bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Comparative population genomics in two sympatric species of *Strophostyles* (Fabaceae) with contrasting
2 life histories
3

4 Sterling A. Herron[1,3], Zachary N. Harris[1], Matthew J. Rubin[2], Allison J. Miller[1,2,3]
5
6 [1]Saint Louis University, Department of Biology, 3507 Laclede Avenue, St. Louis, MO 63103, USA
7 [2]Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA
8 [3]Authors for correspondence (email: amiller@danforthcenter.org; sterling.herron@slu.edu); phone: +1-
9 314-587-1232
10

11 Manuscript received _______; revision accepted _______.
12
13 Running head: Population genomics of *Strophostyles*
14

15 **ABSTRACT**
16

17 PREMISE OF THE STUDY: Life history is an important predictor of population genetic variation,
18 although this link remains unexplored in numerous important plant lineages. One such lineage is the
19 legume genus *Strophostyles*, which contains both annual and herbaceous perennial vines native to eastern
20 North America. Specifically, it remains to be determined whether *Strophostyles* species with different life
21 histories show different patterns of genetic differentiation and diversity, as well as if these species
22 hybridize across their range.
23
24 METHODS: Here we sampled the perennial *Strophostyles helvola* and annual *S. leiosperma* in five sites
25 across a latitudinal transect in the central United States, including three sites where the species occur in

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26 sympatry. Using genotyping-by-sequencing, we identified 5556 polymorphic SNPs across 166
27 individuals.
28
29 KEY RESULTS: There is no evidence that *Strophostyles helvola* and *S. leiosperma* hybridize in the
30 populations examined. Within species, *Strophostyles helvola* (perennial) displays admixture among
31 populations, while *S. leiosperma* (annual) does not, although both species show more genetic diversity
32 among rather than within populations. Patterns of genetic diversity are varied across populations of both
33 species, with both heterozygote excess and deficiency.
34
35 CONCLUSIONS: The complex patterns of genetic differentiation and diversity warrant further
36 investigation of life history and population dynamics in *Strophostyles*, particularly mating system and
37 modes of gene flow. This study exemplifies the diversity of population genetic patterns even within a
38 small genus, and it reinforces the need to characterize such diversity in non-model systems to gain a more
39 complete understanding of how life history contributes to population genetics.
40

41 **Keywords:** crop wild relative; GBS; genotyping-by-sequencing; legume; Leguminosae; life history; life
42 span; Phaseoleae; population genetics; *Strophostyles leiosperma*
43

44 Life history strategy has long been recognized as an important predictor of population genetic variation
45 (Loveless and Hamrick, 1984). In general, annual species are expected to have greater population genetic
46 differentiation (F_{ST} or G_{ST}) and lower within-population genetic diversity (H_S), compared to perennial
47 species (Loveless and Hamrick, 1984; Morishima and Barbier, 1990; Austerlitz et al., 2000). However,
48 these genetic associations with life span may be more directly related to other life history traits which are
49 often associated with life span, particularly mating system (selfing, mixed-mating, or outcrossing;
50 Balfourier et al., 1998; Duminil et al., 2007; Duminil et al., 2009). Selfing or mixed-mating species tend
51 to have greater among-population genetic differentiation and lower within-population genetic diversity as

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52 compared to outcrossing species (Loveless and Hamrick, 1984; Hamrick and Godt, 1996; Duminil et al.,
53 2007; Gamba and Muchhala, 2020). Often annuals are selfing and perennials are outcrossing, although
54 this pattern is not universal, particularly in herbaceous species (Barrett et al., 1996; Barrett and Harder,
55 2017). Considering variability over geographic space is also particularly important as a species' life
56 history may change with differing climates or habitats (Peterson et al., 2016), which can in turn change
57 patterns of population genetic structure (Vest, 2019). Overall, life span alone is not likely to consistently
58 predict population genetic differentiation and diversity, and a broader perspective including other life
59 history traits, the native ecology, and the evolutionary history of the study species and its populations is
60 imperative.

61

62 The legume family (Fabaceae Lindl.) is one of the most economically important and widespread groups
63 of plants, containing ca. 19,500 species, and likewise a striking array of life history diversity (Lewis et al.,
64 2005; Azani et al., 2017). Within Fabaceae, subtribe Phaseolinae Bronn (tribe Phaseoleae Bronn ex DC.)
65 stands out, containing the economically important genera *Phaseolus* L. (common bean) and *Vigna* Savi
66 (cowpea), in addition to 25 other wild genera spread throughout Africa, Asia, Australia, and the Americas
67 (Delgado-Salinas et al., 2011). Despite the economic interest of the subtribe, numerous species,
68 particularly those outside genera containing major crops, remain uncharacterized regarding life history
69 and genetic diversity. Assessment of this diversity will be beneficial in understanding the evolutionary
70 history of these lineages and the potential range of genetic variation in crop wild relatives.

71

72 *Strophostyles* Elliott (fuzzybean) is a monophyletic legume genus within Phaseolinae, consisting of three
73 species of herbaceous vine: *S. helvola* (L.) Elliott, *S. leiosperma* (Torr. & A. Gray) Piper, and *S.*
74 *umbellata* (Muhl. ex Willd.) Britton (Riley-Hulting et al., 2004; Delgado-Salinas et al., 2011). Their
75 native range is centered in the eastern United States, extending from Texas east to Florida, north to the
76 Great Lakes and eastern Canada (USDA, 2018; Pelotto and del Pero Martinez, 1998; Riley-Hulting et al.,
77 2004). *S. helvola* is widely distributed across the central and eastern United States, *S. leiosperma* has the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

78 westernmost distribution and is primarily restricted to a central band from Minnesota south to Texas, and
79 *S. umbellata* is distributed more in the southeastern portion of the U.S. (Riley-Hulting et al., 2004). *S.*
80 *helvola* is known to occur in sympatry with *S. leiosperma* in its western range and *S. umbellata* in its
81 eastern range, but *S. leiosperma* and *S. umbellata* are not known to co-occur (Riley-Hulting et al., 2004).
82 All *Strophostyles* species inhabit mesic environments, including freshwater swamps and flood basins,
83 coastal saltwater areas, sand prairies, and anthropogenically disturbed sites such as ditches and ponds
84 (Riley-Hulting et al., 2004). *S. leiosperma* is however distinct in that it occurs in relatively more open,
85 drier environments (Riley-Hulting et al., 2004). *Strophostyles* is also one of the few Phaseolinae clades
86 that has a broad temperate distribution, and it is the only genus in Phaseolinae to have a distribution
87 centered in the United States (Riley-Hulting et al., 2004). *Strophostyles* is sister to the genus *Dolichopsis*
88 Hassl., which is native to the Gran Chaco in South America, an area that also undergoes freezing
89 temperatures, and its other closest genera, *Macroptilium* (Benth.) Urb., *Mysanthus* G.P. Lewis & A.
90 Delgado, and *Oryxis* A. Delgado & G.P. Lewis, are also native to the neotropics (Riley-Hulting et al.,
91 2004). *Strophostyles* and *Dolichopsis* share keel morphology, where the keel gradually curves to the right
92 of the flower as opposed to a sharp curling of the keel as in many other New World Phaseolinae, and the
93 keel in both genera is also distinctly gibbous (Riley-Hulting et al., 2004). *Strophostyles* species are
94 morphologically distinguished from closely related Phaseolinae genera by their cellular seed coating
95 (which lends buoyancy), divergent stipules, a calyx consisting of four lobes that are attenuate to acute,
96 and persistent floral bracts on the pedicels (Riley-Hulting et al., 2004; Delgado-Salinas et al., 2011).
97 Within *Strophostyles*, species are primarily diagnosed by their flower's keel morphology, although *S.*
98 *leiosperma* in particular possesses other unique characteristics, including distinctly sericeous leaves and
99 pods, and usually smooth seeds (Riley-Hulting et al., 2004). All three *Strophostyles* species are diploid
100 (2n=22; Lackey, 1980; Riley-Hulting et al., 2004; Yatskievych, 2013).
101

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

102 *Strophostyles* species also have potential in agricultural applications. *S. helvola* and *S. leiosperma* have
103 been targeted as a sustainable, native forage and hay supplement for livestock and wildlife in the southern
104 Great Plains region (Muir et al., 2005; Foster et al., 2008), with cultivars of *S. leiosperma* being
105 developed specifically for that purpose (Butler and Muir, 2010). In the wild, *Strophostyles* herbage and
106 seeds are known to be an important food source for wildlife (Bird and Bird, 1931; Wiseman, 1977;
107 Immel, 2001; Gee et al., 2011). Unlike many cultivated Phaseolinae species, *S. helvola* in particular is
108 known for its salinity tolerance in coastal populations (Tsang and Maun, 1999; Zhang et al., 2018;
109 Zuelsdorf, 2018). *Strophostyles* seeds and roots were also historically consumed by indigenous American
110 peoples, highlighting their human palatability (Parker, 1991; Immel, 2001).
111
112 Despite its potential utility and botanical interest, various aspects of *Strophostyles* life history remain
113 unclear. Reported life span of *Strophostyles* species is variable in the literature: *Strophostyles helvola* and
114 *S. leiosperma* have both been reported as either annual or perennial (McGregor et al., 1986; Stubbendieck
115 and Conard, 1989; Isely, 1998; Diggs et al., 1999; Yatskievych, 2013). Nevertheless, the most thorough
116 taxonomic and morphological treatment of the genus to date describes *Strophostyles helvola* as a
117 perennial with a large taproot, and *S. leiosperma* as an annual to short-lived perennial with a small
118 taproot, where the taproot functions as the persisting perennial organ after the annual shoot dies back
119 during the winter (Riley-Hulting et al., 2004; Delgado-Salinas, pers. comm. 2021). Balancing immediate
120 reproduction vs. survival (vegetative allocation) is a major trade-off concept of plant life history theory, is
121 a correlate of plant life span, and is likely to vary by environment (Charnov and Schaffer, 1973;
122 Friedman, 2020; Lundgren and Des Marais, 2020). Specifically, intraspecific variation in life span is
123 associated with environmental disturbance, with annual forms associating with seasonally disturbed sites
124 (often drought) where adult survival is unlikely across seasons or disturbance events, and perennial forms
125 are associated with relatively stable sites where adult survival is likely across seasons (Hall and Willis,
126 2006; Peterson et al., 2016; Monroe et al., 2019).
127

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

128 Mating system is another aspect of life history likely to be an important predictor of genetic diversity
129 patterns in *Strophostyles*. All *Strophostyles* species have overlapping flowering times and attract a range
130 of generalist bee pollinators (particularly *Megachile*; Robertson, 1890; Krombein et al., 1979; Riley-
131 Hulting et al., 2004). *S. helvola* specifically has an observed keel mechanism which serves to retrieve
132 foreign pollen from the floral visitors while also depositing its own pollen on the visitor's thorax via a
133 style-brush, encouraging outcrossing (Foerste, 1885; Robertson, 1890). The pollen deposition mechanism
134 of the other two species is unclear, but this combined with pollinator observations point to at least a
135 partially outcrossing mating system in *Strophostyles* species. *S. leiosperma* is also hypothesized to have at
136 least a partially selfing mating system, based on the observation that closed flowers still produce viable
137 seeds; reproductive assurance through selfing would also be consistent with the annual life span (Riley-
138 Hulting et al., 2004).
139
140 Some phylogenetic and transcriptomic studies have targeted *Strophostyles* and its close relatives, but
141 extensive genomic information is lacking for the genus. Riley-Hulting et al. (2004) used chloroplast and
142 ITS markers in a phylogenetic study to confirm all three *Strophostyles* species as a monophyletic clade, as
143 well as to confirm its phylogenetic placement among a number of neotropical Phaseolinae genera. More
144 recently, Zhang et al. (2018) sequenced the transcriptome of *S. helvola* in order to discover upregulated
145 genes associated with salinity tolerance. However, a whole genome population-level approach has yet to
146 be implemented in *Strophostyles*, which leaves a number of open questions. Specifically, while there is
147 little evidence to date for putative hybridization between species in the genus (Riley-Hulting et al., 2004),
148 this has not been directly investigated in sympatric populations of *Strophostyles* species. Also, while there
149 is preliminary evidence for broad differences in genetic diversity across the range of *Strophostyles* (Riley-
150 Hulting et al., 2004), this has not been shown among multiple individuals within multiple populations.
151 Characterization of population genomics may further reveal signatures of mating system and how patterns
152 of dispersal contribute to gene flow among populations of each *Strophostyles* species.
153

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

154 Here we focus on the sympatric species *Strophostyles helvola* and *S. leiosperma*, in the context of their
155 different life history strategies, in populations across a latitudinal transect from northern Iowa to central
156 Arkansas, USA. We ask the following questions: (1) are *Strophostyles helvola* and *S. leiosperma*
157 genetically distinct, or is there evidence of hybridization between these species when they occur in
158 sympatry; (2) is there genetic differentiation among populations within each species, or is there evidence
159 of admixture among populations; and (3) what are the patterns of genetic diversity within and among
160 populations for each species?
161

162 **MATERIALS AND METHODS**

163

164 ***Sample collection***—This study focused on two of the three *Strophostyles* species: *S. helvola* and
165 *S. leiosperma*, due to their close relationship and sympatric distribution in the Midwest United States
166 (Riley-Hulting et al., 2004). We collected samples from five sites along a latitudinal transect ranging from
167 northern Iowa to central Arkansas, USA (Fig. 1; Table 1; Appendix S1; see Supplemental Data with this
168 article). Collection sites were targeted based on occurrence records for the species (Global Biodiversity
169 Information Facility; www.gbif.org) and by contacting regional botanists. *Strophostyles umbellata* was
170 not collected due to its rarity in the sites sampled (it has not been observed in this region by the authors);
171 while it is reported to occur in this range, it does not appear to occur sympatrically with *S. helvola* and *S.*
172 *leiosperma* at least in the populations observed. In general, *S. umbellata* is known to have more scattered
173 individuals and populations of low density, and thus the likelihood of finding substantial populations is
174 lower (Riley-Hulting et al., 2004).
175

176 From north to south, sites sampled included the Cedar Hills Sand Prairie (Cedar Falls, Iowa, USA; site
177 “IA”), Frost Island Conservation Area and Iliniwek Village State Historic Site (Wayland, Missouri, USA;
178 site “IV-MO”), Shaw Nature Reserve (Gray Summit, Missouri, USA site “SNR-MO”), Sand Pond
179 Conservation Area (Neelyville, Missouri, USA; site “SP-MO”), and Burns Park (North Little Rock,

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

180 Arkansas, USA; site "AR") (Fig. 1; Table 1; Appendix S1). Habitat types were generally open prairie,
181 open grassland/shrubland, and forest edges, or the edge of ponds (SNR-MO) and along a riverbank (AR)
182 (Appendix S1). It was confirmed with site managers that *Strophostyles helvola* and *S. leiosperma* were
183 not known to be intentionally seeded at any site (Appendix S1). Linear distance between nearest
184 collection sites ranged from 219 to 256 km, and mean altitude at each site ranged from 76 to 284 meters
185 above sea level (Fig. 1; Table 1). Plant populations at each site were defined as a group of 15 or more
186 individuals of one species within the same 10 square km area. *Strophostyles helvola* and *S. leiosperma* co-
187 occurred in the three northernmost sites (IA, IV-MO, and SNR-MO), while only *S. helvola* was present at
188 SP-MO and AR. At each site, we sampled 17-29 individual plants per species, and in total, we sampled
189 110 individuals of *S. helvola* and 56 individuals of *S. leiosperma* (Table 1). We collected fresh, young
190 leaf tissue and stored it in paper envelopes, which were sealed in a ziploc bag filled with dry silica for
191 desiccation. Individuals of the same species were collected at least ~3 m apart. Species were generally
192 identified by their distinct papilionoid flower morphology: *S. helvola* exhibits a prominent and distinctly
193 curved keel, while *S. leiosperma* exhibits a small, only slightly curved keel that is mostly hidden by the
194 wing petals (Fig. 1; Riley-Hulting et al., 2004). In the absence of flowers (rare), these species can also be
195 distinguished on leaf and pod morphology. *S. helvola* has ovate, sparsely haired, often lobed leaflets,
196 while *S. leiosperma* has lanceolate, sericeous, unlobed leaflets (rarely having shallow lobes; Riley-
197 Hulting et al., 2004). *S. helvola* also has large, sparsely haired pods (30-96 mm long) and seeds with an
198 accessory cellular coating, while *S. leiosperma* has smaller, distinctly sericeous pods (12-41 mm long)
199 with smaller, usually smooth seeds (Riley-Hulting et al., 2004). For each plant collected, we recorded the
200 latitude, longitude, and elevation, and photographed most plants to document identifying features. For
201 each population, we also collected two herbarium voucher specimens per species, including the entire
202 aboveground and most of the belowground portion of the plant when possible. We deposited the vouchers
203 at the Missouri Botanical Garden (St. Louis, Missouri, USA; Table 1). The map of population sampling
204 was created using the R packages 'ggmap' and 'ggplot2' (Kahle and Wickham, 2013; Wickham, 2016; R
205 Core Team, 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

206

207 ***Genomic DNA extraction*—**We extracted genomic DNA from the leaves of 166 plants collected
208 in the field (Appendix S1). When available, 55-60 mg of dried leaf tissue per sample was ground into
209 powder using liquid nitrogen and a mortar and pestle, or by freezing the tissue in a tube with liquid
210 nitrogen, inserting 0.9-3.175 mm steel ball bearings, and oscillating using a Rech Mixer Mill 400 at 30 Hz
211 for 2 min, or until a powder. DNA was extracted using an E.Z.N.A.® Plant DNA Kit (Omega Bio-tek,
212 Norcross, Georgia, USA) according to the product label, with a few modifications: the initial incubation
213 in P1 buffer was increased to 45-60 min during which samples tubes were vortexed (instead of inverting),
214 and the centrifuge speed and sometimes duration was increased to allow better tissue and DNA pelleting
215 as well as movement of liquid through the DNA mini column. DNA concentration was quantified using a
216 Qubit® dsDNA BR Assay Kit and an Invitrogen™ Qubit™ 3 Fluorometer (Thermo Fisher Scientific,
217 Waltham, Massachusetts, USA). To increase concentration of some samples, we vacufuged them at 30°C
218 and re-eluted them. We retained samples with > 10 ng/µL DNA concentration for sequencing. Eluted
219 DNA was stored at 4°C until shipped for genomic processing (with dry ice).

220

221 ***Sequencing*—**We submitted all samples to the University of Minnesota Genomics Center
222 (Minneapolis, Minnesota, USA), where they were sequenced using genotyping-by-sequencing (GBS), a
223 reduced-representation whole genome method (Elshire et al., 2011; Poland et al., 2012). Dual-indexed
224 GBS libraries were created using a double digest with the two enzymes MspI and PstI on 100 ng of DNA.
225 Samples were treated with T4 ligase and phased adaptors with TGCA and CG overhangs. Libraries were
226 pooled and size selected for 300-744 bp segments, and the pool was diluted to 1.5 nM and sequenced on a
227 half-lane of an Illumina NovaSeq 6000 FlowCell with single-end 1x100-bp reads. This generated 54,142
228 Mb total reads (average 326 Mb per sample), with mean quality score ≥Q30 for all libraries (Appendix
229 S1).

230

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

231 ***Alignment and SNP calling*—**Both ends of sequence reads were trimmed for quality at a
232 minimum Phred score of 30 using bbduk (Bushnell, 2014; qtrim=rl, trimq=30). We aligned quality-
233 filtered reads to the *Phaseolus vulgaris* Andean landrace 'G19833' v.2.1 reference genome
234 (https://phytozome-next.jgi.doe.gov/info/Pvulgaris_v2_1), which was downloaded from the Ensembl
235 Genomes website (https://ensemblgenomes.org/; Yates et al., 2020; accessed Feb 26, 2020), using
236 Burrows-Wheeler Alignment (BWA MEM) with default settings (bwa v0.7.17-r1188; Li and Durbin,
237 2009). In the absence of a *Strophostyles* genome, the *Phaseolus vulgaris* genome was used as a reference
238 genome due to its long history of improvement and high coverage of the genome. Variant calling used
239 bcftools v1.9, with default settings, and genotypes were filtered at a minimum read depth of 10 with
240 vcftools v0.1.16 (--minDP 10; Danecek et al., 2011; all sites passed this filter). Biallelic SNPs were called
241 with PLINK v1.90b4 (Chang et al., 2015). After evaluating different minor allele frequencies and missing
242 genotype rates, a conservative approach was applied: we filtered at a minor allele frequency of 0.025 and
243 a maximum missing genotype rate per SNP of 10% (--biallelic-only --snps-only --maf 0.025 --geno 0.1).
244 Less conservative filtering approaches increased the number of SNPs retained but did not change the
245 broad genetic structure trends or the conclusions. Ultimately, 1,584,797 SNPs were called in total. 5556
246 SNPs and all 166 samples were retained following filtering.
247
248 ***Population genetic differentiation and diversity*—**We implemented fastSTRUCTURE (Raj et al.,
249 2014) to characterize the population genetic structure of all samples, testing values of K clusters ranging
250 from 1 to 8 (total population number considering both species). We then used the chooseK.py function in
251 fastSTRUCTURE to determine the optimal K value based on two criteria: (1) model complexity which
252 maximized marginal likelihood, and (2) the number of model components used to explain structure in the
253 data. K=6 clusters was optimal based on the number of model components, while K=7 was optimal based
254 on maximized marginal likelihood. The seventh cluster group for the K=7 model has an extremely rare
255 representation such that effectively only 6 clusters emerge, so we defer to the original K=6 model for
256 further discussion. Population structure plots were generated using the R packages 'ggplot2' and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

257 'pophelper' (Francis, 2017). We also implemented principal component analysis (PCA) of genetic
258 variation using PLINK (--pca 166) and visualized the PCA using 'ggplot2' with the EIGENVEC output
259 from PLINK.
260
261 For calculations of population statistics, our filtered SNP variant call format file was converted to a
262 GENEPOP text file using PGDSpider2 (Lischer and Excoffier, 2012), which was then converted to a
263 GENIND object in R (package 'adegenet'; Jombart, 2008) and reformatted to a GENEPOP object
264 ('graph4lg' package; Savary, 2020) usable in Genodive v.3.04 (Meirmans, 2020), with the modification
265 of changing missing values to '0000' in order to be interpreted by Genodive as diploid. All further
266 analyses were completed in Genodive. Pairwise F_{ST} values between all populations within species were
267 calculated and P-values generated from 1000 permutations. Isolation by distance was also tested for each
268 species using a Mantel test. Briefly, the test assesses correlation between the geographic distance matrix
269 (based on the average of all sample coordinates for each population of each species, respectively) and the
270 pairwise genetic distance matrix $[F_{ST}/(1\text{-}F_{ST})]$ between populations within a species, with 1000
271 permutations. Genetic diversity metrics were also calculated per population for each species, including
272 the observed heterozygosity (H_O), within-population expected heterozygosity (H_S; Nei's gene diversity;
273 Nei, 1987), and inbreeding coefficient (F_{IS}). The deviation of genotype frequencies within each
274 population from the expectation under Hardy-Weinberg equilibrium (random mating) was calculated
275 using the least squares method (from analysis of molecular variance, AMOVA; Excoffier et al., 1992)
276 with 1000 permutations. An AMOVA was also calculated using the Infinite Allele Model with 1000
277 permutations, in order to dissect the percentage of genetic variance within individuals, among individuals
278 nested within populations, among populations nested within species, and among species.
279

280 RESULTS

281

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

282 **Strophostyles helvola *and* S. leiosperma *are distinct lineages*—***Strophostyles helvola* and *S.*
283 *leiosperma* were found to be genetically distinct from each other within and across collection sites.
284 *Strophostyles helvola* and *S. leiosperma* separated distinctly with K=2 genetic clusters (Fig. 2). Principal
285 component analysis also supported the species separation: PC1, which separates *S. helvola* and *S.*
286 *leiosperma*, explained 45% of the total genetic variation (Fig. 3A). Lastly, the analysis of molecular
287 variance (AMOVA) corroborated this pattern, in that a significant percentage of genetic variation was
288 attributed to among-species variation (62.4%; $F_{CT} = 0.642$, $P = 0.001$; Table 2).
289
290 ***Within-species population genetic structure differs for* Strophostyles helvola *and* S.**
291 **leiosperma—**There were unique patterns of genetic clustering within each species, with *Strophostyles*
292 *helvola* showing evidence of admixture among some populations and *S. leiosperma* showing genetically
293 distinct clusters. For K=6 (optimal K value that explained structure in the data, based on the number of
294 model components), individuals of *S. helvola* assigned to one or more of four clusters, and individuals of
295 *S. leiosperma* assigned to two other distinct clusters (Fig. 2). Within *S. helvola*, individuals from
296 populations IV-MO and SNR-MO showed the strongest population differentiation from other clusters
297 (Fig. 2). SNR-MO was the only population of *S. helvola* with no evidence of admixture, while two
298 individuals of *S. helvola* population IV-MO showed only slight admixture with the IA population (Fig. 2).
299 In contrast, *S. helvola* populations IA, SP-MO, and AR showed more evidence of admixture (Fig. 2). For
300 *S. helvola* population IA, 14/20 individuals' genetic composition assigned only to its own unique genetic
301 cluster, while 6/20 individuals showed admixture, primarily with population IV-MO, but also with SNR-
302 MO and SP-MO/AR (Fig. 2). Most *S. helvola* individuals from population SP-MO and AR have genetic
303 ancestry that assigned predominantly to the same genetic cluster, while the admixture present in both
304 populations assigned primarily to IV-MO (Fig. 2). For population SP-MO specifically, 11/19 individuals
305 assigned only to their own genetic cluster, while the other 8/19 individuals showed signs of admixture,
306 primarily with population IV-MO but also with SNR-MO (Fig. 2). For population AR, 18/29 individuals
307 assigned solely to their own genetic cluster (the same as SP-MO), with the remaining 11/29 again

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

308 showing admixture mainly with population IV-MO but also IA and SNR-MO (Fig. 2). For *S. leiosperma*,
309 only two genetic clusters emerged and there was a distinct lack of admixture. Both the IA and IV-MO
310 populations assigned to the same genetic cluster, with no substantial evidence of shared ancestry with
311 other clusters (Fig. 2). All 37 individuals of the *S. leiosperma* IA and IV-MO populations assigned
312 predominately to their own genetic cluster (Fig. 2). For *S. leiosperma* population SNR-MO, 18/19
313 individuals assigned solely to their own genetic cluster, with only one individual showing slight
314 admixture with the other *S. leiosperma* cluster (Fig. 2). Considering intermediate K values between K=2
315 and K=6, K=3 through K=5 illustrated the stepwise separation of north and south *S. helvola* populations,
316 the two clusters of *S. leiosperma*, and potential mixed ancestry in population IV-MO (Appendix S2). K=7
317 cluster structure (see Methods) was similar to the K=6 pattern, with the main deviations being that *S.*
318 *helvola* populations IV-MO and SP-MO primarily assigned to the same genetic cluster, and that
319 population AR primarily assigned to a different genetic cluster than SP-MO (Appendix S2).
320
321 Fixation index results largely supported the fastSTRUCTURE patterns. *Strophostyles helvola* population
322 SNR-MO showed some of the highest pairwise F_{ST} values with the other *S. helvola* populations (0.359 -
323 0.468) (Table 3). The strong admixture signal of *S. helvola* population IV-MO with IA, SP-MO, and AR
324 was reflected by its lower F_{ST} values when these populations were paired (0.260-0.284) (Table 3). The
325 same predominant genetic cluster assignment of *S. helvola* populations SP-MO and AR was also reflected
326 by their low pairwise F_{ST} (0.294) (Table 3). Consistent with its low levels of admixture, *S. helvola*
327 population IA showed high F_{ST} values paired with SP-MO (0.383) and AR (0.405) (Table 3).
328 Additionally, *Strophostyles leiosperma* cluster separation was reflected in F_{ST} values, where populations
329 IA and IV-MO had a low pairwise F_{ST} of 0.251, while the F_{ST} of *S. leiosperma* population SNR-MO with
330 IA and IV-MO was 0.712 and 0.707, respectively (Table 3). Based on the Mantel test, we also found no
331 evidence of isolation by distance for either species. There was only a slight, nonsignificant positive
332 correlation between genetic [F_{ST}/(1-F_{ST})] and geographic distance for *S. helvola* (R^2=0.020; P=0.362) and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

333 *S. leiosperma* (R^2=0.184; P=0.495). This is consistent with the nonlinear pattern of genetic similarity
334 across geographic space observed in our fastSTRUCTURE and F_{ST} results (Fig. 2; Table 3).
335
336 Principal component analysis also largely supported the intraspecific genetic patterns observed in
337 fastSTRUCTURE and F_{ST} values. PC2, which separated the two clusters of *S. leiosperma*, explained 17%
338 of the total genetic variation (Fig. 3A). *S. helvola* populations remained tightly clustered on both PC1 and
339 PC2, with intraspecific variation only emerging in PC3 and PC4 (Fig. 3B). *S. helvola* individuals with
340 admixture from the AR / SP-MO and IV-MO clusters were intermediate to the non-admixed individuals
341 from these clusters in PC3 and PC4 (Fig. 3B). There was also some overlap of the *S. helvola* SNR-MO
342 cluster with the *S. helvola* AR / SP-MO cluster on PC4 alone, reflecting the slight admixture of SNR-MO
343 in those populations (Fig. 3B). The *S. helvola* IA cluster remained relatively isolated on PC3 and PC4,
344 with the exception of a few admixed individuals approaching the *S. helvola* IV-MO cluster, representing
345 the admixture between those populations (Fig. 3B). *S. leiosperma* clusters all lacked genetic variation on
346 PC3 (F_1=3.11; P=0.08) and PC4 (F_1=1.61; P=0.21) and thus were omitted from Fig. 3B in order to better
347 visualize variation in *S. helvola*.
348
349 The AMOVA further confirmed that, in addition to genetic variation between species, there was a
350 significant level of variation among populations within species (17.7%; F_{SC}=0.471; P=0.001), but a lack
351 of significant variation among individuals within populations (-0.1%; F_{IS}=-0.003; P=0.937; Table 2;
352 slight negative deviations of F statistics from 0 can be interpreted as a lack of genetic structure among
353 members of that group, i.e., F_{IS}≈0; Schneider et al., 2000; Meirmans, 2006).
354
355 ***Population genetic diversity patterns are varied***—Both species showed mixed patterns of genetic
356 diversity among their populations. The genotype frequencies of all populations deviated significantly
357 from Hardy-Weinberg equilibrium (P = 0.001-0.003) with the exception of *S. leiosperma* population
358 SNR-MO (P = 0.487), both due to heterozygote excess and deficiency as determined by F_{IS} (Table 4). The

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

359 greatest observed heterozygosity (H_O) was observed in the *S. helvola* IA population (0.109) and the
360 lowest in the *S. helvola* SNR-MO population (0.087; Table 3). *S. leiosperma* had more consistently high
361 H_O across its populations (0.100-0.108) than *S. helvola* (Table 4). In contrast, *S. helvola* populations had
362 more consistently high within-population gene diversity (H_S; expected heterozygosity; Nei, 1987), with
363 the highest value in population IV-MO (0.125), and the lowest value again in population SNR-MO
364 (0.086; Table 4). *S. leiosperma* populations IA and IV-MO had the lowest overall H_S at 0.078 and 0.076
365 respectively, while *S. leiosperma* population SNR-MO was higher at 0.101 (Table 4). Negative values for
366 the inbreeding coefficient (F_{IS}) were observed in *S. helvola* populations IA (-0.133) and SNR-MO (-
367 0.019), and *S. leiosperma* populations IA (-0.387) and IV-MO (-0.310), indicating heterozygote excess
368 (Table 4). F_{IS} was highest in *S. helvola* IV-MO, SP-MO, and AR (0.111-0.250), indicating heterozygote
369 deficiency and possibly higher levels of inbreeding in these populations (Table 4). *S. leiosperma*
370 population SNR-MO had an F_{IS} value of 0, indicating neither heterozygote deficiency nor excess, and
371 thus approximate Hardy-Weinberg equilibrium (Table 4).
372
373 **DISCUSSION**
374
375 We present the first population genetics study of *Strophostyles*, which supports the genetic separation of
376 *S. helvola* and *S. leiosperma* and highlights unique patterns of genetic structuring and diversity among
377 populations within each species.
378
379 ***Separation of species***—This study provides clear evidence for the genetic differentiation of
380 *Strophostyles helvola* and *S. leiosperma*, at least in the range studied here. While these two species are
381 sympatric across much of the central United States, there does not appear to be gene flow between them.
382 This finding is consistent with the clear morphological distinction of *S. leiosperma* from the other two
383 *Strophostyles* species, as well as genetic evidence from chloroplast and ITS markers (Riley-Hulting et al.,
384 2004). This is in spite of similar pollen morphology across *Strophostyles* species, overlapping flowering

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

385 times, generalist pollinators, and *S. helvola* and *S. leiosperma* being sister species (Krombein et al., 1979;
386 Riley-Hulting et al., 2004; own observation). This clear genetic separation in spite of opportunity for
387 hybridization indicates an as yet undetermined reproductive barrier between *S. helvola* and *S. leiosperma*.
388 One possibility is interspecific pollen-rejection during pollen-pistil interaction (Bedinger et al., 2017);
389 however, demonstrating this will require testing under controlled conditions.
390
391 ***Population genetic structure within species*****—**Population genetic structure was found to differ for
392 *Strophostyles helvola* and *S. leiosperma*. *Strophostyles helvola* is a putatively outcrossing species with
393 large, showy flowers and which is particularly common in mesic sites, while *S. leiosperma* is a putatively
394 selfing species with small flowers and which tends to inhabit relatively more open, xeric sites;
395 nevertheless, there is considerable overlap between their distributions (Fig. 1; Riley-Hulting et al., 2004).
396 In *S. helvola*, we found four distinct genetic clusters among the five populations collected, with the two
397 southernmost populations (SP-MO and AR) primarily assigning to the same cluster (Fig. 2). In *S.*
398 *leiosperma*, we found only two genetic clusters, with the two northernmost populations assigning to the
399 same cluster (Fig. 2).
400
401 Evidence of admixture was discovered primarily in three populations of *Strophostyles helvola*: IA, SP-
402 MO, and AR, while the structure patterns of the other genetic clusters were largely unmixed. Given the
403 great distances between collection sites (>200 km) and likely either selfing and/or insect-mediated
404 pollination, it is unlikely that direct pollen flow is occurring across the populations studied (Hamrick,
405 1982; Riley-Hulting et al., 2004). Genetic admixture here could be from migration via seed dispersal and
406 shared ancestry. Similar to the pattern in *S. helvola*, a study in *Capsella* found that in a selfing species,
407 while there was genetic structure between populations, there were also several individuals in most
408 populations with mixed genetic ancestry, which they attribute to the colonization of a few propagules
409 from distant populations and subsequent genetic integration (St. Onge et al., 2011). However, there are
410 also likely several populations of *S. helvola* and *S. leiosperma* interspersed between these localities,

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

411 which may allow indirect gene flow between the populations studied here. Sampling from such
412 interspersed populations at a finer geographic scale could reveal a genetic gradient of admixture produced
413 by gene flow across these more proximate populations, which may thus function as metapopulations.
414 Populations lacking admixture, such as *S. helvola* populations IV-MO and SNR-MO, may indicate that
415 these populations are relatively isolated from such potential gene flow.
416
417 *Strophostyles helvola* may often be water-dispersed, as this species' seeds are known to have a cellular
418 coating which lends them buoyancy in water (Riley-Hulting et al., 2004). The seeds of *Strophostyles*
419 *umbellata*, which have the same coating as *S. helvola*, have been observed to be buoyant in water for
420 three weeks (Erickson and Young, 1995). The most proximate potential route of dispersal in our study
421 region may therefore be the Mississippi River and its tributaries. However, while most of our sites are
422 near tributaries of the Mississippi River, none of them lie directly downstream of sites at higher latitudes.
423 Specifically, population IA is closest to the Cedar River (~5 km); IV-MO is closest to the Des Moines
424 River (~3 km); SNR-MO is adjacent to the Meramec River (~3 km); and AR is immediately adjacent to
425 the Arkansas River. Population SP-MO is unique in that it is more than 16 km from the nearest river
426 (Current River).
427
428 In contrast, *Strophostyles leiosperma* shows some population differentiation, but no substantive evidence
429 of admixture (Fig. 2). The *S. leiosperma* population SNR-MO showed very high F_{ST} differentiation from
430 *S. leiosperma* populations IA and IV-MO (Table 3). However, *S. leiosperma* samples from SNR-MO did
431 not show qualitatively distinct floral or vegetative morphology compared to that of the other two
432 populations. One explanation is that IA and IV-MO originated from the same ancestral source population
433 or were connected via some route of migration. While F_{IS} in *S. leiosperma* population SNR-MO is 0, the
434 reduced heterozygosity compared to the other two *S. leiosperma* populations may be due to a genetic
435 bottleneck if the population was colonized by only a few founder individuals (Table 4). Unlike *S. helvola*,
436 *S. leiosperma* seeds are generally smooth and not thought to be commonly water dispersed (Riley-Hulting

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

437 et al., 2004). There are no currently recognized infraspecific taxa for *S. leiosperma*, although Riley-
438 Hulting et al. (2004) did find a slight genetic difference (ITS region) between *S. leiosperma* samples from
439 the northernmost / southwestern-most range and samples from Missouri and Arkansas.
440
441 Overall, the greatest amount of genetic variation for both species was found among rather than within
442 populations (Table 2). However, this pattern could be primarily due to the large geographic distances
443 between populations more so than a biological cause, as increasing distance between populations is
444 known to decrease genetic diversity detected within populations and inflate diversity detected among
445 populations (Reisch and Bernhardt-Römermann, 2014). Similarly, the great distances between
446 populations may explain the lack of clear isolation by distance; at the scale of this study, local dispersal
447 and gene flow may not be the predominant drivers of genetic differentiation (Twyford et al., 2020).
448
449 Human-mediated dispersal may also influence *Strophostyles* population dynamics, as evidenced by many
450 herbarium records collected from anthropogenically disturbed sites, such as near railroads and ditches
451 (Riley-Hulting et al., 2004). Riley-Hulting et al. (2004) specifically suggest *Strophostyles* seed dispersal
452 via movement of ballast (gravel/sand track base) during the construction of railways (Riley-Hulting et al.,
453 2004). *Strophostyles helvola* and *S. leiosperma* have repeatedly been observed in railroad ballast in the
454 Midwest (Pepoon, 1927; Gilly and McDonald, 1936; Deam, 1940), likely due to their proclivity for sandy
455 soil. Indeed, the municipalities of the sites SP-MO (Neelyville, Missouri) and AR (Little Rock, Arkansas)
456 were directly connected by the St. Louis, Iron Mountain and Southern Railway as early as the 1870s (St.
457 Louis, 1878; Stepenoff, 1993). In the absence of clear water routes, this may have been a conduit of
458 shared ancestry in *S. helvola* between sites SP-MO and AR. Colonization of railway habitats has been
459 observed in numerous plant species, particularly annual, ruderal species (Mühlenbach, 1979; Arnold,
460 1981; Austin, 2003; Hill and Blaney, 2009). The extent to which water and human dispersal has
461 influenced population structure in *Strophostyles* could be more directly tested by sampling sites along

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

462 major rivers and historic railroads. Rapid human dispersal could also contribute to nonlinear relationship
463 between geographic and genetic distance.
464
465 ***Population genetic diversity within species***—Varied patterns of genetic diversity were also found
466 within each species, which warrants further investigation. Heterozygote excess (negative F_{IS}) was found
467 in both *Strophostyles helvola* and *S. leiosperma*. High heterozygosity in *S. leiosperma* (IA and IV-MO;
468 Table 4) is in contrast to the lack of heterozygosity found by Riley-Hulting et al. (2004) and the
469 hypothesis that *S. leiosperma* is a primarily selfing species. However, Riley-Hulting et al. (2004) only
470 sequenced two markers (trnK and ITS) and not from multiple individuals within populations, so they may
471 have missed this signature of heterozygosity. Two populations of *Strophostyles helvola* also showed
472 heterozygote excess (IA and SNR-MO; Table 4). Heterozygote excess can occur due to self-
473 incompatibility, asexual reproduction, or selection for heterozygous individuals due to inbreeding
474 depression (Stoeckel et al., 2006). This often occurs in longer-lived woody and clonal species (Stilwell et
475 al., 2003; Ge et al., 2005; Stoeckel et al., 2006), although clonality and woodiness is not known to occur
476 in *Strophostyles* (Riley-Hulting et al., 2004). On the other hand, there was also significant heterozygote
477 deficiency among the other *S. helvola* populations IV-MO, SP-MO, and AR, which suggests some degree
478 of inbreeding in these populations (Table 4). Mating system is known to be quite variable within plant
479 species in general and can even vary among conspecific individuals within the same population (Hamrick,
480 1982; Ma et al., 2020). Such mating system variability may help explain the mixed patterns in genetic
481 diversity in *Strophostyles*; testing outcrossing rate and self-compatibility across multiple populations will
482 be needed to confirm this.
483
484 ***Unique genetic patterns in population SNR-MO***—At the Shaw Nature Reserve (SNR-MO), both
485 *Strophostyles helvola* and *S. leiosperma* displayed unique patterns of genetic differentiation and diversity.
486 Both species had distinctly differentiated genetic clusters for SNR-MO, with a lack of admixture with
487 other populations (Fig. 2, 3), and SNR-MO consistently displayed the highest pairwise F_{ST} values with

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

488 other populations (Table 3). Lastly, F_{IS} was close to 0 (Hardy-Weinberg equilibrium) in SNR-MO for
489 both species (Table 4). All of this points to unique population dynamics occurring at SNR-MO, which
490 could be related to the species' unique history of establishment at this site. While no intentional
491 introduction of *Strophostyles* is known at SNR-MO (James Trager, pers. comm. 2021), it is nevertheless
492 possible that an undocumented or unintentional introduction occurred. Such an introduction may have
493 involved only a small number of seeds, which could have induced a genetic bottleneck for both species.
494 The lack of admixture with other populations also suggests that few if any further dispersal events
495 occurred to introduce new genetic variation into this population, leaving it isolated. This possible founder
496 effect is further suggested by the relatively lower level of heterozygosity for *S. leiosperma* in SNR-MO
497 (F_{IS} = 0) compared to the other two populations (F_{IS} = -0.387 to -0.310; Table 4). The geographic origin
498 of the propagules which colonized SNR-MO remains unknown, although the slight representation of the
499 *S. helvola* SNR-MO genetic cluster in the southernmost populations SP-MO and AR suggest a potential
500 origin from the southern range of the species (Fig. 2). Taken together, population SNR-MO serves as an
501 interesting case study of the potential for both species to become isolated and genetically differentiate in
502 certain disconnected populations. This warrants further investigation into the mechanisms responsible for
503 genetic differentiation in this and other isolated populations, as well as whether differentiation was
504 primarily produced by genetic drift or if there is an adaptive component.
505
506 ***Future directions***—Our study is limited in that a large portion of the range of both *S. helvola* and
507 *S. leiosperma* could not be sampled. *Strophostyles leiosperma* extends across much of the central Great
508 Plains region, and *S. helvola* also has a broad distribution across the coastal eastern U.S. (Riley-Hulting et
509 al., 2004); novel genetic diversity likely remains to be discovered in those areas. Thus, it remains to be
510 seen if our patterns of genetic differentiation and diversity apply to a broader sampling of both species, or
511 if they are unique to the region studied here. Significantly high population differentiation found even at
512 this scale suggests considerable genetic variation across the range of *Strophostyles* species. On the other
513 hand, sampling from sites which are more geographically proximate will also be beneficial, in order to

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

514 reveal further evidence of admixture if short-distance pollen transfer and seed dispersal are in fact
515 important mechanisms for gene flow in these species. Ideally, future studies will examine *Strophostyles*
516 genetic diversity across multiple geographic and biological scales. The alignment of both species to the
517 *Phaseolus vulgaris* reference genome will also inherently constrain the amount of genetic variation that
518 can be detected, which can be improved by constructing a de novo genome within the *Strophostyles*
519 genus. Future work should also include the third species in the genus, *Strophostyles umbellata*, which has
520 a separate distribution from the sampling in this study and is known to be strongly perennial (Riley-
521 Hulting et al., 2004). There is also some morphological evidence for hybridization between *S. helvola* and
522 *S. umbellata* (Riley-Hulting et al., 2004).
523
524 Overall, this study reinforces the need to further investigate multiple aspects of *Strophostyles* life history
525 and population biology, particularly life span, mating system, and modes of dispersal, particularly
526 whether these traits differ among populations studied and the ecological correlates of these differences. A
527 precise study on the floral biology of *Strophostyles* species is also warranted, particularly on self-
528 compatibility, interspecific pollen compatibility, anatomical differences in flower morphology, and
529 pollinator interactions, which may all contribute to the observed patterns of genetic diversity and
530 admixture.
531
532 **CONCLUSIONS**
533
534 These data show for the first time the genomic diversity of *Strophostyles* species at the population level.
535 From this, we were able to confirm the genetic separation of *S. helvola* and *S. leiosperma* within the range
536 of this study, and we found complex patterns of genetic structure and diversity within both species. Mixed
537 patterns of admixture and heterozygosity call for a detailed assessment of the life history and reproductive
538 biology of both species, particularly mating system variation, as well as more fine-scale sampling
539 allowing for clarification of local population dynamics. Information gained here demonstrates that there is

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

540 an abundance of genetic diversity across the range of both *S. helvola* and *S. leiosperma*, and much of the
541 distribution of these species remains to be explored. Acquiring genetic information from non-model taxa
542 such as *Strophostyles* species will be critical in discovering novel genetic diversity related to
543 environmental tolerances and other adaptive features that can inform plant evolutionary ecology and crop
544 breeding.
545
546 **ACKNOWLEDGMENTS**
547
548 This research was funded by the Perennial Agriculture Project (Malone Family Land Preservation
549 Foundation and The Land Institute). S.A.H. was supported by a graduate research assistantship from the
550 Saint Louis University Department of Biology. M.J.R. was supported by the Donald Danforth Plant
551 Science Center and the Perennial Agriculture Project. We are grateful to the Department of Energy Joint
552 Genome Institute and collaborators for pre-publication access to the *Phaseolus vulgaris* v.2.1 genome.
553 We thank the following for their assistance in planning population sampling: Claudia Ciotir, Laura Klein,
554 and Joel Swift. We are grateful to the following who assisted in locating collection sites: Steven Cannon,
555 Laura Jackson, Justin Meissen, and Robert S. Wallace. For assistance with permitting and site
556 information, we are indebted to Malissa Briggler, Christopher Crabtree, Ian Hope, Steve Paes, Daryl
557 Smith, and James Trager. For assistance preparing samples and extracting DNA, we thank Niyati Bhakta,
558 Claudia Ciotir, Emma Frawley, Nathan Held, Aidan Leckie-Harre, and Lisa Millar. We thank Donna
559 Herrera at the Missouri Botanical Garden for assistance in depositing voucher specimens. We are grateful
560 to the Carrington, Eveland, Nusinow, Slotkin, and Topp labs (Donald Danforth Plant Science Center) for
561 providing necessary equipment and expertise in processing samples. We also wish to acknowledge the
562 work of REU student Marissa C. Sandoval, for her additional studies on *Strophostyles* germination and
563 herbarium morphological studies. Lastly, we appreciated the time and thoughtful feedback on previous
564 versions of this manuscript from the Miller Lab and from researchers Wendy Applequist, Peter Bernhardt,
565 Gerardo Camilo, Elizabeth Kellogg, Jason Knouft, and Chris Topp.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

566

567 **AUTHOR CONTRIBUTIONS**

568

569 S.A.H. and A.J.M. designed the study, and S.A.H. implemented the research and wrote the manuscript.
570 Z.N.H. and M.J.R. assisted in the computational analysis of genomic data and contributed to the writing
571 of the manuscript.

572

573 **DATA AVAILABILITY**

574

575 All genomic data are deposited in the Sequence Read Archive, and R code is available at the following
576 GitHub page. [Data is in preparation for submission to these permanent archives].

577

578 **SUPPORTING INFORMATION**

579

580 Additional supporting information may be found online in the Supporting Information section at the end
581 of the article.

582

583 **APPENDIX S1.** Excel spreadsheet containing metadata for all samples collected, including location
584 coordinates and details, and genetic quality metrics.

585

586 **APPENDIX S2.** Genetic structure plots with K=2 to K=7 clusters for *Strophostyles helvola* and *S.*
587 *leiosperma*.

588

589 **LITERATURE CITED**

590

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

591 Arnold, R. M. 1981. Population dynamics and seed dispersal of *Chaenorrhinum minus* on railroad cinder
592 ballast. *American Midland Naturalist* 106: 80–91.
593
594 Austerlitz, F., S. Mariette, N. Machon, P. H. Gouyon, and B. Godelle. 2000. Effects of colonization
595 processes on genetic diversity: differences between annual plants and tree species. *Genetics* 154: 1309–
596 1321.
597
598 Austin, K. C. 2003. Botanical processes in urban derelict spaces. Ph.D. dissertation, University of
599 Birmingham, Birmingham, UK.
600
601 Azani, N., M. Babineau, C. D. Bailey, H. Banks, A. R. Barbosa, B. Pinto, J. S. Boatwright, et al. 2017. A
602 new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny.
603 *Taxon* 66: 44–77.
604
605 Balfourier, F., G. Charmet, and C. Ravel. 1998. Genetic differentiation within and between natural
606 populations of perennial and annual ryegrass (*Lolium perenne* and *L. rigidum*). *Heredity* 81: 100–110.
607
608 Barrett, S. C. H., L. D. Harder, and A. C. Worley. 1996. The comparative biology of pollination and
609 mating in flowering plants. *Philosophical Transactions of the Royal Society of London B: Biological*
610 *Sciences* 351: 1271–1280.
611
612 Barrett, S. C. H., and L. D. Harder. 2017. The ecology of mating and its evolutionary consequences in
613 seed plants. *Annual Review of Ecology, Evolution, and Systematics* 48: 135–157.
614
615 Bedinger, P. A., A. K. Broz, A. Tovar-Mendez, and B. McClure. 2017. Pollen-pistil interactions and their
616 role in mate selection. *Plant Physiology* 173: 79–90.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

617

618 Bird, L. G., and R. D. Bird. 1931. Winter food of Oklahoma quail. *Wilson Bulletin* 43: 293–305.

619

620 Bushnell, B. 2014. BBMap: a fast, accurate, splice-aware aligner. No. LBNL-7065E. Lawrence Berkeley
621 National Lab, Berkeley, California, USA.

622

623 Butler, T. J., and J. P. Muir. 2010. 'Rio Rojo' smooth-seeded wild bean, a native annual forage legume.
624 *Journal of Plant Registrations* 4: 103–105.

625

626 Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, 2015. Second-
627 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: s13742-015.

628

629 Charnov, E. L., and W. M. Schaffer. 1973. Life-history consequences of natural selection: Cole's result
630 revisited. *American Naturalist* 107: 791–793.

631

632 Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. Handsaker, et al. 2011.
633 The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

634

635 Deam, C. C. 1940. Flora of Indiana. Wm. B. Burford Printing Co., Indianapolis, Indiana, USA. Also at
636 website https://swbiodiversity.org/.

637

638 Delgado-Salinas, A., M. Thulin, R. Pasquet, N. F. Weeden, and M. Lavin. 2011. *Vigna* (Leguminosae)
639 sensu lato: the names and identities of the American segregate genera. *American Journal of Botany* 98:
640 1694–1715.

641

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 Diggs, G. M., B. L. Lipscomb, and R. J. O'Kennon. 1999. Shinners & Mahler illustrated flora of North
643 Central Texas. Botanical Research Institute of Texas, Fort Worth, Texas, USA.
644
645 Duminil, J., S. Fineschi, A. Hampe, P. Jordano, D. Salvini, G. G. Vendramin, and R. J. Petit. 2007. Can
646 population genetic structure be predicted from life-history traits? *American Naturalist* 169: 662–672.
647
648 Duminil, J., O. J. Hardy, and R. J. Petit. 2009. Plant traits correlated with generation time directly affect
649 inbreeding depression and mating system and indirectly genetic structure. *BMC Evolutionary Biology* 9:
650 177.
651
652 Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011.
653 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:
654 e19379.
655
656 Erickson, D. L., and D. R. Young. 1995. Salinity response, distribution, and possible dispersal of a barrier
657 island strand glycophyte, *Strophostyles umbellata* (Fabaceae). *Bulletin of the Torrey Botanical Club* 122:
658 95–100.
659
660 Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric
661 distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*
662 131: 479–491.
663
664 Foerste, A. F. 1885. The fertilization of the wild bean (*Phaseolus diversifolius*). *American Naturalist* 19:
665 887–888.
666

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

667 Foster, J. L., J. P. Muir, W. C. Ellis, and B. D. Lambert. 2008. Nutritive evaluation of two legumes
668 (*Strophostyles*) supplemented to goats fed a high quality coastal bermudagrass (*Cynodon dactylon*) hay
669 diet. *Texas Journal of Agriculture and Natural Resources* 21: 73–81.
670
671 Francis, R. M. 2017. pophelper: an R package and web app to analyse and visualize population structure.
672 *Molecular Ecology Resources* 17: 27–32.
673
674 Friedman, J. 2020. The evolution of annual and perennial plant life histories: ecological correlates and
675 genetic mechanisms. *Annual Review of Ecology, Evolution, and Systematics* 51: 461–481.
676
677 Gamba, D., and N. Muchhala. 2020. Global patterns of population genetic differentiation in seed plants.
678 *Molecular Ecology* 29: 3413–3428.
679
680 Ge, X. J., M. H. Liu, W. K. Wang, B. A. Schaal, and T. Y. Chiang. 2005. Population structure of wild
681 bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP.
682 *Molecular Ecology* 14: 933–944.
683
684 Gee, K. L., M. D. Porter, S. Demarais, F. C. Bryant, and G. van Vreede. 2011. White-tailed deer: their
685 foods and management, 3rd ed. Samuel Roberts Noble Foundation, Ardmore, Oklahoma, USA.
686
687 Gilly, C., and McDonald, M. 1936. Rare and unusual plants from southeastern Iowa. *Proceedings of the*
688 *Iowa Academy of Science* 43: 143–149.
689
690 Hall, M. C., and J. H. Willis. 2006. Divergent selection on flowering time contributes to local adaptation
691 in *Mimulus guttatus* populations. *Evolution* 60: 2466.
692

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

693 Hamrick, J. L. 1982. Plant population genetics and evolution. *American Journal of Botany* 69: 1685–
694 1693.
695
696 Hamrick, J. L., and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species.
697 *Philosophical Transactions of the Royal Society B: Biological Sciences* 351: 1291–1298.
698
699 Hill, N. M., and C. S. Blaney. 2009. Exotic and invasive vascular plants of the Atlantic maritime ecozone.
700 *In* D. F. McAlpine and I. M. Smith [eds.], Assessment of species diversity in the Atlantic maritime
701 ecozone, 1–18. NRC Research Press, Ottawa, Canada.
702
703 Immel, D. L. 2001. Trailing fuzzybean: *Strophostyles helvula* (L.) Ell. USDA NRCS Plant Guide.
704 Website https://plants.usda.gov/plantguide/pdf/cs_sthe4.pdf.
705
706 Isely, D. 1998. Native and naturalized Leguminosae of the USA. Brigham Young University, Utah, USA.
707
708 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*
709 24: 1403–1405.
710
711 Kahle, D., and H. Wickham. 2013. ggmap: spatial visualization with ggplot2. *R Journal* 5: 144-161.
712
713 Krombein, K., P. Hurd, D. Smith, and B. Burks. 1979. Catalog of Hymenoptera in America north of
714 Mexico, vol. 2. Smithsonian Institution Press, Washington, DC, USA.
715
716 Lackey, J. A. 1980. Chromosome numbers in the Phaseoleae (Fabaceae: Faboideae) and their relation to
717 taxonomy. *American Journal of Botany* 67: 595–602.
718

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

719 Lewis, G., B. Schrire, B. Mackinder, and M. Lock [eds.], 2005. Legumes of the world. Royal Botanic
720 Gardens, Kew, UK.
721
722 Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform.
723 *Bioinformatics* 25:1754–1760.
724
725 Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting
726 population genetics and genomics programs. *Bioinformatics* 28: 298–299.
727
728 Loveless, M. D., and J. L. Hamrick. 1984. Ecological determinants of genetic structure in plant
729 populations. *Annual Review of Ecology and Systematics* 15: 65–95.
730
731 Lundgren, M. R., and D. L. Des Marais. 2020. Life history variation as a model for understanding trade-
732 offs in plant–environment interactions. *Current Biology* 30: R180–R189.
733
734 Ma, Y., S. C. H. Barrett, F. Y. Wang, J. C. Deng, and W. N. Bai. 2020. Do annual and perennial
735 populations of an insect-pollinated plant species differ in mating system? *Annals of Botany* mcaa178.
736
737 McGregor, R. L., T. M. Barkley, R. E. Brooks, and E. K. Schofield. 1986. Flora of the Great Plains.
738 University Press of Kansas, Lawrence, Kansas, USA.
739
740 Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation
741 measure. *Evolution* 60: 2399–2402.
742
743 Meirmans, P. G. 2020. Genodive version 3.0: easy-to-use software for the analysis of genetic data of
744 diploids and polyploids. *Molecular Ecology Resources* 20: 1126–1131.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

745

746 Monroe, J. G., B. Gill, K. G. Turner, and J. K. McKay. 2019. Drought regimens predict life history
747 strategies in *Heliophila*. *New Phytologist* 223: 2054–2062.
748
749 Morishima, H., and P. Barbier. 1990. Mating system and genetic structure of natural populations of wild
750 rice *Oryza rufipogon*. *Plant Species Biology* 5: 31–39.
751
752 Muir, J. P., R. L. Reed, and D. P. Malinowski. 2005. Impact of defoliation on herbage and seed
753 production of *Strophostyles helvula* and *S. leiosperma*. *Native Plants Journal* 6: 123–130.
754
755 Mühlenbach, V. 1979. Contributions to the synanthropic (adventive) flora of the railroads in St. Louis,
756 Missouri, U.S.A. *Annals of the Missouri Botanical Garden* 66: 1–108.
757
758 Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, New York, USA.
759
760 Parker, K. E. 1991. Sponemann Phase archaeobotany. *In* A.C. Fortier, T.O. Maher, and J.A. Williams
761 [eds.], The Sponemann site: the formative emergent Mississippian Sponemann Phase occupations, 377–
762 419. University of Illinois Press, Urbana, Illinois, USA.
763
764 Pelotto, J. P., and M. A. del Pero Martinez. 1998. Flavonoids in *Strophostyles* species and the related
765 genus *Dolichopsis* (Phaseolinae, Fabaceae): distribution and phylogenetic significance. *Sida* 18: 213–222.
766
767 Pepoon, H. S. 1927. Catalog of all plants - native, naturalized and introduced of the fern and seed plants
768 found growing spontaneously in the Chicago area. *In* An annotated flora of the Chicago area. Chicago
769 Academy of Science, Chicago, Illinois, USA.
770

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

771 Peterson, M. L., K. M. Kay, and A. L. Angert. 2016. The scale of local adaptation in *Mimulus guttatus*:
772 comparing life history races, ecotypes, and populations. *New Phytologist* 211: 345–356.
773
774 Poland, J. A., P. J. Brown, M. E. Sorrells, and J. Jannink. 2012. Development of high-density genetic
775 maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:
776 e32253.
777
778 R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical
779 Computing, Vienna, Austria. Website https://www.R-project.org/.
780
781 Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: variational inference of population
782 structure in large SNP data sets. *Genetics* 197: 573–589.
783
784 Reisch, C., and M. Bernhardt-Römermann. 2014. The impact of study design and life history traits on
785 genetic variation of plants determined with AFLPs. *Plant Ecology* 215: 1493–1511.
786
787 Riley-Hulting, E. T., A. Delgado-Salinas, and M. Lavin. 2004. Phylogenetic systematics of *Strophostyles*
788 (Fabaceae): a North American temperate genus within a neotropical diversification. *Systematic Botany*
789 29: 627–653.
790
791 Robertson, C. 1890. Flowers and insects. V. *University of Chicago Press* 15: 199–204.
792
793 Savary, P. 2020. graph4lg: build graphs for landscape genetics analysis. R package version 1.0.1. Website
794 https://CRAN.R-project.org/package=graph4lg.
795

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

796 Schmutz, J., P. E. McClean, S. Mamidi, G. A. Wu, S. B. Cannon, J. Grimwood, J. Jenkins, et al. 2014. A
797 reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*
798 46: 707–713.
799
800 Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN, version 2.000. A software for population
801 genetics data analysis. FAQ list. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
802 Website http://cmpg.unibe.ch/software/arlequin/software/2.000/doc/faq/faqlist.htm.
803
804 St. Louis, Iron Mountain, and Southern Railway Company. 1878. Map of the St. Louis, Iron Mountain
805 and Southern Railway and connections. University of North Texas Libraries, Denton, Texas, USA.
806 Website https://texashistory.unt.edu/ark:/67531/metapth220495/m1/1/.
807
808 St. Onge, K. R., T. Källman, T. Slotte, M. Lascoux, and A. E. Palmé. 2011. Contrasting demographic
809 history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species
810 with different mating systems. *Molecular Ecology* 20: 3306–3320.
811
812 Stepenoff, B. 1993. Butler County, Missouri historical and architectural survey. History and Archives
813 Consultants for Ozark Foothills Regional Planning Commission. Website
814 https://dnr.mo.gov/shpo/survey/BUAS002-R.pdf.
815
816 Stilwell, K. L., H. M. Wilbur, C. R. Werth, and D. R. Taylor. 2003. Heterozygote advantage in the
817 American chestnut, *Castanea dentata* (Fagaceae). *American Journal of Botany* 90: 207–213.
818
819 Stoeckel, S., J. Grange, J. F. Fernández-Manjarres, I. Bilger, N. Frascaria-Lacoste, and S. Mariette. 2006.
820 Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L.
821 *Molecular Ecology* 15: 2109–2118.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

822

823 Stubbendieck, J., and E. C. Conard. 1989. Common legumes of the Great Plains, 1st ed. University of

824 Nebraska Press, Lincoln, Nebraska, USA.

825

826 Tsang, A., and M. A. Maun. 1999. Mycorrhizal fungi increase salt tolerance of *Strophostyles helvola* in

827 coastal foredunes. *Plant Ecology* 144: 159–166.

828

829 Twyford, A. D., E. L. Y. Wong, and J. Friedman. 2020. Multi-level patterns of genetic structure and

830 isolation by distance in the widespread plant *Mimulus guttatus*. *Heredity* 125: 227–239.

831

832 USDA-NRCS. 2018. The PLANTS Database. National Plant Data Team, Greensboro, North Carolina,

833 USA. Website http://plants.usda.gov.

834

835 Vest, K. 2019. Diversity across a latitudinal cline: seasonal timing traits and a cryptic speciation event.

836 M.S. thesis. Binghamton University.

837

838 Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure.

839 *Evolution* 38: 1358–1370.

840

841 Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, New York,

842 USA. Website https://ggplot2.tidyverse.org.

843

844 Wiseman, D. S. 1977. Food habits and weights of bobwhite from northeastern Oklahoma tall grass

845 prairie. *Proceedings of the Oklahoma Academy of Science* 57: 110–115.

846

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

847 Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems
848 of mating. *Evolution* 19: 395–420.
849
850 Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, et al. 2020.
851 Ensembl 2020. *Nucleic Acids Research* 48: D682–D688.
852
853 Yatskievych, G. 2013. Steyermark's flora of Missouri, vol. 3. Missouri Botanical Garden Press, St. Louis,
854 Missouri, USA.
855
856 Zhang, H., C. Zuelsdorf, D. Penneys, S. Fan, J. Kofsky, and B. H. Song. 2018. Transcriptome profiling of
857 a beach-adapted wild legume for dissecting novel mechanisms of salinity tolerance. *Scientific Data* 5:
858 180290.
859
860 Zuelsdorf, C. 2018. Understanding salt adaptation in sand beans (*Strophostyles helvola*). M.S. thesis,
861 University of North Carolina Charlotte, Charlotte, North Carolina. USA.

862 **Table 1.** Summary of population sampling for *Strophostyles helvola* and *S. leiosperma* across sites. *S. leiosperma* was not present at sites SP-MO
863 and AR. Sites are sorted by descending latitude. Lat/long coordinates are the approximate centroid of the primary area of sampling at each site
864 (average of coordinates of all samples), and the elevation is the rounded average among all samples in meters above sea level (masl). Individual
865 sample information is available in Appendix S1.

Location(s)	**Site ID**	**Lat/long**	**Elevation (masl)**	**Species**	**Individuals sequenced**	**Voucher Collection No.**
Cedar Hills Sand Prairie, Cedar Falls, Iowa, USA	IA	42.59764 -92.55197	284	*Strophostyles helvola* *Strophostyles leiosperma*	20 20	SH 3.8; SH 3.28 SH 3.10; SH 3.25
Frost Island Conservation Area & Iliniwek Village State Historic Site, Wayland, Missouri, USA	IV-MO [a]	40.42992 -91.55692	167	*Strophostyles helvola* *Strophostyles leiosperma*	21 17	SH 2.7; SH 2.32 SH 2.6; SH 2.31
Shaw Nature Reserve, Gray Summit, Missouri, USA	SNR-MO	38.48327 -90.81267	183	*Strophostyles helvola* *Strophostyles leiosperma*	21 19	SH 1.30; SH 1.39 SH 1.28; SH 1.33
Sand Pond Conservation Area, Neelyville, Missouri, USA	SP-MO	36.50350 -90.59622	93	*Strophostyles helvola*	19	SH 4.20; SH 4.28
Burns Park, North Little Rock, Arkansas, USA	AR [b]	34.79084 -92.32575	75	*Strophostyles helvola*	29	SH 7.21; SH 7.22

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

866 [a] For IV-MO, the approximate centroid and elevation of sampling from Iliniwek Village State Historic Site was used for this table, since that is
867 where the majority of samples were collected from, while five samples were collected approximately 3 km from there at Frost Island Conservation
868 Area.
869 [b] For AR, the approximate centroid and elevation of sampling from Burns Park was used for this table, since that is where the majority of samples
870 were collected from, while four samples were collected approximately 10 km from there in a gravel ditch in Little Rock, Arkansas.
871
872
873
874
875
876
877
878
879
880
881
882
883

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

884 **Table 2.** Analysis of molecular variance (AMOVA) results from eight populations with 5556 total SNPS for both *Strophostyles helvola* and *S.*
885 *leiosperma*. Here we decompose the genetic variation for individuals nested within populations, populations nested within species, among species,
886 and among individuals. *P*-values are derived from 1000 permutations and thus have a lower threshold of 0.001. The unnested F_{IT} terms is not
887 interpretable and thus not assigned a *P*-value.

Source of variation	**Nested in**	**% Variation**	***F*-stat**	***F*-value**	***P*-value**
Among individuals	Population	-0.1 [a]	F_{IS}	-0.003 [a]	0.937
Among populations	Species	17.7	F_{SC}	0.471	0.001
Among species	---	62.4	F_{CT}	0.624	0.001
Among individuals	---	19.9	F_{IT}	0.801	---

888 [a] Slightly negative values for percent variation and *F* statistics can occur due to a lack of significant genetic structure; since our results show a lack
889 of significance for this source of variation, we interpret these values as equivalent to zero (Schneider et al. 2000; Meirmans 2006).
890
891
892
893
894
895

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

896 **Table 3.** Pairwise F_{ST} values between all combinations of populations within species, including five populations of *Strophostyles helvola*,
897 and three populations of *S. leiosperma*, based on 5556 biallelic SNPs. Populations are arranged from north to south latitudes, first with *S.*
898 *helvola*, then *S. leiosperma*. F_{ST} values are shown below the diagonal, and the *P*-values for the pairwise F_{ST} are above the diagonal. *P*-
899 values are derived from 1000 permutations and thus have a lower threshold of 0.001.

Species	Population	IA	IV-MO	SNR-MO	SP-MO	AR
Strophostyles helvola	IA	---	*0.001	*0.001	*0.001	*0.001
	IV-MO	0.284	---	*0.001	*0.001	*0.001
	SNR-MO	0.412	0.359	---	*0.001	*0.001
	SP-MO	0.383	0.260	0.427	---	*0.001
	AR	0.405	0.277	0.468	0.294	---
Strophostyles leiosperma	IA	---	*0.001	*0.001		
	IV-MO	0.251	---	*0.001		
	SNR-MO	0.712	0.707	---		

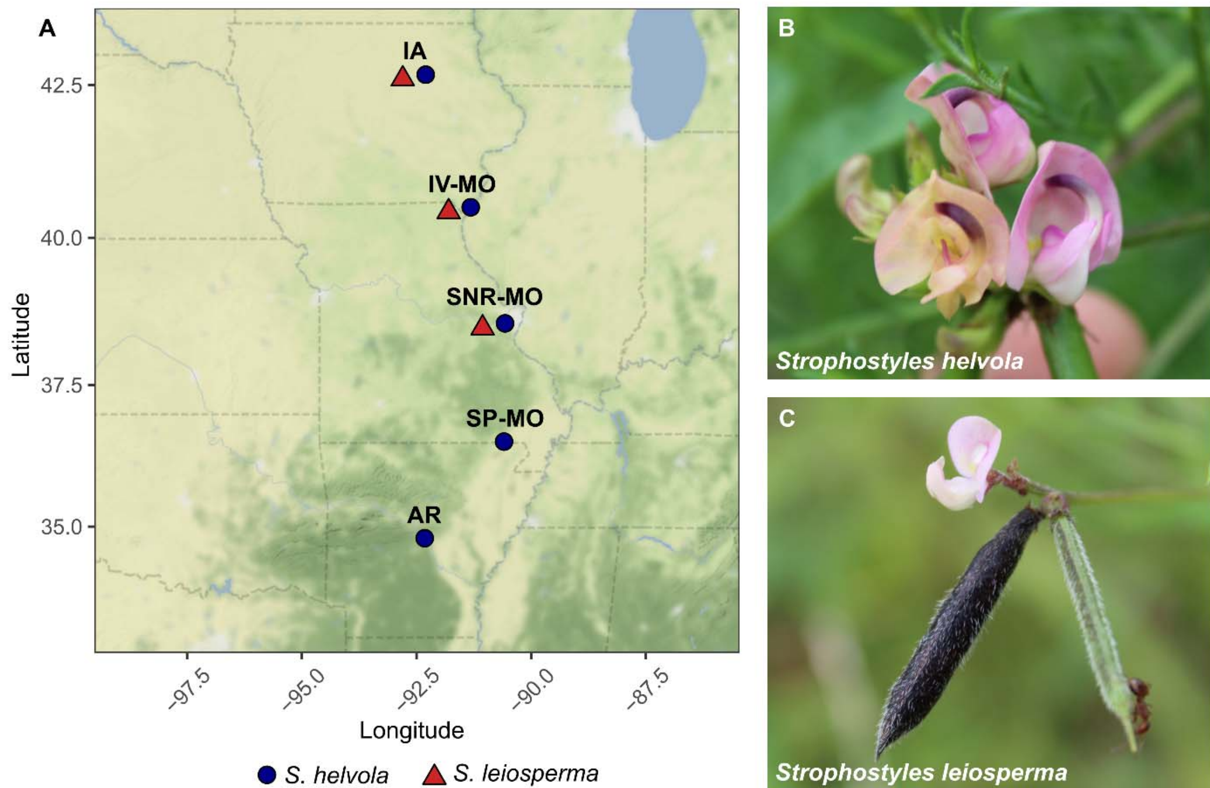
900 * $\alpha < 0.0038$ significance after a Bonferroni correction for multiple tests (13).
901
902
903
904

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

905 **Table 4.** Diversity metrics for each population of *Strophostyles helvola* and *S. leiosperma*. H_O is the observed heterozygosity, H_S is within-
906 population expected heterozygosity (Nei's gene diversity), F_{IS} is the inbreeding coefficient (from an AMOVA), and the HWE *P*-value is for
907 deviation of genotypes from Hardy-Weinberg equilibrium (random mating), where a significant value ($P < 0.05$) indicates significant deviation. *P*-
908 values are derived from 1000 permutations and thus have a lower threshold of 0.001.

Species	Population	H_O	H_S	F_{IS}	HWE *P*-value
Strophostyles helvola	IA	0.109	0.096	-0.133	0.001
	IV-MO	0.099	0.125	0.205	0.001
	SNR-MO	0.087	0.086	-0.019	0.003
	SP-MO	0.091	0.110	0.176	0.001
	AR	0.091	0.102	0.111	0.001
Strophostyles leiosperma	IA	0.108	0.078	-0.387	0.001
	IV-MO	0.100	0.076	-0.310	0.001
	SNR-MO	0.101	0.101	0.000	0.487

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



909

910 **Figure 1.** Diagram of population sampling and photographs of *Strophostyles helvola* and *S. leiosperma*.
911 (A) Shows a map of populations sampled from sites across Iowa, Missouri, and Arkansas, USA. *S.*
912 *helvola* populations are represented by a blue circle and *S. leiosperma* populations are represented by a
913 red triangle. Where the species co-occur at the three northernmost sites, the points are spatially separated
914 for visualization, although they occur at the same coordinates. (B) shows an inflorescence of *S. helvola* at
915 population SNR-MO, and (C) shows an inflorescence of *S. leiosperma* with a flower, unripe pod, and ripe
916 pod, at SNR-MO (not to scale). Note the difference in keel morphology and pubescence on the *S.*
917 *leiosperma* pod. Photo credit: S.A.H.

918

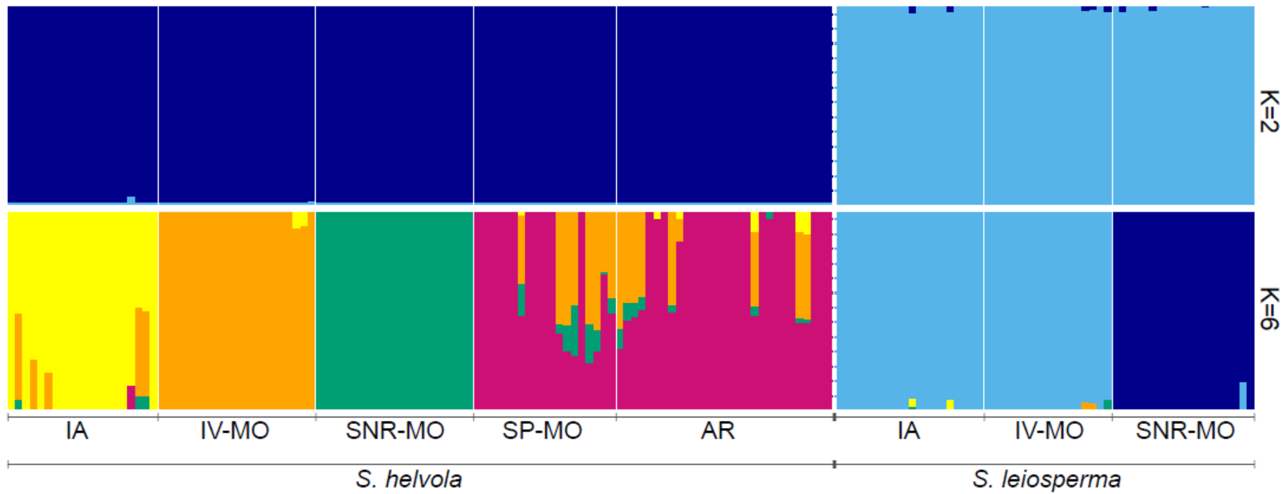
919

920

921

922

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

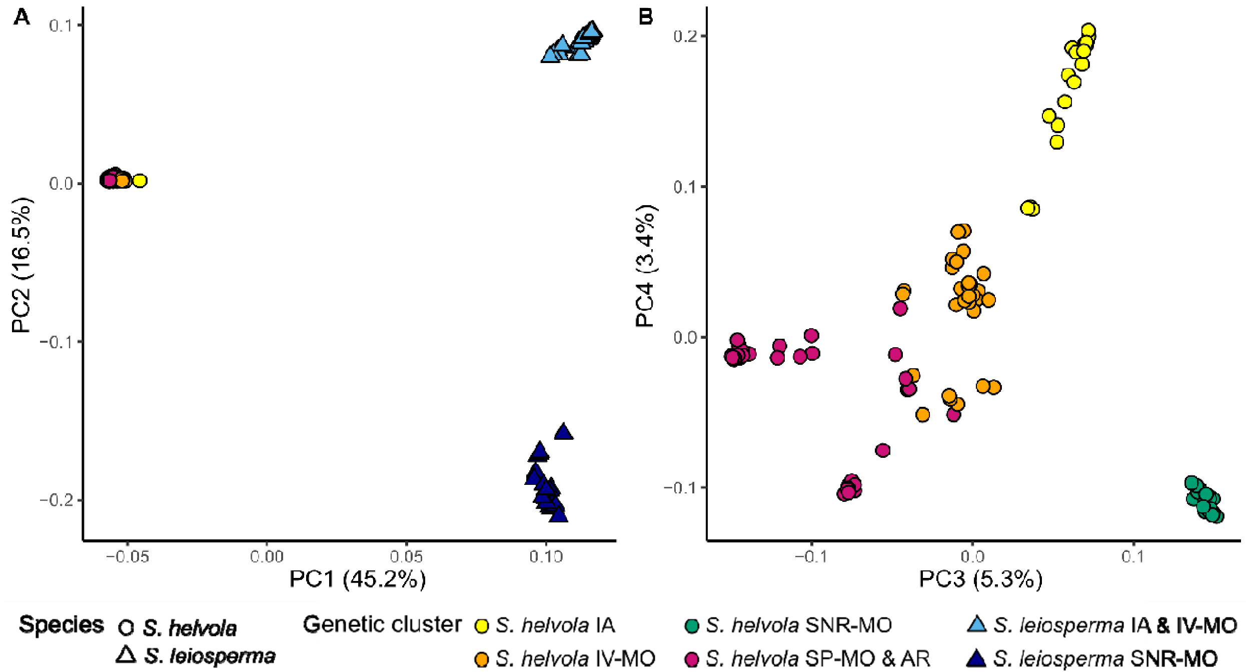


923

924 **Figure 2.** Genetic structure plots with K=2 and K=6 clusters for *Strophostyles helvola* and *S. leiosperma*;
925 K=2 separated the species, and K=6 best explained overall genetic structure in the data. Populations,
926 labeled below the bars, are ordered north to south within species (left to right) and are separated by thin
927 white solid lines; species are separated by a thick white dashed line. Each vertical bar represents a single
928 individual from each population. Each color signifies a unique genetic cluster assignment, with multiple
929 colors within an individual signifying mixed ancestry.

930
931
932
933
934
935
936
937
938
939
940
941

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.08.447599; this version posted June 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



942

943 **Figure 3.** Principal component analysis of SNP data for all populations of *Strophostyles helvola* and *S.*
944 *leiosperma*, where each point represents a single individual; point fill color matches the predominant
945 genetic cluster assignment from Fig. 2 and point shape corresponds to species. Displayed are (A) PC1 and
946 PC2 (both species), and (B) PC3 and PC4 (just *S. helvola*). *S. leiosperma* is omitted from panel (B) due to
947 nonsignificant genetic variation among populations for PC3 and PC4, and in order to better visualize *S.*
948 *helvola* variation.

949

950

951