Shortia brevistyla comb. et stat. nov., (Diaspensiaceae), A narrow endemic from the headwaters of the Catawba River in North Carolina, U. S. A.

L. L. Gaddy[1], T. H. Carter[2], B. Ely[3], S. Sakaguchi[4], A. Matsuo[5] and Y. Suyama[5]

[1]President, terra incognita and Adjunct Faculty,
Department of Biology, University of South Carolina, Columbia, SC, USA
[2]Ph. D. Candidate, Genetics Laboratory, Department of Biology, University of South Carolina
[3]Director, Genetics Laboratory, and Professor of Biology, University of South Carolina
[4]Graduate School of Human and Environmental Studies, Kyoto University, Japan
[5]Graduate School of Agricultural Sciences, Tohoku University, Miyagi, Japan

Correspondence to llgaddy2@gmail.com.

ABSTRACT

Morphological, geographic, and molecular data justify recognition of *Shortia brevistyla* (*Shortia galacifolia* Torr. & A. Gray var. *brevistyla* P.A. Davies) as a distinct species rather than a variety of *Shortia galacifolia*. All known populations of the new *Shortia brevistyla* are found within a 10 km radius on the headwaters of the Catawba River in McDowell County, North Carolina, approximately 100 km northeast of the range of *S. galacifolia. Shortia brevistyla* has significantly smaller flowers, shorter styles, shorter petals, and smaller leaves than *S. galacifolia*. Genetic data indicate that the two species differ at five of 210 nucleotide positions in the ITS1 ribosomal region. Genetic divergence models indicate that the two species diverged approximately 20,000 years ago during the glacial maximum of the Pleistocene. Published on-line www.phytologia.org *Phytologia 101(2): 113-119 (June 21, 2019).* ISSN 030319430.

KEY WORDS: *Shortia brevistyla* (P. A. Davis) Gaddy, *comb. et stat. nov*., Diaspensiaceae, narrow endemic, Catawba River, North Carolina.

Shortia galacifolia Torr. & A. Gray, the only species of *Shortia* heretofore known from North America (Nesom, 2009), was described by Asa Gray from a specimen in Paris collected in 1839 by Andre Michaux from the "Hautes Montagnes de Carolinie" (Williams et al. 2004). Michaux had scribbled on the herbarium sheet that the plant may be a *Pyrola* or a new genus. After extensive searches for the plant in the Carolinas, it was found to occur at low elevations (ca. 300-600 meters) along the Blue Ridge front, not in "high" mountains. Two major locations of the plant have been found in the Carolinas, one in the Keowee-Toxaway drainage of the headwaters of the Savannah River in North and South Carolina, and another on the headwaters of the Catawba River in North Carolina. Interestingly, between the areas where the plant has been found, there is a gap of approximately 100 km where no *Shortia* plants have ever been seen. Davies (1952) named the Catawba drainage plants in North Carolina "var. *brevistyla*," implying that the Keowee-Toxaway plants were "var. *galacifolia*" and concluding (as later did Zahner and Jones 1983) that the type material for *Shortia galacifolia* (in Paris) was collected from plants in the Keowee -Toxaway drainage in South Carolina (Davies 1956). Morphological, geographic, and molecular data indicate that the Catawba shortia - heretofore, *Shortia galacifolia* (Torrey & Gray) var. *brevistyla* P.A. Davies - is justifiably treated at the specific rank. It is morphologically distinct and geographically separated from the Keowee-Toxaway populations of typical *S. galacifolia*, and the two taxa have diverged significantly in nucleotide (ITS1) sequences.

METHODS

Single leaves were removed from individual *Shortia* plants sampled from both Catawba River and Keowee-Toxaway populations. Individual leaves were frozen with liquid nitrogen then ground using SPEX Sample Prep 2010 Grinder (Metuchen, NJ). DNA extractions of the leaves were performed with the Promega Wizard DNA Purification Kit (Promega Madison, WI) according to the manufacturer's instructions. Primers complementary to highly conserved regions of the ribosomal ITS1 region (Cheung 2016) were ordered from Eurofins Genomics (Louisville, KY). The corresponding region of the *Shortia* genome was amplified using PCR with the following conditions after an initial denaturation step of 4 minutes: denaturation at 94°C for 30 seconds, annealing at 55°C for 40 seconds, and elongation at 72°C for 60 seconds. After a final elongation step of 10 minutes at 72°C, a sample of the resulting PCR product was subjected to agarose gel electrophoresis to verify amplification and then sent to Eurofins Genomics for Sanger sequencing. The resulting nucleotide sequences were analyzed utilizing UGENE software (Okonechnikov 2012) and NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Total DNA was extracted from dried leaf samples (ca. 0.5 cm^2) using CTAB (cetyltrimethylammonium bromide) method (Murray and Thompson, 1980) in 1.2mL tubes, after washing the leaf powder for two times with 0.8 mL of a buffer, containing 10% polyethylene glycol, 0.35M solbitol, 50mM Tris-HCl, 0.1% Bovine Serum Albumin and 0.1% β-melcaptoethanol. Genetic polymorphisms were detected with MIG-seq [multiplexed ISSR (inter-simple sequence repeats) genotyping by sequencing] technique, a PCR (polymerase chain reaction)-based procedure for constructing highly reduced representation libraries, involving *de novo* SNP (single nucleotide polymorphism) discovery, and their genotyping using high-throughput sequencing (Suyama and Matsuki, 2015). Briefly, 10 ng of DNA was used to amplify anonymous genomic regions using multiplexed PCR with tailed ISSR primers as the 1st PCR.

Eight primers of the most recommended set of MIG-seq primers (set-1) (Suyama and Matsuki 2015) were employed for the 1st PCR. PCR thermal profile was the same with the original protocol, except that the annealing temperature set to 38ºC for satisfactory amplifications across samples. Subsequently, the 1st PCR products were used as templates for the 2nd PCR (tailed PCR). Using common forward and indexed reverse primers, this step permits the addition of complementary sequences for the binding sites of Illumina sequencing flow cell and index for each sample to the 1st PCR products. Then, the 42 PCR products were purified, fragments in the size range of 350–800 bp were isolated, and their final concentrations were measured by quantitative PCR. Sequencing of the multiplexed library was performed with an Illumina MiSeq Sequencer, using MiSeq Reagent Kit v3 (150 cycle, Illumina).

After read trimming using Trimmomatic v. 0.32 (with the following commands LEADING:19, TRAILING:19, SLIDINGWINDOW:30:20, AVGQUAL:20 and MINLEN:51) (Bolger *et al.*, 2014), Stacks 1.08 (Catchen *et al.*, 2011) was used to process the MIG-seq reads with the following parameter settings; the minimum number of identical reads required to create a stack ($m = 3$), the nucleotide mismatches between loci within a single individual ($M = 2$), and the mismatches between loci when building the catalogue ($n = 1$). The SNP genotype for each individual was exported with a minimum read depth of 10, using the 'populations' command. The genotype data was then processed using PLINK v. 1.07 software (Purcell *et al.*, 2007), and markers with a minor allele frequency < 0.03, missing individual rate > 0.9, missing locus rate > 0.9, and significant deviation from Hardy-Weinberg equilibrium ($P < 0.01$) were filtered out.

Population genetic statistics (the number of polymorphic loci, number of alleles, observed and expected heterozygosity, and inbreeding coefficient) were summarized for each population using GenAlEx 6.5 (Peakall and Smouse, 2012). The significance of Hardy–Weinberg equilibrium (heterozygote deficiency) were examined by exact test implemented in Genepop 4.2 (Rousett, 2008). Population

differentiation was assessed by a summary statistic of F_{ST} and AMOVA (analysis of molecular variance) using GenAlEx 6.5, and visualized by PCA (principal coordinate analysis) using '*adegenet*' package (Jombart, 2008) in R 3.1.0 (R Development Core Team 2014).

Past population demography was modeled by applying a pure isolation model (Hey and Nielsen 2007) to the MIG-seq sequence data. The population model assumed an ancestral population split into two descended populations (i.e. extant var. *galacifolia* and var. *brevistyla*) at a past time (*t*). These populations were allowed to have different parameters of effective population size (θ_{anc}, θ_{glc}, and θ_{brv}). In Bayesian parameter estimation using IMa2 program (Hey and Nielsen, 2007), the maximum priors for each parameter was set as follow: t = 1, $\theta_{anc} = \theta_{glc} = \theta_{brv} = 3$, and under this prior settings, 10 Metropolis-coupled MCMCs with a heating scheme of geometric model (with parameter specifications of -ha 0.99 and -hb 0.75) were generated for 3 million steps after a burn-in period of 1 million steps. Infinite-site model was used to model the sequence evolution of MIG-seq data. To confirm the adequate chain mixing and parameter convergence, independent two runs were performed. After obtaining the posterior parameters, demographic parameters were scaled to absolute divergence time in years and effective population size, by applying the geometric mean of 7.0×10^{-9} (substitutions/site/year) estimated for *Arabidopsis* species (Ossowski et al. 2010) and a generation time of 10 years for *Shortia*.

DESCRIPTION OF SPECIES

SHORTIA BREVISTYLA (P. A. Davies) Gaddy, *comb. et stat. nov*. (Figure 1).

Shortia galacifolia var. *brevistyla* P.A. Davies, Basionym. *Rhodora* 54: 124. 1952.

TYPE: U. S. A. North Carolina, McDowell Co., John's Creek (Fish Hatchery Creek) above the Fish Hatchery, on moist creekbank under *Rhododendron maximum*, 23 March 1951, *P. A. Davies 9149* (holotype: GH!; topotype: USCH!).

Differs from *Shortia galacifolia* Torrey & Gray in that its styles are shorter, its petals are shorter and shallowly toothed, and its leaves are smaller.

Rhizomes: slender, scale-leaved, lignescent. **Stems:** erect, unbranched. **Leaves:** basal, rosulate from rhizome buds, 20-55 mm in length; petiole present; blade orbiculate to elliptic-orbiculate, ovate-oblong, or ovate, margins coarsely crenate-serrate, apex emarginate to truncate, surfaces glabrous, pinnately veined. **Scapes**: bracteate, elongating after flowering. **Inflorescences:** solitary. **Flowers:** sepals distinct; petals 12-17 mm long, corolla rotate to campanulate, lobes white to pink to faded rose-purple, margins slightly toothed; anthers 2-locular, without basal spurs, longitudinally dehiscent; filaments adnate to corolla tube; staminodes present. Styles 6-12 mm in length.

Davies (1952) used style length as the primary character in his description of *Shortia galacifolia* var. *brevistyla.* Although the styles *are* significantly shorter than those of *galacifolia*, styles in *S. brevistyla* populations are often variable in length. And although petals and leaves are significantly shorter in *S. brevistyla,* the best field character, however, is the shallow, blunt teeth on the petals of *S. brevistyla* versus the deeply laciniate teeth on the petals of *S. galacifolia* (Figure 1; Table 1).

Figure 1. Typical flowers of *brevistyla* (left) (2.5x) and *galacifolia* (2x).

Table 1. Morphology of *Shortia galacifolia and Shortia brevistyla.*[1]

SPECIES	FLOWER SHAPE	LEAF LENGTH (MEAN)	PETAL LENGTH	PETAL DISTAL MARGIN	STYLE LENGTH
galacifolia	campanulate	37-82 mm (54)	16-25 mm	deeply laciniate	12-18 mm
brevistyla	rotate to campanulate	20-55 mm (41)	12-15 mm	shallowly toothed	6-12 mm

[1]Data from Davies (1952), Hatley (1977), Nesom (2009), and L. L. Gaddy (field observations: 2018-19).

RESULTS AND DISCUSSION

Two independent analyses of the chromosomal ribosomal ITS1 region were performed using DNA isolated from *Shortia* plants sampled from two different locations within each of the two geographic regions in North Carolina and South Carolina—the Catawba shortia region and the Keowee-Toxaway region—where North American *Shortia* is found. In both sample sets, we obtained 210 base pairs of nucleotide sequence for each sample that was 100% identical for all the North Carolina specimens as well as the *Shortia* ITS1 sequence present in the GenBank database (Table 2). Similarly, the nucleotide sequences of the South Carolina samples were identical, but differed from the North Carolina samples at five nucleotide positions. Therefore, we concluded that the Catawba River drainage North Carolina *Shortia* is genetically distinct from the Keowee-Toxaway (Savannah River) drainage South Carolina *Shortia* and that they have been reproductively isolated from one another long enough for ITS1 regions to diverge at 2% of the nucleotide positions.

Data from the laboratories of Dr. Shota Sakaguchi revealed that after adapter and quality trimming, an average of 191.4k sequence reads were obtained for the individual samples. The reads were then assembled and filtered by Stacks 1.08 and PLINK v. 1.07, which resulted in a genotype matrix consisted of 97 SNP markers. The genotyping rate at these markers was high: 97.1% for *galacifolia* and 99.0% for *brevistyla*. Significantly lower genetic diversity in *brevistyla* is revealed in multiple statistics (Table 2). Among the 92 markers, only 19.6% showed allelic variation in *brevistyla,* while 93.4% were polymorphic in *galacifolia*. The number of alleles (Na = 1.92 in *galacifolia* vs. 1.20 in *brevistyla*) and the heterozygosity index (e.g., H_E = 0.266 in *galacifolia* vs 0.072 in *brevistyla*) were consistently low in *brevistyla.* The inbreeding coefficient values were slightly negative in both species (they are both narrow endemics) but were not significantly deviated from Hardy-Weinberg equilibrium ($P > 0.05$ for both populations) (Figure 2; Table 3).

Table 2. Divergent nucleotide positions in comparison of ITS1 regions in Keowee-Toxaway drainage (SC) and Catawba River drainage (NC) *Shortia* populations.

SC	NC	Nucleotide Number[1]
A	C	73
T	C	84
T	C	100
G	A	158
C	A	159

[1]Sequence number from AY049800, labeled "*S. galacifolia*" ITS1 sequence in Genbank (as it turns out, the Genbank sequence is from *S. brevistyla*).

Table 3. Summary genetic diversity statistics for *galacifolia* and *brevistyla.*[1]

Species		N	P	N_a	H_O	H_E	F_{IS}
galacifolia	average	21.36	0.934	1.92	0.270	0.266	-0.005
(n=22)	SE	0.09	--	0.03	0.018	0.015	0.026
brevistyla	average	19.78	0.196	1.20	0.095	0.072	-0.256
(n=20)	SE	0.04	--	0.04	0.024	0.017	0.034

[1]N: number of genotyped individuals, P: proportion of polymorphic loci, N_a: number of alleles, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{IS}: inbreeding coefficient.

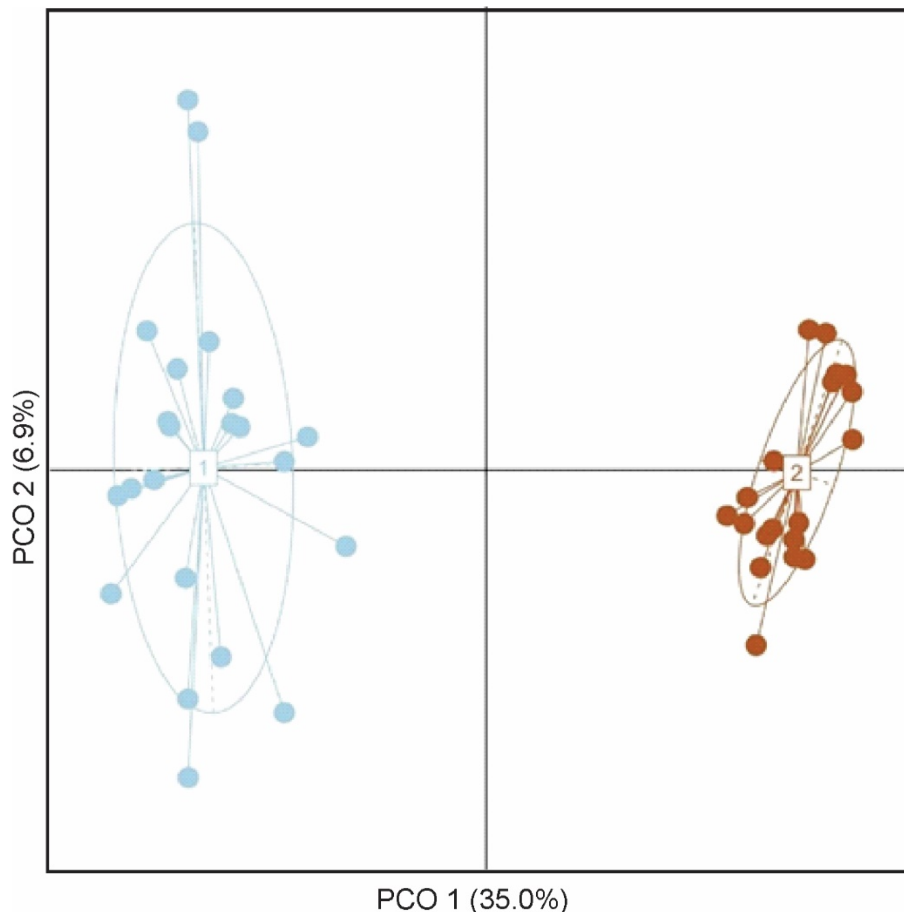

Figure 2. Plot of principal coordinate analysis scores of individual samples: 1-*galacifolia*; 2-*brevistyla.* Numbers on the coordinate axes are the percent variance accounted for by that axis.

The independent two runs using IMa2 program gave highly consistent posterior distributions of demographic parameters, indicating the robustness of our Bayesian analyses. The estimated effective population size of was 5396 (2211, 9641) for *galacifolia* and 619 (265, 2388) for *brevistyla*, respectively [HiPt (95% low, high) values are shown as posterior distribution summary]. The larger population size estimated for the common ancestral population [20609 (15302, 28216)] showed the extant two "species" experienced demographic decline after lineage divergence. The divergence event was estimated to have occurred 20049 (8255, 53070) years before present, a timing roughly corresponding to the last glacial maximum of the Pleistocene.

Like other species of *Shortia* (Gaddy and Nuraliev 2017), both North American species of the genus are narrow endemics with very small ranges and exacting habitats. All the known populations of *S. brevistyla* are all found within 10 km radius in the upper Catawba River drainage just north and northeast of Marion, North Carolina in McDowell County. The more numerous populations of *Shortia galacifolia* all fit into an area within a 15 km radius around its type locality in the Keowee-Toxaway drainage just north of the Lake Jocassee dam (in Oconee County, South Carolina). Finally, the southwestern most *S. brevistyla* population known is about 100 km northeast of the northeastern most *S. galacifolia* population.

Shortia brevistyla appears to have speciated due to genetic drift in isolated populations of *Shortia galacifolia* with low heterozygosity and self-pollinating plants. Estimates, using Bayesian estimation in the IMa2 program, indicate that these two species diverged approximately 20,000 years ago during the time of the glacial maximum (Osssowski et al. 2010).

ACKNOWLEDGEMENTS

Work by the latter three authors was supported by JSPS Core-to-Core Program A. Advanced Research Networks and by JST SICORP program (15656846). We would like Alan Weakley for comments and suggestions that improved the quality of this paper.

LITERATURE CITED

Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. 2011. Stacks:Building and Genotyping Loci De Novo From Short-Read Sequences. G3-Genes Genomes. Genetics 1:171-182.

Cheng, Tao, et al. 2016. Barcoding the Kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources 16.1: 138-149.

Davies, P. A. 1952. Geographical variations in *Shortia galacifolia.* Rhodora 54:121-124.

Davies, P. A. 1956. Type location of *Shortia galacifolia.* Castanea 21:107-113.

Gaddy, L. L. and Maxim S. Nuraliev 2017. *Shortia rotata* (Diapensiaceae), a new species from Vietnam. Wulfenia 24:53–60.

Hatley, J. R. 1977. An analysis of variation in *Shortia galacifolia*. M. S. Thesis. North Carolina State University. Raleigh. 57 p.

Hey, J. and Nielsen, R. 2007. Integration within the Felsenstein equation for improved Markov Chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Sciences 104:2785–2790.

Jombart, T. 2008. Adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Murray, M. G. & Thompson, W. F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8: 4321-4325.

Nesom, G. L. 2009. *Shortia.* In: Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico. 20+ volumes. Oxford University Press. New York and Oxford. Vol. 8, pp. 333-335.
Okonechnikov, Konstantin, et al. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28: 1166-1167.
Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., & Lynch, M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana.* Science 327(5961): 92-94.
Peakall, R. & Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537-2539.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. & Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81: 559-575
R Development Core Team. 2014. R version 3.1.0: A language and environment for statistical computing.
Rousett, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8: 103-106.
Suyama, Y. & Matsuki, Y. 2015. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. Scientific Reports 5: 16963.
Williams, Charlie, E. M. Norman, and G. G. Aymonin. 2004. The type locality of *Shortia galacifolia* T. & G. visited once again. Castanea 69:169-173.
Zahner, R. and Steven M. Jones. 1983. Resolving the type location of *Shortia galacifolia* Torrey & Gray. Castanea 48:163-173.