**POPULATION GENOMICS AND CONSERVATION OF ERIGENIA BULBOSA (APICACEAE), AN EDGE-OF-RANGE SPECIES IN PENNSYLVANIA**

Angela J. McDonnell,1,* Cheyenne L. Moore,† Scott Schuette,‡ and Christopher T. Martine§

*Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022, USA; †Department of Biological Sciences, University of Pittsburgh, 4249 Fifth Avenue, Pittsburgh, Pennsylvania 15260, USA; ‡Pennsylvania Natural Heritage Program, Western Pennsylvania Conservancy, 800 Waterfront Drive, Pittsburgh, Pennsylvania 15222, USA; and §Bucknell University, 1 Dent Drive, Lewisburg, Pennsylvania 17837, USA

Editor: Nicola G. Bergh

**Premise of research.** *Erigenia bulbosa*, or the harbinger of spring, is one of the earliest-blooming wildflowers in eastern North America. As a spring ephemeral of forests and woodlands, it is a common species throughout the Midwest. In Pennsylvania, *E. bulbosa* exhibits an east-west disjunct distribution, where widespread western populations are contiguous with the midwestern range and a handful of smaller populations in the eastern part of the state are restricted to the lower Susquehanna River Valley. The relative isolation of the eastern populations suggests a possible conservation concern; the smaller, less connected populations may be threatened by fluctuations in size and the potential for low genetic diversity. As a consequence, establishing regulatory measures in Pennsylvania has been problematic because of disagreement regarding how to treat *E. bulbosa* during the Pennsylvania Department of Conservation and Natural Resources (PA DCNR) environmental review process. Currently, populations in the east are subject to regulation, while populations in the west are not.

**Methodology.** To better understand the population genetics of the species, we coupled field assessments of *E. bulbosa* with a population genomics approach. We sampled multiple individuals from eight populations and generated a genotyping-by-sequencing data set of single-nucleotide polymorphisms (SNPs) that we used to calculate population statistics (genetic variation \([F_{ST}]\), inbreeding coefficient \([F_{IS}]\), and heterozygosity), estimate population structure (sparse nonnegative matrix factorization), identify clusters of genetically related individuals (discriminant analysis of principal components), estimate a NeighborNet, look at signatures of isolation by distance (IBD), and detect population differentiation (analysis of molecular variance) using the filtered SNP data set.

**Pivotal results.** Our data reveal structure in all populations and suggest genetic isolation between the groups of populations in the eastern and western portions of the state. Genetic groups identified using multivariate methods correspond to populations. Within populations, we estimate low levels of heterozygosity, a significant signature of IBD, and that most of the genetic variation we find is due to differences between populations.

**Conclusions.** Our results indicate that any conservation measures undertaken by the PA DCNR should be applied evenly throughout Pennsylvania. We expect that most populations will continue to be threatened by land use and other development activities. This work illustrates the strength of academic and nonacademic partnerships in fostering outcomes that inform conservation activities for local species of concern.

**Keywords:** genotyping by sequencing, range edge, conservation, rare species, harbinger of spring, inbreeding.

**Online enhancement:** appendix.

**Introduction**

*Erigenia bulbosa* Nutt. belongs to a monotypic genus endemic to the eastern United States. *Erigenia bulbosa* (Michx.) Nutt. occurs in rich, moist, well-drained forests and woodlands in eastern North America, where it is one of the region’s earliest-blooming spring wildflowers (Buddell and Thieter 1985; fig. 1). Two hundred years ago, the species was noted as sometimes emerging through the snow during early March (Nuttall 1818), a tendency that lends support for its most-used common name: the harbinger of spring. While *E. bulbosa* is an early-season bloomer, plants have a slow start to their reproductive phase and may spend 5 or 6 yr in vegetative development before producing their first flowers—often blooming some 6 or 7 yr after germination. The longevity of the swollen underground portion that gives the geophyte its epithet is unknown, but the lag between vegetative and reproductive growth may promote outcrossing. Pollination is likely facilitated by *Adrena* spp. and *Apis mellifera* bees as well as by muscid and syrphid flies (Dailey and Scott 2006). Once reproductively mature, plants produce small umbels of...
diminutive white flowers with bright maroon to dark brown anthers, with the resulting color contrast reflected in another common name: pepper and salt (Holm 1901, 1925; Buddell and Thieret 1985; fig. 2). Erigenia bulbosa is protogynous, a type of floral dichogamy in which the carpels mature before the stamens to minimize self-fertilization (Schlessman et al. 2004). While outcrossing in the species is likely, mature seeds are very small and mostly smooth and do not appear to have dispersal vectors beyond the immediate vicinity of the parent plants (Buddell and Thieret 1985). Recruitment of new individuals in a population appears to be from the successful germination of seeds; the species is not clonal.

Erigenia bulbosa is a relatively common diploid perennial herb (n=11; Constance et al. 1976) throughout the midwestern United States, with the bulk of its distribution centered in the states of Missouri, Illinois, Indiana, and Ohio. It is increasingly rare at its range edges in southeastern Kansas, northeastern Oklahoma, northwestern Georgia, southwestern Virginia, Wisconsin, New York, and Maryland (NatureServe 2019; USDA NRCS 2019). In Pennsylvania, populations of E. bulbosa range in size from fewer than 10 to more than 30,000 individuals, depending on the location and habitat quality. Populations of many species that occur at or near a range edge are often considered unique because of their possible genetic isolation. Thus, the conservation of populations at the periphery of a distribution may be important when they are genetically divergent, which can be an effect of their often isolated nature and compounded by genetic drift and/or natural selection, which may have important implications for the long-term conservation of species (Lesica and Allendorf 1995). Additionally, relative to the core of a species’ range, populations at edges can be susceptible to discontinuous changes in environmental variables that lead to patchiness, both in suitable habitats and in species abundance (Brown 1984; Lawton 1993).

In Pennsylvania, E. bulbosa is at or near its northeastern range edge (fig. 1) and consists of two notably disjunct sets of populations separated by the Ridge and Valley and the Central Appalachians Level III ecoregions (Woods et al. 1999). No populations of E. bulbosa have been found in the Ridge and Valley and the Central Appalachians ecoregions in Pennsylvania despite numerous recent surveys by state botanists in suitable habitats within these ecoregions. Erigenia bulbosa is relatively
widespread and more abundant in the western part of the state, where populations are more or less contiguous with the midwestern distribution. In eastern Pennsylvania, however, populations are restricted to the Northern Piedmont Level III ecoregion (Woods et al. 1999) in the lower Susquehanna River Valley (Kartesz 2015; fig. 3). This distribution has inspired discussion regarding the conservation and management of the species by the Pennsylvania Department of Conservation and Natural Resources (PA DCNR). There are 37 populations totaling approximately 50,000 individuals in western Pennsylvania, where the species has been considered stable, leading conservation practitioners to assign it a low conservation priority. The converse

Fig. 2  *Erigenia bulbosa* in Pennsylvania, morphology and habitat. A, *Erigenia bulbosa* emerging from *Acer* leaf litter with less dissected early leaves. B, Raccoon Creek State Park Wildflower Reserve, Beaver County, western Pennsylvania. C, Collecting *E. bulbosa* and associated ecological data from the Peach Bottom site with Natural Heritage Program botanists Rachel Goad (left) and John Kunsman in York County. D, *Erigenia bulbosa* in flower and fruit at the Raccoon Creek State Park Wildflower Reserve, Beaver County, western Pennsylvania. Photos: A. J. McDonnell and S. Schuette.
is true for the populations from the lower Susquehanna River Valley regions in eastern Pennsylvania, where the Appalachian Mountains serve as a barrier that isolates three populations that total approximately 2000 individuals. Possible fluctuation in population size there poses a higher risk of genetic bottlenecks and/or population extinction, which would make the eastern populations of greater conservation concern. The differences in the populations and the current disjunction within Pennsylvania have led some botanists to recommend separate conservation plans (and consequent regulatory approaches) for plants in the east versus those in the west—an approach that is not only difficult to justify to lawmakers but also potentially difficult for regulatory agencies to enforce.

Erigenia bulbosa has a long history as a regulated species in Pennsylvania, with an official status as state threatened from 1988 until 2007 (Commonwealth of Pennsylvania 1993). Surveys conducted by botanists from the Pennsylvania Natural Heritage Program in western Pennsylvania showed the species to be frequent and abundant in open wooded areas; many populations included thousands of individuals. However, the plant was known in eastern Pennsylvania from four populations that have fluctuated in size since 1980. At present, the populations are relatively small (160–1200 individuals) and occur on steep wooded slopes adjacent to waterways. The Vascular Plant Technical Committee (VPTC), comprising botanists across the state and serving as an advisory body to the PA DCNR, voted in 2008 to downlist E. bulbosa from threatened to rare in Pennsylvania on the basis of an apparent increase in population sizes inferred from field surveys. In conjunction with the decision to downlist the species statewide, the VPTC also voted to recognize the eastern populations as under greater threat than the western populations. While the species remained subject to environmental review when new projects were proposed in habitats where E. bulbosa was found, the western populations were considered
exempt from conservation recommendations. This two-tiered special regulatory consideration has created confusion during environmental review processes and has left the PA DCNR in a difficult position when reasons for the partial regulation for a species in Pennsylvania require justification.

While recently revisiting the regulatory status of *E. bulbosa*, the PA DCNR suggested that it needed more evidence to justify an approach that would lead to differential regulation within the same species. In 2016, the PA DCNR solicited grant proposals to investigate *E. bulbosa* in Pennsylvania. In particular, there is interest in examining whether the eastern populations are geographically and genetically isolated from the western populations and represent a justifiable conservation concern from a regulatory standpoint. To test the hypothesis that the disjunction in the distribution of *E. bulbosa* as observed by conservation practitioners is reflected in its genetic history, we sampled tissues from eight populations throughout Pennsylvania, extracted DNA, and used a genotyping-by-sequencing (GBS) approach to approach to obtain many single-nucleotide polymorphisms (SNPs), a restriction enzyme–based approach appropriate for obtaining many loci from nonmodel organisms that has been used extensively in recent years (Seeb et al. 2011; Peterson et al. 2012; Schilling et al. 2014; Hu et al. 2015; Silliman 2019). We used a filtered SNP data set to estimate population genetic parameters such as genetic variation (*F*$_{ST}$), the inbreeding coefficient (*F*$_{IS}^{'}$), and heterozygosity, visualize the spread of our data using a discriminant analysis of principal components (DAPC), examine population structure using sparse nonnegative matrix factorization (sNMF), infer a population network using a NeighborNet analysis, compare genetic variance within and among groups using an analysis of molecular variance (AMOVA), and examine whether there is a signature of isolation by distance (IBD).

### Material and Methods

Specimens and tissue for DNA extraction were collected in March and April of 2018 from eight sites in Pennsylvania. For each site, population size was assessed, leaves and/or flowers from 10–20 individuals selected from throughout the population were sampled and placed in silica gel, one or two voucher specimens were pressed and dried, and relevant ecological data were collected. All collecting was done under PA DCNR Wild Plant Management Permit numbers 18-001, 18-046, and 18-739. Voucher specimens were deposited in the Wayne E. Manning Herbarium at Bucknell University (BUPL). Because the species is a geophyte and has a tuber, we assumed that our aboveground counts were a reliable estimate of the number of individuals.

DNA was extracted from silica-dried tissue using either a modified CTAB method (Doyle and Doyle 1987) or the FastDNA kit (MP Biomedicals, Santa Ana, CA). DNA samples were quantified using the Qubit dsDNA BR Assay Kit on a Qubit version 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA), and 2–3 μL of DNA was visualized on 1% agarose gels run at 100 V for 1.5 h. To ensure the digestibility of our extracts, 10% of the samples were digested with EcoRI-HF (New England BioLabs, Ipswich, MA) and visualized on a 1% agarose gel. Genomic DNA was shipped to the University of Wisconsin Biotechnology Center lab (http://www.biotech.wisc.edu/services/dnaseq), where additional enzyme testing, library preparation, and Illumina sequencing were completed.

Fragment analyses of five different enzyme combinations revealed that the ApeKI enzyme showed the greatest activity and produced a range of fragment sizes in sufficient abundance for the *Erigenia* DNA samples (see Nguyen et al. 2018). Therefore, a one-enzyme GBS approach was used (Elshire et al. 2011). Total genomic DNA was digested using ApeKI, libraries were prepared, quantified, and pooled, and 150-bp paired-end sequencing was done on a NovaSeq 6000 instrument (Illumina, San Diego, CA). All raw sequencing data are available online via the National Center for Biotechnology Information website, BioProject number PRJNA545957.

Raw data were sorted and de novo assembled using ipyrad software version 0.7.30 (Eaton 2014) with the following settings: datatype = pairgbs, max_low_qual_bases = 5, clust_thresh old = 0.85, max_barcode_mismatches = 0, filter_adapters = 2, max_alleles_consens = 2, min_samples_locus = 4, and max_SNPs_locus = 20. The variant call format (VCF) file output from ipyrad was then filtered using VCFtools 0.1.16 (Danecek et al. 2011) using the following flags: --maf 0.05, --max-alleles 2, --min-alleles 2, --min-meanDP 5, --min-meanDP 3, --max-missing 0.6, --recode, and --remove-indels. After filtering, two samples with 80% or more missing data were removed.

Most analyses used R software (ver. 3.6.0) and packages (R Core Team 2019). First, we explored the qualities of our filtered VCF file using tidyverse ( Wickham et al. 2019), including read quality, read depth per site, missing data per site, allele frequencies, read depth per individual, missing data per individual, and heterozygosity and inbreeding estimates. Some descriptive statistics, including *F*$_{ST}$ (Weir and Cockerham 1984) and *F*$_{IS}$ for each population, were also calculated using the packages dartR (Gruber et al. 2018) and hierfstat (Goudet 2005). The packages pegas (Paradis 2010) and adegenet (Jombart and Mougin 2015) were used to conduct Bartlett’s tests to compare variances. The data were transformed using principal components analysis (PCA), and then clusters were identified using discriminant analysis (DA). DAPC requires the user to define the number of PCs retained in the analysis, so we tested between three and 120 PCs and used the xvalDapc function to cross-validate and identify the optimal number of PCs that best fit our data.

The LEA package (Franchot and François 2015) was used to infer individual admixture coefficients using 100 replicates of 1000 iterations. The snmf function estimates admixture coefficients between 10 and 30 times faster than other methods using sNMF to calculate individual admixture coefficients over a range of *K* values (here, 1–8). We determined the number of ancestral populations through a comparison of the cross-entropy values for each value of *K*.

We used the NeighborNet algorithm (Bryant and Moulton 2004) to estimate cyclic splits and visualize relationships within and among the sampled individuals. A network was generated using SplitsTree5 version 5.0.0_alpha (Huson 1998; Huson and Bryant 2006) using the filtered SNP VCF file, which was first converted to a phylip file using the vcfl2phylip.py script (Ortiz 2019).
The K2P model (Kimura 1980) was used to obtain a distance matrix, and the splits network algorithm (Dress and Huson 2004) was used to estimate a splits network.

Finally, we used dartR and ade packages to conduct Mantel tests to look for IBD and the poppr package to conduct an AMOVA (Excoffier et al. 1992; Excoffier and Smouse 1994). Plots were generated using the package ggplot2 (Wickham 2016); all R code used is available at http://www.github.com/cheyennelmoore/Erigenia-Analyses.

Results

A total of 118 samples of *Erigenia bulbosa* from Pennsylvania were successfully sequenced and included in downstream analyses. The average number of reads obtained per sample was 2.3 million, most of which passed quality filtering (table 1). Assembly of the data with ipyrad resulted in 237,361 total loci. Of these, 99,719 were retained after filtering within ipyrad and were used to generate a concatenated matrix that included 439,647 biallelic SNPs. Additional filtering of the concatenated matrix using VCFtools produced our final data set of 14,350 biallelic SNPs for 118 samples. The filtered SNPs have an average read depth per site of 18.9 (minimum, 8.2; maximum, 969.6) and average missing data per sample of 29% (minimum, 0%; maximum, 40%). These characteristics, along with read depth per individual, percent of missing data per site, observed heterozygosity, the inbreeding coefficient, and minor allele frequency, were plotted and are available as an appendix (available online).

Estimates of pairwise *F*_{ST} show that across the range, populations are considerably well differentiated (fig. 4); our global estimate of *F*_{ST} for the sampled populations is 0.518. When populations are considered in a spatial context, the data follow a pattern corresponding to roughly two regions; one corresponds to the east, and these are well separated; these correspond to our eight sampled populations. Overall, the DAPC analyses estimated in-group admixture coefficients, much like a STRUCTURE analysis (Pritchard et al. 2000). The *k*-means step, along with the cross-entropy criterion, identified clusters in the data and uncovered that *K* = 8 best describes the data; the eight groups found correspond to the sampled populations. Within each group, there are low to moderate levels of admixture (fig. 5).

We used two approaches to infer population structure by determining the number of groups or populations observed in our data: sNMF and DAPC. The sNMF analyses estimated individual admixture coefficients, much like a STRUCTURE analysis (Pritchard et al. 2000). The *k*-means step, along with the cross-entropy criterion, identified clusters in the data and uncovered that *K* = 8 best describes the data; the eight groups found correspond to the sampled populations. Overall, the DAPC reveals a split between east and west, and these are separated by the vertical axis in the DAPC plot (fig. 6), with the left side

| Population name                          | No. individuals included in analyses | Pennsylvania locality | Voucher (herbarium code) or survey no./voucher (herbarium code) | Average no. high-quality reads per sample | Average no. loci in assembly |
|------------------------------------------|-------------------------------------|-----------------------|---------------------------------------------------------------|------------------------------------------|------------------------------|
| Peach Bottom                             | 20                                  | York County           | McDonnell #366 (BUPL)                                         | 1.83 million                             | 19,493                      |
| York Furnace                             | 14                                  | York County           | McDonnell #367 (BUPL)                                         | 2.40 million                             | 21,141                      |
| Safe Harbor                              | 9                                   | York County           | McDonnell #368 (BUPL)                                         | 1.90 million                             | 21,907                      |
| Braddock’s Trail Park                    | 15                                  | Westmoreland County  | Schuette #F18SCH03/                                          | 1.91 million                             | 21,884                      |
|                                          |                                     |                       | Schuette 2095 (BUPL)                                          |                                         |                              |
| Cedar Creek Park                         | 15                                  | Westmoreland County  | Schuette #F18SCH02/                                          | 2.41 million                             | 21,921                      |
|                                          |                                     |                       | Schuette 2094 (BUPL)                                          |                                         |                              |
| Slippery Rock Creek Natural Area         | 15                                  | Butler County         | Schuette #F18SCH05/                                          | 2.31 million                             | 27,596                      |
| Raccoon Creek State Park Wildflower Reserve | 15                               | Beaver County         | Schuette #F18SCH01/                                          | 2.41 million                             | 22,287                      |
|                                          |                                     |                       | Isaac 10420 (CM)                                              |                                         |                              |
| Ryerson Station State Park               | 15                                  | Greene County         | Schuette #F18SCH04/                                          | 3.23 million                             | 32,956                      |
|                                          |                                     |                       | Schuette 2096 (BUPL)                                          |                                         |                              |
Fig. 4  Heat map of pairwise genetic variation ($F_{ST}$) values (Weir and Cockerham 1984) for eight populations of *Erigenia bulbosa* (*n* = 118) using 14,350 single-nucleotide polymorphisms. Populations sampled from eastern Pennsylvania include Peach Bottom, York Furnace, and Safe Harbor. Populations sampled from western Pennsylvania include Braddock’s Trail Park, Cedar Creek Park, Slippery Rock Creek Natural Area, Raccoon Creek State Park Wildflower Reserve, and Ryerson Station State Park.

Table 2  Descriptive Statistics for Each Population, including Observed Heterozygosity ($H_O$), Expected Heterozygosity ($H_E$), Bartlett’s $K^2$, and Inbreeding Coefficient ($F_{IS}$) Values

| Population                          | $H_O$ | $H_E$ | Bartlett’s $K^2$ | $F_{IS}$ |
|-------------------------------------|-------|-------|-------------------|----------|
| Peach Bottom                        | .032  | .126  | 8189.6*           | .749     |
| York Furnace                        | .037  | .103  | 4109.2*           | .635     |
| Safe Harbor                         | .027  | .062  | 3076.3*           | .558     |
| Braddock’s Trail Park               | .095  | .219  | 2720.5*           | .565     |
| Cedar Creek Park                    | .053  | .145  | 4495.7*           | .633     |
| Slippery Rock Creek Natural Area   | .089  | .232  | 7087.1*           | .617     |
| Raccoon Creek State Park Wildflower Reserve | .029  | .108  | 6815.9*           | .735     |
| Ryerson Station State Park          | .080  | .221  | 3416.9*           | .640     |

Note. The Bartlett’s $K^2$ test statistic compares the variances between $H_O$ and $H_E$.
* $P$ is significant at $\alpha = 0.05$. 
of the graph containing the easternmost populations, while the right side contains populations sampled from the west. The analysis draws inertia ellipses for each found group when more than one DA is retained; these represent the relative position of each group along the two selected axes. While there is separation between populations in the eastern and western portions of the state and most populations are separated from each other, there is overlap between individuals from the Peach Bottom and York Furnace sites in the east, which may suggest a retention of shared ancestry indicative of previous connectivity between these neighboring populations. Alternatively, it may be suggestive of a founder event after a prior dispersal from one of these populations to the other.

Patterns of common ancestry are corroborated by the NeighborNet plot (fig. 7), which shows that the Peach Bottom and York Furnace samples share many edges. The NeighborNet also shows strong separation by geography; populations in the western portion of the state are on the left side of the graph, while populations from the east are found on the right side. The parallel edges at the center of the graph suggest that no recent gene flow has occurred between populations in the east and west. The lack of parallel edges near the tips within each region suggests that gene flow among populations has not occurred recently.

Our tests for IBD among all populations were significant (999 replicates; Mantel statistic $r$: 0.944, $P = 0.01$), which suggests that local genetic variation has accrued in the sampled populations. The AMOVA, which is based on differences between pairs of different haplotypes, uncovers some of the genetic diversity within and between populations. We found a much larger proportion of between-population (53.82%) than within-population (29.77%) variance, which is summarized from all SNPs in our filtered data set.

**Discussion**

Pennsylvania *Erigenia* populations are each genetically unique and appear to have much lower than expected levels of heterozygosity and high levels of inbreeding as measured by $F_{IS}$, regardless of their differences in census size. As a species existing at the edge of its range, *Erigenia bulbosa* may be impacted by numerous factors...
previously identified in the literature; edge-of-range taxa often inhabit sites that are ecologically marginal (Abeli et al. 2014), show decreases in seed production (Jump and Woodward 2003), and experience increased impacts due to climate change (Rehm et al. 2015). The central marginal hypothesis (Antonovics et al. 2002; Eckert et al. 2008) suggests that populations on the edges of a distribution might also be expected to exhibit lower levels of genetic diversity and higher measures of genetic differentiation, which is likely related to historical genetic drift, founder, inbreeding, and/or bottleneck events.

The life history and biology of *E. bulbosa* may also support highly structured, separately evolving populations. Individual plants are slow to mature, growing vegetatively until producing their first flowers 6 or 7 yr after germination (Buddell and Thie-ret 1985). Flowers are protogynous, with carpels receptive before the maturation of the anthers (Schlessman et al. 2004), which appears to support outcrossing. After maturity, *E. bulbosa* blooms for only a short period of time in the early spring of each year, and it is unknown how many flowering seasons individuals undergo. After pollination, the plants produce small capsules with tiny seeds that appear to have limited dispersal capability on the basis of our observations during leaf and specimen collection. Limited pollen and seed dispersal within plant populations likely facilitates the establishment of inbreeding and genetic substructuring through genetic drift in the absence of selection (Tero et al. 2005). Limitations to gene flow may be especially problematic for *E. bulbosa* in eastern Pennsylvania, where only four populations containing around 2000 individual plants have been relocated by recent surveys. Even in western Pennsylvania, where the species is found in 37 populations totaling approximately 50,000 individuals, our results indicate that possible genetic limitations could be further exacerbated by the threats noted above. It is possible that within populations, we sampled a wide range of age cohorts that may include close relatives. While we made an effort to sample individuals from throughout each site, sampling of close relatives might lead to an overestimate of $F_{IS}$. We do not know enough about the life span of *E. bulbosa*, but this aspect of its biology may certainly impact our results and would be worthy of future investigation.

Inbreeding within a population can decrease genetic diversity and impacts the effective population size (Charlesworth 2009), or the number of individuals from an ideal population that have the same genetic response to random processes as the real population size over time (Ellstrand and Elam 1993). While the inbreeding coefficients we estimated suggest that the amount
of heterozygosity relative to that of the ancestral population appears to be very low, the $F_{IS}$ values alone do not necessarily indicate the presence of inbreeding depression, where low heterozygosity is shown to be correlated with low fitness; inbreeding depression in *Erigenia* should also be a focus of future research. However, in combination with AMOVA, which tells us that most of the genetic variation in our data set occurs between populations and is a significant signal of the isolation of populations by distance, we speculate that these populations are still at risk from future threats. The lack of corridors and current population connection in combination with threats from land use and other development activities that increase the likelihood of population fragmentation may potentially have negative impacts on population sizes that in turn may result in decreased genetic diversity (Waples 2010).

These disjunct sets of populations are separated by rather extensive landscape-level changes, including the Ridge and Valley and the Central Appalachians ecoregions; no populations of *E. bulbosa* have been found between the east and west sites, in the Ridge and Valley and the Central Appalachians ecoregions, during recent surveys. Sites harboring eastern populations of *E. bulbosa* have already experienced disruptions from the damming of waterways, a process that has led to the burial of floodplain habitats. Field surveys in these floodplains failed to relocate historical populations that are now considered extirpated from these areas (fig. 3, inset). Likewise, some extant populations occur on privately held land that is actively used and is also under pressure from development.

Potential negative effects from genetic drift coupled with existing low heterozygosity and low dispersal capabilities lead us to speculate that *E. bulbosa* might not successfully adapt to future land use changes and, presumably, the effects of ongoing climate change (Dawson et al. 2011). Additional studies are needed to determine the effective population size of *E. bulbosa* so that appropriate conservation measures are implemented to ensure that the species persists in its natural habitat. We believe that all populations in the state warrant the same conservation consideration. This would entail making a change to the current two-tiered special regulatory consideration, which requires partial regulation for the populations in the east and exempts western populations from conservation recommendations. Our suggestion helps to eliminate confusion about the species during environmental review processes and removes the PA DCNR from the difficult position of justifying partial regulation. Furthermore, it is supported by our findings; within populations, we estimate low levels of heterozygosity, a significant signature of IBD, and

![NeighborNet network as estimated by SplitsTree5. Populations sampled from western Pennsylvania include Braddock’s Trail Park, Cedar Creek Park, Slippery Rock Creek Natural Area, Raccoon Creek State Park Wildflower Reserve, and Ryerson Station State Park. Populations sampled from eastern Pennsylvania include York Furnace, Peach Bottom, and Safe Harbor.](image_url)
have found that most of the genetic variation is due to differences between populations. Evenly applied conservation measures throughout Pennsylvania will be important to protect populations from changes in land use, development, and other threats.

Our study is among the first in the region to use population-scale genomic data to help inform conservation priorities. Similar results were found for the wetland sedge Scirpus ancrostaetus Schuyler, also in Pennsylvania, where plants sampled from the same wetland exhibited low heterozygosity and little genetic variation (Cipollini et al. 2017). Genetic diversity in populations of rare plants is considered an important advantage in responding to environmental changes and should be considered a primary goal for conservation (Frankham 2005; Cipollini et al. 2017). Likewise, this information must be included in conservation status assessments so that the appropriate steps are taken to ensure their long-term viability. While the NatureServe Conservation Rank Calculator does include introduced genetic material as a threat to a species, the use of population genetic information in the status rank calculators and rank assessments is not yet considered in the threats assessment (NatureServe 2020). We hope that our work helps to highlight the need for conservation status assessments to include population genetic data that elicit informed conservation planning for rare species.

Our collaborative team of agency and academic botanists has shown that while *E. bulbosa* in Pennsylvania comprises geographically and genetically distinct units, all populations appear equal in deserving some level of conservation consideration.

There are benefits when academic and nonacademic partners work together to foster outcomes that inform conservation actions affecting local species of concern. In this particular case, we took advantage of the opportunity to combine a field-based understanding of the statewide distribution and status of *E. bulbosa* with newer techniques in population genomics. This approach has led to an applied conservation science recommendation and generated opportunities for early-career researchers in areas related to both academic and nonacademic conservation biology. Meaningful advances in native plant conservation efforts in Pennsylvania and everywhere likely hinge on the cultivation of similar partnerships.

**Acknowledgments**

This project was funded by a grant from the PA DCNR Wild Resource Conservation Program (WRCP-17571) and the David Burpee Endowment at Bucknell University. Rachel Goad and John Kunzman provided invaluable field assistance in locating and collecting populations in eastern Pennsylvania. Dr. Maile Neel at University of Maryland herbarium (MARY) granted access to specimens from the eastern part of the range. We acknowledge the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing GBS and Illumina sequencing services. We thank Dr. Alistair Potts and two anonymous reviewers for comments and suggestions that greatly improved this work.

**Literature Cited**

Abeli T, R Gentili, A Mondoni, S Orsenigo, G Ross 2014 Effects of marginality on plant population performance. J Biogeogr 41:239–249.

Antonovics J, TJ Newman, BJ Best 2002 Spatially explicit studies on the ecology and genetics of population margins. Pages 97–117 in J Silvertown, J Antonovics, eds. Integrating ecology and evolution in a spatial context. Blackwell, Oxford.

Brown JH 1984 On the relationship between abundance and distribution of species. Am Nat 124:255–279.

Bryant D, V Moulton 2004 Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265.

Buddell GF, JW Thieret 1985 Notes on *Erigenia bulbosa* (Apiaceae). *Bartonia* 51:69–76.

Charlesworth B 2009 Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205.

Cipollini K, P Lavretsky, D Cipollini, J Peters 2017 Genetic population structure across the range of endangered northeastern bulrush *Scirpus ancrostaetus*. Int J Plant Sci 178:67–78.

Commonwealth of Pennsylvania 1993 Conservation of Pennsylvania native wild plants. Chapter 82 *in* Pennsylvania code: title 25. Environmental Resources, Harrisburg, PA.

Constance L, TI Chuang, CR Bell 1976 Chromosome numbers in Umbelliferae V. Am J Bot 63:608–625.

Dailey TB, PE Scott 2006 Spring nectar sources for solitary bees and flies in a landscape of deciduous forest and agricultural fields: production, variability, and consumption. J Torrey Bot Soc 133:535–547.

Danecek P, A Auton, G Abecasis, CA Albers, E Banks, MA DePristo, RE Handsaker, et al 2011 The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Dawson TP, ST Jackson, JJ House, IC Prentice, GM Mace 2011 Beyond predictions: biodiversity conservation in a changing climate. Science 332:53–58.

Doyle J, JL Doyle 1987 Genomic plant DNA preparation from fresh tissue: the CTAB method. Phytochem Bull 19:11.

Dress AM, DH Huson 2004 Constructing splits graphs. Trans Comput Biol Bioinform 1:109–115.

Eaton DAR 2014 PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849.

Eckert CG, KE Samis, SC Lougheed 2008 Genetic variation across species’ geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol 17:1170–1188.

Ellstrand NC, DR Elam 1993 Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242.

Elsheire RJ, JC Glaubitz, Q Sun, JA Poland, E Buckler, SE Mitchell 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379.

Excoffier L, PE Smouse 1994 Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* 136:343–359.

Excoffier L, PE Smouse, JM Quattro 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Frankham R 2005 Genetics and extinction. *Biol Conserv* 126:131–140.

Frichet E, O François 2015 LEA: an R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929.

Goudet J 2005 HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186.
Gruber B, PJ Umack, OF Berry, A Georges 2018 dartR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18:691–699.

Holm T 1901 *Erigenia bulbosa* Nutt.: a morphological and anatomical study. Am J Sci 11:63–72.

——— 1925 Hibernation and rejuvenation, exemplified by North American herbs. Am Midl Nat 9:439–512.

Hu Z, B Mbacké, R Perumal, MC Guèye, O Sy, S Bouchet, PV Vara Prasad, GP Morris 2015 Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): comparative analysis of global accessions and Senegalese landraces. BMC Genomics 16:1048.

Huson DH 1998 SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.

Huson DH, D Bryant 2006 Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267.

Jombart T 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jump AS, FI Woodward 2003 Seed production and population density decline approaching the range-edge of *Cirsium* species. New Phytol 160:349–358.

Kartesz JT 2015 BONAP North American plant atlas. Biota of North America Program, Chapel Hill, NC. http://bonap.net/napa.

Kimura M 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Lawton JH 1993 Range, population abundance and conservation. Trends Ecol Evol 8:409–413.

Lesica P, FW Allendorf 1995 When are peripheral populations valuable for conservation? Conserv Biol 9:753–760.

NatureServe 2019 Comprehensive report: *Erigenia bulbosa*. http://explorer.natureserve.org/servlet/NatureServe?searchName=Erigenia+bulbosa.

——— 2020 Conservation rank calculator, version 3.2. http://www.natureserve.org/conservation-tools/conservation-rank-calculator.

Nguyen TK, J Yu, H Choi, B In, J Lim 2018 Optimization of genotyping-by-sequencing (GBS) in chrysanthemums: selecting proper restriction enzymes for GBS library construction. Hortic Sci Technol 36:108–114.

Nuttall T 1818 The genera of North American plants: and a catalogue of the species, to the year 1817. D Heartt, Philadelphia.

Ortiz EM 2019 vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. http://doi.org/10.5281/zenodo.2540861.

Paradis E 2010 pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420.

Peterson BK, JN Weber, EH Kay, HS Fisher, HE Hockstra 2012 Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Pritchard JK, M Stephens, P Donnelly 2000 Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team 2019 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/.

Rehm EM, P Olivas, J Stroud, KJ Feeley 2015 Losing your edge: climate change and the conservation value of range-edge populations. Ecol Evol 5:4315–4326.

Schilling MP, PG Wolf, AM Duffy, HS Rai, CA Rowe, BA Richardson, KE Mock 2014 Genotyping-by-sequencing for *Populus* population genomics: an assessment of genome sampling patterns and filtering approaches. PLoS ONE 9:e95292.

Schlessman MA, FR Barrie, BE van Wyk, PM Tilney 2004 Protogyne in Apiaceae, subfamily Apioideae: systematic and geographic distributions, associated traits, and evolutionary hypotheses. S Afr J Bot 70:475–487.

Seeb JE, G Carvalho, L Hauser, K Naish, S Roberts, LW Seeb 2011 Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol Ecol Resour 11:1–8.

Siliman K 2019 Population structure, genetic connectivity, and adaptation in the Olympia oyster (*Ostrea lurida*) along the West Coast of North America. Evol Appl 12:923–939.

Tero N, J Aspi, P Siikamäki, A Jäkäläniemi 2005 Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). Heredity 94:478–487.

USDA NRCS (Natural Resources Conservation Service) 2019 The PLANTS database. National Plant Data Team, Greensboro, NC. http://plants.usda.gov.

Waples RS 2010 Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Mol Ecol Resour 10:785–796.

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

Wickham H, M Averick, J Bryan, W Chang, L D'Agostino McGowan, R Françoise, G Grolemund, et al 2019 Welcome to the tidyverse. J Open Source Softw 4:1686.

Wood AJ, JM Omernik, DD Brown 1999 Level III and IV ecoregions of Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Corvallis, OR.