**NOVEL MICROSATellite DEVELOPMENT AND CHARACTERIZATION FOR Phacelia formosula (Hydrophyllaceae)**

**James P. Riser II**, **Anna L. Schwabe**, **and Jennifer Ramp Neale**

*Denver Botanic Gardens, 909 York Street, Denver, Colorado 80206 USA*

- *Premise of the study:* Microsatellite primers were developed to characterize genetic diversity and structuring in the genus *Phacelia* (Hydrophyllaceae) and to further conservation efforts for *P. formosula*.
- *Methods and Results:* Fifteen novel microsatellite primers were developed for *P. formosula*. These were characterized for genetic variation in three separate *P. formosula* populations. Two to nine alleles were found per locus. Overall observed heterozygosity and expected heterozygosity ranged from 0.000 to 0.800 and 0.000 to 0.840, respectively. Additionally, these loci were successfully amplified and showed polymorphism in *P. gina-glenneae* and a potential new *Phacelia* species.
- *Conclusions:* These microsatellite markers will be useful in assessing genetic diversity, structuring, and gene flow within and among populations of the rare *P. formosula*, in addition to related *Phacelia* species. These markers will provide important genetic data needed for appropriate conservation and management of these rare plants.

**Key words:** Colorado; conservation genetics; Hydrophyllaceae; microsatellite; Phacelia; Phacelia formosula.

*Phacelia* Juss. (Hydrophyllaceae) is a speciose genus with approximately 167 (USDA NRCS, 2017) species in the United States, predominantly in western states. We follow the taxonomy of the Boraginales Working Group (Luebert et al., 2016) in conserving *Phacelia* within the family Hydrophyllaceae as opposed to in a subfamily of the Boraginaeae (e.g., APG IV, 2016). In addition to the ongoing debate regarding the status of the Hydrophyllaceae, this maintains agreement with the recent *Flora of Colorado* (Ackerfield, 2015). Given the number of species in both the genus and family, we expect that these markers will have broad applicability for conservation and population-level studies. Additionally, there are many rare and locally endemic species in *Phacelia* (34 species with a G1 or G2 rank; NatureServe, 2017). Whereas previous population genetic studies in *Phacelia* used cpDNA (Levy et al., 1996) or allozymes (Levy and Neal, 1999), we developed the first primers specifically for population-level assessments in the genus.

*Phacelia formosula* Osterh. (North Park phacelia) is a rare endemic found only in the North Park basin in Jackson County, Colorado, USA. Within this area, *P. formosula* is found in scattered small populations restricted to soils derived from the Coalmont Formation (U.S. Fish and Wildlife Service, 2011). An understanding of the genetic diversity and distribution of *P. formosula* would be extremely useful in guiding management and conservation actions. Currently, these data are lacking for *P. formosula*, as well as other *Phacelia* species.

Here we report the development and characterization of 15 novel microsatellite loci for *Phacelia*, all of which were tested for polymorphism in *P. formosula*. Additionally, we cross-amplified these loci in a presumably closely related species, *P. gina-glenneae* N. D. Atwood & S. L. Welsh, and in a recently discovered population of uncertain specific status (*Phacelia* sp. in Table 1).

**METHODS AND RESULTS**

Microsatellite development using DNA extracted from silica-dried *P. formosula* leaf tissue was conducted by Ecogenics GmbH (Balgach, St. Gallen, Switzerland). Microsatellite content of the genomic DNA fragments was enriched via biotin-labeled tetranucleotide (GTAT, GATA, AAAC, and AAAG; Roche 454 platform [Basel, Basel-Stadt, Switzerland] with GS FLX Titanium reagents) and dinucleotide (CT and GT; Illumina MiSeq platform [San Diego, California, USA] using the Nano 2 × 250 version 2 format) repeats using magnetic streptavidin beads. The enrichments were multiplexed with additional species and produced libraries with 4264 and 13,858 reads (respectively), which were assessed for microsatellites using Primer3 (Rozen and Skaletsky, 1999). The tetranucleotide reads averaged 415 bp in length with 151 reads containing a tetra- or trinucleotide microsatellite insert ≥ 6 repeat units. The dinucleotide reads averaged 402 bp in length, and 1502 reads contained a dinucleotide microsatellite insert ≥10 repeat units. Suitable primer design was possible in 83 of the tetranucleotide reads and 918 of the dinucleotide reads, of which 24 and 16 reads (respectively) were tested for functionality and polymorphism in seven samples using the methods of Schuelke (2000). The assessment resulted in 40 loci, which were then narrowed to 15 by the authors based on multiplex potential. These loci were multiplexed in two panels for data collection and analyses (Table 1).

For locus amplification within populations, total genomic DNA was extracted from silica-dried leaf tissue at Denver Botanic Gardens using the Omega E.Z.N.A. DNA Mini Kit (Omega Bio-tek, Norcross, Georgia, USA; short protocol with both elution steps). Amplification was carried out at the Nevada Genomics Center (Reno, Nevada, USA) using two PCR panels with different...
### Table 1. Characteristics of 15 microsatellite loci developed in *Phacelia formosula* from Jackson County, Colorado, USA, including annealing temperature for two separate PCR panels and individual primer concentrations.

| Locus       | Primer sequences (5′–3′) | Repeat motif | $T_a$ (°C) | Primer concentration (μM) | Fluorescent label | Allele size range (bp) | GenBank accession no. |
|-------------|--------------------------|--------------|------------|---------------------------|------------------|------------------------|-----------------------|
| Phafor_00006| F: GAGTTTCGGAGAACAGTGC   | (TATG)$_b$   | 63.8       | 0.75                      | 6-FAM           | 189–197                | KP281305              |
|             | R: CTGGTGATCCGTCATCAAAGC |              |            |                           |                  |                        |                       |
| Phafor_00246| F: CCGCCTACTCTTTCTGACAG | (TACA)$_{10}$ | 63.8       | 2                         | NED             | 203–222                | KP281306              |
|             | R: AGCAATTGGCATTAGGCAGGC |              |            |                           |                  |                        |                       |
| Phafor_00567| F: AAAACGACATCCATCTTTTG  | (ATAC)$_{11}$ | 62         | 2                         | VIC             | 222–260                | KY442304              |
|             | R: TCCGCTAATGACGCATCGTG  |              |            |                           |                  |                        |                       |
| Phafor_00650| F: ATACACAGACAAGCCGAACTTC | (TGTA)$_b$   | 63.8       | 2                         | PET             | 194–222                | KP281307              |
|             | R: TTACCGCTTCAACACCAAC   |              |            |                           |                  |                        |                       |
| Phafor_00745| F: AGGGCTGACACATGACTCTTC | (TTCT)$_{8}$ | 63.8       | 2                         | NED             | 253–271                | KP281308              |
|             | R: GCACCCCGTGTGGTACG     |              |            |                           |                  |                        |                       |
| Phafor_01477| F: GCAAACATGAACTATACGCC | (CA)$_{17}$  | 62         | 2                         | 6-FAM           | 83–96                  | KP281309              |
|             | R: GTGATTCAATGCACTACATGG |              |            |                           |                  |                        |                       |
| Phafor_01499c| F: TGCAAAAGAGAATTCATCAAC | (GA)$_{14}$  | 62         | 2                         | PET             | 228–240                | KP281303              |
|             | R: AGGCCCTATCCTGTCCACATC |              |            |                           |                  |                        |                       |
| Phafor_01817c| F: ATCCGGTGACAGATGCCG   | (AG)$_{14}$  | 63.8       | 1.5                       | VIC             | 110–133                | KY442307              |
|             | R: GCCCCCTGGGAGAACAAAAG  |              |            |                           |                  |                        |                       |
| Phafor_02245| F: TCCAAGTCTAGCGAGTCCG  | (ATC)$_b$    | 62         | 4                         | 6-FAM           | 208–223                | KP281310              |
|             | R: TCCGATTACTGAGCTACTCAAG|              |            |                           |                  |                        |                       |
| Phafor_02638| F: GATGCGCGATCTGCGTGG   | (TATG)$_{12}$ | 62         | 4                         | NED             | 220–249                | KP281312              |
|             | R: CAATAGAAGGACAGTCCACC |              |            |                           |                  |                        |                       |
| Phafor_02824| F: CCCTCCTGTAGCTTCGATG  | (TATC)$_{18}$ | 63.8       | 4                         | 6-FAM           | 262–288                | KY442308              |
|             | R: CAAACTGCTGAAATGCACC  |              |            |                           |                  |                        |                       |
| Phafor_03037| F: AACATCTCGATCCGACCG   | (TATC)$_{17}$ | 63.8       | 1.5                       | VIC             | 231–239                | KP281311              |
|             | R: TGGCTGACAGTAGTGGTA   |              |            |                           |                  |                        |                       |
| Phafor_03754| F: GGGTTAGTACGATGAGGATCG | (TATG)$_{b}$ | 62         | 2                         | NED             | 165–189                | KP281313              |
|             | R: ATTGACTTCCATCATGGAGT  |              |            |                           |                  |                        |                       |
| Phafor_05461c| F: TTTGCTGACAGCAGCAAGAG | (GT)$_{12}$  | 62         | 2                         | 6-FAM           | 155–167                | KP281302              |
|             | R: AAGAGCTGAGGAGAAGAG   |              |            |                           |                  |                        |                       |
| Phafor_13597s| F: TGGTTGTCAGAGATGCAGGCG | (CT)$_{14}$  | 62         | 1                         | VIC             | 106–140                | KP281304              |
|             | R: ATTGGAAGTGGACTAGAACGA |              |            |                           |                  |                        |                       |

*Note: $T_a$ = annealing temperature.*
annealing temperatures to maximize multiplexing effectiveness. Final PCR volume was 11 μL: 1 μL aliquot of panel mix (containing: forward primers [labeled with a universal M13 tail: 5’-TGTAAACGACGCTACGATCA-3’], reverse primers [primer concentration varied by loci, see Table 1], a fluorescently labeled [6-FAM, NED, PET, or VIC] 5’ tag, and 20 ng of DNA template) and 10 μL of QIAGEN Multiplex PCR Mastermix (QIAGEN, Hilden, Germany). Reaction conditions were as follows: an initial 15-min 95°C denaturing step; followed by 40 amplification cycles of 95°C for 30 s, 62°C (panel 1) or 63.8°C (panel 2) for 45 s, and 72°C for 45 s; followed by a final elongation step at 72°C for 30 min using a GeneAmp 9700 thermocycler (Applied Biosystems, Carlsbad, California, USA). PCR products were diluted to an appropriate concentration determined by PicoGreen dilution tests. One microliter of diluted PCR product was added to 10 μL of HiDi Formamide with the size standard GeneScan 500 LIZ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and 7 μL of molecular-grade water followed by electrophoresis on an ABI Prism 3730 DNA Analyzer (Applied Biosystems) at the Nevada Genomics Center. Genotype data were visualized and fragment sizes scored at Denver Botanic Gardens using Geneious version 6.0.6 (Kearse et al., 2012).

GenAIEx version 6.3 (Peakall and Smouse, 2006) was used to calculate observed heterozygosity (H_o) and expected heterozygosity (H_e) and to test for deviation from Hardy–Weinberg equilibrium (HWE). GENEPOP (Raymond and Rousset, 1995; Rousset, 2008) was used to test for linkage disequilibrium between any two loci. ML-Multilocus homozygous and heterozygous states were allowed to pass using the default settings.

All microsatellite loci were variable and polymorphic in all three of the P. formosula populations. The number of alleles per locus ranged from two to nine (JC1, N = 30), from two to seven (JC2, N = 30), and from two to eight (JC3, N = 30). At JC1, 30 of 35 loci ranged from 0.133 to 0.800 and 0.124 to 0.840, respectively. At JC2, 28 of 35 loci ranged from 0.033 to 0.700 and 0.064 to 0.707, respectively. At JC3, 24 of 35 loci ranged from 0.000 to 0.700 and 0.064 to 0.707, respectively. At JC3, the H_e ranged from 0.000 to 0.700 and the H_o ranged from 0.067 to 0.742 (Table 2). Three loci (Phafor_00567, Phafor_02245, and Phafor_02824) at the JC1 population and four loci at both the JC2 population (Phafor_00567, Phafor_01499c, and Phafor_02245) and the JC3 population (Phafor_02245, and Phafor_03754) and the JC3 population (Phafor_00246, Phafor_00567, Phafor_01817c, and Phafor_02245) showed significant deviation from Hardy–Weinberg equilibrium for HWE (Table 2). After Bonferroni correction, no evidence of significant linkage disequilibrium was detected. Heterozygote deficiencies, possibly indicating the presence of null alleles, were detected for three loci (Phafor_00567, Phafor_02245, and Phafor_03754) and the JC3 population (Phafor_00246, and Phafor_01817c and Phafor_02245) at the JC1 population and four loci at both the JC2 population (Phafor_00567, Phafor_01477, Phafor_01817c, Phafor_02245, and Phafor_13597s), and seven loci at the JC3 population (Phafor_00246, Phafor_00567, Phafor_00745, Phafor_02245, Phafor_02824, and Phafor_05461c). All 15 microsatellite loci were successfully cross-amplified in both P. gina-glenaeae and a Phacelia population of uncertain specific status.

CONCLUSIONS

The novel microsatellite markers described here are the first developed not only for Phacelia, but also for the Hydrophyllaceae. These markers will be valuable for investigating population genetic structure in Phacelia and potentially other genera within Hydrophyllaceae. Knowledge of genetic diversity present within and among the scattered populations of the rare P. formosula will be used to better manage the known populations to ensure their future persistence. Additionally, these markers will be useful for assessing genetic diversity in a newly discovered population of Phacelia that is morphologically similar to P. formosula but occurs in different habitats more than 40 km away. Investigating potential gene flow between this new population and existing P. formosula populations will be helpful in inferring its specific status. Cross-amplification in P. gina-glenaeae demonstrates the utility of these markers in assessing genetic diversity in other species of Phacelia. Our results indicate the presence of potential null alleles. Several methods (Chapuis and Estoup, 2007) and programs can be used to detect and account for null alleles in population-level analyses (such as Kalinowski and

http://www.bioone.org/loi/apps

3 of 4
Taper, 2006; van Oosterhout et al., 2006), and we encourage their use with these markers. These microsatellite markers constitute a valuable tool for fine-scale genetic investigations in the genus Phacelia, as well as conservation of rare Phacelia species.

LITERATURE CITED

ACKERFIELD, J. 2015. Flora of Colorado. Botanical Research Institute of Texas, Fort Worth, Texas, USA.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society 181: 1–20.

CHAPUIS, M. P., AND A. ESToup. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24: 621–631.

KALNOWSKI, S., AND M. TAPER. 2006. Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conservation Genetics 7: 991–995.

KEARSE, M., R. MOBB, A. WILSON, S. STONES-HAVAS, M. CHEUNG, S. STURROCK, S. BUXTON, ET AL. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (Oxford, England) 28: 1647–1649.

LEVY, F., AND C. L. NEAL. 1999. Spatial and temporal genetic structure in chloroplast and allozyme markers in Phacelia dubia implicate genetic drift. Heredity 82: 422–431.

LEVY, F., J. ANTONOVICS, J. E. BOYNTON, AND N. W. GILLHAM. 1996. A population genetic analysis of chloroplast DNA in Phacelia. Heredity 76: 143–155.

LUEBERT, F., L. CECCHI, M. W. FROBEICH, M. GOTTSCHLING, C. M. GUILLIAMS, K. E. HASENSTAB-LEHMAN, H. H. HILGER, ET AL. 2016. Familial classification of the Boraginales. Taxon 65: 502–522.

NATUREServe. 2017. NatureServe Explorer: An online encyclopedia of life [web application]. Version 7.0. NatureServe, Arlington, Virginia, USA. Website http://explorer.natureserve.org [accessed 15 February 2017].

PEAKALL, R., AND P. E. SMouse. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288–295.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248–249.

ROUSSET, F. 2008. GENEPOP’007: A complete reimplementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8: 103–106.

ROZEN, S., AND H. SKALETSKY. 1999. Primer3 on the WWW for general users and for biologist programmers. In S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments: A poor man’s approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology 18: 233–234.

USDA, NRCS. 2017. The PLANTS Database. National Plant Data Team, Greensboro, North Carolina, USA. Website http://plants.usda.gov [accessed 17 February 2017].

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.

VAN OOSTERHOUT, D., D. WEETMAN, AND W. F. HUTCHINSON. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Molecular Ecology Notes 6: 255–256.

APPENDIX 1. Voucher and general location information for five Phacelia populations used in this study.

| Species                  | Population | Locality       | N  | Voucher (Accession no.) |
|--------------------------|------------|----------------|----|-------------------------|
| Phacelia formosula Osterh. | JC1        | Jackson County, Colorado, USA | 30 | M. Islam 1487 (KHD00062092) |
| Phacelia sp.              | JC2        | Jackson County, Colorado, USA | 30 | N. D. Atwood 33622 (BRY-V 0050698) |
|                          | JC3        | Jackson County, Colorado, USA | 30 | N. D. Atwood 33558 (BRY-V 0050700) |
| Phacelia gina-glenneae N. D. Atwood & S. L. Welsh | LC        | Larimer County, Colorado, USA | 30 | M. Islam 1489 (KHD00062091) |
|                          | GC         | Grand County, Colorado, USA | 30 | M. Islam 12-271 (KHD00051791) |

N = number of individuals sampled.

*Detailed location information has been omitted due to the protected status of these species.

One voucher was collected from each sampled population. Vouchers were deposited at the Kathryn Kalmbach Herbarium (KHD), Denver Botanic Gardens, Denver, Colorado, USA, or the S. L. Welsh Herbarium (BRY), Brigham Young University, Provo, Utah, USA.

http://www.bioone.org/loi/apps