1 TITLE: Target enrichment and extensive population sampling help untangle the recent, rapid
2 radiation of *Oenothera* sect. *Calylophus*
3
4 RUNNING HEAD: Cooper et al., Phylogenomics of *Oenothera* sect. *Calylophus*
5
6 AUTHORS:
7
8 Benjamin J. Cooper[1,2,*], Michael J. Moore[3], Norman A. Douglas[4], Warren L. Wagner[5], Matthew
9 G. Johnson[1,6], Rick P. Overson[1,7], Angela J. McDonnell[1], Rachel A. Levin[8], Robert A. Raguso[9],
10 Hilda Flores Olvera[10], Helga Ochoterena[10], Jeremie B. Fant[1,2], Krissa A. Skogen[1,2], Norman J.
11 Wickett[1,2,*]
12
13 [1] The Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden,
14   1000 Lake Cook Rd., Glencoe, IL 60022, USA
15 [2] Northwestern University, Program in Plant Biology and Conservation, O.T. Hogan Hall, Room
16   6-140B, 2205 Tech Drive, Evanston, IL 60208, USA
17 [3] Oberlin College, Department of Biology, 119 Woodland St., Oberlin, OH 44074, USA
18 [4] Department of Biology, University of Florida, Gainesville, FL 32611, USA
19 [5] Department of Botany, MRC-166, Smithsonian Institution, PO Box 37012, Washington, DC
20   20013-7012, USA
21 [6] Department of Biological Sciences, Texas Tech University, Box 43131 Lubbock, TX 79409,
22   USA
23 [7] School of Sustainability, Arizona State University, PO Box 875502, Tempe, AZ 85287-5502,
24   USA
25 [8] Department of Biology, Amherst College, 25 East Drive, Amherst, MA, 01002, USA
26 [9] Department of Neurobiology & Behavior, Cornell University, 215 Tower Road, Ithaca, NY
27   14853, USA
28 [10] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México,
29   Mexico City, Mexico
30
31 *Authors for correspondence: nwickett@chicagobotanic.org, benjamin.cooper06@gmail.com

32   ABSTRACT

33   *Oenothera* sect. *Calylophus* is a North American group of 13 recognized taxa in the

34   evening primrose family (Onagraceae) with an evolutionary history that may include

35   independent origins of bee pollination, edaphic endemism, and permanent translocation

36   heterozygosity. Like other groups that radiated relatively recently and rapidly, taxon boundaries

37   within *Oenothera* sect. *Calylophus* have remained challenging to circumscribe. In this study, we

38   used target enrichment, flanking non-coding regions, summary coalescent methods, tests for

39   gene flow modified for target-enrichment data, and morphometric analysis to reconstruct

40   phylogenetic hypotheses, evaluate current taxon circumscriptions, and examine character

41   evolution in *Oenothera* sect. *Calylophus*. Because sect. *Calylophus* comprises a clade with a

42   relatively restricted geographic range, we were able to extensively sample across the range of

43   geographic and morphological diversity in the group. We found that the combination of exons

44   and flanking non-coding regions led to improved support for species relationships. We

45   reconstructed potential hybrid origins of some accessions and note that if processes such as

46   hybridization are not taken into account, the number of inferred evolutionary transitions may be

47   artificially inflated. We recovered strong evidence for multiple origins of the evolution of bee

48   pollination from ancestral hawkmoth pollination, the evolution of edaphic specialization on

49   gypsum, and permanent translocation heterozygosity. This study applies newly emerging

50   techniques alongside dense infraspecific sampling and morphological analyses to effectively

51   address a relatively common but recalcitrant problem in evolutionary biology.

52

53   *Keywords*.– Gypsum Endemism, Onagraceae, *Oenothera* sect. *Calylophus*, Pollinator Shift,

54   Recent Radiation, Phylogenomics, Target Enrichment

2

55

56   INTRODUCTION

57   The challenges of reconstructing species histories for groups that arose through recent,

58   rapid radiations are well established. Phylogenetic signal can be obscured by processes such as

59   incomplete lineage sorting (ILS) and gene flow (Maddison and Knowles 2006; Knowles and

60   Chan 2008; Christie and Knowles 2015), resulting in short branch lengths and conflicting gene

61   tree topologies. Consequently, approaches that use few loci or concatenation may fail to resolve

62   the most accurate species tree (Eckert and Carstens 2008; Leaché et al. 2014; Xi et al. 2014;

63   Giarla and Esselstyn 2015). This may be particularly common in plants that are thought to have

64   experienced rapid or recent radiation with ongoing hybridization and high levels of ILS. The

65   application of target enrichment methods that efficiently sequence hundreds of nuclear loci,

66   coalescent-based phylogenetic methods that account for ILS and gene flow, and extensive

67   sampling of morphologically diverse populations across the geographic range should allow for

68   more accurate representations of phylogenetic relationships (Maddison and Knowles 2006;

69   Knowles and Chan 2008; Knowles 2009; Mamanova et al. 2010; Lemmon et al. 2012; Straub et

70   al. 2012; Bryson et al. 2014; Weitemier et al. 2014; Mandel et al. 2014; Stephens et al. 2015;

71   Johnson et al. 2016).

72   *Oenothera* sect. *Calylophus* currently comprises seven species (thirteen taxa) with a

73   center of diversity in western Texas, southern New Mexico, and north-central Mexico (Fig. 1;

74   Towner 1977; Turner and Moore 2014; Wagner 2021). Previous analyses suggest that *Oenothera*

75   sect. *Calylophus* forms a well-supported, morphologically coherent clade with a relatively

76   restricted geographic range (Towner 1977; Levin et al. 2004; Wagner et al. 2007; Turner and

77   Moore 2014; Wagner 2021). However, as with other groups that have experienced rapid

3

78   radiations, taxon boundaries within *Oenothera* sect. *Calylophus* have been challenging to define,

79   likely due to phenomena such as overlapping morphological boundaries, ongoing introgression,
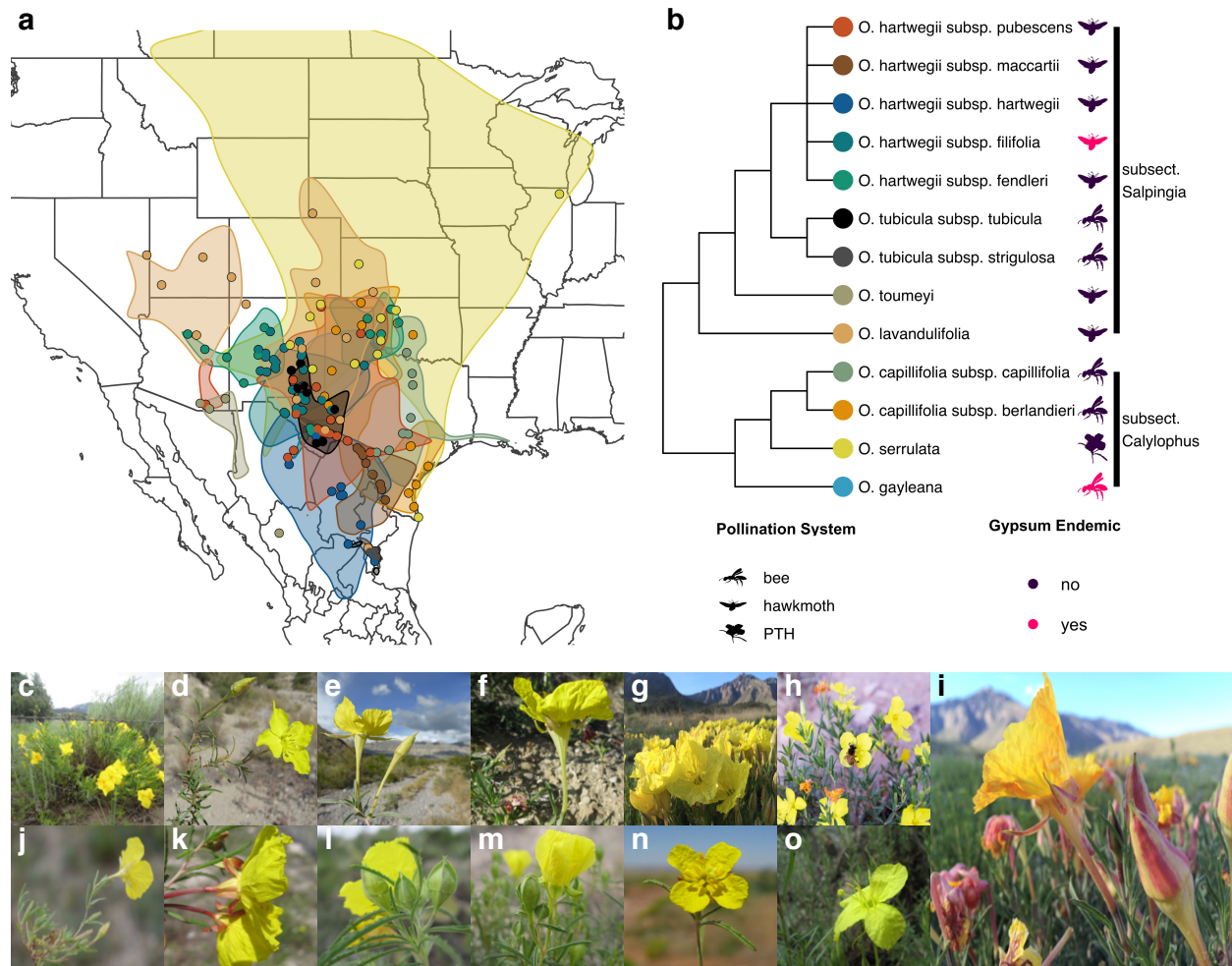
80   and incomplete lineage sorting.

Figure 1

81   In the most comprehensive study of the group to date, Towner (1977) circumscribed taxa

82   using morphology, breeding system, geography, and ecology, but it was noted (and our field

83   observations confirm) that characters often overlap among taxa (Towner 1977). Taxa within

84   *Oenothera* sect. *Calylophus* are divided into two easily recognizable subsections: subsect.

85   *Salpingia* and subsect. *Calylophus* (Towner 1977; Wagner et al. 2007). Pollination varies

4

86 between the two subsections; flowers of subsect. *Salpingia* are adapted to vespertine pollination

87 by hawkmoths, except for *O. tubicula*, which opens in the morning and is primarily pollinated by

88 bees (Towner 1977), while taxa in subsect. *Calylophus* are predominantly bee-pollinated

89 (Towner 1977). Taxa in both subsections have geographic ranges that partially (or even largely)

90 overlap, resulting in occasional morphologically intermediate populations (Towner 1977).

91 Although confounding for morphological-based analyses, this observed pattern of reticulation is

92 consistent with a recent, rapid radiation occurring in parallel with climatic fluctuations and

93 increasing aridity in the region since the Pleistocene (Raven 1964; Towner 1977; Nason et al.

94 2002; Katinas et al. 2004). Hawkmoth pollination, which is ancestral in Onagraceae and common

95 in *Oenothera* sect. *Calylophus*, is known to result in long-distance pollen movement (Stockhouse

96 1973; Skogen et al. 2019). Therefore, gene flow may have been extensive over the evolutionary

97 history of hawkmoth-pollinated taxa, increasing the chances that processes such as historical

98 introgression may obscure phylogenetic signal in extant plants (Elrich and Raven 1969). With a

99 phylogenomic approach that samples hundreds of nuclear loci, we may better illuminate both the

100 history of these species and the key evolutionary processes related to speciation in this group.

101   Understanding speciation in angiosperms remains a fundamental question in evolutionary

102 biology (Barrett et al. 1996; Rajakaruna 2004; van der Niet et al. 2006; Wilson et al. 2007;

103 Crepet and Niklas 2009; Peakall et al. 2010; Xu et al. 2011; Van der Niet and Johnson 2012;

104 Boberg et al. 2014). *Oenothera* section *Calylophus* has an evolutionary history that likely

105 involves changes in reproductive system (pollinator functional group, breeding system) and

106 edaphic endemism. For example, there are thought to be two independent shifts between

107 pollinators from hawkmoth to bee pollination (Towner 1977; Fig. 1b), despite many studies in

108 other plant groups showing a directional bias in shifts from bee to hummingbird or hawkmoth

109   pollination (Barrett et al. 1996; Wilson et al. 2007; Thomson and Wilson 2008; Tripp and Manos

110   2008; Barrett 2013). However, pollinator shifts that do not follow this sequence may be more

111   likely when the extent of trait divergence and specialization does not completely inhibit

112   secondary pollinators such as bees, as has been suggested in *Oenothera* sect. *Calylophus*

113   (Stebbins 1970; Tripp and Manos 2008; Van Der Niet et al. 2014). Shifts to autogamy are also

114   frequent across angiosperms and in Onagraceae alone there are an estimated 353 shifts to modal

115   autogamy (i.e. mostly autogamous; Raven 1979). *Oenothera* sect. *Calylophus* also includes at

116   least one autogamous species, *O. serrulata*, which exhibits permanent translocation

117   heterozygosity, a phenomenon in which all chromosomes are translocationally heterozygous

118   (PTH; Towner 1977) (Fig. 1b). While the evolution of PTH has been assessed in molecular

119   phylogenetic analyses across Onagraceae (Johnson et al. 2009; Hollister et al. 2019), no study to

120   date has examined this transition in a well-sampled clade with extensive population sampling.

121   Lastly, abiotic ecological factors such as edaphic specialization are also known to drive

122   speciation in some groups (Rajakaruna 2004; van der Niet et al. 2006), including in *Oenothera*

123   sect. *Pachylophus* (Patsis et al. in press). For example, serpentine endemics represent ~10% of

124   the endemic flora in California even though serpentine soils account for about 1% of terrestrial

125   habitat in the state (Brady et al. 2005). Similarly, the Chihuahuan Desert is comprised of

126   numerous isolated islands of gypsum outcrops and current estimates suggest that at least 235 taxa

127   from 36 different plant families are gypsum endemics (Moore and Jansen 2007; Moore et al.

128   2014). It is suspected that gypsum endemism has also evolved independently in *Oenothera* sect.

129   *Calylophus* at least twice (Towner 1977; Turner and Moore 2014; Fig. 1b). Ultimately, to

130   understand the role that these transitions have played in shaping the diversity of *Oenothera* sect.

131   *Calylophus*, a robust phylogeny is required.

132   Here, we use target enrichment, summary coalescent methods, and morphometric

133 analyses to reconstruct a phylogenetic relationships, leveraging this information to re-examine

134 previous taxonomic concepts, test for instances of hybridization, and resolve the history of

135 pollinator shifts, PTH, and gypsum endemism in *Oenothera* sect. *Calylophus*. Target enrichment

136 is a cost-effective method for sequencing hundreds of loci across a high volume of samples,

137 producing highly informative datasets for phylogenetics (Lemmon et al. 2012; Straub et al. 2012;

138 Mandel et al. 2014; Weitemier et al. 2014; Heyduk et al. 2015; Stephens et al. 2015; Johnson et

139 al. 2016). While target enrichment is generally designed to capture coding regions, a significant

140 proportion of flanking non-coding regions can be recovered (the "splash-zone"; Weitemier et al.

141 2014). The inclusion of non-coding regions may be particularly informative for recent radiations,

142 since these regions are less constrained by selective pressures and may contain on average more

143 informative sites at shallower time scales (Folk et al. 2015). We included these flanking non-

144 coding regions in our sequence alignments to evaluate their impact on reconstructing lower-order

145 relationships. Importantly, we sampled extensively, including individuals from numerous

146 populations across the geographic and morphological ranges of all thirteen taxa in the section

147 (Fig. 1a). This study presents an example of how combining these molecular techniques with

148 dense sampling and morphological analysis can be used to effectively address a common

149 problem in evolutionary biology.

150

151 MATERIALS AND METHODS

152   A total of 194 individuals spanning the geographic, morphological, and ecological ranges

153 of all 13 recognized taxa in *Oenothera* sect. *Calylophus* [following Towner (1977) and Turner

154 and Moore (2014)] were included in this study (Fig. 1a, S1, S2) along with 8 outgroups

155   representing other major sections of *Oenothera* (*Eremia, Gaura, Kneiffia, Lavauxia, Oenothera,*

156   *Pachylophus*, and *Ravenia*) and other genera (*Chylismia* and *Eulobus*) in Onagraceae (S1, S2).

157   DNA was extracted from fresh, silica-dried leaf tissue (S3). PTH status was confirmed or

158   reassessed for individuals in subsect. *Calylophus* by assessing pollen fertility, when flowers were

159   present, using a modified Alexander stain (Alexander 1969, 1980; S3).

160   Target nuclear loci for enrichment were determined by clustering transcriptome

161   assemblies of *Oenothera serrulata* (1KP accession *SJAN*) and *Oenothera capillifolia* subsp.

162   *berlandieri* (1KP accession *EQYT*). Starting with the 956 phylogenetically informative

163   *Arabidopsis* loci identified by Duarte et al. (2010; S3), we identified 322 homologous, single-

164   copy loci in our clusters and used these in the probe design process. Libraries were enriched for

165   these loci using the MyBaits protocol (Arbor Biosciences, Ann Arbor, MI, USA) and sequenced

166   on an Illumina MiSeq (2 x 300 cycles, v3 chemistry; Illumina, Inc., San Diego, California,

167   USA). Raw reads have been deposited at the NCBI Sequence Read Archive (BioProject

168   PRJNA544074; See S3 for details). Reads were trimmed using Trimmomatic (Bolger et al. 2014;

169   S3) and trimmed, quality-filtered reads were assembled using HybPiper (Johnson et al. 2016).

170   From the assembled loci, we produced two datasets: **"exons",** constiting of exon-only

171   alignments, and **"supercontig",** consisting of alignments containing both the exon alignment and

172   flanking non-coding regions (the "splash-zone" per Weitemier 2014 and reconstructed using

173   supercontigs produced by HybPiper). We used these two datasets to test the most recent

174   taxonomic circumscription of the group with several methods: (1) phylogenetic inference of

175   concatenated alignments (two analyses: exons and supercontigs) using RAxML (Stamatakis

176   2014), (2) ASTRAL-II (Mirarab and Warnow 2015; Sayyari and Mirarab 2016) species tree

177   inference (two analyses: exons and supercontigs), (3) SVD Quartets (Chifman and Kubatko

8

178   2014, 2015) species tree inference (one analysis: supercontigs), (4) Phyparts (Smith et al. 2015;

179   one analysis: supercontigs), (5) IQtree (Minh et al. 2018) with both gene and site concordance

180   factors (one analysis: supercontigs).

181   We used HyDe (Blischak et al. 2018) to test for putative hybrid origins of selected taxa

182   and accessions by calculating D-Statistics (Green et al. 2010) for a set of hypotheses (S3). To

183   further characterize population-level processes or genetic structure within sect. *Calylophus*, we

184   extracted and filtered SNPs by mapping individual reads against reference supercontigs (see

185   https://github.com/lindsawi/HybSeq-SNP-Extraction) and used Discriminant Analysis of

186   Principal Components (Jombart et al. 2010) as implemented in the R package *adegenet* (Jombart

187   2008) and the snmf function in the LEA package (Frichot and François 2015) in R (R Core

188   Team, 2020; S3).

189   We evaluated current taxonomic concepts and patterns of morphological variation by

190   measuring character states for vegetative and floral structures that have been used historically to

191   discriminate taxa in *Oenothera* sect. *Calylophus* (Towner 1977): plant height, leaf length (distal),

192   leaf width (distal), leaf length/width ratio (distal), leaf length (basal), leaf width (basal), leaf

193   length/width (basal), sepal length, and sepal tip length (S3). Measurements were made for 125 of

194   the sequenced samples (S11); we were unable to measure all traits for 73 samples because we did

195   not have access to the herbarium vouchers, or the trait of interest was not captured on the

196   voucher, therefore some samples were dropped from the analysis due to missing values. Finally,

197   the number of transitions and inferred ancestral conditions of reproductive system were mapped

198   onto an ASTRAL species tree, with individuals grouped into species, using the stochastic

199   mapping function in the R package *phangorn* version 2.5.5 (Schliep 2011; S3).

200

201   RESULTS AND DISCUSSION

202   *Sequencing and Phylogenetic Results*

203   Sequencing resulted in a total of 80,273,296 pairs of 300-bp reads with an average of

204   625,323 reads per sample. Following quality filtering, assembly and alignment, we recovered

205   204 loci that were present in at least 70% of the samples. Across all datasets and analyses,

206   *Oenothera* sect. *Calylophus* was monophyletic. At the subsection level there was strong

207   agreement in topology between concatenation and coalescent-based trees. For example, subsect.

208   *Calylophus* was recovered as sister to *Oenothera* subsect. *Salpingia* (minus *O. toumeyi*) with

209   strong support across all analyses and *O. toumeyi* [considered by Towner (1977) to be in subsect.

210   *Salpingia*; Fig. 2, S4-7] was recovered as sister to subsect. *Calylophus* across all trees, with

211   strong bootstrap support. Within subsect. *Calylophus* there was poor resolution for currently

212   recognized taxa in all analyses, whereas taxon relationships were better resolved in subsect.

213   *Salpingia* (Fig. 2, S4-7). With coalescent-based tree reconstruction, most taxa *sensu* Towner

214   (1977) were recovered as monophyletic with moderate to strong support (Fig. 2, S6-8). In

215   contrast, both the exon and supercontig concatenation trees recovered most currently recognized

216   taxa as non-monophyletic (S4, S5). Given that concatenation has been shown to produce

217   incorrect topologies in the presence of high ILS (Roch and Steel 2015) and that *Oenothera* sect.

218   *Calylophus* underwent recent radiation, we believe the paraphyly of taxa in both concatenation

219   trees might be artifactual. We therefore interpret relationships based on our coalescent-based

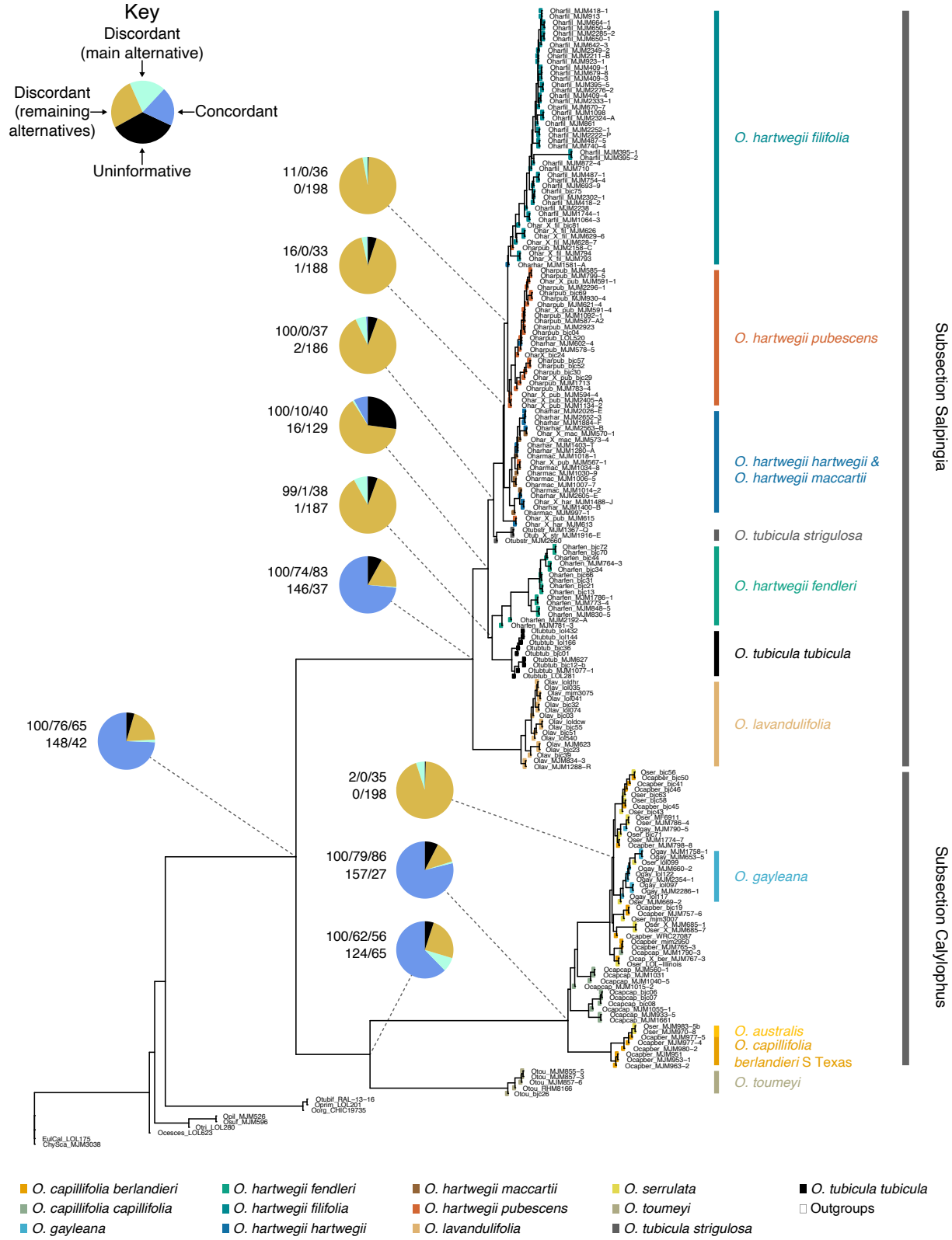220   trees, which comprise the focus for the remainder of the paper (Fig. 2, S6-8).

10

Figure 2

221   To understand whether summary coalescent relationships display a consistent signal

222   across the genome, we quantified gene tree and site concordance using Phyparts and IQtree. We

223   found that gene tree concordance was highest at the deepest nodes at the species level where we

224   expected less ILS and more time between speciation events (Fig. 2, S8, S9). Correspondingly,

225   gene tree concordance was lowest at the subspecies level where increased sharing of ancestral

226   alleles and ongoing gene flow are more likely (Fig. 2, S8, S9). For example, within *O. hartwegii*

227   less than 1% of genes were concordant for bifurcations representing all currently recognized taxa

228   at the subspecies level (Fig. 2, S8, S9). For species-level nodes with high support and high gene

229   tree concordance, site concordance was also high; for example, *O. lavandulifolia* had high

230   bootstrap support in summary coalescent trees (BS = 100), high gene tree concordance (Phyparts

231   = 94% concordance, gCF = 74), and high site concordance (sCF = 83; Fig. 2, S9h). For

232   subspecies with high support, but low gene tree concordance, site concordance was moderate.

233   For example *O. hartwegii* subsp. *fendleri* had high bootstrap support (BS = 99), low gene tree

234   concordance (Phyparts = <1% concordance, gCF = 1), and moderate site concordance (sCF = 38;

235   Fig. 2, S9b). For subspecies that were monophyletic in our coalescent-based trees, but that had

236   low support and low gene tree concordance, site concordance was moderate with an average of

237   35% of sites in agreement for taxa at these nodes (S9[c,f,g,v]). For example *O. hartwegii* subsp.

238   *filifolia* had low bootstrap support, low gene tree concordance (Phyparts = <1% concordance,

239   gCF = 0), and moderate site concordance (sCF = 36; Fig. 2, S9c). This is an important finding

240   because while species tree relationships can be obscured by ILS, site concordance factors, which

241   may be less constrained and less subject to ILS at shallower evolutionary timescales, provide a

242   key alternative method of support (Minh et al. 2018).

12

243 In general, topologies of exon and supercontig datasets were similar, with no major

244 differences in clade membership, but the inclusion of the flanking non-coding regions increased

245 support at shallow nodes in our trees. However, this trend was not universal. For example, in

246 subsect. *Salpingia*, using the supercontig dataset decreased support slightly for one taxon (*O.*

247 *hartwegii* subsp. *filifolia*), and in subsect. *Calylophus* it led to paraphyly of another (*O.*

248 *capillifolia* subsp. *capillifolia*). For six other taxa, our results showed that using supercontigs

249 increased bootstrap support. Therefore, these results demonstrated a net benefit of including

250 flanking non-coding regions for resolving relationships among closely related taxa.

251

252 *Hybridization and Geneflow*

253 Using concatenated loci from the supercontig dataset, we used HyDe (Blischak et al.

254 2018) to test for signals of hybridization. We used 552,521 sites and tested 22 hypotheses for

255 either individuals or groups suspected to be of hybrid origin based on field observations of

256 morphological intermediacy, geographic location, and topological position in our coalescent-

257 based trees and found evidence of hybridization in three individuals representing two taxa, both

258 in subsect. *Salpingia*. The highest signal of hybridization, with a gamma value ($\hat{\gamma}$) of 0.947

259 suggesting more historic gene-flow, was observed in one of three sampled individuals of *O.*

260 *tubicula* subsp. *strigulosa* (*MJM1916.E*). This involved admixture between *O. tubicula* subsp.

261 *tubicula* and the clade consisting of *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp.

262 *maccartii* (Z-score = 5.585, p-value = 0.000, $\hat{\gamma}$ = 0.947; Fig. 3a, S10). We also detected

263 significant levels of hybridization, with $\hat{\gamma}$ ranging from 0.332 to 0.338 suggesting more

264 contemporary gene-flow, in two individuals in *O. hartwegii* subsp. *pubescens*, *BJC29* (Z-score =

265 2.378, p-value = 0.009, $\hat{\gamma}$ = 0.338) and *MJM594* (Z-score = 2.094, p-value = 0.018, $\hat{\gamma}$ = 0.332).

266   This more recent gene flow involved admixture between *O. hartwegii* subsp. *pubescens* and the

267   clade consisting of *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii* (Fig. 3a,

268   S10).

269   The finding that one individual of *O. tubicula* subsp. *strigulosa* may be of hybrid origin is

270   consistent with gene flow between *O. tubicula* subsp. *strigulosa* and its sister taxon *O. tubicula*

271   subsp. *tubicula* (sensu Towner 1977). In our coalescent-based analyses, the two subspecies of *O.*

272   *tubicula* were not recovered as sister taxa, and this relationship was strongly supported (S9[j-k]).

273   If the two *O. tubicula* taxa arose independently, this would support the hypothesis that bee

274   pollination arose in *Oenothera* sect. *Calylophus* independently three times. However, while the

275   summary coalescent analyses we utilized to estimate phylogenies accounted for ILS in tree

276   estimation, they did not account for gene flow (Meng and Kubatko 2009; Gerard et al. 2011;

277   Kubatko and Chifman 2019). Our HyDe results may support the hypothesis that *O. tubicula*

278   subsp. *strigulosa* has experienced gene flow from two closely related taxa, and may have hybrid

279   origins resulting from crossing between *O. tubicula* subsp. *tubicula* and *O. hartwegii* subsp.

280   *hartwegii* (Fig. 3a, S10). This is consistent with Towner's interpretation of this taxon, which he

281   hypothesized may represent a stabilized derivative of introgression between *O. tubicula* subsp.

282   *tubicula* and *O. hartwegii* subsp. *hartwegii* (Towner 1977). Therefore, the placement of *O.*

283   *tubicula* subsp. *strigulosa* as sister to the rest of the *O. hartwegii* species complex in our trees

284   may result from past gene flow and hence may not represent independent origins of bee

285   pollination in subsect. *Salpingia*. These results underscore the importance of explicitly including

286   tests for hybridization in phylogenetic studies. In the case of these data, estimating a species tree

287   given a set of gene trees within a coalescent framework without considering other non-ILS

288   sources of signal conflict could artificially inflate the number of inferred evolutionary transitions.
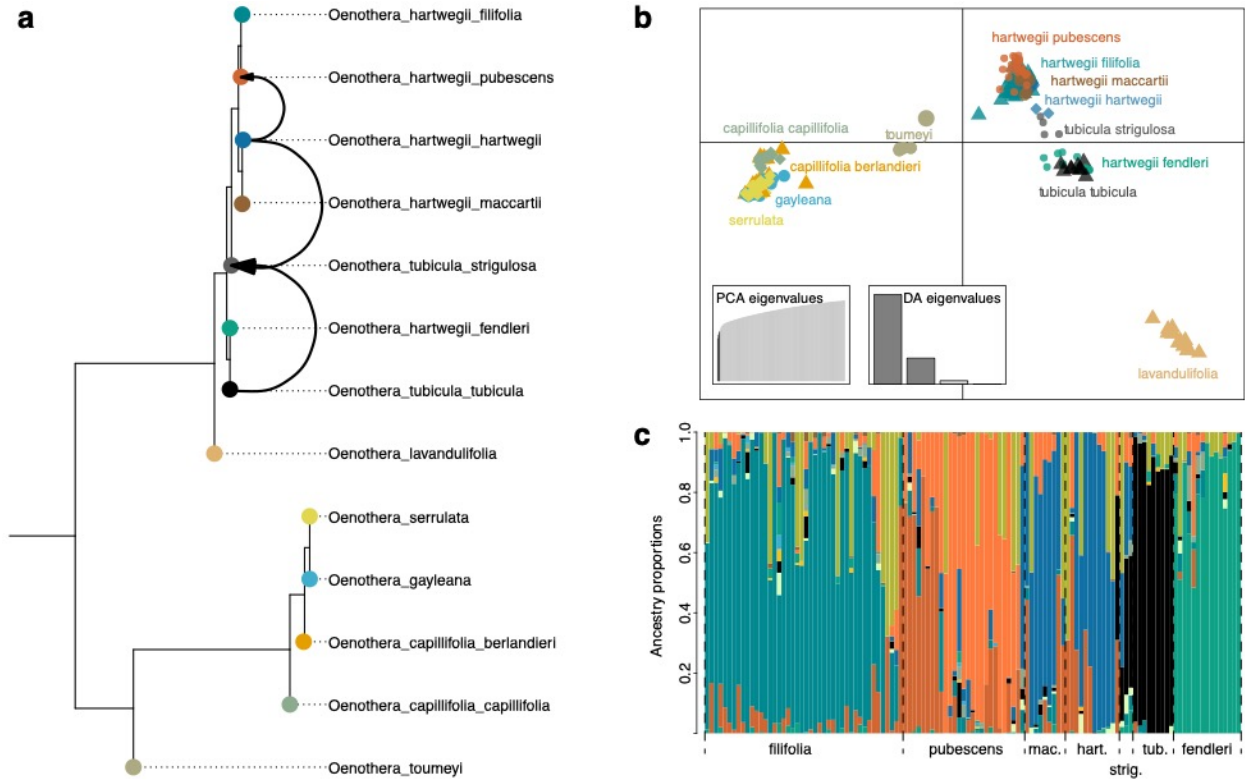
14

Figure 3

Our HyDe results also suggest that at least some of the morphological intermediacy and overlap among taxa in the group is due to continued, or at least recent, gene flow. For example, both *O. hartwegii* subsp. *pubescens* individuals that are inferred to have significant levels of admixture were collected from morphologically intermediate populations of *O. hartwegii* subsp. *pubescens* and *O. hartwegii* subsp. *hartwegii*. In addition, *O. hartwegii* subsp. *hartwegii* was a parent in all three hybridization events (S10). Thus, gene flow may explain this taxon's non-monophyly in our summary coalescent results. However, despite the often confounding patterns of overlapping morphological variation among closely related taxa in subsect. *Salpingia*, this pattern does not necessarily appear to be the result of admixture, as many of the tests for hybridization based on field observations were not significant (Fig 3a, S10). What is also clear from these results is that much like collecting hundreds of nuclear genes provides a more

15

300   nuanced picture of phylogenetic signal and taxon relationships, our results show that collecting

301   multiple individuals from across the geographic and morphological ranges is necessary for a

302   more complete picture of relationships among closely related taxa.

303   After filtering, we extracted a set of 9,728 single nucleotide polymorphisms (SNPs) from

304   both coding and non-coding regions. A Discriminant Analysis of Principal Components (DAPC;

305   Fig. 3b) using these data clearly distinguishes *Oenothera* subsect. *Salpingia* from *O.* subsect.

306   *Calylophus*, with *O. toumeyi* intermediate between the two, which is consistent with the

307   phylogenetic results presented here. Additionally, the DAPC identifies *O. lavandulifolia* as a

308   distinct genetic cluster from the remaining taxa in subsection *Salpingia*. The overlap between

309   taxa, for example between the remaining taxa in subsection *Salpingia*, is consistent with the high

310   levels of gene tree discordance identified by PhyParts (Fig. 3a). For this latter group of taxa, we

311   computed estimates of ancestry coefficients using snmf, which suggests a substantial amount of

312   shared ancestral polymorphisms while also showing some evidence of clear genetic structure

313   among taxa (Fig. 3c). Consistent with the phylogenetic analyses, there does not appear to be any

314   clear genetic distinction between *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp.

315   *maccartii*, whereas *O. hartwegii* subsp. *fendleri*, *O. tubicula* subsp. *tubicula,* and *O. hartwegii*

316   subsp. *filifolia* appear to be largely distinct.

317

318   *Morphological Analysis*

319   We conducted morphometric Principal Components Analysis (PCA) to determine if

320   morphological patterns were consistent with phylogentic results and to examine if specific

321   characters could be used to diagnose taxa as circumscribed by our phylogenetic analysis. Towner

322   (1977) observed overlapping and confounding patterns of morphological variation among taxa

16

323   within subsections, particularly within the *O. hartwegii* species complex. Despite this, because

324   some taxa (e.g., *O. hartwegii* subsp. *fendleri*) were strongly supported by our summary

325   coalescent trees we expected that they would be well distinguished in morphometric analysis.
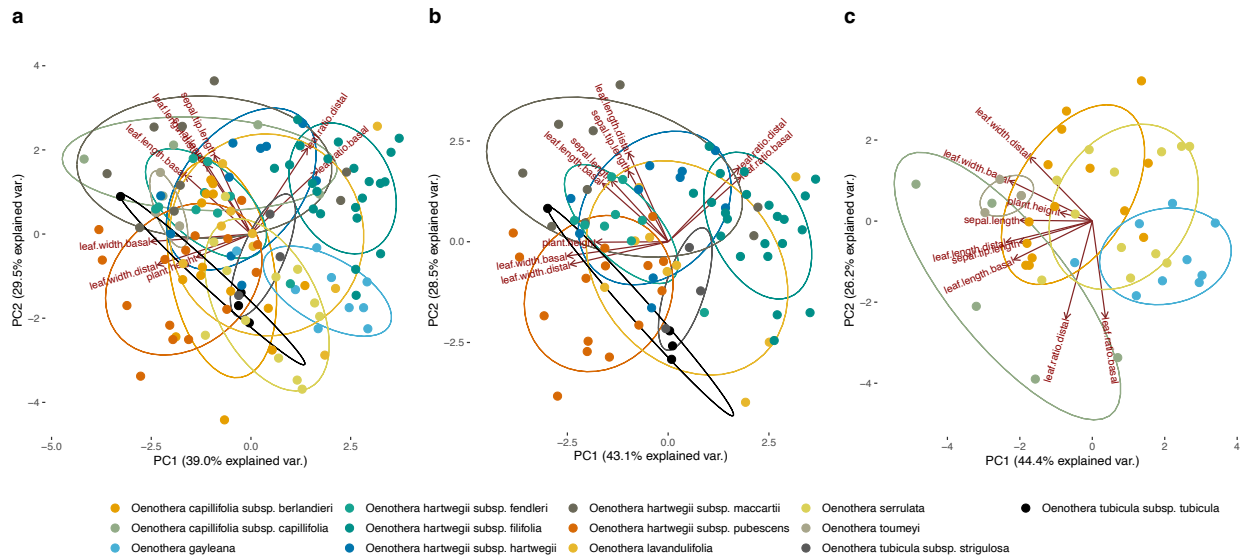

Figure 4

326   The main traits that separated taxa in subsect. *Salpingia* were leaf traits and plant size, while

327   in subsect. *Calylophus* the main traits that separated taxa were sepal traits. In subsect. *Salpingia*,

328   PC1 accounted for 43.1% of variance in PCA, while PC2 accounted for 28.5% (Fig. 4b).

329   Morphological characters most associated with PC1 were leaf width (distal and basal), plant

330   height, and leaf length/width ratio (distal and basal). Those associated with PC2 were leaf length

331   (distal and basal), sepal length, and sepal tip length (Fig. 4b). In subsect. *Calylophus*, PC1

332   accounted for 44.4% of explained variance and PC2 accounted for 26.2% (Fig. 4c). The

333   characters most associated with PC1 in subsect. *Calylophus* include leaf length (distal and basal),

334   sepal length, and sepal tip length. Those most associated with PC2 were leaf length/width ratio

335   (distal and basal) and distal leaf width (Fig. 4c).

17

336   Our results support Towner's understanding of taxon boundaries by underscoring previous

337   difficulties in identifying individuals in this group based on morphology (Towner 1977). We

338   found substantial overlapping morphological variation among currently recognized taxa in both

339   subsections, though some taxa exhibited better grouping than others. The amount of overlap

340   between taxa was not a function of the strength of tree support for a given taxon in our summary

341   coalescent results. For example, *O. hartwegii* subsp. *fendleri*, a taxon that was well supported in

342   our summary coalescent trees, exhibited some of the highest degree of overlap with other taxa in

343   PCA space. Conversely, both *O. hartwegii* subsp. *filifolia* and *O. hartwegii* subsp. *pubescens*,

344   two taxa that formed poorly supported clades in our trees, formed clusters on the outer edges of

345   PCA space and had less overlap than other taxa (Fig. 4b). Interestingly, *O. hartwegii* subsp.

346   *hartwegii*, the taxon that was identified as a parent in all three instances of admixture in our

347   HyDe analysis, also overpas morphologically with most other taxa in subsect. *Salpingia* (Fig.

348   4b). This is not surprising given that it is widely distributed in northern Mexico and western

349   Texas and frequently comes into contact with related taxa resulting in sympatric populations and

350   occasional morphologically intermediate populations.

351

352   *Implications for Reproductive Systems and Edaphic Endemism*

353   Our results show that shifts from hawkmoth to bee pollination likely occurred twice in

354   sect. *Calylophus* (S3) and thus may be more common in *Oenothera* than previously thought. The

355   strongly supported sister relationship of *O. toumeyi* to remaining subsect. *Calylophus* in our

356   summary coalescent results is consistent with two independent shifts to bee pollination, once in

357   the ancestor of subsect. *Calylophus*, and another in subsect. *Salpingia* on the branch leading to

358   *O. tubicula* (Fig. 2). Independent shifts to bee pollination from hawkmoth pollination are perhaps

18

359   not surprising considering that within sect. *Calyophus*, hawkmoth-pollinated floral forms exhibit

360   variation between populations in hypanthium length and diameter and do not prevent occasional

361   pollination by bees (Lewis 2015; Towner 1977). Hawkmoth-pollinated taxa in sect *Calylophus*

362   exhibit vespertine anthesis, which separates them temporally from diurnal bees, but variation in

363   the timing of anthesis is also common between populations (Towner 1977), and hawkmoths are

364   documented to vary greatly in abundance spatiotemporally (Miller 1981; Campbell et al. 1997;

365   Artz et al. 2010). In cases of pollen limitation, night-blooming plants benefit from bimodal

366   pollination between moths and bees by acquiring pollinator assurance against yearly variation or

367   local extinction of specific pollinators, as shown for other *Oenothera* species (Barthell and

368   Knops 1997; Artz et al. 2010), *Lonicera japonica* (Miyake and Yahara 1998) and for night-

369   blooming *Ancistrophora* cacti (Schlumpberger et al. 2009).

370   Additionally, it has been shown that florivore-mediated selection drives floral trait shifts

371   in sect. *Calylophus* towards bee-pollinated floral forms (Jogesh et al. 2017; Bruzzese et al. 2019).

372   Variation in reproductive traits that allows some continued pollination by bees provides an

373   alternative mode of pollen transfer and may represent a mechanism for ensuring pollination.

374   While studies have shown that premating barriers contribute greatly to reproductive isolation

375   (Stanton et al. 2016), our results show that multiple, independent shifts from hawkmoth to bee

376   pollination and associated morphological changes, such as the shorter floral tube length of bee

377   pollinated flowers, may occur in sect. *Calylophus*, and hence may not be a particularly reliable

378   character for diagnosing taxa in this group.

379   Stochastic mapping (supplemental) suggests that there are multiple origins of permanent

380   translocation heterozygosity (PTH) in sect. *Calylophus* While ring chromosomes are common

381   and found in all taxa in sect. *Calylophus*, PTH is currently known from only one taxon, *O.*

19

382   *serrulata*. Because neighboring populations of *O. serrulata* and its putative progenitor O.

383   *capillifolia* subsp. *berlandieri* often resemble each other phenetically, Towner (1977)

384   hypothesized that *O. serrulata* may have originated multiple times through independent origins

385   of translocation heterozygosity in different geographic regions, and may be best recognized as "a

386   complex assemblage of populations having a common breeding system." However, this has

387   never before been explored in a phylogenetic context, nor has it been clearly demonstrated with

388   phylogenetic studies in Onagraceae. In our summary coalescent trees, all currently recognized

389   taxa in subsect. *Calylophus* were paraphyletic and *O. serrulata* was scattered throughout the

390   subsection (Fig. 2, S6, S7). Although support values are not always high for the positions of

391   various individuals of *O. serrulata*, there is at least one well defined, well supported split among

392   populations of *O. serrulata*. In our summary coalescent trees, the two *O. serrulata* accessions

393   from south Texas (*MJM970* & *MJM983*) grouped with other south Texas populations of *O.*

394   *capillifolia* subsp. *berlandieri* with generally strong support (Fig. 2, S6, S7). This relationship

395   was supported in PCA space as well, where *MJM983* was morphologically more similar to the

396   south Texas *O. capillifolia* subsp. *berlandieri* accessions than to other *O. serrulata* (Fig 4C). Our

397   results are therefore consistent with an independent origin of PTH in coastal Texas populations

398   of *O. serrulata*, demonstrating at minimum two origins of PTH (see *Taxonomic Implications*

399   below). However we cannot rule out other independent origins of PTH. Given the prevalence of

400   translocations among partial sets (i.e. not all seven) of homologous chromosomes in *O.* section

401   *Calylophus* (Towner 1977), and if this could be considered an intermediate step towards

402   "complete" PTH, perhaps it is not surprising to reconstruct multiple origins. Our results suggest

403   that a more complete assessment of the extent and distribution of this phenomenon in section

404   *Calylophus* is needed.

20

405   Independent origins of gypsum endemism in sect. *Calylophus* are also supported by our

406   analyses (supplemental). Edaphic specialization is a fundamental driver of speciation in plants

407   and contributes greatly to endemism and species diversity in areas with geologically distinct

408   substrates such as gypsum and serpentine outcrops (Kruckeberg 1984; Anacker et al. 2011;

409   Cacho and Strauss 2014; Moore et al. 2014). To date, two gypsum endemic taxa have been

410   described in sect. *Calylophus*, one in each subsection: *O. hartwegii* subsp. *filifolia*, which is

411   relatively widespread on gypsum in New Mexico and trans-Pecos Texas and only rarely

412   sympatric with other taxa, and the recently described *O. gayleana*, which is found in

413   southeastern New Mexico and adjacent western Texas, with disjunct populations in northern

414   Texas and western Oklahoma (Turner and Moore 2014). Despite low support and low gene tree

415   congruence in our analyses, the two gypsum endemic taxa had moderate sCF support (Fig. 2,

416   S9), much like other taxa with similarly low support and high levels of discordance. In

417   addition,while both gypsum endemics overlapped with other taxa in the morphometric analysis,

418   they occupied morphological extremes in PCA space (Fig. 4). Given that other well-supported

419   taxa also overlap morphologically, it is perhaps not surprising that the two gypsum endemic taxa

420   are not more differentiated from other taxa morphologically. Perhaps the strongest evidence in

421   our data for their recognition as distinct taxa is that we found no evidence of admixture between

422   either of these gypsum endemic taxa and other closely related taxa (Fig. 3a, S10).

423

424   *Taxonomic Implications*

425   The most consequential taxonomic result that arises from our analyses is the position of

426   *O. toumeyi*, a member of subsect. *Salpingia* as circumscribed by Towner (1977). In the present

427   study, this species is resolved as sister to subsect. *Calylophus* with strong support, rendering

21

428   subsect. *Salpingia* paraphyletic (Fig. 2, S9). Towner (1977) grouped *O. toumeyi* with *O.*

429   *hartwegii* due to similar floral and bud characters including large flowers and long floral tubes

430   suggestive of hawkmoth pollination, and rounded buds with long, free sepal-tips. Because the

431   breeding system is a defining difference in the current circumscription between the two

432   subsections in sect. *Calylophus*, this result supports abandoning subsections altogether in sect.

433   *Calylophus*.

434   Within subsect. *Salpingia* our results also suggest the need for revision. While our

435   phylogenetic analyses strongly support the current circumscription of *O. lavandulifolia* (sensu

436   Towner 1977) as a distinct species within subsect. *Salpingia* (Fig. 2, S9), the relationships of the

437   other two species *O. hartwegii* and *O. tubicula* are less clear. Towner (1977) differentiated these

438   two species by the breeding system and grouped the five subspecies of *O. hartwegii* together

439   based on a pattern of reticulate and intergrading variation in which taxa were distinguished from

440   one another by often slight differences in pubescence and leaf shape. Our morphometric analysis

441   confirmed this pattern; however, our phylogenetic results indicated that one taxon, *O. hartwegii*

442   subsp. *fendleri*, shares a closer relationship with the bee pollinated *O. tubicula* subsp. *tubicula*

443   than other taxa in the hawkmoth pollinated *O. hartwegii* species complex (Fig. 2, S6, S7). This

444   relationship was strongly supported and renders *O. hartwegii*, according to the current

445   circumscription, paraphyletic (Towner 1977). Based on strong phylogenetic support for this

446   clade, and its strong morphological distinctiveness as described by Towner (1977), we suggest

447   that *O. hartwegii* subsp. *fendleri* be elevated to the species rank along with both subspecies of *O.*

448   *tubicula* which were equally well supported in phylogenetic analysis and are geographically

449   isolated. Furthermore, our results support a the possible elevation to species rank for *O.*

450   *hartwegii* subsp. *filifolia*. While this taxon was poorly supported in our summary coalescent trees

451   (Fig. 2, S6, S7), we found no evidence of hybridization between this taxon and other closely

452   related taxa. In addition, *O. hartwegii* subsp. *fillifolia* is restricted to gypsum. Therefore, we

453   believe that the ecological distinctiveness and lack of gene flow of *O. hartwegii* subsp. *filifolia*

454   with other taxa in the *O. hartwegii* species complex warrants its elevation as a distinct species. In

455   light of these changes, and to maintain consistency in classification in the subsection, we feel

456   that despite the evidence of hybridization of *O. hartwegii* subsp. *pubescens* with *O. hartwegii*

457   subsp. *hartwegii*, it possesses a morphological distinctiveness that is supported by our

458   phylogenetic results. We therefore recommend *O. hartwegii* subsp. *pubescens* be elevated to the

459   species level, while *O. hartwegii* subsp. *hartwegii* and *O. hartwegii* subsp. *maccartii* be retained

460   as is, forming a polytypic species with two subspecies.

461   In contrast to the relatively clear divisions among taxa in subsect. *Salpingia,* none of the

462   four currently recognized taxa in the subsect. *Calylophus* were consistently recovered as

463   monophyletic. For example, *O. capillifolia* subsp. *capillifolia* was monophyletic in our exon-

464   only summary coalescent tree, but not in the "supercontig" tree, and *O. capillifolia* subsp.

465   *berlandieri* and *O. serrulata* were scattered throughout sect. *Calylophus* in both trees, perhaps

466   suggesting widespread gene flow and/or multiple origins of PTH (Fig. 2, S6). Importantly, our

467   results suggest that the circumscription of *O. gayleana* sensu Turner and Moore (2014) should be

468   amended. Specifically, we find that the populations of subsect. *Calylophus* from northern Texas

469   and western Oklahoma that were assigned to *O. gayleana* by Turner and Moore (2014; *MJM790-*

470   *5, BJC71*) may instead may belong to *O. serrulata* based on both their phylogenetic positions

471   (Fig. 2, S6, S7) and reduced pollen fertility (S12). These north Texas/western Oklahoma

472   populations seem to represent slightly narrower-leaved individuals of *O. serrulata*, which is a

23

473 common inhabitant of the extensive gypsum outcrops of this area (although it is not restricted to

474 gypsum there).

475 Finally, our results highlight an unrecognized cryptic taxon within *O. capillifolia* formed

476 by southern Texas coastal populations currently recognized as *O. capillifolia* subsp. *berlandieri*.

477 Towner (1977) described *O. capillifolia* as a polytypic species with two well-differentiated

478 morphological races. Though he noted the geographic and cytological distinction of the southern

479 Texas coastal populations of *O. capillifolia* subsp. *berlandieri*, these populations were included

480 in *O. capillifolia* subsp. *berlandieri* primarily because of completely overlapping morphological

481 variation. In our results, this cryptic clade of southern Texas coastal populations of *O. capillifolia*

482 subsp. *berlandieri* is the most phylogenetically well supported clade in subsect. *Calylophus* and

483 therefore may warrant taxonomic distinction based on our data (Fig. 2, S9r). Similarly, the

484 southern Texas coastal populations of *O. serrulata*, which is likely an independent origin of PTH

485 derived from this cryptic southern Texas coastal clade of *O. capillifolia* subsp. *berlandieri*, are

486 ecologically distinctive and geographically disjunct from other *O. serrulata* (occurring in coastal

487 dunes, unlike other populations in western Texas, Oklahoma, and northern Texas). In the past

488 they were considered distinctive enough to be described as a species, *Calylophus australis*

489 (Towner & Raven 1970). However, Towner (1977) later combined this species with *O. serrulata*

490 based on his decision to treat all PTH populations as *O. serrulata*. Combined with our results

491 here and the ecogeographic distinctiveness consistent with an independent origin of PTH in

492 coastal Texas, we believe this taxon also warrants recognition as a second PTH species in

493 *Oenothera* sect. *Calylophus*.

494

495 CONCLUSIONS

24

496   Here we describe a robust example of resolving a recent, rapid radiation using multiple

497   sources of evidence: (1) extensive sampling from populations throughout the geographic and

498   morphological range, (2) target enrichment for hundreds of nuclear genes, (3) the inclusion of

499   flanking non-coding regions, (4) gene tree-based hybridization inference, (5) SNPs extracted

500   from target enrichment data, and (6) morphometrics. Our results indicate that in recently radiated

501   species complexes with low sequence divergence and/or high levels of ILS that could be an

502   intractable problem with traditional loci, the use of targeted enrichment in addition to flanking

503   non-coding regions provides a net benefit and is essential to recover species-level resolution. Our

504   results also underscore the importance of summary coalescent methods and evaluating gene tree

505   discordance for resolving historical relationships in recalcitrant groups. By explicitly testing for

506   hybridization using gene tree approaches, we also demonstrate that the estimated number of

507   character state transitions may be artifactually inflated if hybridization is not taken into account.

508   This, in combination with morphometrics, provided key evolutionary insights where

509   relationships in summary coalescent methods may be obscured by gene flow. Importantly, our

510   study uncovers strong evidence for multiple origins of biologically important phenomena,

511   including the evolution of bee pollination, PTH, and edaphic specialization. Consequently,

512   *Oenothera* sect. *Calylophus* might represent a powerful system for understanding these

513   phenomena, especially with future genome sequence data.

514

515   DATA AVAILABILITY

516   Illumina reads generated for this study are available at the NCBI Sequence Read Archive under

517   BioProject PRJNA544074. Assembled exon and supercontig sequences, multiple sequence

518   alignments, and SNP files are available at http://dx.doi.org/10.5061/dryad.[NNNN].

519

26

541 The Shaw Fellowship, New Mexico Native Plant Society, American Society of Plant

542 Taxonomists, and the Society of Herbarium Curators.

543

544 REFERENCES

545

546 Alexander M.P. 1969. Differential Staining of Aborted and Nonaborted Pollen. Stain Technol.
547   44:117–122.
548 Alexander M.P. 1980. A Versatile Stain for Pollen Fungi, Yeast and Bacteria. Stain Technol.
549   55:13–18.
550 Anacker B.L., Whittall J.B., Goldberg E.E., Harrison S.P. 2011. Origins and consequences of
551   serpentine endemism in the California flora. Evolution (N. Y). 65:365–376.
552 Artz D.R., Villagra C. A., Raguso R. A. 2010. Spatiotemporal variation in the reproductive
553   ecology of two parapatric subspecies of *Oenothera cespitosa* (Onagraceae). Am. J. Bot.
554   97:1498–510.
555 Barrett S., Harder L., Worley A. 1996. The comparative biology of pollination and mating in
556   floweing plants. Philos. Trans. R. Soc. B Biol. Sci. 351:1271–1280.
557 Barrett S.C.H. 2013. The evolution of plant reproductive systems: how often are transitions
558   irreversible? Proc R Soc B. 280:20130913.
559 Barthell J. F., Knops J. M. 1997. Visitation of evening primrose by carpenter bees: evidence of
560   a" mixed" pollination syndrome. Southwest. Nat. 86-93.
561 Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A python package for
562   genome-scale hybridization detection. Syst. Biol. 67:821–829.
563 Boberg E., Alexandersson R., Jonsson M., Maad J., Ågren J., Nilsson L.A. 2014. Pollinator
564   shifts and the evolution of spur length in the moth-pollinated orchid *Platanthera bifolia*.
565   Ann. Bot. 113:267–275.
566 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
567   sequence data. Bioinformatics. 30:2114–2120.
568 Brady K.U., Kruckeberg A.R., Bradshaw H.D. 2005. Evolutionary ecology of plant adaptation to
569   serpentine soils. Annu. Rev. Ecol. Evol. Syst. 36:243–266.
570 Bruzzese D.J., Wagner D.L., Harrison T., Jogesh T., Overson R.P., Wickett N.J., Raguso R.A.,
571   Skogen K.A. 2019. Phylogeny, host use, and diversification in the moth family Momphidae
572   (Lepidoptera: Gelechioidea). PLoS One. 14:e0207833.
573 Bryson R.W., Linkem C.W., Dorcas M.E., Lathrop A., Jones J.M., Alvarado-Díaz J., Grünwald
574   C.I., Murphy R.W. 2014. Multilocus species delimitation in the *Crotalus triseriatus* species
575   group (serpentes: Viperidae: Crotalinae), with the description of two new species. Zootaxa.
576   3826:475–496.
577 Cacho N.I., Strauss S.Y. 2014. Occupation of bare habitats, an evolutionary precursor to soil
578   specialization in plants. Proc. Natl. Acad. Sci. 111:15132–15137.
579 Campbell D.R.., Waser N.M.., Melendez-Ackerman E.J.. 1997. Analyzing pollinator-mediated
580   selection in a plant hybrid zone: Hummingbird visitation patterns on three spatial scales.
581   Am. Nat. 149:295–315.
582 Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model.
583   Bioinformatics. 30:3317–3324.
584 Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the

27

585   coalescent model with time-reversible substitution processes, site-specific rate variation,
586   and invariable sites. J. Theor. Biol. 374:35–47.

587   Christie M.R., Knowles L.L. 2015. Habitat corridors facilitate genetic resilience irrespective of
588   species dispersal abilities or population sizes. Evol. Appl. 8:454-463.

589   Crepet W.L., Niklas K.J. 2009. Darwin's second "abominable mystery": Why are there so many
590   angiosperm species? Am. J. Bot. 96:366–81.

591   Duarte J.M., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J.,
592   dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in *Arabidopsis,*
593   *Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels.
594   BMC Evol Biol. 10:61.

595   Eckert A.J., Carstens B.C. 2008. Does gene flow destroy phylogenetic signal? The performance
596   of three methods for estimating species phylogenies in the presence of gene flow. Mol.
597   Phylogenet. Evol. 49:832–842.

598   Ehrlich, P. R., & Raven, P. H. (1969). Differentiation of populations. Science. 165:1228–1232.

599   Folk R.A., Mandel J.R., Freudenstein J. V. 2015. A protocol for targeted enrichment of intron-
600   containing sequence markers for recent radiations: A phylogenomic example from
601   *Heuchera* (Saxifragaceae). Appl. Plant Sci. 3:1500039.

602   Frichot E, Mathieu F, Trouillon T, Bouchard G and François O (2014) Fast and efficient
603   estimation of individual ancestry coefficients. Genetics. 196:973-983.

604   Gerard D., Gibbs H.L., Kubatko L. 2011. Estimating hybridization in the presence of
605   coalescence using phylogenetic intraspecific sampling. BMC Evol. Biol. 11:291.

606   Giarla T.C., Esselstyn J.A. 2015. The challenges of resolving a rapid, recent radiation: Empirical
607   and simulated phylogenomic of Philippine shrews. Syst. Biol. 64(5):727-740.

608   Green R., Krause J., Briggs A., Rasilla Vives M., Fortea Pérez F. 2010. A draft sequence of the
609   neandertal genome. Science. 328:710–722.

610   Heyduk K., Trapnell D.W., Barrett C.F., Leebens-mack J.I.M. 2016. Phylogenomic analyses of
611   species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. Biol.
612   J. Linn. Soc. 117(1):106-120.

613   Hollister J.D., Greiner S., Johnson M.T.J., Wright S.I. 2019. Hybridization and a loss of sex
614   shape genome-wide diversity and the origin of species in the evening primroses (Oenothera,
615   Onagraceae). New Phytol. 224:1372–1380.

616   Jogesh T., Overson R.P., Raguso R.A., Skogen K.A. 2017. Herbivory as an important selective
617   force in the evolution of floral traits and pollinator shifts. AoB Plants. 9(1):plw088.

618   Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett
619   N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-
620   throughput sequencing reads using target enrichment. Appl. Plant Sci. 4:1600016.

621   Johnson M.T.J., Smith S.D., Rausher M.D. 2009. Plant sex and the evolution of plant defenses
622   against herbivores. Proc. Natl. Acad. Sci. 106:18079–18084.

623   Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers.
624   Bioinformatics 24:1403-1405.

625   Jombart T, Devillard S and Balloux, F (2010). Discriminant analysis of principal components: a
626   new method for the analysis of genetically structured populations. BMC Genetics. 11:94.

627   Katinas L., Crisci J., Wagner W., Hoch P. 2004. Geographical diversification of tribes
628   Epilobieae, Gongylocarpeae, and Onagreae (Onagraceae) in North America, based on
629   parsimony analysis of endemicity and track compatibility analysis. Ann. Missouri Bot.
630   91:159–185.

28

Knowles L.L. 2009. Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. Syst. Biol. 58:463–467.

Knowles L.L., Chan Y.-H. 2008. Resolving species phylogenies of recent evolutionary radiations. Ann. Missouri Bot. Gard. 95:224–231.

Kruckeberg A. 1984. California Serpentines: Flora, Vegetation, Geology, Soils, and Management Problems. Berkeley: Univ of California Press.

Kubatko L.S., Chifman J. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC Evol. Biol. 19:1–13.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61:727–744.

Levin R., Wagner W., Hoch P. 2004. Paraphyly in Tribe Onagreae : Insights into Phylogenetic Relationships of Onagraceae Based on Nuclear and Chloroplast Sequence Data. Syst. Bot. 29:147–164.

Lewis, Emily. 2015. Differences in Population Genetic Structure of Hawkmoth and Bee-Pollinated Species of *Oenothera* (Onagraceae) Are More Pronounced at a Landscape Scale. Northwestern University Libraries. Masters Thesis. Northwestern University.

Maddison W.P., Knowles L. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. Evol. Biol. 55:21–30.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., Turner D.J. 2010. Target-enrichment strategies for next- generation sequencing. Nat. Methods. 7:111–118.

Mandel J.R., Dikow R.B., Funk V. a, Masalia R.R., Staton S.E., Kozik A., Michelmore R.W., Rieseberg L.H., Burke J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. Appl. Plant Sci. 2:1–6.

Meng C., Kubatko L.S. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. Theor. Popul. Biol. 75:35–45.

Miller R.B.. 1981. Hawkmoths and the Geographic Patterns of Floral Variation in *Aquilegia caerulea*. Evolution (N. Y). 35:763–774.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37(9):2727-2733.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.

Miyake T., Yahara T. 1998. Why does the flower of *Lonicera japonica* open at dusk?. Can. J. Bot. *76*(10):1806-1811.

Moore M., Jansen R. 2007. Origins and biogeography of gypsophily in the Chihuahuan Desert plant group *Tiquilia* Subg. subg. *Eddya*. Syst. Bot. 32:392–414.

Moore M.J., Mota J.F., Douglas N.A., Flores Olvera H., Ochoterena H. 2014. Ecology assembly evolution gypsophile floras. In: Rajakaruna N., Boyd R.S., Harris T.B., editors. Plant Ecology and Evolution in Harsh Environments. New York: Nova Science Publishers, Inc. p. 97–128.

Nason J.D., Hamrick J.L., Fleming T.H. 2002. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. Evolution 56:2214–2226.

29

677 Van der Niet T., Johnson S.D. 2012. Phylogenetic evidence for pollinator-driven diversification
678   of angiosperms. Trends Ecol. Evol. 27:353–361.
679 van der Niet T., Johnson S.D., Linder H.P. 2006. Macroevolutionary data suggest a role for
680   reinforcement in pollination system shifts. Evolution 60:1596–1601.
681 Van Der Niet T., Peakall R., Johnson S.D. 2014. Pollinator-driven ecological speciation in
682   plants: New evidence and future perspectives. Ann. Bot. 113:199–211.
683 Peakall R., Ebert D., Poldy J., Barrow R.A., Francke W., Bower C.C., Schiestl F.P. 2010.
684   Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually
685   deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. New Phytol.
686   188:437–50.
687 Rajakaruna N. 2004. The Edaphic Factor in the Origin of Plant Species. Int. Geol. Rev. 46:471–
688   478.
689 Raven P.H. 1964. Catastrophic Selection and Edaphic Endemism. Evolution 18:336–338.
690 Raven P.H. 1979. A survey of reproductive biology in Onagraceae. New Zeal. J. Bot. 17:575–
691   593.
692 R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for
693   Statistical Computing, Vienna, Austria.
694 Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned
695   sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.
696 Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from
697   quartet frequencies. Mol. Biol. Evol. 33(7):1654-68.
698 Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27(4):592-593.
699 Schlumpberger B. O., Cocucci A. A., Moré M., Sérsic A. N., Raguso R. A. 2009. Extreme
700   variation in floral characters and its consequences for pollinator attraction among
701   populations of an Andean cactus. Ann. Bot. *103*(9):1489-1500.
702 Skogen K.A., Overson R.P., Hilpman E.T., Fant J.B. 2019. Hawkmoth pollination facilitates
703   long-distance pollen dispersal and reduces isolation across a gradient of land-use change.
704   Ann. Missouri Bot. Gard. 104:495–511.
705 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets reveals
706   conflict, concordance, and gene duplications with examples from animals and plants. BMC
707   Evol. Biol. 15:150.
708 Stamatakis A. 2014. Stamatakis - 2014 - RAxML version 8 a tool for phylogenetic analysis and
709   post-analysis of large phylogenies. 2010–2011.
710 Stanton K., Valentin C.M., Wijnen M.E., Stutstman S., Palacios J.J., Cooley A.M. 2016.
711   Absence of postmating barriers between a selfing vs. Outcrossing chilean *Mimulus* species
712   pair. Am. J. Bot. 103:1030–1040.
713 Stebbins G.L. 1970. Adaptive Radiation of Reproductive Characteristics in Angiosperms, I:
714   Pollination Mechanisms. Annu. Rev. Ecol. Syst. 1:307–326.
715 Stephens J.D., Rogers W.L., Heyduk K., Cruse-Sanders J.M., Determann R.O., Glenn T.C.,
716   Malmberg R.L. 2015. Resolving phylogenetic relationships of the recently radiated
717   carnivorous plant genus *Sarracenia* using target enrichment. Mol. Phylogenet. Evol. 85:76–
718   87.
719 Stockhouse R.E.I. 1973. Biosystematic Studies of *Oenothera* L. Subgenus *Pachylophus*. Univ.
720   Microfilm. Ph.D. thesis. Colorado State University.
721 Straub S.C.K., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the
722   tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot.

723   99:349–64.
724 Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
725   Version 4. Sinauer Associates, Sunderland, Massachusetts.
726 Thomson J.D., Wilson P. 2008. Explaining Evolutionary Shifts between Bee and Hummingbird
727   Pollination: Convergence, Divergence, and Directionality. Int. J. Plant Sci. 169:23–38.
728 Towner H.F., Raven P.H. 1970. A new species and some new combinations in *Calylophus*
729   (Onagraceae). Madrono. 20: 241-245.
730 Towner H.F. 1977. The Biosystematics of *Calylophus* (Onagraceae). Ann. Missouri Bot. Gard.
731   64:48–120.
732 Tripp E.A., Manos P.S. 2008. Is floral specialization an evolutionary dead-end? Pollination
733   system transitions in *Ruellia* (Acanthaceae). Evolution 62:1712–1737.
734 Turner B.L., Moore M.J. 2014. *Oenothera gayleana* (*Oenothera* sect. *Calylophus*, Onagraceae),
735   a new gypsophile from Texas, New Mexico, and Oklahoma. Phytologia.org. 96:200–206.
736 Wagner W., Hoch P., Raven P. 2007. Revised Classification of the Onagraceae. Syst. Bot.
737   Monogr. 83:1–240.
738 Wagner W.L. 2021. In press. In: Flora of North America Editorial Committee, eds. 1993+.
739   Flora of North America North of Mexico, 20+ vols. New York and Oxford: Oxford
740   University Press.
741 Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston A.
742   2014. Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant
743   Phylogenomics. Appl. Plant Sci. 2:1400042.
744 Wilson P., Wolfe A.D., Armbruster W.S., Thomson J.D. 2007. Constrained lability in floral
745   evolution: Counting convergent origins of hummingbird pollination in Penstemon and
746   Keckiella. New Phytol. 176:883–890.
747 Xi Z., Liu L., Rest J.S., Davis C.C. 2014. Coalescent versus Concatenation Methods and the
748   Placement of *Amborella* as Sister to Water Lilies. Syst. Biol. 63:919–932.
749 Xu S., Schlüter P.M., Scopece G., Breitkopf H., Gross K., Cozzolino S., Schiestl F.P. 2011.
750   Floral isolation is the main reproductive barrier among closely related sexually deceptive
751   orchids. Evolution. 65:2606–20.

752 FIGURE LEGENDS

753

754 **Figure 1.** (a) Range map of all taxa in *Oenothera* sect. *Calylophus* (based on Towner 1977).
755 Sampling locations of leaf tissue samples (points; color corresponds to taxa in cladogram to the
756 right [Figure 1b]) and estimated taxon ranges (polygons; colors correspond to Figure 1b)
757 proposed by Towner (1977) and Turner and Moore (2014). (b) Estimated cladogram of
758 *Oenothera* sect. *Calylophus* sensu Towner (1977) and Turner and Moore (2014). Symbols to the
759 right of tip labels signify pollination system (bee, hawkmoth, or Permanent Translocation
760 Heterozygosity [PTH]) and the symbol color specifies whether a given taxon is a gypsum
761 endemic (purple = no, pink = yes). Photo panels: (c) *Oenothera hartwegii* subsp. *fendleri* (d)
762 *Oenothera hartwegii* subsp. *filifolia* (e) *Oenothera hartwegii* subsp. *hartwegii* (f) *Oenothera*
763 *hartwegii* subsp. *maccartii* (g) *Oenothera hartwegii* subsp. *pubescens* (h) *Oenothera tubicula*
764 subsp. *tubicula* (i) *Oenothera lavandulifolia* (j) *Oenothera tubicula* subsp. *strigulosa* (k)
765 *Oenothera capillifolia* subsp. *berlandieri* (l) *Oenothera capillifolia* subsp. *capillifolia* (m)
766 *Oenothera gayleana* (n) *Oenothera serrulata* (o) *Oenothera toumeyi*

767
768 **Figure 2.** ASTRAL-II summary coalescent tree constructed using the "Supercontig" dataset with
769 100 bootstraps. At relevent nodes, piecharts represent Phyparts analysis (blue = concordant,
770 green = most frequent conflict, yellow = all other conflict, black = uninformative gene trees), top
771 row of support values are bootstrap values from ASTRAL-II, and gCF and sCF from IQTree
772 (BS/gCF/sCF), bottom two support values are number of concordant gene trees for the node and
773 total number of gene trees minus the number of concordant gene trees at that node
774 (concord/discord). Colored tip points correspond to taxon designation.
775
776 **Figure 3.** (a) Summary of HyDe Analysis annotated on ASTRAL-III species tree constructed
777 using the "Supercontig" dataset; black arrows represent direction of admixture detected by HyDe
778 anaylsis. (b) Discriminant Analysis of Principal Components based on a filtered set of SNPs
779 extracted from the entire supercontig dataset , and (c) sNMF plot of inferred ancestry coefficients
780 using the same set of filtered SNPs but limited to subsection *Salpingia* (minus *O. lavandulifolia*)
781
782 **Figure 4.** Morphometric Principal Components Analysis (PCA) using 9 morphological
783 characters (plant height, leaf length [basal and distal], leaf width [basal and distal], leaf
784 length/width ratio (basal and distal), and sepal tip length) for (a) Section *Calylophus* (b)
785 Subsection *Salpingia* without *O. toumeyi*, and (c) Subsection *Calylohus* with *O. toumeyi*
786 included.

787 SUPPLEMENTAL MATERIAL

788 S1.
789 See excel table "S1 Accessions and Seq Stats"
790

791 S2.
792 Number of leaf tissue accessions sequenced from each taxon

Taxon	No. of Accessions	
O. lavandulifolia	16	794
O. tubicula subsp. *strigulosa*	3	795
O. tubicula subsp. *tubicula*	9	796
O. hartwegii subsp. *fendleri*	15	797
O. hartwegii subsp. *filifolia*	44	798
O. hartwegii subsp. *hartwegii*	12	799
O. hartwegii subsp. *maccartii*	9	800
O. hartwegii subsp. *pubescens*	26	801
O. hartwegii	1	802
O. toumeyi	5	803
O. capillifolia subsp. *berlandieri*	17	804
O. capillifolia subsp. *capillifolia*	11	805
O. gayleana	9	806
O. serrulata	17	807
Outgroups:		808
O. pilosella	1	809
O. organensis	1	810
O. primiveris	1	811
O. tubifera	1	812
O. triloba	1	813
O. cespitosa subsp. *cespitosa*	1	814
O. suffrutescens	1	815
Chylismia scapoidea subsp. *scapoidea*	1	816
Eulobus californicus	1	817
Total	203	818

819 S3.
820 MATERIALS AND METHODS
821 *Taxon Sampling, DNA Extraction, and Determination of PTH*
822 A total of 194 individuals spanning the geographic, morphological, and ecological ranges
823 of all 13 recognized taxa in *Oenothera* sect. *Calylophus* [following Towner (1977) and Turner
824 and Moore (2014)] were included in this study (Fig. 1a, S1, S2) along with eight outgroups
825 representing other major sections of Oenothera (*Eremia*, *Gaura*, *Kneiffia*, *Lavauxia*, *Oenothera*,
826 *Pachylophus*, and *Ravenia*) and other genera (*Chylismia* and *Eulobus*) in Onagraceae (S1, S2).
827 All leaf tissue samples were collected from individuals in the field between 2007 and 2015 and
828 voucher specimens were deposited at the United States National Herbarium (US), with

33

829 duplicates in most cases at either the Nancy Rich Poole Herbarium (CHIC) or the George T.
830 Jones Herbarium at Oberlin College (OC; S1). DNA was extracted from fresh, silica-dried leaf
831 tissue using either (1) a modified CTAB protocol (Doyle 1987), (2) the Nucleon PhytoPure DNA
832 extraction kit (GE Healthcare Life Sciences, Pittsburgh, Pennsylvania, USA), or (3) a modified
833 CTAB and silicon dioxide purification protocol (Doyle 1987; Sharma and Purohit 2012; See S13
834 for details) followed by passing any extractions retaining a brown or yellow coloration through a
835 Qiagen Qiaquick PCR spin column for additional purification according to the manufacturer's
836 protocol (Qiagen, Venlo, Netherlands). The third DNA extraction method was used for difficult
837 to extract, polysaccharide-rich leaf tissue samples that yielded gooey, discolored DNA following
838 initial extraction. PTH status was determined for individuals in subsect. *Calylophus* using floral
839 morphology and/or when flowers were present by assessing pollen fertility using a modified
840 Alexander stain, as PTH taxa have a demonstrated 50% reduction in pollen fertility (Towner
841 1977). For accessions identified as either *O. capillifolia* subsp. *berlandieri* or *O. serrulata* that
842 had sufficient pollen available, pollen was removed from flowers and stained using a modified
843 Alexander stain (Alexander 1969, 1980). Accessions with less than 50% viable pollen were
844 assigned to *O. serrulata*, the only currently recognized PTH taxon in subsect. *Calylophus*. Pollen
845 count data are provided in Supplement 12 (See S12 for details).
846
847 *Bait Design, Library Construction, Target Enrichment, and Sequencing*
848 We targeted 322 orthologous, low-copy nuclear loci determined by clustering
849 transcriptomes of *Oenothera serrulata* (1KP accession *SJAN*) and *Oenothera berlandieri* (1KP
850 accession *EQYT*) to select a subset of the 956 phylogenetically informative *Arabidopsis* loci
851 identified by Duarte et al. (2010). Transcriptomes of two *Oenothera* species, *O. serrulata* and *O.*
852 *berlandieri*, were assembled and optimal isoforms were filtered for the longest reading frame
853 using custom Perl scripts. Assembled transcripts were aligned as amino acids to the 956 TAIR
854 loci of *Arabidopsis* in TranslatorX (Abascal et al. 2010). This alignment identified 956
855 orthologous sequences, from which 322 loci were randomly selected. BLAST searches of amino
856 acid sequences from these loci were carried out to ensure orthology between the transcript loci
857 and the *Arabidopsis* TAIR locus. The bait set was designed from these 322 loci, which were
858 selected from both *O. serrulata* and *O. berlanderi* sequences. A set of 19,994 120-bp baits tiled
859 across each locus with a 60 base overlap (2x tiling) was manufactured by Arbor Biosciences
860 (formerly MYcroarray, Ann Arbor, Michigan, USA). Sequencing libraries for 67 samples were
861 prepared with the Illumina TruSeq Nano HT DNA Library Preparation Kit (San Diego,
862 California, USA) following the manufacturer's protocol, except using half volumes beginning
863 with the second addition of Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen,
864 Carlsbad, CA, USA). DNA samples were sheared using a Covaris M220 Focused-Ultrasonicator
865 (Covaris, Woburn, Maryland, USA) to a fragment length of ~550 bp (for an average insert size
866 of ~420 bp). The remaining 134 libraries were constructed by Rapid Genomics (Gainesville,
867 Florida, USA), with custom adapters. The Illumina i5 and i7 barcodes were used for all libraries.
868 Libraries were enriched for these loci using the MyBaits protocol (ArborBiosciences
869 2016) with combined pools of libraries totaling 1.2 μg of DNA (12 libraries/pool at 100
870 ng/library). Libraries with less than 100 ng of total recovered DNA were pooled together in
871 equimolar concentrations using available product, resulting in some pools with less than 1.2 μg
872 of DNA. The smallest successful pool contained four samples with 6 ng of library each.
873 Hybridization was performed at 65°C for approximately 18 hours. The enriched libraries were
874 reamplified with 14 to 18 PCR cycles and a final cleanup was performed using a Qiagen

875   QiaQuick PCR cleanup kit following the manufacturer's protocol to remove bead contamination
876   (Qiagen, Venlo, Netherlands). DNA concentrations were measured using a Qubit 2.0
877   Fluorometer (Life Technologies, Carlsbad, California, USA) and molarity was measured on an
878   Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). A final
879   cleaning step using Dynabeads MyOne Streptavidin C1 magnetic beads was performed on pools
880   with adapter contamination as detected on the Bioanalyzer. Pools were sequenced in four runs at
881   equimolar ratios (4 nM), on an Illumina MiSeq (2 x 300 cycles, v3 chemistry; Illumina, Inc., San
882   Diego, California, USA) at the Pritzker DNA lab (Field Museum, Chicago, IL, USA). This
883   produced 80,273,296 pairs of 300-bp reads. Reads were demultiplexed and adapters trimmed
884   automatically by Illumina Basespace (Illumina 2016). Raw reads have been deposited at the
885   NCBI Sequence Read Archive (BioProject PRJNA544074).
886
887   *Quality Filtering, Assembly and Alignment*
888   A summary of read quality from each sample was produced using FastQC
889   (http://www.bioinformatics.babraham.ac.uk/people.html), which revealed read-through adapter
890   contamination in many of the poorer quality samples. To remove read-through contamination
891   and filter for quality, reads were trimmed for known Illumina adapters using Trimmomatic
892   (Bolger et al. 2014) with the following settings:
893   ILLUMINACLIP:<illumina_adapters.fasta>:2:30:10 LEADING:10 TRAILING:10
894   SLIDINGWINDOW:4:20 MINLEN:20. Trimmed, quality-filtered reads were assembled using
895   the HybPiper pipeline (Johnson et al. 2016) with default settings, followed by the intronerate.py
896   script, to produce both exons and the "splash zone" flanking non-coding region-containing
897   supercontigs. Only pairs with both mates surviving trimming and quality filtering were used for
898   HybPiper.
899   To compare the influence of "splash-zone" non-coding regions, two sets of alignments
900   were created: (1) exons alone, and (2) coding sequences plus the "splash-zone" (Hereby referred
901   to as supercontigs). For multiple sequence alignments of exons alone, protein and nucleotide
902   sequences assembled in HybPiper were gathered into fasta files by gene. For protein sequences
903   only, stop codons were changed to "X" using a sed command-line regular expression to facilitate
904   alignment, and sequences were aligned using MAFFT with settings: --auto --adjustdirection --
905   maxiterate 1000 (Katoh et al. 2002). Aligned protein sequences were then used to fit unaligned
906   nucleotide sequences into coding frame alignments using pal2nal with default settings (Suyama
907   et al. 2006). In-frame, aligned DNA sequences were trimmed to remove low-coverage positions
908   and sequences composed only of gaps using TrimAl with the automated setting, which is
909   optimized for maximum likelihood analyses (Gutíerrez et al. 2009). For supercontigs, nucleotide
910   sequences assembled using HybPiper were gathered into fasta files by gene, gene names were
911   removed from fasta headers using a command-line regular expression, and sequences were
912   aligned in MAFFT with settings: --auto --adjustdirection --maxiterate 1000 (Katoh et al. 2002).
913   Reverse compliment tags ("_R_") were removed from taxon names using a command-line
914   regular expression, and sequences were trimmed using TrimAl with previously listed settings
915   optimized for maximum likelihood analyses (Gutíerrez et al. 2009). To minimize the effects of
916   missing data on phylogenetic analyses, accessions with < 50% of loci passing quality filtering
917   were removed, and genes that were recovered across < 70% of the total remaining samples were
918   also removed. Following quality filtering, we recovered 204 high quality loci (present in at least
919   70% of samples) and an average of 625,323 reads per sample (S1). Final assemblies are available
920   at Dryad (http://dx.doi.org/10.5061/dryad.[NNNN]). Bioinformatics pipelines and analyses were

921   run at the high-performance computing cluster ("Treubia") at the Chicago Botanic Garden unless
922   otherwise specified.
923
924   *Phylogenetic Reconstruction*
925   We conducted phylogenetic analyses using two strategies for each set of alignments.
926   Alignments were concatenated and analyzed using maximum likelihood (ML) in RAxML
927   (Stamatakis 2014; hereafter referred to as "concatenation"), whereas coalescent-based analyses
928   were conducted using ML gene trees in ASTRAL-II (Mirarab and Warnow 2015; Sayyari and
929   Mirarab 2016) and using unlinked SNPs in SVDquartets (Chifman and Kubatko 2014, 2015)
930   implemented in PAUP* beta version 4.0a168 (Swofford 2003; S3). In concatenation analyses,
931   after aligning each gene separately in MAFFT, genes were concatenated, partitioned, and
932   maximum likelihood trees were reconstructed in RAxML Version 8 (Stamatakis 2014) using the
933   GRTCAT model with 100 "rapid-boostrapping" psuedoreplicates and *Chylismia scapoidea* as
934   the outgroup, on the CIPRES Science Gateway computing cluster (Miller et al. 2010). For
935   coalescent analyses, individual gene trees were first estimated using RAxML Version 8
936   (Stamatakis 2014), with 100 "rapid-bootstraping" psuedo-replicates and settings: -p 12345 -x
937   12345 -N 100 -c 25 -f a -m GTRCAT –s, and *Chylismia scapoidea* as the outgroup. Gene trees
938   based on supercontigs were not partitioned by codon position. Coalescent-based analyses of
939   accessions were conducted in ASTRAL-II (Mirarab and Warnow 2015; Sayyari and Mirarab
940   2016) with default settings using the best RAxML gene trees and their associated bootstrap files
941   as input, and in SVD Quartets (Chifman and Kubatko 2014, 2015) with default settings using an
942   inframe aligned supermatrix with all 204 loci and supercontigs. ASTRAL-II and SVD Quartets
943   analysis was performed with 100 multi-locus bootstraps.
944   To assess concordance among gene trees and provide additional support complementary
945   to bootstrap values, we conducted two additional analyses. First, we assessed raw gene tree
946   concordance using Phyparts (Smith et al. 2015). Prior to running Phyparts, nodes with < 33%
947   support in the supercontig RAxML gene trees were collapsed using the sumtrees command in
948   Dendropy (Sukumaran 2010). These gene trees were then re-rooted using *Chylismia scapoidea*
949   as the outgroup and ASTRAL-II was rerun using these collapsed, re-rooted gene trees as the
950   input files. Pie charts showing gene tree discordance were generated and overlaid on the
951   resulting ASTRAL-II tree using the PhypartsPiecharts script
952   (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts). Phyparts piecharts
953   and gene tree concordance values were also added to Figure 1 by importing the two data files
954   produced by the Phypartspiecharts.py into R and manually matching them to key nodes on our
955   ASTRAL-II supercontig tree using *ggtree* version 1.14.6 (G Yu, DK Smith, H Zhu, Y Guan
956   2017). We also generated gene and site concordance factors for our ASTRAL-II tree constructed
957   using supercontigs in IQTree v1.7-beta16 (Minh et al. 2018). IQtree calculates the gene
958   concordance factor (gCF) and accounts for incomplete taxon coverage among gene trees and
959   therefore may provide a more accurate representation of agreement among gene trees than other
960   methods. In addition to gCF, IQTree calculates the site concordance factor (sCF), which is
961   defined as the percentage of decisive nucleotide sites supporting a specific node (Minh et al.
962   2018). We used the RAxML gene trees produced for the ASTRAL-II supercontig analysis, and
963   supercontig alignments themselves, as the inputs for IQtree. For computing sCF, we randomly
964   sampled 100 quartets around each internal node. Finally, we mapped gCF and sCF values to the
965   ASTRAL-II supercontig tree produced in our previous summary coalescent analysis (Fig. 2). All

966 phylogenetic trees, with the exception of the full Phyparts picharts tree, were visualized using the
967 R package *ggtree* version 1.14.6 (G Yu, DK Smith, H Zhu, Y Guan 2017).
968
969 *Ancestral State Reconstruction*
970   To infer ancestral conditions and the number of transitions in reproductive system, we
971 used the *phangorn* (Schliep 2011) package in R. First, a Coalescent-based species tree with
972 accessions grouped into taxa using a mapping file was estimated in ASTRAL-III (Zhang et al.
973 2018) with default settings using the best RAxML gene trees and their associated bootstrap files,
974 from the supercontig alignments, as input. Next we time calibrated the ASTRAL-III species tree
975 to 1 million years based on estimates of other taxa in the genus (Evans et al. 2009) using the
976 makeChronosCalib function in the *ape* (Paradis et al. 2004) package in R, and estimated an
977 ultrametric tree using the chronos function in *ape* (Paradis et al. 2004) with settings: lambda = 1,
978 model = "relaxed". Finally, we performed marginal reconstruction of ancestral character states
979 using the maximum likelihood method using the optim.pml and ancestral.pml functions in the
980 *phangorn* (Schliep 2011) package in R.
981
982 *Testing for Hybrid Orgins with HyDe*
983   To test for putative hybrid origins of selected taxa, we used HyDe (Blischak et al. 2018)
984 to calculate D-Statistics (Green et al. 2010) for a set of hypotheses (S10). Briefly, HyDe
985 considers a four-taxon network of an outgroup and a triplet of ingroup populations to detect
986 hybridization from phylogenetic invariants that arise under the coalescent model with
987 hybridization. Introgression between P3 and either P1 or P2 influences the relative frequencies of
988 ABBA and BABA, and the D-statistic measures the imbalance between these frequencies. We
989 tested the triplets in (S10) and set *Chylismia scapoidea* as the outgroup. We considered
990 hypothesis tests significant at an overall $\alpha < 0.05$ level with estimates of γ between 0 and 1. Z-
991 scores greater than 3 are generally interpreted as strong evidence of introgression.
992
993 *Population-level Analysis*
994   To further characterize population-level processes or genetic structure within sect.
995 *Calylophus*, we extracted and filtered SNPs by mapping individual reads against reference
996 supercontigs (see https://github.com/lindsawi/HybSeq-SNP-Extraction). To account for
997 duplicates arising from PCR during HybSeq in SNP calling and filtering, first we selected the
998 sample with the highest target recovery rate and sequencing depth as a target reference sequence
999 (Oenothera_capillifolia_berlandieri_bjc19) and gathered supercontigs for this individual into a
1000 single target FASTA file. We then ran BWA (Li and Durbin 2009) to align sequences, Samtools
1001 'index' (Danecek et al. 2011) and GATK CreateSequenceDictionary (Poplin et al. 2017),
1002 respectively, on the resulting target FASTA file. Next we ran a custom script
1003 "variant_workflow.sh" using both read files from each *Calylophus* sample as input to create a
1004 vcf file for each sample. SNP's were called for each individual using GATK (Poplin et al. 2017)
1005 and the vcf file from each sample as input. The resulting vcf file created in the previous step was
1006 filtered to remove indels using GATK and the original target FASTA file as input, and then
1007 filtered again based on read mapping and quality with GATK VariantFiltration with settings: --
1008 filterExpression "QD < 5.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum
1009 < -8.0" (Poplin et al. 2017). Finally, we generated a reduced SNP file in FastStructure format
1010 using PLINK (Purcell et al. 2007) to remove SNPs that did pass filter using the command: plink -
1011 -vcf-filter --vcf Pachylophus.filtered.snps.vcf --const-fid --allow-extra-chr --geno --make-bed --

1012 recode structure. Finally we used Discriminant Analysis of Principal Components (Jombart et al.
1013 2010) as implemented in the R package *adegenet* (Jombart 2008) and the snmf function in the
1014 LEA package (Frichot and François 2015) in R (R Core Team, 2020).
1015
1016 *Morphological Measurements and Analysis*
1017   To assess taxon boundaries and patterns of morphological variation, we measured
1018 character states for the following key morphological structures that have been used historically to
1019 discriminate taxa in sect. *Calylophus* (Towner 1977): plant height, leaf length (distal), leaf width
1020 (distal), leaf length/width ratio (distal), leaf length (basal), leaf width (basal), leaf length/width
1021 (basal), sepal length, and sepal tip length. Measurements were made with digital calipers when
1022 possible, or with a standard metric ruler and dissecting scope, from voucher specimens of nearly
1023 all sampled populations of sect. *Calylophus* included in our molecular phylogenetic analyses.
1024 Measurements are provided in S11 and have been deposited at Dryad
1025 (http://dx.doi.org/10.5061/dryad.[NNNN]). Morphological measurements were log transformed
1026 using the R base function 'log' (R Core Team 2018) prior to Principal Components Analysis
1027 (PCA), which was conducted in R using the *stats* package version 3.7 and the function
1028 'prcomp'(R Core Team 2018). All 'NA' values were omitted from analysis. Plots of PCA results
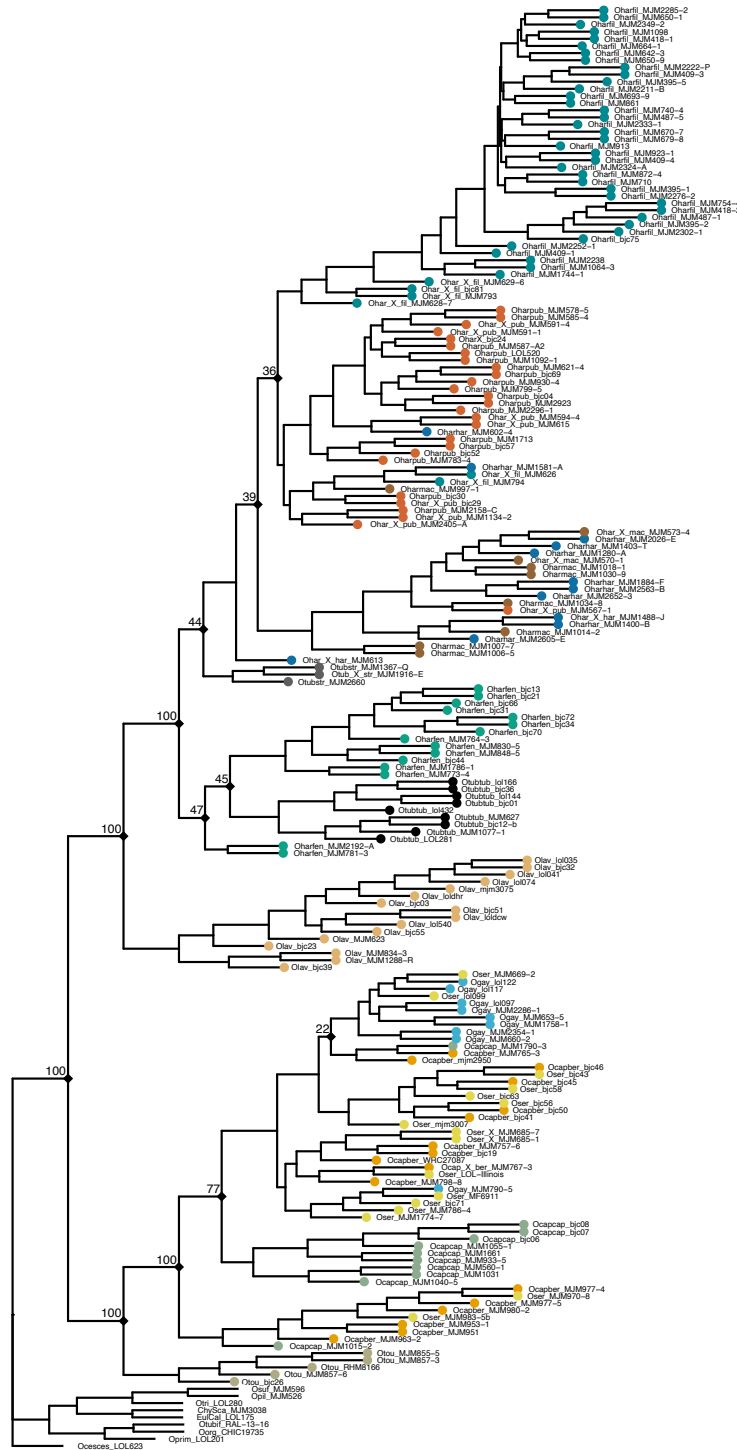1029 were visualized using the *ggplot2* package in R (Wickham 2016).
1030

1031 REFERENCES
1032 Abascal F., Zardoya R., Telford M.J. 2010. TranslatorX: Multiple alignment of nucleotide
1033   sequences guided by amino acid translations. Nucleic Acids Res. 38:7–13.
1034 Alexander M.P. 1969. Differential Staining of Aborted and Nonaborted Pollen. Stain Technol.
1035   44:117–122.
1036 Alexander M.P. 1980. A Versatile Stain for Pollen Fungi, Yeast and Bacteria. Stain Technol.
1037   55:13–18.
1038 ArborBiosciences. 2016. In-Solution Sequence Capture for Targeted High-Throughput
1039   Sequencing: Version 3.02. .
1040 Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A python package for
1041   genome-scale hybridization detection. Syst. Biol. 67:821–829.
1042 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
1043   sequence data. Bioinformatics. 30:2114–2120.
1044 Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model.
1045   Bioinformatics. 30:3317–3324.
1046 Chifman J., Kubatko L. 2015. Identifiability of the unrooted species tree topology under the
1047   coalescent model with time-reversible substitution processes, site-specific rate variation,
1048   and invariable sites. J. Theor. Biol. 374:35–47.
1049 Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E.,
1050   Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R. 2011. The variant call format
1051   and VCFtools. Bioinformatics. 27:2156–2158.
1052 Doyle J.J. 1987. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue.
1053   Phytochem Bull.:11–15.
1054 Duarte J.M., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J.,
1055   dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in Arabidopsis,
1056   Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels.
1057   BMC Evol Biol. 10:61.

1058 Evans M.E.K., Smith S. a, Flynn R.S., Donoghue M.J. 2009. Climate, niche evolution, and
1059   diversification of the "bird-cage" evening primroses (Oenothera, sections Anogra and
1060   Kleinia). Am. Nat. 173:225–40.
1061 G Yu, DK Smith, H Zhu, Y Guan T.L. 2017. ggtree: an R package for visualization and
1062   annotation of phylogenetic trees with their covariates and other associated data. Methods
1063   Ecol. Evol.:28–36.
1064 Green R., Krause J., Briggs A., Rasilla Vives M., Fortea Pérez F. 2010. A draft sequence of the
1065   neandertal genome. Science (80-.). 328:710–722.
1066 Gutíerrez S.C., Martínez J.M.S., Gabaldón T. 2009. TrimAl : a tool for automatic alignment
1067   trimming. Bioinformatics.:1–2.
1068 Illumina. 2016. BaseSpace Sequence Hub. Available from https://basespace.illumina.com/.
1069 Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett
1070   N.J. 2016. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-
1071   Throughput Sequencing Reads Using Target Enrichment. Appl. Plant Sci. 4:1600016.
1072 Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple
1073   sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
1074 Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
1075   Bioinformatics. 25:1754–1760.
1076 Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for
1077   inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010.
1078 Minh B.Q., Hahn M.W., Lanfear R. 2018. New methods to calculate concordance factors for
1079   phylogenomic datasets. bioRxiv.:487801.
1080 Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many
1081   hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.
1082 Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R
1083   language. Bioinformatics. 20:289–290.
1084 Poplin R., Ruano-Rubio V., DePristo M.A., Fennell T.J., Carneiro M.O., Auwera G.A. Van der,
1085   Kling D.E., Gauthier L.D., Levy-Moonshine A., Roazen D., Shakir K., Thibault J.,
1086   Chandran S., Whelan C., Lek M., Gabriel S., Daly M.J., Neale B., MacArthur D.G., Banks
1087   E. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples.
1088   bioRxiv.:201178.
1089 Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar
1090   P., De Bakker P.I.W., Daly M.J., Sham P.C. 2007. PLINK: A tool set for whole-genome
1091   association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.
1092 R Core Team. 2018. R: A language and environment for statistical computing. .
1093 Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from
1094   quartet frequencies. Mol. Biol. Evol.:msw079.
1095 Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.
1096 Sharma P., Purohit S.D. 2012. An improved method of DNA isolation from polysaccharide rich
1097   leaves of Boswellia serrata Roxb. Indian J. Biotechnol. 11:67–71.
1098 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets reveals
1099   conflict, concordance, and gene duplications with examples from animals and plants. BMC
1100   Evol. Biol. 15:150.
1101 Stamatakis A. 2014. Stamatakis - 2014 - RAxML version 8 a tool for phylogenetic analysis and
1102   post-analysis of large phylogenies. :2010–2011.
1103 Sukumaran J. and M.T.H. 2010. DendroPy: A Python library for phylogenetic computing.

1104 Bioinformatics.:1569–1571.

1105 Suyama M., Torrents D., Bork P. 2006. PAL2NAL: Robust conversion of protein sequence
1106 alignments into the corresponding codon alignments. Nucleic Acids Res. 34:609–612.

1107 Towner H.F.. 1977. The Biosystematics of Calylophus (Onagraceae). Ann. Missouri Bot. Gard.
1108 64:48–120.

1109 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

1110 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree
1111 reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153.

1112

1113   S4.

1114
1115   Concatenation tree constructed using the exon-only dataset and 100 bootstraps. Bootstrap values
1116   indicated at relevant nodes.

1117   S5.

1118
1119   Concatenation tree constructed using the supercontig dataset and 100 bootstraps. Bootstrap
1120   values indicated at relevant nodes.

1121   S6.

1122
1123   ASTRAL-II summary coalescent tree constructed using the exon-only dataset and 100
1124   bootstraps. Bootstrap values indicated at relevant nodes.
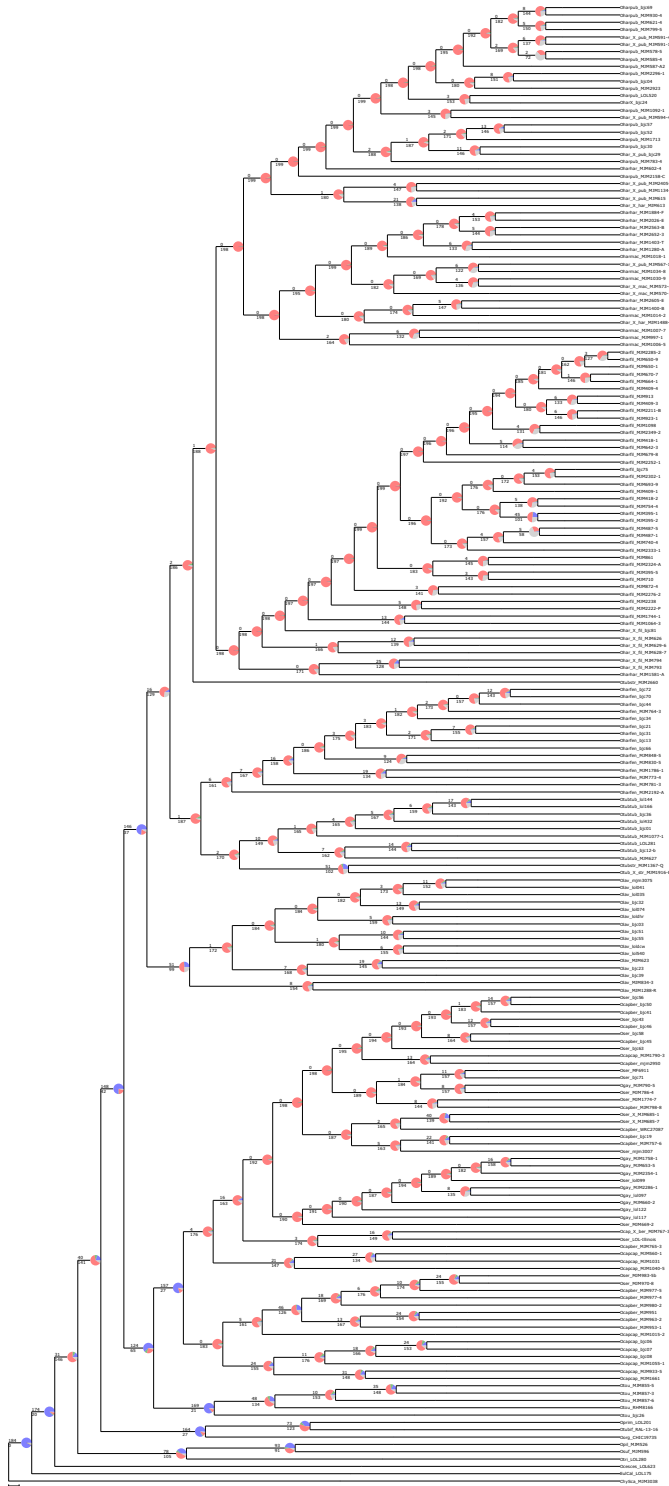
1125   S7.

1126
1127   SVD Quartets summary coalescent tree constructed using the supercontig dataset with 100
1128   bootstraps. Bootstrap values indicated at relevant nodes.

1129

1130   S8.

1131
1132   Phyparts piecharts ASTRAL-II tree constructed using the supercontig dataset. Piechart colors
1133   correspond to: blue = concordant, green = top alternative bipartition, red = all other alternative
1134   bipartitions, black = uninformative for that node.

45

1135 S9.

1136 Summary of support values for current and proposed taxonomic treatments, by analysis. 'e'

1137 signifies trees based on exon-only data, "e+i" signifies trees based on supercontigs. '*p'* indicates

1138 paraphyletic, unsupported taxon treatment according to tree topology.

	Taxon	Concat. e	Concat. e+i	ASTRAL-II e	ASTRAL-II e+i	SVDQ e+i	Phyparts	gCF iQtree e+i	sCF iQtree e+i
a.	*Oenothera hartwegii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
b.	*Oenothera hartwegii* subsp. *fendleri* (Towner 1977)	<50	*p*	81	99	*p*	1/ 188	1	38
c.	*Oenothera hartwegii* subsp. *filifolia* (Towner 1977)	*p*	*p*	<50	<50	<50	1/ 190	0	36
d.	*Oenothera hartwegii* subsp. *hartwegii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
e.	*Oenothera hartwegii* subsp. *macartii* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	
f.	*Oenothera hartwegii* subsp. *hartwegii* + subsp. *macartii*	*p*	*p*	<50	<50	<50	0/ 200	0	33
g.	*Oenothera hartwegii* subsp. *pubescens* (Towner 1977)	*p*	*p*	<50	<50	<50	0/ 200	0	36
h.	*Oenothera lavandulifolia* (Towner 1977)	100	*p*	100	100	100	147/37	74	83
i.	*Oenothera tubicula* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*1/ 188*	*p*	*p*
j.	*Oenothera tubicula* subsp. *tubicula* (Towner 1977)	*p*	*p*	81	99	<50	2/ 171	1	38
k.	*Oenothera tubicula* subsp. *strigulosa* (Towner 1977)	*p*	*p*	81	100	<50	*p*	0	37
l.	*Oenothera fendleri* + *Oenothera tubicula* subsp. *tubicula*	*p*	*p*	100	100	<50	16/129	10	40
m.	*Oenothera toumeyi* (Towner 1977)	100	100	100	100	100	125/65	62	56
n.	*Oenothera toumeyi* + subsect. *Calylophus*	100	100	100	100	100	149/42	76	65
o.	*Oenothera capillifolia* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
p.	*Oenothera capillifolia* subsp. *capillifolia* (Towner 1977)	*p*	*p*	<50	*p*	<77	*p*	*p*	*p*
q.	*Oenothera capillifolia* subsp. *berlandieri* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
r.	*Oenothera capillifolia* subsp. *berlandieri* (New Taxon; South Texas coastal populations)	*p*	*p*	<50	100	100	*157/27*	79	86
s.	*Oenothera serrulata* (Towner 1977)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
t.	*Oenothera 'australis'* (New Taxon; South Texas *O. serrulata*)	*p*	100	*p*	*p*	*p*	10/175	*p*	*p*
u.	*Oenothera gayleana* (Turner & Moore 2014)	*p*	*p*	*p*	*p*	*p*	*p*	*p*	*p*
v.	*Oenothera gayleana* (Revised Taxon)	*p*	*p*	*p*	<50	<50	0/ 193	0	35

1139

1140   S10.
1141   List of admixture hypotheses tested, Zscore, P-value and Gamma results from HyDe analysis.
1142   Each row represents a triplet set that was tested consisting of a putative hybrid individual and
1143   two parent groups; Parent 1 and Parent 2 (see "HyDe Group" Appendix I for group
1144   membership).
1145

Putative hybrid individuals	Parent 1 Group	Parent 2 Group	Zscore	P	Gamma
Ohar_X_fil_MJM629.6	CoreOharfil	CoreOharpub	-99999.9	1	-0.049
Ohar_X_fil_MJM628	CoreOharfil	CoreOharpub	-99999.9	1	-0.057
Ohar_X_fil_MJM626	CoreOharfil	CoreOharpub	-99999.9	1	-0.001
Ohar_X_fil_MJM793	CoreOharfil	CoreOharpub	-99999.9	1	-0.057
Ohar_X_fil_MJM794	CoreOharfil	CoreOharpub	-99999.9	1	-0.034
Ohar_X_fil_BJC81	CoreOharfil	CoreOharpub	-99999.9	1	-0.042
Oharhar_MJM1581	CoreOharpub	Oharhar/harmac	-99999.9	1	1.706
Ohar_X_pub_BJC29	CoreOharpub	Oharhar/harmac	2.378	0.009**	0.338
Oharpub_BJC30	CoreOharpub	Oharhar/harmac	-99999.9	1	-0.215
Oharpub_MJM2158.C	CoreOharpub	Oharhar/harmac	1.113	0.133	0.229
Ohar_X_pub_MJM1134.2	CoreOharpub	Oharhar/harmac	0.043	0.483	0.015
Ohar_X_pub_MJM2405.A	CoreOharpub	Oharhar/harmac	0.882	0.189	0.158
Ohar_X_pub_MJM594.4	CoreOharpub	Oharhar/harmac	2.094	0.018*	0.332
Ohar_X_pub_MJM615	CoreOharpub	Oharhar/harmac	0.413	0.340	0.096
Oharmac_MJM997.1	CoreOharpub	Oharhar/harmac	1.563	0.059	0.224
Ohar_X_har_MJM613	CoreOharpub	Oharhar/harmac	-99999.9	1	-0.495
Otub_X_str_MJM1916.E	Otubtub	Oharhar/harmac	5.585	0.000***	0.947
Oharfen_MJM2192.A	CoreOharfen	CoreOharfil	0.305	0.380	0.001
Oharfen_MJM781.3	CoreOharfen	CoreOharpub	0.025	0.490	9.69E-05
Ocapcap_MJM1015.2	CoreOcapcap	S. TX Ocapber	0.467	0.320	0.003
Ocap_X_ber_MJM767.3	central OK Ocapber	Pecos River/Southern Plains	-99999.9	1	-0.003

•   Indicates value meets significance threshold of $P < 0.05$
**   Indicates value meets significance threshold of $P < 0.001$
*** Indicates value meets significance threshold of $P < 0.0001$

1146

1147   S11.
1148   See excel table "S11 Morphometric Data".

1149   S12.
1150   See excel table "S12 Pollen Counts".
1151

47

S13.

Oenothera HybSeq CTAB Silica DNA extractions
b. cooper 11/10/2015

Adapted from JJSA Protocol, J. Fant's Protocol and Sharma and Puohit (2012) for Silica-dried Leaves of *Oenothera* sect. *Calylophus*

SAMPLE PREPARATION & GRINDING

1. Warm CTAB buffer to 65°C in water bath in the hood and pour liquid nitrogen into dipping canister (You will only need a few inches of liquid in the Dewar, make sure to pour it back into the large holding canister when you are done to prevent evaporation).

2. Put a small volume of 1:1 sterilized sand and 3 metal/ceramic beads into centrifuge tubes, then put tubes in freezer (-20) until envelopes are ready.

3. Keeping envelopes/leaf tissue on ice to prevent thawing; take envelopes and centrifuge tubes out of freezer and weigh/estimate small amount of dried tissue (e.g., 10-50 mg of leaves), put in labeled microfuge tubes. After each tube is filled, place tube in freezer.
   a. Make sure leaf tissue is broken up into small pieces when adding to ensure full grinding.
   b. Keeping tissue frozen prevents phenolic compounds from oxidizing, and enzymes from activating and breaking down DNA
   c. Follow DNA free lab protocols using bleach to prevent contamination/cross contamination of samples.

4. Mix 45 (= .3%) µl 2-mercaptoethanol to 15 ml CTAB buffer in small beaker (enough for ~15 samples...double for 30 samples, etc.)
   a. Mercaptoethanol must be measured and added in the fume hood.
   b. CTAB is a detergent that lyses (breaks down) the cell membranes. The PVP in the buffer helps bind the polysaccharides and might co-ppt

5. Dip tubes in Liquid Nitrogen (or store in -80 for ~30 minutes) and **immediately** Grind tissue **thoroughly** in Fast Prep machine (thorough grinding may require 2-4 cycles, re-dip tubes in liquid nitrogen after 2 cycles to ensure no thawing occurs).

6. **Immediately** after grinding add 600 µl of CTAB/mercaptoethanol buffer to each tube
   a. **BE CAREFUL.** It is important to prevent cross-contamination when opening tubes. **Powdered tissue can spread easily through the air or on your fingers.** Make sure open tubes are spaced apart from one another on the tray; open each tube away from other open tubes/caps; and place caps a safe distance from one another. Make sure leaf powder is not stuck to your gloves before touching consecutive tubes/caps.
   b. Grind on FastPrep for one cycle

EXTRACTION AND ISOLATION OF DNA (once leaf tissue is well-ground)

1. Incubate tubes for 60 minutes at 60°C on Thermoblock or water bath
   a. Shake tubes several times ~ every 10 minutes during incubation (for compacted samples, vortex on high setting to ensure mixing)
   b. When incubation is complete, turn Thermoblock down to 37°C

2. Add 600 µL chloroform-isoamyl (24:1) to tube and vortex on High setting to ensure thorough mixing.
   a. **BE CAREFUL.** It is important to prevent cross-contamination at this step when opening/handling tubes. Follow steps outlined previously.
   b. Chlorophyll and other pigments get transferred into the CI layer because pigments are non-polar and dissolve into the highly non-polar chloroform. The DNA remains in the aqueous layer.

3. Spin for 10 min at 9,000 RPM.
   a. After spinning, the upper layer of each tube should have a clear liquid: this is where your DNA is.

4. Transfer top (clear) liquid to a new (labeled) 1.5 ml eppindorf tube. **Do not be Greedy**, leaving some liquid in tube is fine – the important thing is not to pipette **ANY** interphase material.
   a. A yield of ~400-500 µL of supernatant is common.
   b. If you accidently disturb the interphase, or believe that you have pipetted interphase, return supernatant to microfuge tube and re-centrifuge for 10 min at 9000 rpm. Repeat step 4.

DNA PRECIPITATION

1. Add 1/10 volume 3M sodium acetate and ½ volume 5M NaCl (i.e. 500ul of solution add 250 ul of NaCl, 50 ul of Na-Acetate).

2. Add 2/3rd volume of cold isopropanol (to previous volumes = 500 ul isop)
   a. Isoproponal is heavy alcohol so allows DNA ppt with lower volume

48

1207   3.   Mix & store in -20°C (freezer) for 1-2 hours
1208   *(NOTE: POTENTIAL STOPPING POINT FOR A DAY OR MORE…SAMPLES CAN STAY AT -20°C FOR DAYS)*
1209   4.   Spin for 3 min at **10K-12K RPM.**
1210   5.   Decant the supernatant & drain; make sure pellet stays at bottom of tube.
1211   a.   Pellets are usually white but sometimes they can even be brown, this is not necessarily a problem.
1212   6.   Wash pellet with 500ul of 70% EtOH & flick tubes to dislodge pellet from bottom, then vortex on high to
1213   clean.
1214   7.   Spin for 3 min at 10K-12K RPM.
1215   8.   Decant the supernatant making sure pellet stays at bottom of tube
1216   9.   Leave tubes open and place tubes on Speedvac (vacuum on -2.5, heat = low, spin) for ~20 minutes or until
1217   all EtOH has evaporated.
1218   a.   If pellets are not dry after 20 minutes, check frequently to prevent over drying which makes the pellet
1219   difficult to re-suspend.
1220   b.   Alternatively, can air dry on thermoblock
1221   *(NOTE: POTENTIAL STOPPING POINT …SAMPLES CAN STAY AT 20°C FOR DAYS – place in cupboard but cover with kimwipe)*
1222   FINAL CLEANING & RESUSPENSION OF DNA
1223   1.   Add 500 ul of 0.5 mL High Salt TE Buffer and close lids
1224   2.   Incubate tubes on Thermoblock at 37-50°C (45) until pellet dissolves (15-30 minutes);
1225   a.   Vortex frequently if this is taking a long time
1226   *(NOTE: POTENTIAL STOPPING POINT FOR A DAY OR MORE…SAMPLES CAN STAY AT -20°C FOR DAYS)*
1227   3.   Add 3 volumes of binding buffer (3 volumes for every 1 volume DNA pellet = ~100-150 ul BB), **let stand for
1228   20 minutes**
1229   4.   Centrifuge at 550 g for 10 min
1230   5.   transfer supernatant to clean 2ml eppi tube leaving any colored or gelatinous precipitate behind.
1231   6.   Add 300 µl of silica suspension and mix for 30 minutes by regular and frequent gentle inversion
1232   7.   Centrifuge at 550 g for 10 minutes, discard supernatent
1233   8.   Re-suspend silica pellet in 1.5 mL of wash buffer 1
1234   9.   Centrifuge at 3000 g for 15 seconds, decant supernatent
1235   10. Re-suspend silica pellet in 1.5 mL of wash buffer 2
1236   11. Centrifuge at 3000 g for 15 seconds, decant supernatant, dry pellet completely on speed vac
1237   12. suspend silica pellet in 300 ul TE buffer, incubate at 50° mixing regularly by vortex
1238   13. centrifuge at 11600 g for 1 min. then transfer supernatant into clean 1.5 mL eppi tube.
1239   14. Precipitate DNA by adding 50 ul of 3M sodium Acetate and 500 uL 100% Etoh
1240   15. Store at -20° C for at least 2 hours
1241   *(NOTE: POTENTIAL STOPPING POINT FOR A DAY OR MORE…SAMPLES CAN STAY AT -20°C FOR DAYS)*
1242   16. Centrifuge at 11600 g for 5 min, discard supernatant, dry pellet on speed vac
1243   17. Dissolve in 50 uL 1X TE buffer (it may not look like there is anything in the tube at this point, but there is
1244   lots of pure DNA in there! Add the TE buffer)
1245   18. Store in -20°C (freezer)
1246   19. In rare circumstances that an extraction still has coloration at this stage (yellow or brown usually) you can
1247   now put it through a Qiagen Qiaquick spin column or other proprietary cleanup column to further purify
1248   the DNA.
1249
1250   REAGENTS (recipes below) and SUPPLIES
1251   •   CTAB buffer (in glass container in refrigerator)
1252   •   Liquid Nitrogen
1253   •   2-mercaptoethanol (in fume hood with gloves)
1254   •   24:1Chloroform:isoamyl alcohol - glass container in fumehood use gloves and keep in hood -(Or 25:24:1
1255   phenol-chloroform:isoamyl)
1256   •   5M Nacl
1257   •   3M M sodium acetate
1258   •   Isopropanol and 70% and 100% EtOH
1259   •   High Salt TE buffer
1260   •   Binding Buffer

1261 • Silica Matrix
1262 • Wash Buffer 1
1263 • Wash Buffer 2
1264 • 1X TE buffer
1265 • 1.5 ml eppi tubes
1266 • 2 ml eppi tubes
1267

1268 **CTAB BUFFER** [100 mM Tris-HCl (pH=8.0), 1.4 M NaCl, 20 mM EDTA, 2% (or 4%) CTAB
1269 (hexadecyltrimethylammonium bromide), 2% PVP-40 (polyvinylpyrollidone, m.w. 40000), + 0.3% β-
1270 mercaptoethanol (add in fume hood in a beaker to a premeasured volume of CTAB buffer required for the number
1271 of extractions planned plus one).]
1272
1273 For 200 mL CTAB buffer 2% (4%):
1274 20 ml 1M Tris-Cl
1275 1.4 M NaCl (16.36 g NaCl)
1276 • The addition of NaCl at concentrations higher than 0.5 M, along with CTAB, is known to remove
1277 polysaccharides during DNA extraction
1278 8 ml 0.5 M EDTA
1279 4 g CTAB (8 g for 4%)
1280 4 g PVP-40
1281 ~120 ml dH2O, then fill to 200 mL
1282
1283 **TE buffer**: [10 mM Tris-HCl (pH=8.0), 1 mM EDTA.]

1284 10 mL 1 M Tris-HCl
1285 2 ml 0.5 M EDTA
1286 982 ml dH2O
1287
1288 **High Salt TE Buffer**: [10 mM Tris-HCl (pH=8.0), 1 mM EDTA, 1M NaCl.]
1289 10 ml 1 M Tris-HCl
1290 2 ml 0.5 M EDTA
1291 200 ml 5M NaCl
1292 788 ml dH2O
1293
1294 **Binding Buffer**: [50mM Tris (pH7.5), 6M NaClO4, 1mM EDTA]
1295 For 200 ml:
1296 10mL Tris
1297 168.552 g NaClO4
1298 400 ul EDTA
1299 top off to 200 mL with dH20
1300
1301 **Wash Buffer 1**: [3 volumes binding buffer, 1 volume water]
1302 For 60 samples:
1303 90 ml Binding Buffer
1304 30 ml dH20
1305
1306 **Wash Buffer 2**: [1 volume 40 mM Tris (pH 8.0), 4 mM EDTA, 0.8 M NaCl, 1 Volume ethanol]
1307 For 60 samples:
1308 1.8 ml 1M tris
1309 360 ul .5M EDTA
1310 7.2 ml 5M NaCL

1311   35.64 ml dh20
1312   45 ml 100% etoh
1313
1314   **24:1 Chloroform: Isoamyl**
1315   240ml Chloroform
1316   10ml Isoamyl alcohol
1317
1318   25:24:1 Chloroform: Isoamyl Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA. – Purchase to ensure right pH
1319   250ml Phenol
1320   240ml Chloroform
1321   10ml Isoamyl alcohol
1322
1323   **Silica suspension (use lab grade silicon dioxide 99%)**

1. Suspend silicon dioxide powder in ~20 volumes of water (50 ml : 1L)
    a. use magnetic stirring rod to bring silica into uniform suspension
2. Decant suspension into clean beaker leaving behind heavy sediment at bottom (discard this)
3. Allow suspension to settle by gravity for 2-3 hours
4. Decant and discard supernatant with fine material keeping silicon dioxide that has settled
5. Re-suspend silicon dioxide in dH20, repeating step until fine particles removed (2-3X total is good)
    a. use magnetic stirring rod to bring silica into uniform suspension
6. Re-suspend silicon dioxide in dH20 and transfer to 15 ml falcon tubes
7. Collect silicon dioxide by low speed centrifugation (up to 3200 rpm for 3 min)
8. Add ~15 ml of 10% bleach solution to each aliquot and mix frequently by inversion/vortexing for 15-30 minutes
9. Re-suspend diatomite in dH20 by vortexing, then centrifuge at low speed for 3 minutes to pellet
10. Discard supernatent
11. Repeat step 9 and 10 two more times
12. Re-suspend in dH20 and autoclave
13. Re-suspend silica dioxide pellet in 1 volume PCR grade DNA/RNA free H2O
14. Aliquot into 1.5 ml eppi tubes and store at 4° C
15. Vortex thoroughly to bring silica into uniform suspension before using

51

1361   S14.
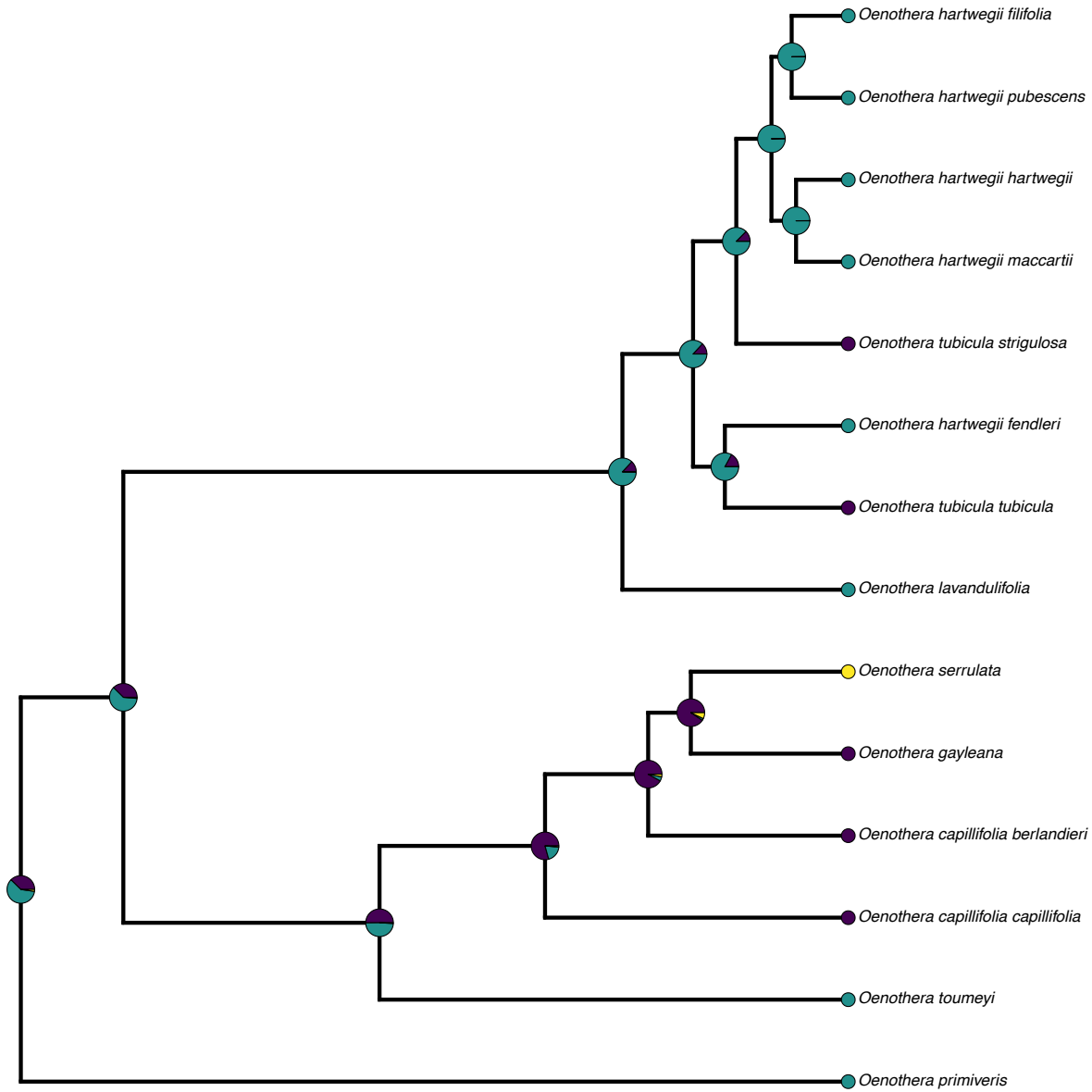
1362
1363   Ancestral State Reconstruction of reproductive system in sect. *Calylophus* using supercontigs
1364   and accessions grouped into taxa. Pie-charts on nodes represent likelihood of ancestral
1365   reproductive system at each node (teal = hawkmoth pollination, purple = bee pollination, yellow
1366   = PTH).
1367