M., Bonnin, I., Olivieri, I., Prosperi, J. & Ronfort, J. 2007. Effective population size associated with self-fertilization: lessons from temporal changes in allele frequencies in the selfing annual *Medicago truncatula*. *J. Evol. Biol.* **20**: 2349–2360.

Stebbins, G. 1974. *Flowering Plants: Evolution Above the Species Level*. Harvard University Press, Cambridge, MA, USA.

Takebayashi, N. & Morrell, P.L. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am. J. Bot.* **88**: 1143–1150.

Tatarenkov, A., Gao, H., Mackiewicz, M., Taylor, D., Turner, B. & Avise, J. 2007. Strong population structure despite evidence of recent migration in a selfing hermaphroditic vertebrate, the mangrove killifish (*Kryptolebias marmoratus*). *Mol. Ecol.* **16**: 2701–2711.

Trouvé, S., Degen, L., Renaud, F. & Goudet, J. 2003. Evolutionary implications of a high selfing rate in the freshwater snail *Lymnaea truncatula*. *Evolution* **57**: 2303–2314.

Viard, F., Bremond, P., Labbo, R. & Justy, F. 1996. Microsatellites and the genetics of highly selfing populations in the freshwater snail *Bulinus truncatus*. *Genetics* **142**: 1237–1247.

Weir, B. & Cockerham, C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Willis, J. 1999. The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution* **53**: 1678–1691.

Winton, L., Hansen, E. & Stone, J. 2006. Population structure suggests reproductively isolated lineages of *Phaeocryptopus gaeumannii*. *Mycologia* **98**: 781–791.

Wright, S., Kalisz, S. & Slotte, T. 2013. Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. B* **280**: 20130133.

Zhang, Z.-Q. & Li, Q.-J. 2008. Autonomous selfing provides reproductive assurance in an alpine ginger *Roscoea schneideriana* (Zingiberaceae). *Ann. Bot.* **102**: 531–538.

Zhang, C., Zhou, G.-Y., Yang, Y.-P. & Duan, Y.-W. 2014. Better than nothing: evolution of autonomous selfing under strong inbreeding depression in an alpine annual from the Qinghai-Tibet plateau. *J. Syst. Evol.* **52**: 363–367.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Summary demographic data for S_1 plants used in this study by year and environment.

Table S2 Lineage frequencies of S_1 plants used in this study by year and environment.

Table S3 Genetic diversity summary data by locus for *Lobelia inflata* sampled from the Petawawa Research Forest.

Table S4 Pairwise comparisons testing pairwise linkage disequilibrium between the 22 polymorphic loci.

Table S5 Observed alleles at 22 microsatellite loci and flower colour phenotype for 21 P_1 individuals.

Figure S1 Mean monthly temperature (°C) at Petawawa by year from 2008–10.

Figure S2 Maximum monthly temperature (°C) at Petawawa by year from 2008–10.

Figure S3 Mean monthly precipitation (in mm) at Petawawa by year from 2008–10.

Data deposited at Dryad: doi: 10.5061/dryad.p2c0b

Received 17 April 2015; revised 9 August 2015; accepted 17 August 2015