Phacelia formosula (Boraginaceae) genetic research: Executive Summary of Research conducted by Denver Botanic Gardens

Authors James P. Riser II, Jennifer Ramp Neale, Michelle DePrenger-Levin, Melissa B. Islam, Anna L. Schwabe. Denver Botanic Gardens Sept 26, 2019

In 1919, George Osterhout described *Phacelia formosula* Osterh. (North Park phacelia) from Jackson County in northern Colorado, USA (Osterhout 1919). *P. formosula* is endemic to the North Park basin where it occurs in small scattered populations restricted to soils derived from the Coalmont Formation at elevations from about 2400 to 2500 m (Fig. 1) (U.S. Fish and Wildlife Service (USFWS), 1986; 2011). North Park phacelia is a biennial plant that only produces flowering plants from the previous year's vegetative rosettes, thus seed production and rosette survival are essential to its long-term persistence. The species is therefore sensitive to habitat disturbance and climate change (drought in particular) resulting in yearly fluctuations in the number of flowering individuals. *P. formosula* was listed as endangered under the Endangered Species Act in 1982 (47 FR 38540-38543). The species faces threats due to its restriction to highly erodible soils, which are readily utilized by off-highway vehicles. Additional threats stem from coal, oil, and gas exploration, and livestock grazing within *P. formosula* habitat. Known occurrences are distributed across Bureau of Land Management (BLM), US Fish and Wildlife Service (USFWS Arapaho National Wildlife Refuge), and private lands (Colorado Natural Heritage Program; CNHP 2014).

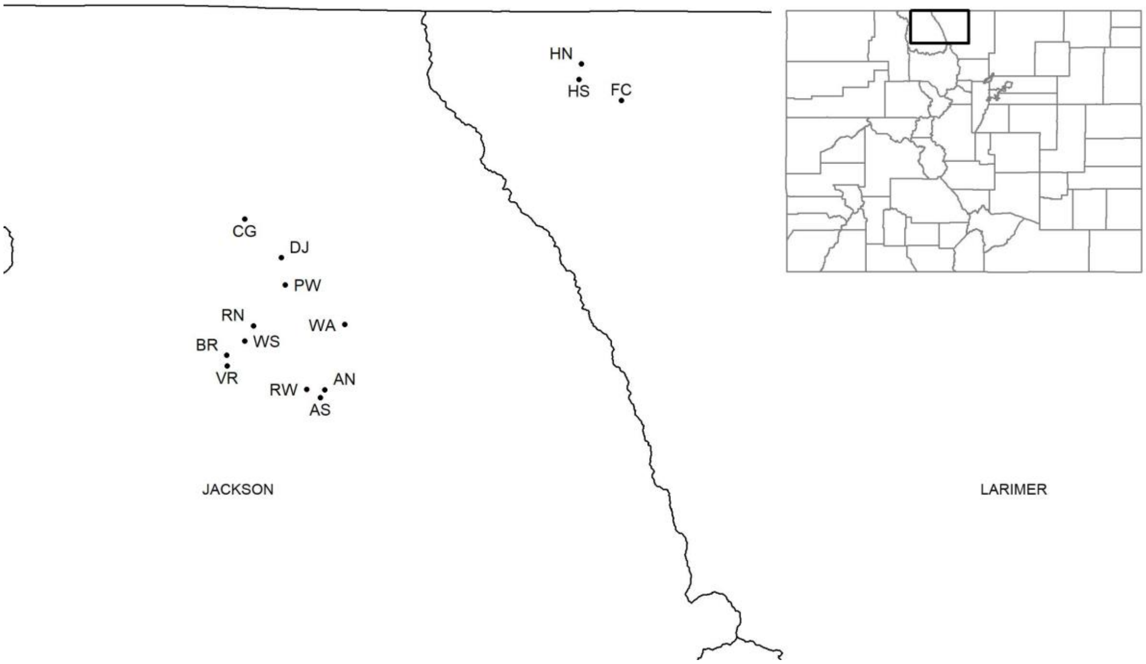

Figure 1. Map of northern Colorado (USA) showing locations of 14 populations spanning the known range of *Phacelia formosula* in Jackson and Larimer Counties. Population codes follow Table 1.

In 2004, two occurrences of a morphologically similar *Phacelia* were discovered approximately 35 km from the North Park plants in adjacent Larimer County (Fig. 1) (USFWS 2011). These two *Phacelia* populations in the Laramie River Valley are separated from the North Park plants by the Medicine Bow Mountain range (CNHP 2014). The Larimer County plants grow on soils derived from the Niobrara Formation, not the Coalmont Formation as seen with *P. formosula* in Jackson County. It is unclear whether this represents a new species or a range extension of *P. formosula*. In 2013, Atwood and Welsh referred to these populations as "*P. formosula* var. *schulyii;*" however, this name was never formally published (Atwood and Welsh, 2013).

Given the morphological similarities between these populations of *Phacelia* in northern Colorado, questions have arisen as to whether or not these populations are taxonomically distinct. Here we use multiple lines of evidence to assess the taxonomic identity of Jackson County *P. formosula* and the adjacent Larimer County *Phacelia* populations. The distinctiveness of the Larimer County populations from the Jackson County populations is assessed using morphological, genetic, and ecological characters. Specifically, we ask, is there significant morphological, genetic, and/or ecological differentiation between the sampled geographic regions to warrant the recognition of the Larimer County *Phacelia* populations as a distinct species or distinct at any taxonomic level?

Methods

Sample collection:

Fieldwork to collect genetic samples from previously identified *Phacelia* populations in northern Colorado occurred in July of 2009. Leaf tissue samples from each of 30 individuals were collected at 11 original element occurrence sites in Jackson County representing the eight then known *P. formosula* populations (Fig. 1). Three additional collection sites were added from adjacent Larimer County (30 individuals/site) representing two *Phacelia* occurrences of uncertain taxonomic affinity. At each sampled individual:

-GPS coordinates were taken.

-Leaf tissue for genetic analysis was silica dried in the field prior to DNA extraction at Denver Botanic Gardens.

Morphology:

Fieldwork to collect morphological trait data was conducted in August - September of 2014. Nine morphological traits were measured and leaves collected for morphometric analysis from 30 individuals at four sites representing the range of the populations sampled for genetic assessment (site codes: FC, RN, BR, and AS; Table 1). Measurements were taken in the field using digital calipers.

A single leaf was removed from each sampled individual and placed flat in a labeled coin envelope, which was then dried in silica gel and imaged for landmark analysis.

Where needed, voucher specimens were collected and are accessioned at the Kathryn Kalmbach Herbarium (KHD: Denver Botanic Gardens).

Gaps in Continuous characters across Geography (GCG, Zapata and Jiménez (2012) was used to assess discontinuities or gaps in continuous morphological traits between geographic regions (i.e., Jackson County and Larimer County) and among populations within a region.

Geometric morphometrics (GMM) was used to examine variation in size and lobes of *Phacelia* leaves via a combination of Cartesian coordinates of 21 selected leaf landmarks and Procrustes analysis (Viscosi and Cardini 2011).

Population Genetics:
Total genomic DNA from all 420 silica-dried samples was extracted at Denver Botanic Gardens. Twelve microsatellite loci were amplified via PCR for all 14 sampled *Phacelia* sites (Riser et al. 2017). Genetic diversity analyses were conducted using GenAlEx v. 6.5 (Peakall and Smouse, 2006; 2012).

Analyses of Geographic Structuring:
Genetic data were examined for geographic structuring using four analyses: discriminant analysis of principal components (DAPC, implemented in R package ADEGENET 1.3-0, (Jombart et al. 2010), Bayesian clustering (STRUCTURE v.2.3.4, Pritchard *et al*., 2000; Falush *et al*. 2007), boundary detection (BARRIER 2.2, Manni *et al*., 2004), and Analysis of Molecular Variance (ARLEQUIN 3.5. (Schneider et al. 2000). Results were visualized in CLUMPAK (Kopelman et al. 2015).

Genetic sequence analysis:
Using standard PCR techniques and primers (ITS4, ITS5) we amplified the nrITS region for a subset of 25 individuals from Jackson (18), Larimer (5), and Grand County (2). *Phacelia* sequences from Gilbert (2005; S1341_gilbert_ITS2005_treebase.nex) were retrieved from GenBank and added to the sequences we generated to assess relationships.

Ecological Niche analysis:
Ecological differences among *Phacelia* populations were assessed through an ecological niche analysis. Bioclimatic variables (Fick and Hijmans, 2017) and soil characteristics were examined to determine if populations between the two counties fall within a different environmental niche.

RESULTS

Sample collection:
Leaf tissue samples were collected for DNA extraction and subsequent analyses from 420 individual plants at 14 populations of *Phacelia* growing in Jackson and Larimer counties in northern Colorado. Following microsatellite amplification, individuals with >30% missing data were removed resulting in 395 retained

samples (Table 1). Morphological data were collected from 120 individuals at four populations (FC, RN, BR, and AS). Sixty-five leaves were measured for the leaf morphometric analysis.

Table 1. Fourteen populations of *Phacelia* collected for morphological and genetic evaluation from two counties in Colorado, USA.

Population (Code)	$N_C/N_A/N_M/N_L$	**County**	**Voucher (Accession no.)**
Hohnholz NE (HN)	30/11/-/-	Larimer	*N. D. Atwood 33568* (BRY-V 0050691)
Hohnholz SW (HS)	30/30/-/-	Larimer	*R. Hartman 69019* (RM 772392)
Forrester Creek (FC)	30/24/30/9	Larimer	*M. Islam 1489* (KHD00062091)
California Gulch (CG)	30/30/-/-	Jackson	*M. Islam 12-524* (KHD00055193)
Diamond J NE (DJ)	30/30/-/-	Jackson	*N. D. Atwood 33622* (BRY-V 0050698)
Powerline W (PW)	30/30/-/-	Jackson	*M. Islam 12-519* (KHD00055087)
Research Natural Area (RN)	30/30/30/13	Jackson	*M. Islam 1491* (KHD00062089)
Weather Station (WS)	30/30/-/-	Jackson	*M. McGuire 5* (BRY-V 0050695)
Walden Airport (WA)	30/30/-/-	Jackson	*N. D. Atwood 1977* (BRY-V 0050703)
Brownlee (BR)	30/30/30/22	Jackson	*M. Islam 1486* (KHD00062093)
Verner (VR)	30/30/-/-	Jackson	*N. D. Atwood 33564* (BRY-V 0050701)
Rockwell (RW)	30/30/-/-	Jackson	*N. D. Atwood 33558* (BRY-V 0050700)
Arapaho NWR North (AN)	30/30/-/-	Jackson	*M. Islam 1487* (KHD00062092)
Arapaho NWR South (AS)	30/30/30/21	Jackson	*M. Islam 1487* (KHD00062092)

$N_C/N_A/N_M/N_L$= number of samples collected/number of samples analyzed following microsatellite amplification/number of samples analyzed for overall morphology/ number of samples analyzed for leaf landmark morphology

Morphology:
We measured four populations of *Phacelia*: three from Jackson County (AS, BR, RN) and one from Larimer County (FC). These samples did not clearly separate in multivariate morphological space. We had previously determined that a frequency cutoff of 0.2 captured the morphological gap between the Jackson and Larimer County populations and the putatively closely related *P. gina-glenneae*. Only one of the sampled populations (AS) showed enough significant morphological distinctiveness from the other three populations to infer a gap between them. However, this was one of the Jackson County populations and not the outlying Larimer County population.

Leaf Morphometrics:
The GMM analysis examining variation in leaf size and lobing using 21 leaf landmarks detected no significant difference between either the Jackson versus Larimer county populations nor within either county.

Population Genetics:
Twelve microsatellite loci were successfully amplified in 395 individuals across all 14 sampled populations of *Phacelia* from northern Colorado. All 14 *Phacelia*

populations maintain high levels of population genetic diversity. Mean percent polymorphic loci for all populations was 97%. All 12 microsatellite loci were polymorphic at 10 populations and >80% polymorphic at the other four populations. Across all populations, a total of 104 alleles were scored.
The mean number of different alleles per population was 3.6 and the mean number of effective alleles was 2. Overall mean observed heterozygosity was 0.341 (range: 0.177 to 0.475). The overall mean expected and mean unbiased expected heterozygosities were 0.405 (range: 0.306 to 0.525) and 0.413 (range: 0.311 to 0.534), respectively.

The within-Jackson County analysis showed that considerable genetic diversity exists within the Arapaho National Wildlife Refuge (ANWR) plus nearby Rockwell *Phacelia* populations (populations RW+AN+AS) as compared to the rest of the county. Mean numbers of alleles per population were similar between ANWR (4.8) and the rest of Jackson County (5.8) as were the mean number of effective alleles per population (2.7 and 2.3, respectively). ANWR populations had higher mean observed heterozygosities (0.446) than the rest of the Jackson County populations (0.294). Likewise, mean expected and mean unbiased expected heterozygosities were higher at the three ANWR populations (0.551 and 0.554, respectively) than for the rest of the Jackson County *Phacelia* populations (0476. and 0.477, respectively). All loci were polymorphic in both groupings.

The coefficient of inbreeding was calculated for all 14 individual *Phacelia* populations as well as comparisons between Jackson and Larimer County populations and Jackson County versus the Arapaho National Wildlife Refuge populations. In all cases inbreeding was indicated with fewer heterozygotes observed than expected. Values ranged from a low of 0.055 to a high of 0.468. Jackson County had a higher inbreeding coefficient (0.387) than Larimer County (0.250) in the comparison between populations in the two counties. The Arapaho National Wildlife Refuge populations had a lower inbreeding coefficient (0.191) compared to the remainder of the Jackson County populations (0.383).

Pairwise proportions of genetic variation (F_{ST}) among the 14 *Phacelia* populations (via AMOVA) ranged from a low of 0.026 to a high of 0.567. All values were significantly different from zero (P<0.001). Likewise, the lowest N_m was 0.191 and the highest was 9.301.

Twenty-five private alleles were found across all 14 *Phacelia* populations. However, the private alleles were not evenly distributed among the populations: six populations had no private alleles while Forrester Creek (FC)in Larimer County had 10. The number of private alleles between counties was similar (Jackson County = 29, Larimer = 24). Within the Jackson County populations, the three ANWR populations had twice as many (22) private alleles as compared to the rest of the sites combined (11).

Analyses of Geographic Structuring:

The discriminant analysis of principle components (DAPC) was run for three clustering scenarios testing 2, 3, and 4 clusters. For each scenario, 25 PCs and 10 discriminant functions were retained. In the two cluster scenario, individuals were assigned to the two clusters according to the discriminant functions at an average of 98.5% (group 1: 98.7%, group 2: 98.4%). In the three cluster scenario, assignment was at an average of 99.5% (group 1: 99%, group 2: 100%, group 3: 100%), and with the four cluster scenario at an average of 98.8% (group 1: 98.6%, group 2: 100%, group 3: 100%, group 4: 97.3%). High proportions of assignment indicate very low admixture among groups.

The Bayesian cluster assessment (STRUCTURE) supports 2 as the most likely number of clusters with additional moderate support for 4 clusters. The 2 clusters do not segregate by county, but rather the three populations in Larimer County cluster with the ANWR populations. In the four clusters scenario the ANWR and Larimer County populations fall out as distinct clusters and the populations within Jackson County segregate into two clusters.

The boundary detection (BARRIER) analysis indicated genetic boundaries between all the sampled *Phacelia* populations. Barriers between populations ranged from one (PW-RN and RN-WS) to 12 (e.g., AS-RW, CG-RN, and WA-PW) within Jackson County. Larimer County populations (HN, HS, and FC) were separated by 11 or 12 inferred barriers. Twelve inferred barriers separated the Jackson and Larimer county populations.

The analysis of molecular variance (AMOVA) assesses genetic variation in a hierarchical structure. When assessing all 14 populations as a single group, the among population variation is 33.57%. When assessing variation among counties, 26.17% of the variation is distributed among county with 19.63% distributed among populations within county. In testing the K=2 scenario from the DAPC and STRUCTURE results, 18.82% of the variation is distributed among the two groups (Larimer + ANWR vs remainder of Jackson) group, and 21.23% is among populations within group. When assessing the K=4 scenario as identified in the DAPC and STRUCTURE results, 23.22% of the variation is distributed among groups and 13.51 among populations within group. All assessments were significant at $p<0.05$.

Genetic sequence analysis:
All nrITS sequences were identical except for one base pair in one of the California Gulch (CG) individuals. All populations from the two counties form a polytomy and are not distinguishable based on sequence differences.

Climate envelope analysis:
Repeated randomization (Schoener's D and the I statistic) to create a null distribution of niche overlap between species distribution models comparing points from the two counties (niche equivalency test from R package dismo; Warren et al. 2008) show that the actual niche overlap between models of each county fall within

the distribution of randomized niche overlap values and thus do not significantly differ.

DISCUSSION

While initial assessments of morphology from Dwayne Atwood suggested that the populations in Larimer County were distinct and should be described as a distinct taxon, further assessment combining morphological, ecological and genetic data do not support two distinct taxa for the sampled populations across the two counties.

All analyses failed to support significant differences between the two counties assessed (Fig. 2). While there is some genetic structuring to the data there is as much variation among groups in Jackson County as is detected when comparing Jackson and Larimer Counties to each other. Specifically:

- Overall morphology and leaf morphometric analyses failed to find any differences between Jackson and Larimer county populations. One site in Jackson County (AS) was significantly different from the others.

- Population genetic assessments showed high levels of genetic diversity within populations. Additionally, there is just as much diversity between populations as there is between the two counties. Especially, the Arapaho National Wildlife Refuge populations including Rockwell, which were as different from the other Jackson county populations as they were from the Larimer County populations.

- The DAPC analysis failed to show a Jackson versus Larimer county grouping. Some populations within Jackson County grouped with Larimer County populations in the 2-cluster analysis while the 3- and 4-cluters analyses showed no county-based patterns.

- Bayesian clustering analysis supported groupings of 2 or 4 clusters, favoring the 2-cluster scheme over the 4-cluster one. Again, no county-based pattern was found.

- Inferred barriers to gene flow were just as strong between populations within each county as between the two counties.

- All hierarchical comparisons are statistically significant when assessing strength of population structure through analysis of variance. While there is significant genetic structure among counties, there is as much variation among groups when parsed either by individual populations or by the groupings resulting from the cluster analyses. The within population variation is high meaning that each population houses high levels of unique genetic variation.

- Genetic sequence data did not distinguish any groupings, instead placing all sampled *Phacelia* populations into an unresolved polytomy.

- No ecological differences were detected between the habitats with the ecological niche overlap analysis.

Other than a geographic discontinuity between these *Phacelia* populations in Jackson and Larimer counties, there is no evidence to support the recognition of Larimer County populations as distinct in any way from *Phacelia formosula*. However, the considerable genetic variation maintained within the individual populations, particularly at FC, RW, AN, and AS, argues for the importance of protecting all *Phacelia formosula* populations.

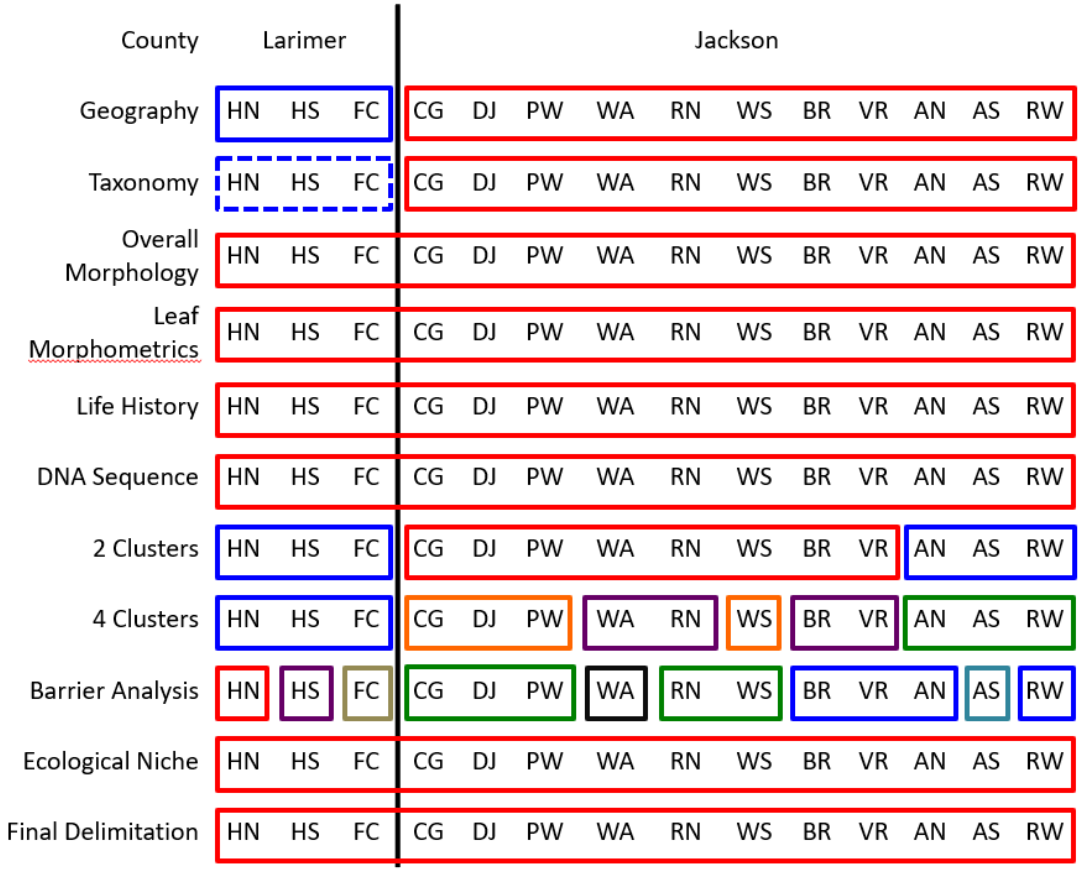

Figure 2. Schematic representation of multiple groupings and analyses of taxonomic distinctness in *Phacelia* populations from two counties in northern Colorado, USA. The 2 Cluster and 4 Cluster groupings represent STRUCTURE K-values.

REFERENCES

Atwood, N. D. and S. L. Welsh. 2013. New plant taxa from Colorado, New Mexico, and Arizona. Western North American Naturalist 73(1): 113–115

Colorado Natural Heritage Program. 2014. Biodiversity Tracking and Conservation System. Colorado State University, Ft. Collins, CO. Data exported April 2014.

Falush D, M. Stephens, J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7:574–578.

Fick S.E. and R.J. Hijmans. 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology.

Gilbert C., J. Dempcy, C. Ganong, R. Patterson, G.S. Spicer. 2005. Phylogenetic relationships within *Phacelia* subgenus *Phacelia* (Hydrophyllaceae) inferred from nuclear rDNA ITS sequence data. Systematic Botany 30(3): 627-634.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11**.**

Kopelman, N.M. J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15(5): 1179-1191.

Manni F., E. Guérard, E. Heyer. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by “Monmonier’s algorithm.” Human Biology 76:173–190.

Osterhout, G. E. 1919. Additions to the flora of Colorado. Bulletin of the Torrey Botanical Club Vol. 46(2): 53-56.

Peakall R., and P.E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295.

Peakall R. and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28:2537-2539.

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

Riser J.P., A.L. Schwabe, J.R. Neale. 2017. Novel microsatellite development and characterization for *Phacelia formosula* (Hydrophyllaceae). Applications in Plant Science 5(7): 1700030.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software program for population genetic data analysis. Genetics and Biometry Laboratory, Geneva.

U.S. Fish and Wildlife Service. 2011. 5-Year review of *Phacelia formosula* (North Park phacelia), December 2011. U.S. Fish and Wildlife Service, Western Colorado Field Office, Grand Junction, Colorado, USA.

U.S. Fish and Wildlife Service. 1986. North Park Phacelia (*Phacelia formosula*) Recovery Plan. U.S. Fish and Wildlife Service, Denver, CO. 28 pp.

Viscosi, V., and A. Cardini. 2011. Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners. Plos One 6:20.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62:2868-2883.

Zapata, F., and I. Jimenez. 2012. Species Delimitation: Inferring Gaps in Morphology across Geography. Systematic Biology 61:179-194.