

# Analysis of mating system and genetic structure in the endangered, amphicarpic plant, Lewton's polygala (*Polygala lewtonii*)

Joel F. Swift<sup>1</sup> · Stacy A. Smith<sup>2</sup> · Eric S. Menges<sup>2</sup> · Burgund Bassüner<sup>1</sup> · Christine E. Edwards<sup>1</sup>

Received: 1 June 2015 / Accepted: 20 June 2016 / Published online: 4 July 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** *Polygala lewtonii* is a federally endangered, amphicarpic plant with a mixed mating system and three types of flowers: (1) aboveground, chasmogamous flowers (i.e., open-pollinated; CH), (2) aboveground, cleistogamous flowers (i.e., closed, selfing; CL) and (3) CL flowers on belowground stems (amphicarpy). Aboveground seeds are ant-dispersed, whereas belowground seeds are spaced across the length of the rhizome. Here, we collected individuals of *P. lewtonii* at both range-wide and fine geographic scales and genotyped them at 11 microsatellite loci. We analyzed patterns of genetic diversity and structure to understand: (1) the predominant mating system (selfing or outcrossing), (2) the movement of pollen and seeds across the landscape, and (3) the optimal strategy to conserve the full range of genetic variation. *P. lewtonii* reproduces predominantly by selfing or bi-parental inbreeding, but reproduction occurred through each of the three flower types. Some individuals produced by selfing/inbreeding were tightly clustered spatially, and were likely produced either by belowground flowers or by aboveground flowers with limited seed dispersal. Other selfed/inbred individuals were spatially separated (maximum of 15 m), and were likely produced by aboveground flowers

followed by seed dispersal by ants. Fine-scale patterns of genetic structure indicate that some gene flow is occurring among aboveground CH flowers but both pollen and outcrossed seeds are moving limited distances (maximum of 0.5 km). Because genetic variation is structured at a fine spatial scale, protecting many populations is necessary to fully conserve the genetic variation in *P. lewtonii*. Conservation seed banking, if accompanied by research on seed germination requirements, may also contribute to the effective protection of genetic variation in *P. lewtonii*.

**Keywords** Amphicarpy · Ant dispersal · Conservation · Genetic structure · Mixed mating system · Microsatellite · Self-fertilization

## Introduction

Plants have developed a variety of reproduction strategies that allow them to pass on genetic material, the goal of every organism, even in the face of changing climates and environmental stress. Angiosperm species demonstrate a wide range of mating systems that range from ~100 % outcrossing to ~100 % selfing (Schemske and Lande 1985; Husband and Schemske 1996). The rate of outcrossing and selfing is strongly affected by floral morphology (Brunet and Eckert 1998), and in particular, the degree to which flowers open for pollination.

Many flowers are chasmogamous (CH), which refers to flowers that fully open and can be pollinated by a variety of agents, such as wind, bats, birds, and insects (Heinrich 1975), thereby promoting outcrossing among genetically different individuals of a species. Outcrossing can be advantageous because recombination may yield new combinations of genotypes and random mating may

**Electronic supplementary material** The online version of this article (doi:10.1007/s10592-016-0860-3) contains supplementary material, which is available to authorized users.

✉ Christine E. Edwards  
Christine.Edwards@mobot.org

<sup>1</sup> Center for Conservation and Sustainable Development, Missouri Botanical Garden, PO Box 299, St. Louis, MO 63166, USA

<sup>2</sup> Plant Ecology Program, Archbold Biological Station, 123 Main Dr., Venus, FL 33960, USA

increase heterozygosity, which masks deleterious recessive alleles, resulting in individuals with greater fertility or biomass (Schemske and Pautler 1984; Meyer et al. 2004; Birchler et al. 2010). However, because outcrossing is dependent on the presence of both conspecifics and a mechanism to pollinate the plant, reproduction by outcrossing is not guaranteed. Flowers may also be cleistogamous (CL), which refers to flowers that remain closed and are obligately self-fertilized (Darwin 1877; Lord 1981). Selfing provides reproductive assurance regardless of whether conspecifics and an available mechanism to pollinate the plant are present (Stebbins 1957; Jain 1976; Barrett and Shore 1987; Lloyd 1992; Kalisz et al. 2004; Schueller 2004; Busch 2005; Herlihy and Eckert 2005). Because each selfed offspring inherits 100 % of its genes from the mother plant, if the mother is well adapted to her environment and the environment is stable, this will lead to high fitness (Solbrig 1976; Holsinger 1991). However, selfed offspring have low genetic diversity, which may limit their ability to withstand fluctuating environmental conditions (Solbrig 1976; Holsinger 2000). Selfed offspring may also suffer from inbreeding depression, in which offspring are homozygous for deleterious recessive alleles and consequently express a phenotype with low vigor (Brown and Munday 1982; Clay and Antonovics 1985; Schemske and Lande 1985; Charlesworth and Charlesworth 1987; Husband and Schemske 1996; Berg and Redbo-Torstensson 1999; Eckert 2000; Kalisz et al. 2004). In a predominantly selfing species, selection may eliminate offspring with low vigor, thereby purging deleterious recessive alleles from the population (Charlesworth and Charlesworth 1987; Swindell and Bouzat 2006), minimizing subsequent inbreeding depression.

Some plant species produce more than one type of flower, maintaining both open-pollinated CH flowers and self-pollinated CL flowers on the same plant (Schemske 1978; Lord 1981; Jasieniuk and Lechowicz 1987; Berg and Redbo-Torstensson 1999), or a so-called “best of both worlds” mating system (Cruden and Lyon 1989), which can provide reproductive assurance regardless of whether conditions are conducive to cross pollination (Kalisz and Vogler 2003). Some of these plant species have a very rare breeding system termed amphicarpy, in which they produce flowers both aboveground and belowground (Cheplick 1987; Kaul et al. 2000), and thus produce two types of seeds (i.e., aboveground and belowground). The two types of seeds may differ in size, dispersal pattern, nutrient availability, germination requirements, genetic composition, and ecological role (Clay and Antonovics 1985; Cheplick 1987, 1994; Clavijo 1995; Liang et al. 2009; Sadeh et al. 2009). In amphicarpic plants, the aboveground flowers produce seeds (by either selfing or outcrossing) that can be dispersed across the landscape.

The belowground flowers are self-fertilized and produce fruits that are dispersed only as far as flowers are spaced along the rhizome from the parent plant (Weiss 1980; Kaul et al. 2000; Barker 2005). Seeds produced below the ground may be a means to regenerate quickly in the absence of competition after a fire (Cheplick 1987). Belowground seeds are often exceptionally well provisioned by the parent plant and produce progeny with higher fitness in the local habitat relative to aboveground seeds (Cheplick 2005), especially in stressful environments. Amphicarpy is very rare, occurring in only around 36 plant species in ten families (Kaul et al. 2000). A few of these amphicarpic plant species produce three types of flowers: both CH and CL flowers aboveground and CL flowers belowground. This unusual mating system is found in *Amphicarpaea bracteata* (Joseph Trapp and Hendrix 1988; Baskin and Baskin 2014), *A. edgeworthii* (Zhang et al. 2006; Baskin and Baskin 2014), *Polygala polygama* (Holm 1929; Cheplick 1987), and *Polygala lewtonii* (Weekley and Brothers 2006).

The mating system and seed dispersal strategy of a plant species strongly affect the partitioning of genetic variation within and among populations (Loveless and Hamrick 1984; Hamrick and Godt 1989, 1996; Sun 1999; Holsinger 2000), which has important implications for protecting the genetic diversity of a species of conservation concern. A plant species that is predominantly selfing and has limited seed dispersal will have little genetic diversity maintained within populations and most genetic diversity partitioned among populations (Brown 1979; Hamrick 1982; Cole and Biesboer 1992; Sun 1999; Holsinger 2000; Siol et al. 2008; Dick et al. 2014) as such, because each population holds a unique set of alleles, it is necessary to protect many populations at a fine geographic scale to fully protect genetic diversity (Furches et al. 2009). In contrast, a plant species that is predominantly outcrossing and has both pollinators and seed dispersers that effectively move genetic material across the landscape will maintain most of its genetic diversity within populations and little variation will be partitioned among populations (Brown 1979; Hamrick 1982). In this case, conservation of a few larger populations will likely fully protect the genetic diversity of a species.

Because designing an effective strategy to conserve the full range of genetic variation in an endangered species depends on how its genetic variation is partitioned within and among populations, it is important to assess patterns of genetic structure in each species of interest (Holsinger and Gottlieb 1991). Assessing patterns of genetic structure is particularly important in an endangered species with a mixed mating system, as it may be unclear whether inbreeding or outcrossing is the predominant mode of reproduction, thereby making it difficult to infer how genetic

variation will be partitioned within and among populations based on mating system alone (Jain 1976; Travis et al. 2004). This is likewise important in species that have more than one type of seed or dispersal strategy, as it may be unclear which is predominantly employed by the species, limiting the inferences that can be made about the structuring of genetic variation based on dispersal strategy.

*Polygala lewtonii* Small (Polygalaceae) is a federally endangered perennial herb that has three flower types: both aboveground and belowground self-pollinated CL flowers as well as aboveground, open pollinated CH flowers. The aboveground CH flowers of *P. lewtonii* are insect-pollinated. All three flower types produce seeds that are similar in appearance but may differ in size, and all have an elaiosome that attracts ants, which are the primary seed dispersers of the aboveground seeds of this species. Field observations have found that ants may disperse aboveground seeds (which may be produced by either selfing or outcrossing) many meters away from the maternal plant. There is no evidence that the seeds produced by belowground CL flowers are accessed or dispersed by ants; they are therefore spaced from the parental stem only as far as a rhizome travels belowground and are likely to germinate very close to the maternal plant (Weekley and Brothers 2006).

*Polygala lewtonii* is endemic to five counties in central Florida, USA, where it occupies xeric, pyrogenic, upland habitats both on the Lake Wales Ridge and on the Mount Dora Ridge in the Ocala National Forest. Populations are exclusive to yellow sands (Menges et al. 2007b) in longleaf pine/wiregrass sandhill and oak hickory scrub. This habitat has experienced widespread land conversion to agriculture and residential housing, and many patches of remaining habitat have become degraded due to fire suppression (USFWS 1999; Weekley et al. 2008), thereby causing severe population declines in *P. lewtonii*. Currently, around 35 % of the extant populations of *P. lewtonii* are unprotected, many of which will likely be lost as development and fire suppression continues (USFWS 2009). Inadequate prescribed fire and land management practices may also threaten populations in some protected areas (USFWS 2009). Given that this species has already experienced declines and will likely face additional losses of populations in the future, questions have arisen about the most effective way to adequately protect as much of its genetic variation as possible in this highly threatened species. Very little is known about the relative rates of reproduction via selfing and outcrossing in this species and how this affects the partitioning of genetic variation within and among populations. This information is important for designing conservation plans that will effectively conserve as much of the full range of genetic variation in this endangered species as possible.

In this study, our goal was to use genetic data to provide insight into the reproductive biology of *P. lewtonii*. We used 11 microsatellite markers to assess patterns of genetic diversity and structure within and among populations of *P. lewtonii* on both range-wide (across Central Florida) and local scales ( $\sim 0.28$  km<sup>2</sup> area). Specifically, our goals were to: (1) use levels of genetic diversity to understand the predominant mating system (i.e., selfing or outcrossing) employed by *P. lewtonii*, (2) use patterns of genetic structure to understand both the migration of pollen and selfed/outcrossed seeds across the landscape, and (3) use these data to help develop strategies to effectively conserve the full range of genetic variation in *P. lewtonii*.

## Materials and methods

### Study species

*Polygala lewtonii* is a small ( $\leq 20$  cm), short-lived (up to 10 years), perennial herb that produces one to several annual stems (Weekley and Menges 2012). Open, CH flowers occur on terminal racemes and are insect-pollinated. CH flowers are bisexual, showy, colored bright pink to purple, open from February through April, and have high fruit set (71.2 %) (Weekley and Brothers 2006). Each CH flower is open for 2–4 days (Weekley and Brothers 2006). As in other members of its genus, delayed selfing has been documented in CH flowers; however, this mechanism has been shown to provide minimal reproductive assurance in *P. lewtonii* (Weekley and Brothers 2006). Small, self-pollinating, CL flowers occur in the leaf axils of aboveground racemes and along belowground stems, and are generally produced from June to January. Sexual maturation may occur as quickly as 1 year after germination, with CH flowers typically preceding both forms of CL flowers (Koontz et al. 2015; Weekley and Menges 2012). Fruits of belowground CL flowers are heavier and are produced later in life than aboveground fruits (Koontz et al. in preparation). The fruit of all three types of flowers is a dehiscent, two-seeded capsule and contains an elaiosome, a fleshy appendage thought to attract ants. Ants have been observed to vigorously collect aboveground seeds, transport them to the nest, and remove elaiosomes (Menges and Weekley 2003). Germination was found to be independent of elaiosome presence in one experimental study (Menges et al. 2014).

Fire appears to be essential to *P. lewtonii*, as it kills individuals but promotes seedling recruitment, possibly through smoke cues (Lindon and Menges 2008; Weekley and Menges 2012). *P. lewtonii* responds to fire through mass germination and seedling recruitment from a persistent soil seed bank (Weekley and Menges 2012). In the

absence of fire, populations may persist though seedling recruitment pulses that coincide with high rainfall pulses during wet winters (Weekley and Menges 2012). Seeds are capable of forming a long-lived, persistent seedbank (Weekley and Brothers 2006).

#### Sample collection and DNA extraction

The samples for this study were collected at two spatial scales. To assess range-wide genetic structure, we collected samples from each of three publicly protected lands between 2008 and 2009, including the Ocala National Forest (ONF), Scrub Point Preserve (SPP), and Lake Wales Ridge State Forest (LWR; Fig. 1). These three sites were chosen because they, together with one additional sampling locality in Carter Creek (described in the following paragraph), span the full geographic range of the species (Fig. 1). At each site, we sampled leaf material from 24 individuals of *Polygala lewtonii*, with sampled individuals spaced several meters apart whenever possible to ensure that different individuals were likely to be sampled. This sampling allowed insights into gene flow on a range-wide scale.

To assess fine-scale patterns of genetic structure within a population, we sampled 288 individuals within a study area at Carter Creek Lake Wales Ridge Wildlife and Environmental Area in spring 2014 (CC; Figs. 1, 2a). We chose a spatially stratified sampling scheme to randomly sample individuals in this population at a fine spatial scale. The study area was defined by compiling all known locations (GPS points) for *P. lewtonii* at CC. Within the defined study area (estimated range is  $\sim 0.28$  km<sup>2</sup> area), we created four blocks to spatially structure our leaf collections across the entire population. To establish collection plots, we generated random points buffered by 10 m in ArcGIS version 10.1 for each of the four blocks. In the field we navigated to each point, in a random order within each block, until we established eight collection plots per block. To establish a collection plot, we required a minimum of nine *P. lewtonii* individuals with >20 leaves to be within a 5 m radius of the randomly generated collection point. Centered on each collection point, we established a circular collection plot with a minimum of 1 m radius (range 1–4 m radius) separated by a minimum of 10 m between collection plots in the same block, with the closest plots in neighboring blocks separated by larger distances (mean = 350 m, range 140–630 m between neighboring blocks). Within each collection plot, we counted all *P. lewtonii* individuals falling within the plot boundaries. We then randomly selected nine *P. lewtonii* individuals to collect leaves for genetic analysis and recorded the stage class of each sampled individual (seedling, vegetative, flowering).

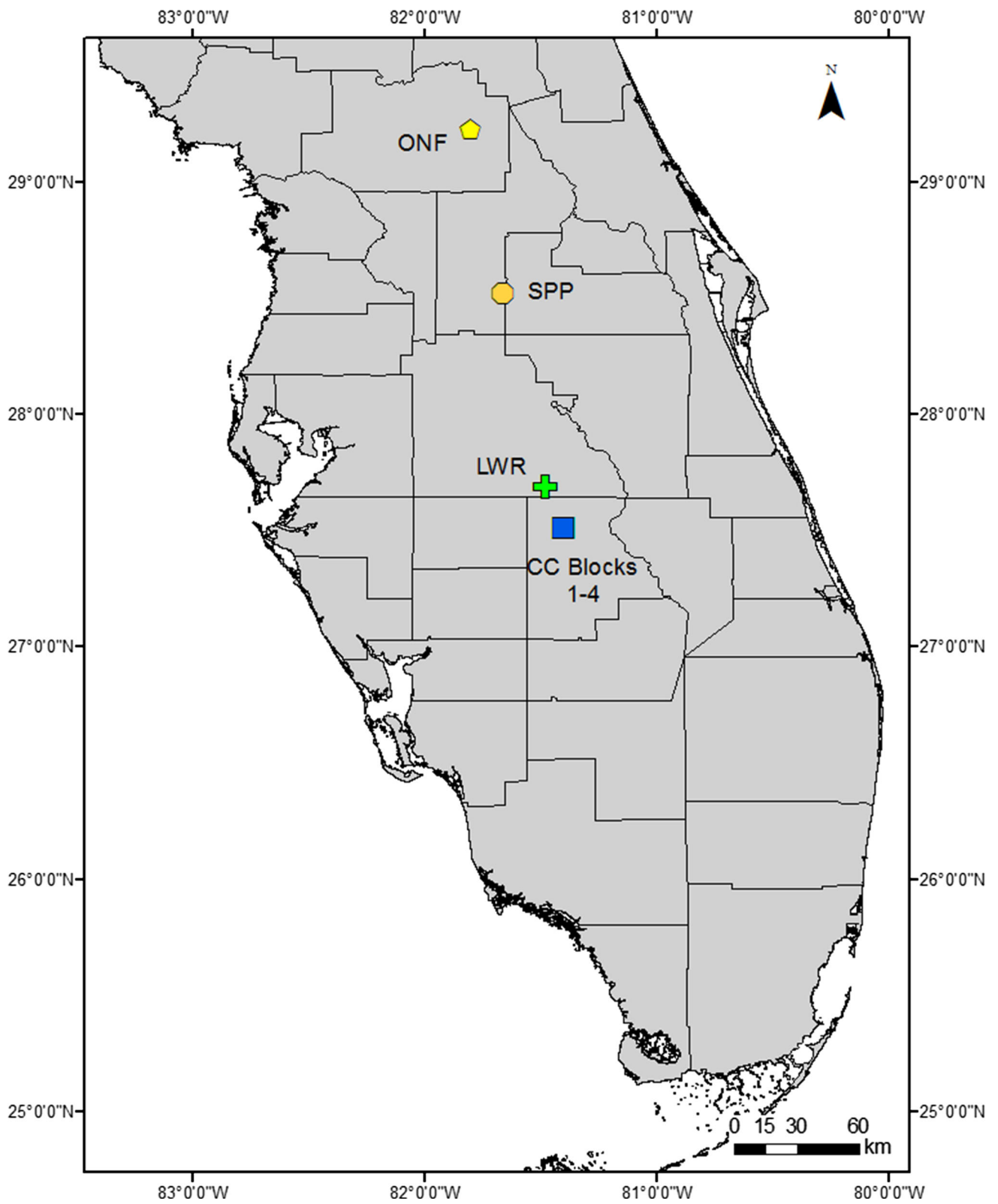
This spatially stratified sampling scheme was designed to understand the structuring of genetic variation at an

extremely fine scale, with distances within and among plots and blocks designed to sample at the range of distances that could be potentially traveled by pollinators and selfed and outcrossed seeds. Plants sampled within plots were spaced an average of 1.34 m (range 0.04–6.9 m), designed to capture the minimum distance that a seed may travel, either because it was produced aboveground but not ant dispersed, or because it was produced belowground and could only be spaced as far as a rhizome can grow. The distance between plots (minimum of 10 m) was designed to capture seed dispersal distances that could be traveled by aboveground, ant-dispersed seeds. The distances between blocks (mean = 0.64 km, range 0.14–1.22 km) were designed to capture the larger distances that pollinators may travel; blocks were spaced at varying distances from each other to capture the range in potential distances that pollinators may travel. Seed dispersal between blocks is unlikely because of the short distances traveled by ants. By analyzing patterns of genetic diversity and structure in populations sampled at this fine scale, our goal was to use the spatial spread of genotypes and genetic clusters to provide some insight into the patterns of reproduction by the three different flower types in *P. lewtonii*.

In total, 288 unique *P. lewtonii* leaf samples were collected in CC, while an additional 72 samples were collected in the range-wide sampling, resulting in a total of 360 samples. DNA was extracted from each sample using a cetyltrimethyl ammonium bromide (CTAB) protocol (Doyle and Doyle 1987), which was modified by using smaller volumes and adding an additional wash step with 95 % ethanol.

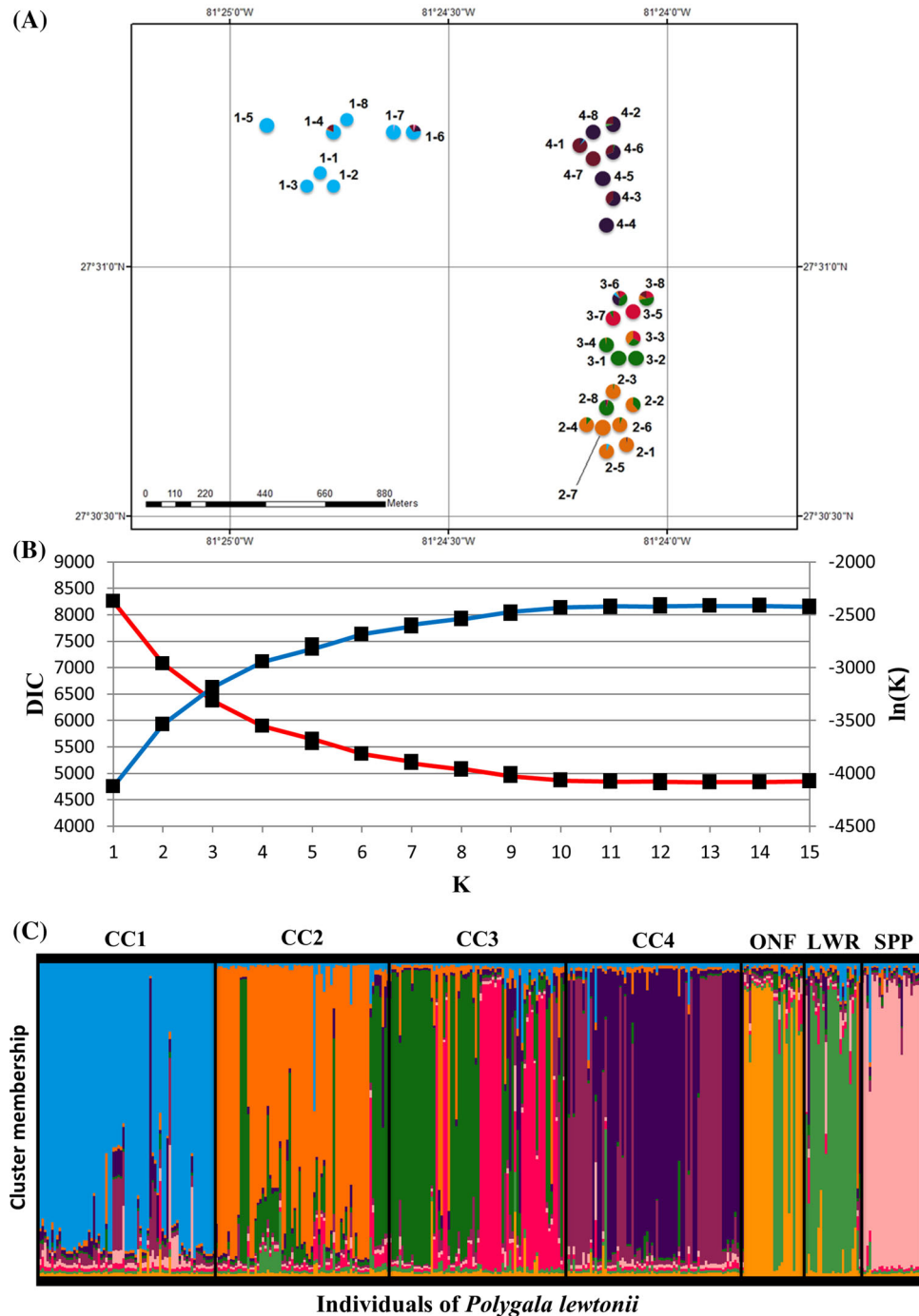
#### Microsatellite marker development and genotyping

To identify microsatellites, we carried out shotgun sequencing of genomic DNA of one individual of *P. lewtonii* using an Illumina MiSeq. Library preparation was carried out using Nextera DNA sample prep kits and Nextera index kits (Illumina). The indexed sample of *P. lewtonii* was pooled with samples from other studies and sequenced using  $2 \times 150$  bp paired-end reads. We trimmed low-quality bases of the resulting reads and assembled the reads *de novo* into contigs using the Medium sensitivity/fast setting using Geneious version R7 (Biomatters Inc.). We searched contig consensus sequences for di-, tri- and tetra-nucleotide microsatellite repeats using MSAT-COMMANDER version 1.04 (Faircloth 2008). Polymerase chain reaction (PCR) primers were designed from a subset of these sequences using the default settings in PRIMER3 (Rozen and Skalesky 1999; Faircloth 2008). We added an M13 tag (5'-CACGACGTTGTAAAACGAC-3') to the 5' end of each forward primer to employ a universal dye-labeling approach (Boutin-Ganache et al. 2001).



**Fig. 1** Map of range-wide collections of *P. lewtonii* in central Florida. Within the CC site, we carried out fine-scale sampling, with plants grouped into plots within four blocks (see Fig. 2a)





**Fig. 2** Results of InStruct analyses. **a** Plot-level map of *P. lewtonii* plots within Carter Creek (CC), with the name and location of each plot indicated. Colored pie charts for each plot correspond to color of the clusters assigned by InStruct and the percent membership in each cluster as presented in 2c. **b** Plot of  $-\ln$  likelihood ( $\ln(K)$ ; blue line) and deviance information criterion (DIC; red line) for each value of

$K$ . Values begin to plateau at  $K = 9$ . **c** Population structure inferred by InStruct for 360 *P. lewtonii* individuals from all seven blocks at the optimal value of  $K$ ,  $K = 9$  (number of clusters). The blocks are separated by black lines. Membership in clusters (different colors) is represented by the proportion of the bar in the color. (Color figure online)

We tested 86 primers for amplification using genomic DNA from two samples of *P. lewtonii*. PCR amplifications were performed in 10  $\mu$ L reactions containing 0.5 U of

GoTaq Flexi DNA polymerase (Promega), 1  $\times$  Promega Colorless GoTaq Flexi Buffer, 1.5 mM  $MgCl_2$ , 4.5 pmol each of the reverse primer and one of four fluorescently

labeled M13 primers (6-FAM, VIC, NED, or PET; Applied Biosystems), 0.18 pmol of the M13-tagged forward primer, and 0.5 mM of each dNTP. PCR temperature cycling conditions were as follows: (1) 3 min at 94 °C, (2) denaturation for 30 s at 94 °C, (3) annealing for 30 s at 52 °C, (4) extension for 45 s at 72 °C, (5) 35 repetitions of steps 2–4, and (6) a final elongation at 72 °C for 20 min. We diluted PCR products 1:10 and sent them for genotyping on an ABI3730xl DNA Analyzer, with all genotyping carried out by loading 1 µL of the diluted PCR product, 9.0 µL of formamide, and 0.5 µL LIZ 500 size standard (Applied Biosystems). All genotyping was carried out at the DNA analysis facility at Science Hill at Yale University. In this initial screening, loci that amplified in both individuals of *P. lewtonii* were selected for further testing in 16 individuals of *P. lewtonii* (four individuals from each of four populations) from across the geographic range of the species (Fig. 1). We selected 11 loci with the most consistent amplification, the highest levels of polymorphism, and most easily scored peaks.

To assess patterns of genetic structure in the full sample of *P. lewtonii* individuals (Table 1), we genotyped each of the 360 individuals at the 11 dinucleotide repeat microsatellite loci. In PCR, four fluorescently labeled M13 primers, each labeled with a differently colored dye, were used to label separate loci, which were then pooled and run as described above. Fragment analysis and scoring were carried out using automated fragment scoring panels developed for each locus in GeneMarker version 1.6 (Soft Genetics LLC), and then the data were checked manually. For three plots, we were unable to successfully obtain a

genotype for all individuals at one locus, particularly for loci PL-18 and PL-82.

**Data analysis**

To analyze patterns of genetic diversity across both broad and fine geographic scales, we analyzed data with individuals grouped into populations at both the block level (i.e., including the three range-wide populations and the four blocks at CC; seven total) and plot level (i.e., the 32 plots within the four blocks at CC). We tested for linkage disequilibrium (LD) between pairs of loci in each population and for deviations from Hardy–Weinberg equilibrium (HWE) at each locus/population combination using Fisher’s exact tests in Genepop version 4.2 (Rousset 2008). Diversity indices and summary statistics were estimated in both plots and blocks using Microsat Analyser version 4.05 (Dieringer and Schlotterer 2003) and FSTAT version 2.9.3 (Goudet 2002) diversity indices included observed and expected heterozygosity, number of alleles, allelic richness (El Mousadik and Petit 1996), and inbreeding coefficient ( $F_{IS}$ ). For allelic richness, rarefaction was used to account for differences in sample size. Because null alleles can cause large heterozygote deficiencies and inbreeding coefficients, we used INEST version 2.0 (Chybicki and Burczyk 2009) to test for the presence of null alleles. INEST employs a population-inbreeding model to measure the frequency of null alleles at each locus while simultaneously calculating the inbreeding coefficient ( $F_{IS}$ ) within each population. We employed the Bayesian MCMC approach with 200,000 cycles, keeping every 200th update,

**Table 1** Primer sequences (forward and reverse) and characteristics of 11 novel microsatellite loci developed for *Polygala lewtonii*, including repeat motif, allelic size range as determined in 16

individuals (four individuals from each of four populations of *Polygala lewtonii*; see Table 2), and GenBank Accession Number

Locus name	Forward primer sequence	Reverse primer sequence	Repeat motif	Size range	GenBank accession#
PL73	ACAGAGACGGTGGCTCTAGATTTTC <sup>a</sup>	AAACCATGCGCCTGACAAC	AT(8)	146–158	KR154467
PL40	TGGCTCTATCTAAATACGTGC <sup>a</sup>	TCCATCTGCGTTTAAATGTTAATC	AT(10)	165–177	KR154463
PL56	GCAGCGGAGAACAACATATAGC <sup>a</sup>	ACGTCTAGATGTACAGGCAC	AG(18)	194–222	KR154466
PL80	CCGTAAGTACTCCAATGCACG <sup>a</sup>	GAGAACTGCACACTCGCC	AG(10)	125–129	KR154471
PL82	AGGAACACATTACGTATAGCTAGC <sup>a</sup>	GTATTCAGCAGACCTCAAG	AT(8)	165–177	KR154472
PL74	CGTTGTCTTCTCAAACAAATCAGC <sup>a</sup>	CTGCACTACAAACTGATCGGG	AC(9)	171–177	KR154468
PL18	TGCCCAAATATACCAACCAAGC <sup>a</sup>	TTGGCATATTTGTGTCTGGAC	AT(9)	218–240	KR154462
PL76	TCTCCGACTCAATCAGATTCCC <sup>a</sup>	ATGAACCGGTAGTGTGGCTG	CT(10)	142–146	KR154469
PL48	TGCTCCTACCTTGCTTGAAG <sup>a</sup>	AATTCAGTTCATTGTCATTGAGAG	AT(9)	167–179	KR154464
PL79	CGACGTTTCATGCTTTCAACAG <sup>a</sup>	CTGTCTCCCAACCTACACCC	AG(8)	123–125	KR154470
PL54	TTCCAAGAATTCGGGACACC <sup>a</sup>	ACCTGAGTCACTGGAAAGG	AG(8)	132–134	KR154465

<sup>a</sup> An M13 tag (CACGACGTTGTAAAACGAC) added to 5’ end of primer for amplification with a universal, fluorescently labeled M13 primer

and a burn-in period of 20,000 cycles to calculate the percent of null alleles and a revised  $F_{IS}$  ( $F_{IS}^B$ ) in each population.

To investigate patterns of genetic structure, we used Microsat Analyser version 4.05 to calculate pairwise  $F_{ST}$  for all possible pairs of blocks and plots, with 100,000 permutations to assess significance. Because  $F_{ST}$  values of highly variable markers, such as microsatellites, are constrained by population-level homozygosity levels and thus do not reach their maximum at 1 (see Hedrick 2005 for a more detailed explanation), we also calculated a standardized measure of pairwise population differentiation,  $G'_{ST}$  (Hedrick 2005) using the program Genodive version 2.0b23 (Meirmans 2006). We used sequential Bonferroni corrections for all tests involving multiple comparisons (Rice 1989).

To investigate patterns of genetic structure in *P. lewtonii* without a priori grouping of individuals into populations, we initially analyzed our data using STRUCTURE version 2.3.4 (Pritchard et al. 2010). Despite multiple analyses with long burn-ins and run lengths, we were unable to obtain consistent groupings across independent runs at each  $K$ . The inability to obtain consistent clusters in STRUCTURE is likely because of significant inbreeding in *P. lewtonii* (see results) one of the primary assumptions of STRUCTURE is outcrossing and the program is unable to distinguish between non-random mating as a result of inbreeding or population structure (Gao et al. 2007).

We next used InStruct (Gao et al. 2007), a Bayesian program that estimates patterns of genetic structure without a priori grouping individuals into populations, which differs from STRUCTURE in that it eliminates the assumption of Hardy–Weinberg equilibrium within the cluster and jointly estimates the selfing rate and population structure. We analyzed patterns of genetic structure in the entire data set of 360 individuals using ten independent chains of the MCMC algorithm at each  $K$  from 1 to 15. InStruct was run with the default settings, allowing for admixture, with population information used only to organize individuals in figures to understand how population designations corresponded to patterns of genetic structure. After preliminary analyses to determine the adequate burn-in and number of iterations, we used a burn-in of 500,000 iterations and a run length of 1,000,000 iterations for each chain. To ensure convergence and repeatability, we examined the groupings across all runs at each  $K$  in CLUMPAK (Kopelman et al. 2015) to ensure that the results of analyses were fully repeatable (Gilbert et al. 2012). To determine the optimal value of  $K$ , we used the deviance information criterion (DIC) as reported by InStruct (Gao et al. 2011) and also plotted the  $-\ln$  likelihood values from InStruct to understand how they changed with increasing values of  $K$ . Following guidelines for selecting the value of  $K$  in the STRUCTURE documentation (Pritchard et al. 2010), when several values

of  $K$  had similar  $-\ln$  likelihoods, we visually inspected the results to determine whether increasing the values of  $K$  revealed clear patterns of genetic structure; we then selected the smallest value of  $K$  with clear patterns of genetic structure. To understand how grouping of individuals affects estimates of genetic diversity and structure, we also calculated each measure of genetic diversity and genetic structure, with individuals grouped into the genetic clusters determined by InStruct (see “Results” section).

Additionally, we analyzed whether populations exhibited isolation by distance using Mantel tests (1967). We calculated a pairwise genetic distance matrix in Genodive using Nei’s (1978) genetic distances and a pairwise geographic distance matrix using the geographic distance calculator (Ersts 2015). We carried out analyses with individuals grouped by blocks or plots using a standard Mantel test with 10,000 permutations in Genodive.

To provide insight into dispersal distances and the hierarchical partitioning of genetic variation, we carried out an analysis of molecular variance, or AMOVA (Excoffier et al. 1992) in Arlequin version 3.11 (Excoffier et al. 2005). We used a locus-by-locus AMOVA with significance ascertained using 10,000 permutations. We carried out AMOVA with individuals grouped two ways. First, we investigated the amount of variance attributable to the ‘among blocks,’ ‘among plots within blocks,’ and ‘among individuals within plots’ using data from all 288 *Polygala lewtonii* samples from CC. Only the CC samples were used for this analysis because they were spatially structured to capture distances likely traveled by pollinators and seed dispersal (see “Sample collection and DNA extraction” section above). Second, we conducted an AMOVA using all 360 individuals of *P. lewtonii*, with individuals grouped into the nine clusters determined by InStruct (see “Results” section).

## Results

### Microsatellite primer development

The MiSeq run resulted in 245,323 paired-end reads, which were assembled *de novo* into 97,377 contigs. MSATCOMMANDER identified 591 unique microsatellites in contig consensus sequences, from which PRIMER3 identified 86 unique microsatellite primers. Of these, 24 microsatellite loci amplified in the initial screening of two individuals, and 15 loci were polymorphic in a sample of 16 individuals from four populations, 11 loci were selected to genotype all individuals in this study (Table 1). After sequential Bonferroni analysis, analyses of linkage disequilibrium found no evidence of significant linkage between any two loci.



**Levels of genetic diversity in populations sampled at wide and narrow geographic scales**

Analyses of levels of genetic diversity at the range-wide scale, with individuals grouped into populations according to collection location at SPP, LWR, ONF, and CC1-4 (Table 2), revealed high levels of inbreeding. The mean number of alleles across all loci in each population ranged from 2.455 to 4.091 (overall mean = 3.23; Table 2). Allelic richness averaged across the 11 loci for each population ranged from 2.001 to 3.321 (overall mean = 2.52; Table 2). The  $H_E$  (expected heterozygosity) averaged across all loci in each block ranged from 0.208 to 0.516 (overall mean = 0.348; Table 2) and the  $H_O$  (observed heterozygosity) averaged across loci ranged from 0.03 to 0.129 (Overall mean = 0.066; Table 2). The populations showed deviations from Hardy–Weinberg equilibrium, with every locus in every block showing significant heterozygote deficiencies after sequential Bonferroni correction. When averaged across loci,  $H_O$  was smaller than  $H_E$  in every population; consequently, the  $F_{IS}$  (inbreeding coefficient) averaged across all loci in each population ranged from 0.633 to 0.915 (mean = 0.811, Table 2), suggesting high levels of inbreeding. The percentage of null alleles in each block averaged across all loci ranged from 0.067 to 0.205 (mean = 0.115; Table 2). No single locus consistently had a large frequency of null alleles across all populations. Even when null alleles were taken into account, the inbreeding coefficient averaged across all loci in each population ( $F_{IS}^B$ ) remained high, ranging from 0.610 to 0.879 (mean = 0.763; Table 2), such that the interpretation was the same for analyses that did and did not account for null alleles. Results of analyses of genetic structure were also highly similar regardless of whether null alleles were taken into account (analyses not shown).

Because the interpretation of results was the same, results of analyses that accounted for null alleles are not further reported.

Analyses of genetic diversity at the plot scale at CC also revealed high levels of inbreeding. The mean number of alleles per plot averaged across loci ranged from 1 to 2.091 (mean = 1.455; Table S1). The  $H_E$  per plot averaged across loci ranged from 0 to 0.294 (mean = 0.115; Table S1), while the  $H_O$  per plot averaged across loci ranged from 0 to 0.125 (mean = 0.036; Table S1). The plots showed deviations from Hardy–Weinberg equilibrium, with 24 of 32 plots showing significant heterozygote deficiencies across loci after sequential Bonferroni correction.  $H_O$  was less than or equal to  $H_E$  in 30 of 32 plots, and consequently,  $F_{IS}$  values averaged across loci in the plots were generally strongly positive, averaging 0.714 (range –0.093 to 1.00; Table S1).

**Population structure**

Analyses of genetic structure at the range-wide scale, which included individuals sampled across two of Florida’s central ridges, revealed considerable structuring of genetic variation. In analyses of pairwise  $F_{ST}$  and  $G'_{ST}$  with individuals grouped by blocks, significant pairwise  $F_{ST}$  values were found between all possible pairs of blocks after sequential Bonferroni correction (range = 0.181–0.486), even between the geographically proximal blocks sampled at CC (Table 3).  $F_{ST}$  values generally increased with geographic distance between the sampling locations. Contrary to expectations, the geographically proximal ONF and SPP blocks had a larger pairwise  $F_{ST}$  than the geographically more distant LWR and ONF blocks ( $F_{ST}$  = 0.415 and 0.191, respectively; Table 3). Pairwise  $G'_{ST}$  values showed similar patterns as pairwise  $F_{ST}$ , although  $G'_{ST}$  values were

**Table 2** Collection information and measures of genetic diversity for each block of *P. lewtonii* used within the study

Block abbreviation	Collection locality	County in Florida	Lat, long	$H_O$	$H_E$	$F_{IS}$	% null alleles	$F_{IS}^B$	$A$	$A_R$	$n$
CC1	Carter Creek-Block 1	Highlands	27.519924, –81.413333	0.035	0.208	0.832	0.142	0.732	2.727	2.001	72
CC2	Carter Creek-Block 2	Highlands	27.510796, –81.401603	0.064	0.330	0.806	0.205	0.729	3.182	2.326	72
CC3	Carter Creek-Block 3	Highlands	27.513677, –81.401779	0.035	0.411	0.915	0.188	0.879	3.000	2.521	72
CC4	Carter Creek-Block 4	Highlands	27.520696, –81.403335	0.030	0.253	0.884	0.084	0.862	3.364	2.188	72
ONF	Ocala National Forest	Marion	29.230831, –81.805428	0.129	0.516	0.757	0.065	0.698	3.818	3.175	24
LWR	Lake Wales Ridge State Forest	Polk	27.691486, –81.480602	0.067	0.451	0.853	0.053	0.836	4.091	3.321	24
SPP	Scrub point preserve	Lake	28.524682, –81.664801	0.100	0.270	0.633	0.067	0.610	2.455	2.118	24
	Mean values across the blocks			0.066	0.348	0.811	0.115	0.763	3.234	2.521	

$H_O$  Observed heterozygosity,  $H_E$  expected heterozygosity,  $F_{IS}$  inbreeding coefficient, percent null alleles,  $F_{IS}^B$  inbreeding coefficient taking into account null alleles,  $A$  mean number of alleles per block across loci,  $A_R$  allelic richness mean across loci analyzed for each of the blocks separately,  $n$  number of plants from each block included in the present analysis

**Table 3** Pairwise  $F_{ST}$  (upper diagonal) and  $G'_{ST}$  (lower diagonal) values for the seven sampling blocks of *P. lewtonii* included within the study

Population and block	CC1	CC2	CC3	CC4	ONF	LWR	SPP
CC-block 1	–	0.511*	0.446*	0.356*	0.540*	0.419*	0.431*
CC-block 2	0.699	–	0.181*	0.416*	0.387*	0.318*	0.486*
CC-block 3	0.645	0.290	–	0.345*	0.299*	0.247*	0.401*
CC-block 4	0.462	0.586	0.516	–	0.421*	0.296*	0.363*
Ocala National Forest	0.741	0.624	0.532	0.600	–	0.191*	0.415*
Lake Wales Ridge	0.555	0.496	0.434	0.407	0.373	–	0.313*
Scrub point preserve	0.539	0.710	0.644	0.488	0.671	0.493	–

\* Indicates significant pairwise  $F_{ST}$  comparisons ( $P < 0.05$ )

higher, ranging from 0.290 to 0.741 (Table 3). Consistent with the patterns found using  $F_{ST}$  and  $G'_{ST}$ , Mantel tests with individuals grouped into blocks revealed a significant correlation between geographic and genetic distance ( $r = 0.473$ ,  $P = 0.025$ ), indicating isolation by distance.

Within the CC sampling site, analyses revealed strong patterns of genetic structure at the plot scale. The average pairwise  $F_{ST}$  between plots within a block was 0.50 (range =  $-0.09$ – $0.94$ ; Table S2), whereas the average pairwise  $F_{ST}$  between plots in different blocks was 0.67 (range =  $0.20$ – $1.00$ ; Table S2). Most pairwise  $F_{ST}$  values were found to be significant after sequential Bonferroni correction.  $G'_{ST}$  values show similar patterns of genetic structure, with the values ranging from 0.1 to 1 (Table S2). The magnitude of pairwise  $F_{ST}$  and  $G'_{ST}$  values between plots sampled at a fine geographic scale generally mirror the patterns found in the range-wide analysis, with pairwise  $F_{ST}$  generally increasing with geographic distance. Mantel tests with individuals grouped into plots revealed a highly significant correlation ( $r = 0.556$ ,  $P < 0.001$ ) between geographic and genetic distance, indicating a strong signal of isolation by distance.

InStruct analyses were used to assess the patterns of genetic structure without a priori assignment of individuals into clusters while simultaneously measuring the selfing rate in each cluster. The DIC approach in InStruct estimated the

**Table 4** Self-fertilization rates for the nine clusters determined by InStruct (see Fig. 2)

InStruct cluster	Self-fertilization rate
1	0.795
2	0.801
3	0.839
4	0.844
5	0.887
6	0.891
7	0.922
8	0.925
9	0.933
Mean	0.871

optimal value of  $K$  as  $K = 12$ , whereas the  $-\ln$  likelihood and DIC values plateaued at  $K = 9$  (Fig. 2b). Examination of results at  $K = 9$  revealed clear assignment of individuals to clusters (Fig. 2c), whereas the patterns of genetic structure at values larger than  $K = 9$  revealed many individuals with roughly equal assignments to several clusters and overall weak patterns of genetic structure, indicating that these value of  $K$  were too high. We thus selected  $K = 9$  as the optimal value of  $K$ . At  $K = 9$ , the rate of self-fertilization in each cluster ranged from 0.795 to 0.933 (overall mean = 0.871; Table 4). At  $K = 9$  there was strong geographic structuring of the genetic variation, with each cluster containing geographically grouped individuals (Fig. 2a, c). Individuals in the four blocks at CC were grouped into six InStruct clusters; the first cluster contained most individuals from CC1, the second contained most individuals from CC2, the third and fourth clusters each contained around half of the individuals from CC3, and the fourth and fifth clusters each contained around half of the individuals from CC4 (Fig. 2a, c). The remaining three clusters each contained most of the individuals of from one of the three populations sampled at a range-wide scale (Fig. 2c).

**Table 5** Results of analyses of genetic diversity with individuals grouped into InStruct clusters

InStruct cluster	$H_O$	$H_E$	$F_{IS}$	$A$	$A_R$	$n$
1	0.034	0.209	0.606	2.636	2.069	71
2	0.073	0.312	0.708	3.545	2.490	64
3	0.034	0.318	0.804	2.727	2.264	51
4	0.031	0.301	0.647	2.091	1.946	24
5	0.013	0.204	0.565	2.364	2.039	25
6	0.034	0.206	0.486	2.455	1.933	53
7	0.143	0.474	0.618	3.727	3.230	20
8	0.073	0.477	0.708	4.273	3.473	27
9	0.104	0.273	0.472	2.545	2.196	25

Clusters are arranged in left–right order as shown in Fig. 2c. (i.e., cluster 1 = blue/CC1, cluster 2 = orange/CC2, cluster 3 = dark green/CC3, cluster 4 = bright pink/CC3, cluster 5 = maroon/CC4, cluster 6 = purple/CC4, cluster 7 = yellow orange/ONF, cluster 8 = light green/LWR and cluster 9 = pink/SPP)

**Table 6** Pairwise  $F_{ST}$  (upper diagonal) and  $G'_{ST}$  (lower diagonal) values among the nine genetic clusters determined by InStruct

InStruct cluster	1	2	3	4	5	6	7	8	9
1	–	0.533*	0.565*	0.473*	0.417*	0.432*	0.561*	0.435*	0.431*
2	0.718	–	0.263*	0.419*	0.479*	0.452*	0.458*	0.340*	0.508*
3	0.756	0.387	–	0.338*	0.537*	0.483*	0.415*	0.307*	0.515*
4	0.601	0.612	0.498	–	0.523*	0.443*	0.413*	0.295*	0.481*
5	0.522	0.674	0.746	0.701	–	0.350*	0.450*	0.340*	0.370*
6	0.546	0.618	0.655	0.573	0.441	–	0.492*	0.342*	0.452*
7	0.747	0.709	0.653	0.667	0.655	0.661	–	0.220*	0.447*
8	0.593	0.536	0.497	0.502	0.524	0.474	0.424	–	0.317*
9	0.540	0.727	0.738	0.680	0.490	0.579	0.692	0.514	–

Clusters are arranged in left–right order as shown in Fig. 2c. (i.e., cluster 1 = blue/CC1, cluster 2 = orange/CC2, cluster 3 = dark green/CC3, cluster 4 = bright pink/CC3, cluster 5 = maroon/CC4, cluster 6 = purple/CC4, cluster 7 = yellow orange/ONF, cluster 8 = light green/LWR and cluster 9 = pink/SPP)  
 \* Indicates significant pairwise  $F_{ST}$  comparisons ( $P < 0.05$ )

**Table 7** Analysis of molecular variance (AMOVA) results with individuals grouped either by block and plot or by InStruct clusters

Source of variation	Sum of squares	Variance components	Percentage variation	Fixation indices
Individuals by block				
Among blocks	293.3	0.686	37.0	$F_{ST}$ : 0.696
Among plots within blocks	282.4	0.603	32.6	$F_{SC}$ : 0.517
Within plots	266.7	0.563	30.4	$F_{CT}$ : 0.370
Organized by InStruct clusters				
Among clusters	668.3	1.237	44.1	$F_{ST}$ : 0.441
Within clusters	908.4	1.567	55.9	

All levels are statistically significant ( $P < 0.00001$ )

**Patterns of genetic diversity and structure with individuals grouped by InStruct clusters**

To understand how the grouping of individuals affects estimates of genetic diversity and structure, we also conducted analyses of genetic diversity and structure using individuals grouped by InStruct clusters, which largely showed similar results to those generated with individuals grouped into blocks. Overall, estimates of  $H_E$ ,  $H_O$ ,  $A$ , and  $A_R$  were generally similar to those found with individuals grouped by blocks (Table 5). The one exception is values of  $F_{IS}$  were smaller in some InStruct clusters than those found when individuals were grouped by blocks, indicating that population structure may have artificially increased  $F_{IS}$  in some blocks. Although some values differed between the two grouping methods,  $F_{IS}$  values were still high when grouped according to InStruct clusters (0.47–0.80) and indicate significant inbreeding. For analyses of pairwise  $F_{ST}$  and  $G'_{ST}$  with individuals grouped by InStruct clusters, results were also highly similar to those found when individuals were grouped into blocks. Pairwise  $F_{ST}$  values for all possible pairs of

InStruct clusters were significant after sequential Bonferroni correction, ranging from 0.220 to 0.565 (Table 6), and pairwise  $G'_{ST}$  values showed similar patterns as pairwise  $F_{ST}$  but their magnitudes were higher because  $G'_{ST}$  is scaled to a maximum of 1, ranging from 0.387 to 0.756 (Table 6).  $F_{ST}$  and  $G'_{ST}$  values generally increased with geographic distance between the clusters.

An AMOVA investigated the partitioning of genetic variation at the block, plot, and the individual level within the spatially stratified CC sampling site. About 37 % of variation was attributable to the ‘among blocks’ level ( $F_{ST} = 0.696$ ,  $P < 0.0001$ ; Table 7), 33 % was attributable to the ‘among plots within blocks’ level ( $F_{SC} = 0.517$ ,  $P < 0.0001$ ; Table 7), and 30 % of the variation was attributable to the ‘within plots’ level ( $F_{CT} = 0.370$ ,  $P < 0.0001$ ; Table 7). Another AMOVA investigated the partitioning of genetic variation of the entire data set, with individuals grouped using the genetic clusters as defined by InStruct. Approximately 44 % of the variation was partitioned among clusters, while 56 % was partitioned within clusters ( $F_{ST} = 0.441$ ,  $P < 0.0001$ ; Table 7).

## Discussion

Because *P. lewtonii* has an unusual and complex mating system, our goals were to use levels of genetic diversity to understand the predominant mating system (i.e., selfing or outcrossing) employed by *P. lewtonii* and to use patterns of genetic structure to understand the migration of pollen and selfed and outcrossed seeds across the landscape. We sampled individuals at CC using a spatially stratified sampling scheme. Plants within plots were sampled at an average pairwise distance of 1.34 m, designed to represent the minimum range that a seed produced by belowground flowers or an undispersed seed produced aboveground would be spaced from a parent plant. The distance between plots (10 m or greater) was designed to capture seed dispersal distances traveled by aboveground, ant-dispersed seeds. Blocks were spaced at varying distances (range 0.14–1.22 km) to reflect the range of possible distances traveled by pollinators.

Analyses of genetic diversity revealed significant deficiencies in observed heterozygosity and large inbreeding coefficients. These analyses support the conclusion that *P. lewtonii* reproduces predominantly via self-fertilization and inbreeding. Examination of genotypes within plots often found tightly spatially clustered groups of individuals that were almost 100 % homozygous and highly genetically similar to each other. All of these highly similar, homozygous individuals were produced by selfing or biparental inbreeding, and their close geographic proximity to each other was likely produced by seeds being dispersed extremely close to the parent plant. Unfortunately, because individuals produced by selfing/inbreeding above and belowground would be genetically indistinguishable, it is impossible to determine whether these tight clusters were produced above or belowground from the genetic approach used in this study. In addition, some nearly identical individuals were also spaced across larger geographic distances, such as across neighboring plots; these individuals were likely produced in aboveground flowers via self-fertilization or bi-parental inbreeding and were subsequently dispersed by ants across distances ~10–15 m. These dispersal distances, estimated by genetic analysis, are consistent with field observations of the distances that ants disperse seeds (averaging about 140 cm with an observed maximum of 727 cm (Menges et al. 2007a)).

Even though a large proportion of reproduction likely occurred via self-fertilization, Bayesian analyses of genetic structure clustered individuals from several groups of neighboring plots, suggesting that some gene flow is occurring among aboveground CH flowers from these groups of plots. However, the small spatial scale of genetic structuring of these clusters indicates that outcrossing is

occurring across very limited distances. For the individuals sampled at CC, which were collected within a small geographic area (~0.28 km<sup>2</sup>; Fig. 2c), we identified six genetically and geographically differentiated clusters. Mantel tests also revealed isolation by distance across both the plot and block scales, indicating that gene flow is spatially limited across fine geographic scales. The fact that individuals are genetically differentiated even between geographically proximal blocks indicates that most pollination is occurring within a range between 15 and 550 m (i.e., representing the minimum and maximum distances between plots within each cluster). Ants primarily disperse the seeds produced by aboveground CH flowers short distances, and these genetic results also suggest that pollinators generally do not transfer pollen across large geographic distances.

At the range-wide scale, genetic analyses showed increasing pairwise  $F_{ST}$  and  $G'_{ST}$  with increasing geographic distance. The exception was the low genetic divergence between the ONF and LWR populations, which was unexpected given that these populations are not most geographically proximal to each other (i.e., the SPP is located between ONF and LWR) and occur on different ridges. One possible explanation for these patterns is a long-distance dispersal event that moved genetic material between the populations. If this is the case, given the large distances between these populations, it is likely that this dispersal event was human-mediated. Another possibility is that this is an artifact of limited spatial sampling, and that increasing our sampling to include additional populations located between those sampled in the study might provide finer-scale resolution of range-wide patterns of genetic structure and genetic connectivity, possibly revealing routes by which these individuals were dispersed.

Results of this study indicate that the majority of reproduction in *P. lewtonii* occurs via inbreeding; the fact that less reproduction occurs via outcrossing may be caused by several factors, several of which may occur during pollination. Given the close geographic proximity of many closely related plants, it is possible that bi-parental inbreeding or pollen transfer between stems on the same plant (i.e., geitonogamy) is limiting the outcrossing rate of CH flowers. This hypothesis could be tested by genotyping the offspring produced by CH flowers to measure the number of outcrossed versus inbred offspring. Another possible cause for low rates of outcrossing could be low pollinator visitation rates; a previous study of the pollination biology of aboveground CH flowers in *P. lewtonii* found only 0.3 visits per flower per hour (Weekley and Brothers 2006). One possible explanation for such low pollinator visitation rates could be that the primary

pollinator has experienced population declines, geographic range shifts, or has gone extinct, leading to a shortage of pollinators for the number of flowers open at one time. Even with low pollinator visitation rates, CH flowers had high rates of fruit set (71.2 %; Weekley and Brothers 2006).

In addition to factors that limit outcrossing during pollination, another explanation for low rates of reproduction by outcrossing is factors affecting the performance of outcrossed seeds and seedlings. For example, seeds/seedlings may have lower survival rates than those produced by selfing, possibly because seeds produced aboveground (including all outcrossed seeds) are poorly provisioned or experience environmental conditions that are less suitable for germination than seeds produced belowground. Previous research has shown that fruits produced by belowground selfing are often exceptionally well-provisioned by the parent plant and produce progeny with higher fitness in the local habitat (Cheplick 2005); if belowground seeds are superior, this would result in a larger proportion of reproduction occurring by selfing. Small-scale germination tests also found that seeds produced by the CH flowers had very low germination rates (2–20 %) (Weekley and Brothers 2006; Lindon and Menges 2008), but it is unclear whether this is because only the outcrossed seeds produced by aboveground CH flowers have low germination rates because they are inferior, or because all types of seeds produced by *P. lewtonii* have low germination rates (i.e., due to inbreeding depression or environmental factors). Another possible cause of low outcrossing rates is that the seeds produced aboveground (including all outcrossed seeds) may experience less optimal environmental conditions for germination than those produced belowground. For example, belowground seeds may experience less variable soil moisture levels than those produced aboveground and may therefore have more consistent seed germination rates. Field experiments showed high seedling recruitment after a fire (Lindon and Menges 2008; Weekley and Menges 2012), and it is also possible that fire has different effects on the germination of aboveground and belowground seeds; if this is the case, then we would expect to see differences in outcrossing rates in years that experience fire in relation to those that do not. Further research is necessary to identify the relative importance of these factors in influencing the reproductive success by both selfing and outcrossing in *P. lewtonii*.

Although our study was designed to provide some indication of the reproduction by the three types of flowers in *P. lewtonii*, only partial inference about whether a seed is produced belowground or aboveground was possible using the genetic approach employed in this study. The existence of heterozygous, outcrossed individuals and genetic differentiation across the landscape provides

evidence of reproduction by aboveground CH flowers. Highly homozygous individuals that are genetically similar to each other but spaced >3–4 m apart provides evidence of selfing aboveground followed by seed dispersal by ants. Tight clusters of highly homozygous individuals within a 1–2 m radius provide evidence for selfing with virtually no seed dispersal, but we are unable to distinguish whether those seeds were produced aboveground or belowground. Unfortunately, because individuals produced aboveground or belowground by selfing should be identical and genetically indistinguishable, we are unable to think of a genetic approach to confidently distinguish the origin of a selfed individual in a wild population. Furthermore, simply getting an estimate of the amount of reproduction occurring belowground in this federally endangered species is challenging because excavating the rhizomes of wild plants would be highly destructive. One possibility would be to grow experimental plants in sterile soil (to ensure no seeds were in the seed bank), bag or remove aboveground flowers and seeds, and then excavate the roots of some plants or simply observe how many seedlings are produced from the soil. Although the timing of excavation would be tricky and this approach would likely miss some seeds, this would still likely be the best way to provide an estimate of the seed production belowground. The seeds resulting from excavations could then be subjected to further experiments to understand their germination rates.

### Conservation implications

*Polygala lewtonii* is listed as federally endangered (USFWS 1999) and is restricted to a small portion of the Lake Wales Ridge and the Mount Dora Ridge in the Ocala National Forest. Degradation and loss of suitable habitat from fire suppression and land conversion for agriculture and residential development (USFWS 1999) caused past population declines in *P. lewtonii*, and these factors still threaten many existing populations. Because of these ongoing threats to the persistence of many populations, recovery efforts are underway for this species, and these efforts would benefit from knowledge of the partitioning of genetic variation to ensure that the maximum amount of genetic variation is protected in this species. AMOVA analyses in *P. lewtonii* revealed that the majority of genetic variation was partitioned among blocks and plots, indicating that genetic differences exist between groups of individuals sampled at a fine geographic scale (i.e., between individuals 10 m apart). Because genetic diversity is structured at such a fine geographic scale, the loss of all plants within an area the size of a plot could cause the loss of a unique portion of the genetic variation of the species. The loss of an entire site (from either land conversion or fire suppression) would result in huge losses of unique



genetic variation. Thus, the best way to protect the full range of genetic diversity in this species would be to publicly protect and effectively manage each site that harbors *P. lewtonii*. However, only some of the populations are publicly protected, and acquisition of additional populations for protected areas is complicated by historic fragmentation of upland habitat and limited funds for conservation.

In addition to in situ conservation, a complementary approach to protect the full range of genetic variation is through long-term storage of seeds in conservation seed banks, which would ensure that genetic variation is protected in the event that populations are lost. Currently, however, conservation seed banking may be inefficient due to low germination rates observed in previous seed germination experiments (Lindon and Menges 2008). Additionally, the survival rate of seeds under long-term storage is unknown. Thus, important areas of further study in this species are to understand if any factors limit reproduction in aboveground flowers, as these will most likely be collected for conservation seed banks, as well as assessing seed survival in long term-storage and determining the optimal conditions (e.g., smoke, heat, scarification) for breaking dormancy and promoting seed germination. Even though seed banking may be inefficient, it is important for ongoing conservation efforts to focus on seed banking or other ex situ conservation measures in populations of *P. lewtonii* that currently exist in unprotected or poorly managed areas, as extirpation would likely cause the irrecoverable loss of a unique subset of genetic variation in the species. Within each site, seeds should be collected from individuals at a fine geographic scale (i.e., every 10–20 m) to capture as much genetic diversity as possible. Subsequent conservation seed banking efforts should focus on collecting from as many protected sites as possible to ensure against stochastic extinction events.

Conservation efforts may also be aided by additional genetic research on the reproductive biology of *P. lewtonii*. For example, given that seed germination rates are very low (2–20 %; Weekley and Brothers 2006; Lindon and Menges 2008), an experiment to quantify how outcrossing/inbreeding affects seed viability and germination would be useful. To do this, seeds could be harvested from aboveground CH flowers and genotyped to quantify the percentage of seeds produced by outcrossing versus inbreeding to understand the overall outcrossing rate. Comparing this number to the average percentage of germinants produced by outcrossing versus inbreeding would indicate whether low seed germination rates are caused by inbreeding depression. Additional experiments to evaluate the performance and fitness of inbred versus outcrossed individuals would also help determine whether inbreeding depression is negatively affecting populations. If we find strong evidence of inbreeding

depression, then improvements in outcrossing rates could be accomplished by transplanting individuals between genetically different populations. Additionally, genotyping seedlings to understand the rates of outcrossing versus inbreeding in years with and without fire would help to determine if environmental differences are affecting the relative germination rates of aboveground versus belowground seeds. If aboveground seeds germinate more readily after a fire, this would result in an increase in the frequency of outcrossed individuals in a population; such research may reveal an additional means by which to promote increases in the outcrossing rate and concomitantly, levels of genetic diversity in *P. lewtonii*.

**Acknowledgments** The authors would like to thank Dr. David Bogler for logistical support during the Missouri Botanical Garden Research Experience for Undergraduates program; the DNA analysis facility at Science Hill at Yale University for processing genetic samples; Carl Weekley, Stephanie Koontz, Jamie Peeler, Devon Picklum, David Zaya, Jennifer Navarra, Sarah Hicks, Lauren Sullivan, John Lowell, Kate Prengaman, and David Horton for assistance with field work; and David Boshier and two anonymous reviewers for comments on previous versions of this manuscript. This work was supported by the Florida Department of Agriculture and Consumer Services Division of Plant Industry (Grant number 020159) and the Research Experiences for Undergraduates Program of National Science Foundation (Grant number DBI-1157030).

## References

- Barker NP (2005) A review and survey of basicarpy, geocarpy, and amphicarpy in the African and Madagascan flora. *Ann Mo Bot Gard* 92:445–462
- Barrett SCH, Shore JS (1987) Variation and evolution of breeding systems in the *Turnera ulmifolia* L. complex (Turneraceae). *Evolution* 41:340–354
- Baskin C, Baskin J (2014) Seeds: ecology, biogeography, and evolution of dormancy and germination, 2nd edn. Academic/Elsevier, San Diego
- Berg H, Redbo-Torstensson P (1999) Offspring performance in three cleistogamous *Viola* species. *Plant Ecol* 145:49–58
- Birchler JA, Yao H, Chudalayandi S, Vaimanb D, Veitac RA (2010) Heterosis. *Plant Cell* 22:2105–2112
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M-13 tailed primers improve the readability and usability of microsatellite analyses performed with two different allele sizing methods. *Biotechniques* 31:6–8
- Brown AHD (1979) Enzyme polymorphism in plant populations. *Theor Popul Biol* 15:1–42
- Brown AHD, Munday J (1982) Population-genetic structure and optimal sampling of land races of barley from Iran. *Genetica* 58:85–96
- Brunet J, Eckert CG (1998) Effects of floral morphology and display on outcrossing in Blue Columbine, *Aquilegia caerulea* (Ranunculaceae). *Funct Ecol* 12:596–606
- Busch JW (2005) The evolution of self-compatibility in the geographically peripheral populations of *Leavenworthia alabamica* (Brassicaceae). *Am J Bot* 92:1503–1512
- Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst* 18:237–268

- Cheplick GP (1987) The ecology of amphicarpic plants. *Trends Ecol Evol* 2:97–101
- Cheplick GP (1994) Life history evolution in amphicarpic plants. *Plant Species Biol* 9:119–131
- Cheplick GP (2005) Biomass partitioning and reproductive allocation in the invasive, cleistogamous grass *Microstegium vimineum*: influence of the light environment. *Torrey Bot Soc* 132:214–224
- Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered* 100:106–113
- Clavijo ERDE (1995) The ecological significance of fruit heteromorphism in the amphicarpic species *Catananche lutea* (Asteraceae). *Int J Plant Sci* 156:824–833
- Clay K, Antonovics J (1985) Demographic genetics of the grass *Danthonia spicata*: success of progeny from chasmogamous and cleistogamous flowers. *Evolution* 39:205–210
- Cole CT, Biesboer DD (1992) Monomorphism, reduced gene flow, and cleistogamy in rare and common species of *Lespedeza* (Fabaceae). *Am J Bot* 79:567–575
- Cruden RW, Lyon DL (1989) Facultative xenogamy: examination of a mixed mating system. *Evol Ecol Plants* 171:208
- Darwin C (1877) The different forms of flowers on plants of the same species. John Murray, London
- Dick CA, Herman JA, O'Dell RE, Lopez-Villalobos A, Eckert C, Whittall JB (2014) Cryptic genetic subdivision in the San Benito evening primrose (*Camissonia benitensis*). *Conserv Genet* 15:165–175
- Dieringer D, Schlotterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* 3:167–169
- Doyle J, Doyle J (1987) A rapid isolation procedure for small quantities of fresh tissue. *Phytochem Bull* 19:11–15
- Eckert CG (2000) Contributions of autogamy and geitonogamy to self-fertilization in a mass-flowering, clonal plant. *Ecology* 81:532–542
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet* 92:832–839
- Ersts PJ (2015) Geographic distance matrix generator (version 1.2.3)
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resour* 8:92–94
- Furches MS, Wallace LE, Helenurm K (2009) High genetic divergence characterizes populations of the endemic plant *Lithophragma maximum* (Saxifragaceae) on San Clemente Island. *Conserv Genet* 10:115–126
- Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–1651
- Gao H, Bryc K, Bustamante CD (2011) On identifying the optimal number of population clusters via the deviance information criterion. *Plos One* 6:e21014
- Gilbert KJ, Andrew RL, Bock DG, Franklin MT, Kane NC, Moore JS, Moyers BT, Renaut S, Rennison DJ, Veen T, Vines TH (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Mol Ecol* 21:4925–4930
- Goudet J (2002) FSTAT (version 2.9.3.2) a computer program to calculate F-statistics. *J Hered* 86:485–486
- Hamrick JL (1982) Plant population genetics and evolution. *Am J Bot* 69:1685–1693
- Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown ADH, Clegg MT, Kahler AL, Weir BS (eds) *Plant population genetics, breeding and germplasm resources*. Sinauer, Sunderland, pp 44–63
- Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond* 351:1291–1298
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59:1633–1638
- Heinrich B (1975) Energetics of pollination. *Ecol Syst* 6:139–170
- Herlihy CR, Eckert CG (2005) Evolution of self-fertilization at geographical range margins? A comparison of demographic floral, and mating system variables in central vs. peripheral populations of *Aquilegia canadensis* (Ranunculaceae). *Am J Bot* 92:744–751
- Holm T (1929) Morphology of North American species of *Polygala*. *Bot Gaz* 88:167–185
- Holsinger KE (1991) Mass-action models of plant mating systems: the evolutionary stability of mixed mating systems. *Am Nat* 138:606–622
- Holsinger KE (2000) Reproductive systems and evolution in vascular plants. *Proc Natl Acad Sci USA* 97:7037–7042
- Holsinger KE, Gottlieb LD (1991) Conservation of rare and endangered plants: Principles and prospects. In: Falk DA, Holsinger KE (eds) *Genetics and conservation of rare plants*. Oxford University Press, New York, pp 195–208
- Husband BC, Schemske DW (1996) Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* 50:54–70
- Jain SK (1976) The evolution of inbreeding in plants. *Annu Rev Ecol Syst* 7:469–495
- Jasieniuk M, Lechowicz MJ (1987) Spatial and temporal variation in chasmogamy and cleistogamy in *Oxalis montana* (Oxalidaceae). *Am J Bot* 74:1672
- Joseph Trapp E, Hendrix SD (1988) Consequences of a mixed reproductive system in the hog peanut, *Amphicarpaea bracteata*, (Fabaceae). *Oecologia* 75:285–290
- Kalish S, Vogler DW (2003) Benefits of autonomous selfing under unpredictable pollinator environments. *Ecology* 84:2928–2942
- Kalish S, Vogler DW, Hanley KM (2004) Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature* 430:884–887
- Kaul V, Koul AK, Sharma MC (2000) The underground flower. *Curr Sci* 78:39–44
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191
- Koontz SM, Weekley CW, Crate SJH, Menges ES (2015) Patterns of chasmogamy and cleistogamy by age and season in rare *Poylgala*. Botany 2015 Conference website. <http://2015.botany.conference.org/engine/search/index.php?func=detail&aid=170>. Accessed 30 July 2016
- Liang ZC, Huang P, Yang J, Rao GY (2009) Population divergence in the amphicarpic species *Amphicarpaea edgeworthii* Benth. (Fabaceae): microsatellite markers and leaf morphology. *Biol J Linn Soc* 96:505–516
- Lindon HL, Menges E (2008) Effects of smoke on seed germination of twenty species of fire-prone habitats in Florida. *Castanea* 73:106–110
- Lloyd DG (1992) Self- and cross-fertilization in plants II. The selection of self-fertilization. *Int J Plant Sci* 153:370–380
- Lord EM (1981) Cleistogamy: a tool for the study of floral morphogenesis, function and evolution. *Bot Rev* 47:421–449
- Loveless M, Hamrick J (1984) Ecological determinants of genetic structure in plant populations. *Annu Rev Ecol Syst* 15:65–95

- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60:2399–2402
- Menges ES, Weekley CW (2003) Continuation of demographic research on four state-listed Lake Wales Ridge endemic plants. Final report to the Division of Plant Industry
- Menges ES, Weekley CW, Clarke GL, Smith SA (2007a) Further demographic research on five state-listed Lake Wales Ridge endemic plants. Final report to the Division of Plant Industry
- Menges ES, Weekley CW, Hamz  SI, Pickert RL (2007b) Soil preferences for listed plants on the Lake Wales Ridge in Highlands County, Florida. *Fla Sci* 70:24–39
- Menges ES, Smith SA, Koontz SM (2014) Conservation research on state-listed plants endemic to the Lake Wales Ridge. Final report to endangered and threatened plant conservation grants program
- Meyer RC, Torjek O, Becher M, Altmann T (2004) Heterosis of biomass production in *Arabidopsis*. Establishment during early development. *Plant Physiol* 134:1813–1823
- Nei M (1978) Estimation of Average Heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
- Pritchard JK, Wen Z, Falush D (2010) Documentation for structure software: Version 2.3 1–39
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43:223–225
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8:103–106
- Rozen S, Skalesky H (1999) Primer3 on the WWW for general users and biologist programmers
- Sadeh A, Guterman H, Gersani M, Ovadia O (2009) Plastic bet-hedging in an amphicarpic annual: an integrated strategy under variable conditions. *Evol Ecol* 23:373–388
- Schemske DW (1978) Evolution of reproductive characteristics in *Impatiens* (Balsaminaceae): the significance of cleistogamy and chasmogamy. *Ecology* 59:596–613
- Schemske DW, Lande R (1985) The evolution of self-fertilization and inbreeding depression in plants II: empirical observations. *Soc Study Evol* 39:41–52
- Schemske DW, Pautler LP (1984) The effects of pollen composition on fitness components in a neotropical herb. *Oecologia* 62:31–36
- Schueler SK (2004) Self-pollination in island and mainland populations of the introduced hummingbird-pollinated plant, *Nicotiana glauca* (Solanaceae). *Am J Bot* 91:672–681
- Siol M, Prosperi JM, Bonnin I, Ronfort J (2008) How multilocus genotypic pattern helps to understand the history of selfing populations: a case study in *Medicago truncatula*. *Heredity* 100:517–525
- Solbrig OT (1976) On the relative advantages of cross- and self-fertilization. *Ann Missouri Bot Gard* 63:262–276
- Stebbins LG (1957) Self fertilization and population variability in the higher plants. *Am Nat* 91:337–354
- Sun M (1999) Cleistogamy in *Scutellaria indica* (Labiatae): effective mating system and population genetic structure. *Mol Ecol* 8:1285–1295
- Swindell WR, Bouzat JL (2006) Selection and inbreeding depression: effects of inbreeding rate and inbreeding environment. *Evolution* 60:1014–1022
- Travis SE, Proffitt CE, Ritland K (2004) Population structure and inbreeding vary with successional stage in created *Spartina alterniflora* marshes. *Ecol Appl* 14:1189–1202
- USFWS (1999) South florida multi-species recovery plan. U.S. Fish and Wildlife Service, Atlanta
- Weekley CW, Brothers A (2006) Failure of reproductive assurance in the chasmogamous flowers of *Polygala lewtonii* (Polygalaceae), an endangered sandhill herb. *Am J Bot* 93:245–253
- Weekley CW, Menges ES (2012) Burning creates contrasting demographic patterns in *Polygala lewtonii* (Polygalaceae): a cradle-to-grave analysis of multiple cohorts in a perennial herb. *Aust J Bot* 60:347–357
- Weekley CW, Menges ES, Pickert RL (2008) An ecological map of Florida's Lake Wales Ridge: a new boundary delineation and an assessment of post-columbian habitat loss. *Biol Sci* 71:45–65
- Weiss PW (1980) Germination, reproduction, and interference in the amphicarpic annual *Emex spinosa* (L.) Campd. *Oecologia* 45:244–251
- Zhang Y, Yang J, Rao G (2006) Comparative study on the aerial and subterranean flower development in *Amphicarpaea edgeworthii* Benth. (Leguminosae: papilionoideae), an amphicarpic species. *Int J Plant Sci* 167:943–949