

1 **Exploring the possible role of hybridization in the evolution of photosynthetic pathways in**  
2 ***Flaveria* (Asteraceae), the prime model of C<sub>4</sub> photosynthesis evolution**

3

4

5 Diego F. Morales-Briones<sup>1\*</sup> and Gudrun Kadereit<sup>1\*</sup>

6

7 <sup>1</sup> Princess Therese von Bayern chair of Systematics, Biodiversity and Evolution of Plants,  
8 Ludwig-Maximilians-Universität München, Munich, Germany.

9

10 **\* Correspondence:**

11 Gudrun Kadereit

12 [g.kadereit@biologie.uni-muenchen.de](mailto:g.kadereit@biologie.uni-muenchen.de)

13

14 Diego F. Morales-Briones

15 [dfmoralesb@gmail.com](mailto:dfmoralesb@gmail.com)

16

17

18 **Abstract**

19 *Flaveria* (Asteraceae) is the prime model for the study of C<sub>4</sub> photosynthesis evolution and seems  
20 to support a stepwise acquisition of the pathway through C<sub>3</sub>-C<sub>4</sub> intermediate phenotypes, still  
21 existing in *Flaveria* today. Molecular phylogenies of *Flaveria* based on concatenated data  
22 matrices are currently used to reconstruct the complex sequence of trait shifts during C<sub>4</sub>  
23 evolution. To assess the possible role of hybridization in C<sub>4</sub> evolution in *Flaveria*, we re-  
24 analyzed transcriptome data of 17 *Flaveria* species to infer the extent of gene tree discordance  
25 and possible reticulation events. We found massive gene tree discordance as well as reticulation  
26 along the backbone and within clades containing C<sub>3</sub>-C<sub>4</sub> intermediate and C<sub>4</sub>-like species. An  
27 early hybridization event between two C<sub>3</sub> species might have triggered C<sub>4</sub> evolution in the genus.  
28 The clade containing all C<sub>4</sub> species plus the C<sub>4</sub>-like species *F. vaginata* and *F. palmeri* is robust  
29 but of hybrid origin involving *F. angustifolia* and *F. sonorensis* (both C<sub>3</sub>-C<sub>4</sub> intermediate) as  
30 parental lineages. Hybridization seems to be a driver of C<sub>4</sub> evolution in *Flaveria* and likely  
31 promoted the fast acquisition of C<sub>4</sub> traits. This new insight can be used in further exploring C<sub>4</sub>  
32 evolution and can inform C<sub>4</sub> bioengineering efforts.

33

34 **Keywords:** C<sub>3</sub>-C<sub>4</sub> intermediates; C<sub>4</sub> photosynthesis evolution; Hybridization; HyDe analysis;  
35 RNA-Seq; Reticulate evolution; Species network analysis.

36

37

38

## 39 INTRODUCTION

40 The detection of gene tree discordance is common in the phylogenomic era. Discordance can be  
41 the product of multiple processes and is commonly attributed to either incomplete lineage sorting  
42 (ILS) and/or hybridization (Pamilo and Nei, 1988; Doyle, 1992; Galtier and Daubin, 2008).  
43 Hybridization is a fundamental process in the evolution of animals, plants, and fungi (Giraud et  
44 al., 2008; Schwenk et al., 2008; Soltis and Soltis, 2009; Payseur and Rieseberg, 2016), and  
45 methods to investigate hybridization in a phylogenetic context recently have been developed  
46 greatly. These include methods that estimate phylogenetic networks while accounting for ILS  
47 and hybridization simultaneously (e.g., Solís-Lemus and Ané, 2016; Wen et al., 2018), and  
48 methods that detect hybridization based on site patterns or phylogenetic invariants (e.g., (Green  
49 et al., 2010; Durand et al., 2011; Kubatko and Chifman, 2019). The current ease to produce  
50 phylogenomic data sets and the availability of new analytical methods facilitate the exploration  
51 of reticulate evolution in any clade across the Tree of Life, including those that have particular  
52 significance as model lineages, such as the flowering plant genus *Flaveria* (Asteraceae) for the  
53 study of C<sub>4</sub> photosynthesis evolution.

54 *Flaveria* belongs to the sunflower tribe Heliantheae (Anderberg et al., 2007). According  
55 to the most recent revision by Powell (1978), *Flaveria* includes 21 morphologically rather  
56 similar species distributed mainly in southern USA and northern Mexico, with few species  
57 occurring in the Caribbean and South America. The two weedy and self-compatible C<sub>4</sub> species,  
58 *F. trinervia* and *F. bidentis*, have been introduced almost worldwide  
59 (<https://powo.science.kew.org/>). Species of *Flaveria* usually show scattered occurrences in  
60 unconnected, localized populations near rivers, creeks, irrigation canals, fields, roadsides, and  
61 ponds, often on saline or gypseous soils (Powell, 1978). They are either robust shrubs or

62 herbaceous perennials, or annuals (mainly the C<sub>4</sub> species). The genus stands out in Asteraceae  
63 for its reduced floral features and reduced and secondarily aggregated capitula (Anderberg et al.,  
64 2007). Reduction is most evident in *F. trinervia* (C<sub>4</sub>) and aggregation of capitula mimicking a  
65 single capitulum in *F. anomala* (C<sub>3</sub>-C<sub>4</sub>). *Flaveria* is consistently diploid (see Powell, 1978 and  
66 ref. therein; only exceptions are some tetraploid populations of *F. pringlei*) with a haploid  
67 chromosome number of  $n = 18$ . Artificial hybridization among 16 species of *Flaveria* was  
68 successful, and F1 hybrids could be obtained in most species' combinations (see Table 1 in  
69 Powell, 1978). F2 and backcross crosses also resulted in offspring in a high number of  
70 combinations, but only four of these were fertile. Powell (1978) excluded frequent natural  
71 hybridization in *Flaveria* mainly because of geographical isolation.

72 C<sub>4</sub> photosynthesis in *Flaveria* was first recognized by Smith and Turner (1975), and the  
73 presence of C<sub>3</sub>-C<sub>4</sub> intermediate species was first noted by Brown (pers. comm. in Powell, 1978),  
74 first verified by Apel and Maass (1981) and then studied in detail biochemically in four species  
75 by Ku et al. (1983) and Nakamoto et al. (1983). Numerous publications characterizing the  
76 physiology and biochemistry of C<sub>3</sub>-C<sub>4</sub> intermediate species of *Flaveria* followed (Ku et al., 1991  
77 and ref. therein). At the same time crossing experiments of C<sub>3</sub> and C<sub>4</sub> *Flaveria* species as well as  
78 backcrosses or crosses between C<sub>3</sub>-C<sub>4</sub> intermediate species revealed the transfer of C<sub>4</sub> properties  
79 as well as the simultaneous functioning of C<sub>3</sub> and C<sub>4</sub> pathways in hybrids (see Apel et al., 1988  
80 as an example and Kadereit et al., 2017 for review). Of all genera that contain C<sub>3</sub>-C<sub>4</sub> intermediate  
81 species, *Flaveria* has the highest diversity of C<sub>3</sub>-C<sub>4</sub> phenotypes including C<sub>2</sub> photosynthesis  
82 (Sage et al., 2012; briefly described in Table 1), and arguably is the only lineage that allows to  
83 infer a detailed sequence of increasing C<sub>4</sub>-ness (Sage et al., 2012). Against this background,  
84 *Flaveria* qualified as the model group for the establishment (Monson and Moore, 1989) and

85 subsequent refinement of a model of stepwise acquisition of C<sub>4</sub> photosynthesis through  
86 intermediate steps (Sage et al., 2014 and ref. therein).

87         One challenge of the *Flaveria* model is that C<sub>3</sub>-C<sub>4</sub> intermediate phenotypes might also  
88 have resulted from reticulation during the diversification of the genus, especially when  
89 reproductive barriers are leaky among extant species as soon as they get into contact (Powell,  
90 1978) and hybridization between C<sub>3</sub> and C<sub>4</sub> species is possible. Therefore, a phylogenetic study  
91 exploring the occurrence and location of past reticulation events is needed. Comprehensive  
92 molecular phylogenetic studies of *Flaveria* published so far were either based on few molecular  
93 markers only (McKown et al., 2005), or on concatenated data matrices and inference methods  
94 unable to reveal tree discordance, possible reticulation, or incomplete lineage sorting (Lyu et al.,  
95 2015). Irrespective of this, numerous current studies of evolutionary change during the  
96 establishment of the C<sub>4</sub> pathway rely on the *Flaveria* model (e.g., Lyu et al., 2021; Taniguchi et  
97 al., 2021).

98         The aim of this study is to use available transcriptome data of 17 species of *Flaveria* to  
99 assess the extent of reticulation during the diversification of the genus, and to evaluate these  
100 findings with respect to the evolution of C<sub>4</sub> photosynthesis in the genus and its suitability as  
101 general model of C<sub>4</sub> evolution.

102

103

104

105

106

107 **Table 1.** Photosynthetic types in *Flaveria* according to Sage et al. (2014 and 2018 and ref.  
 108 therein); species names marked \* were not sampled in this study, mesophyll (M), bundle sheath  
 109 (BS), glycine decarboxylase (GDC).

Photosynthetic type	Characteristics	<i>Flaveria</i> species representing this type <sup>1</sup>
C <sub>3</sub>	Photorespiratory cycle operates completely within single M cells, BS cells small <sup>2</sup> with few organelles, veins widely spaced	<i>F. cronquistii</i> <sup>2</sup> , <i>F. mcdougallii</i> *
C <sub>3</sub> proto-kranz	Functionally C <sub>3</sub> , activated BS cells, greater vein density, mitochondria localized at the inner BS wall adjacent to the vasculature	<i>F. pringlei</i> , <i>F. robusta</i>
C <sub>2</sub> Type I	Low or no GDC expression in M cells, high number of centripetally located organelles in the BS cells, CO <sub>2</sub> compensation point reduced in comparison to C <sub>3</sub> , lack of any C <sub>4</sub> cycle	<i>F. angustifolia</i> , <i>F. chlorifolia</i> , <i>F. sonorensis</i>
C <sub>2</sub> Type II	In addition to type I, modest C <sub>4</sub> cycle enhancement	<i>F. anomala</i> , <i>F. floridana</i> , <i>F. linearis</i> *, <i>F. pubescens</i> , <i>F. ramosissima</i>
C <sub>4</sub> -like	Strong C <sub>4</sub> metabolic cycle but also weak C <sub>3</sub> cycle in the M cells	<i>F. brownii</i> , <i>F. palmeri</i> , <i>F. vaginata</i>
C <sub>4</sub>	No C <sub>3</sub> cycle, CO <sub>2</sub> -saturate photosynthesis below 500 ppm CO <sub>2</sub>	<i>F. bidentis</i> , <i>F. campestre</i> *, <i>F. kochiana</i> , <i>F. trinervia</i> , <i>F. australasica</i> ,

110 <sup>1</sup> *Flaveria oppositifolia*\* so far only classified as C<sub>2</sub> without further specification.

111 <sup>2</sup> *Flaveria cronquistii* qualifies as a C<sub>3</sub>+ species (see Sage et al., 2018) because it has larger and  
 112 photosynthetically more active BS cells (McKown and Dengler, 2007).

113

## 114 MATERIAL AND METHODS

### 115 Taxon sampling

116 We included publicly available transcriptomes from 17 species of *Flaveria* (Table S1). In  
117 addition, we included outgroups from four genomes of Asteraceae (*Chrysanthemum*, *Helianthus*,  
118 *Lactuca* and *Stevia*) following (Mandel et al., 2019; Table S1).

119

### 120 Homology and orthology inference

121 Raw read processing, transcriptome assembly, low-quality and chimeric transcript  
122 removal, transcript clustering into putative genes, translation, and final coding sequences (CDS)  
123 redundancy assessment were carried out following Morales-Briones et al. (2021) with minor  
124 modifications as follows. Sequencing errors in raw reads were corrected with Rcorrector (Song  
125 and Florea, 2015) and reads flagged as uncorrectable were discarded. Sequencing adapters and  
126 low-quality bases were removed with Trimmomatic v 0.39 (Bolger et al., 2014). Additionally,  
127 chloroplast and mitochondrial reads were filtered out with Bowtie2 v 2.4.4 (Langmead and  
128 Salzberg, 2012) using publicly available Asterales organelle genomes from the Organelle  
129 Genome Resources database (RefSeq; [Pruitt et al., 2007]; last accessed on June 4, 2021) as  
130 references. Read quality was assessed with FastQC v 0.11.9  
131 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>) and overrepresented sequences were  
132 discarded. *De novo* assembly was carried out with Trinity v 2.13.2 (Grabherr et al., 2011) with  
133 default settings, but without in silico normalization. Assembly quality was assessed with  
134 Transrate v 1.0.3 (Smith-Unna et al., 2016). Low quality and poorly supported transcripts were  
135 removed using individual cut-off values for three contig score components of Transrate: 1)  
136 proportion of nucleotides in a contig that agrees in identity with the aligned read,  $s(\text{Cnuc}) \leq 0.25$ ;

137 2) proportion of nucleotides in a contig that have one or more mapped reads,  $s(\text{Ccov}) \leq 0.25$ ; and  
138 3) proportion of reads that map to the contig in correct orientation,  $s(\text{Cord}) \leq 0.5$ . Furthermore,  
139 chimeric transcripts (*trans*-self and *trans*-multi-gene) were removed following the approach  
140 described in Yang and Smith (2013) using *Helianthus annuus* as the reference proteome, and a  
141 percentage similarity and length cutoff of 30 and 100, respectively. To remove isoforms and  
142 assembly artifacts, filtered reads were remapped to filtered transcripts with Salmon v 1.5.2 (Patro  
143 et al., 2017) and putative genes were clustered with Corset v 1.09 (Davidson and Oshlack, 2014)  
144 using default settings, except that we used a minimum of five reads as threshold to remove  
145 transcripts with low coverage (-m 5). Only the longest transcript of each putative gene inferred  
146 by Corset was retained as suggested in Chen et al. (2019). Filtered transcripts were translated  
147 with TransDecoder v 5.3.0 (Haas et al., 2013) with default settings and the proteomes of  
148 *Arabidopsis thaliana*, *Helianthus annuus*, and *Lactuca sativa* to identify open reading frames.  
149 Finally, coding sequences (CDS) from translated amino acids were further reduced with CD-HIT  
150 v 4.8.1 (-c 0.99; [Fu et al., 2012]) to remove near-identical sequences. Scripts used can be found  
151 at [https://bitbucket.org/yanglab/phylogenomic\\_dataset\\_construction/src/master/](https://bitbucket.org/yanglab/phylogenomic_dataset_construction/src/master/) (Morales-  
152 Briones et al., 2021).

153 Homology inference was done with an all-by-all BLASTN search on CDS with an *E*  
154 value cutoff of 10. BLAST hits were filtered with a minimal hit coverage of 40%. Homolog  
155 groups were clustered with MCL v 14-137 (van Dongen, 2000) using a minimum minus log-  
156 transformed *E* value cutoff of 5 and an inflation value of 1.4, and only clusters with at least 17  
157 taxa were retained. Homolog cluster sequences were aligned using the OMM\_MACSE v 11.05  
158 pipeline (Scornavacca et al., 2019). Alignments were further trimmed to remove columns with  
159 more than 90% missing data using Phyx (Brown et al., 2017). Homolog trees were inferred using



160 RAxML v 8.2.11 (Stamatakis, 2014) with the GTRCAT model and 200 rapid bootstrap (BS)  
161 replicates. Monophyletic and paraphyletic tips of the same species were removed, keeping the tip  
162 with the highest number of characters in the trimmed alignment following Yang and Smith  
163 (2014). Spurious tips were detected and removed using TreeShrink v 1.3.9 (Mai and Mirarab,  
164 2018) with the 'per-gene' mode, a false positive error rate threshold ( $\alpha$ ) of 0.05, and excluding the  
165 outgroups. Trees were visually inspected, and deep paralogs producing internal branch lengths  
166 longer than 0.20 were cut apart retaining subclades with at least 10 taxa to obtain final homolog  
167 trees. Orthology inference was done using the 'monophyletic outgroup' (MO) approach from  
168 Yang and Smith (2014). The MO approach filters for trees that have outgroup taxa being  
169 monophyletic and single-copy, and therefore filters for single- and low-copy genes. This  
170 approach roots the gene tree by the outgroups, traverses the rooted tree from root to tip, and  
171 removes the side with less taxa when gene duplication is detected (Yang and Smith, 2014). If no  
172 taxon duplication is detected in a homolog tree, the MO approach outputs a one-to-one ortholog.  
173 We set all species of *Flaveria* as ingroups, and *Chrysanthemum*, *Helianthus*, *Lactuca*, and *Stevia*  
174 as outgroups, keeping orthologs that included at least 10 taxa.

175

### 176 **Tree inference and detection of gene tree conflict**

177 Sequences from individual orthologs were aligned using the OMM\_MACSE pipeline.  
178 Columns with more than 20% missing data were trimmed with Phyx, and only alignments with  
179 at least 500 characters and all 21 taxa were retained and concatenated. We estimated a maximum  
180 likelihood (ML) tree of the concatenated matrix with IQ-TREE v 2.1.3 (Minh et al., 2020)  
181 searching for the best partition scheme (Lanfear et al., 2012) followed by ML gene tree inference  
182 and 1000 ultrafast bootstrap replicates for clade support. To estimate a coalescent-based species

183 tree, first, we inferred individual gene trees with IQ-TREE using extended model selection  
184 (Kalyaanamoorthy et al., 2017) and 200 non-parametric bootstrap replicates for clade support.  
185 Gene trees were then used to infer a species tree with ASTRAL-III v 5.7.7 (Zhang et al., 2018)  
186 using local posterior probabilities (LPP; Sayyari and Mirarab, 2016) to assess clade support.

187 We explored gene tree discordance by calculating the number of concordant and  
188 discordant bipartitions on each node of the concatenated and ASTRAL trees using Phyparts  
189 (Smith et al., 2015). We used individual gene trees with BS support of at least 50% for each  
190 node. Additionally, to distinguish conflict from poorly supported branches, we carried out a  
191 Quartet Sampling (QS; Pease et al., 2018) analysis using the concatenated matrix with a partition  
192 by gene (--genetrees), the concatenated IQ-TREE and ASTRAL trees, and 1000 replicates. To  
193 further visualize gene tree conflict, we built a cloudogram with the DensiTree function from  
194 Phangorn v 2.7.1 (Schliep, 2011) in R (R Core Team 2021). We first time-calibrated individual  
195 ortholog gene trees, for visualization purposes only, with TreePL v 1.0 (Smith and O'Meara,  
196 2012). The most recent common ancestor (MRCA) of *Helianthus* and *Flaveria* was fixed to 21.5  
197 MYA, and the MRCA of *Flaveria* was fixed to 4.3 MYA based on Mandel et al. (2019).

198

### 199 **Testing for potential reticulation**

200 First, to investigate if gene tree discordance can be explained by ILS alone, we performed  
201 coalescent simulations like Cloutier et al. (2019). An ultrametric tree with branch lengths in  
202 mutational units ( $\mu T$ ) was inferred with PAUP v 4.0a (build 168; Swofford, 2002) by  
203 constraining a ML tree search to the ASTRAL tree and using the concatenated alignment, a  
204 GTRGAMMA model, and enforcing a strict molecular clock. The mutational branch lengths  
205 from the constrained tree and branch lengths in coalescent units ( $\tau = T/4N_e$ ) from the ASTRAL

206 tree were used to estimate the population size parameter theta ( $\Theta = \mu T / \tau$ ; Degnan and Rosenberg,  
207 2009) for internal branches. Terminal branches were set with  $\Theta = 1$ . We then used Phybase v 1.4  
208 (Liu and Yu, 2010), that implements the formula from Rannala and Yang (2003), to simulate  
209 10,000 gene trees using the constraint tree and the estimated theta values. Lastly, we calculated  
210 the distribution of Robinson and Foulds (1981) tree-to-tree distances between the ASTRAL tree  
211 and each original gene tree using Phangorn and compared this with the distribution of tree-to-tree  
212 distances between the ASTRAL tree and the simulated gene trees.

213         To test for potential reticulation, we inferred species networks using maximum pseudo-  
214 likelihood (Yu and Nakhleh, 2015) in PhyloNet v 3.8.2 (Than et al., 2008) with the command  
215 “InferNetworks\_MPL” and using individual ML gene trees as input. We included all 17 species  
216 of *Flaveria* and *Helianthus* as an outgroup (18-taxon data set). Network searches were performed  
217 allowing for up to ten reticulation events and using only nodes in the gene trees that have BS  $\geq$   
218 50%. To find the network with optimal number of reticulations, we plotted the number of  
219 reticulations versus the pseudo-likelihood score to determine when the score stabilizes (Blair and  
220 Ané, 2020). Given that the pseudo-likelihood scores did not stabilize after networks with ten  
221 reticulation events (see results) and to reduce network complexity, we performed one additional  
222 round of searches by removing taxa involved in reticulation from the previous search. We  
223 removed *F. brownii* (C<sub>4</sub>-like) and *F. pringlei* (C<sub>3</sub>; 16-taxon data set). These two species were  
224 inferred to be product of reticulation events in all ten original searches and were not involved in  
225 additional reticulation events (i.e., they are not a parental lineage of other reticulation events; see  
226 results). Network searches for the reduced data set were carried out similarly as with the original  
227 data set.

228

229           Additionally, we tested for hybridization with HyDe (Blischak et al., 2018), which uses  
230 site pattern frequencies (Kubatko and Chifman, 2019) to quantify the hybridization parameter  $\gamma$   
231 between two parental lineages that form a hybrid lineage. We tested all triplet combinations  
232 using the ‘run\_hyde.py’ script, the concatenated alignment, and a mapping file to assign species.  
233 Test significance was assessed with a Bonferroni correction ( $\alpha= 0.05$ ) for the number of  
234 hypothesis tests conducted with estimates of  $\gamma$  between 0 and 1 (Blischak et al., 2018). We  
235 carried out HyDe hybridization test using all species and without *F. pringlei* (C<sub>3</sub>), as it has been  
236 identified as a potential artificial hybrid (Lyu et al., 2015).

237

### 238 **Assessment of whole genome duplication**

239           To investigate potential whole genome duplication, product of reticulation events in  
240 *Flaveria* (see results), we mapped gene duplication events onto the inferred species tree  
241 following (Yang et al., 2018). First, we extracted rooted ingroup clades (orthogroup) from the  
242 final homolog trees by requiring at least 15 taxa and only orthogroups with an average BS  $\geq 50$   
243 were used for mapping. Then gene duplication events were then mapped onto the MRCA on the  
244 species tree when two or more taxa overlapped between the two daughter clades on the rooted  
245 ingroup clade. Each node on a species tree can be counted only once from each gene tree to  
246 avoid nested gene duplications inflating the number of recorded duplications (Yang et al., 2018).  
247 Orthogroup extraction and mapping were carried out using the scripts “extract\_clades.py” and  
248 “map\_dups\_mrca.py” from <https://bitbucket.org/blackrim/clustering> (Yang et al., 2018).

249

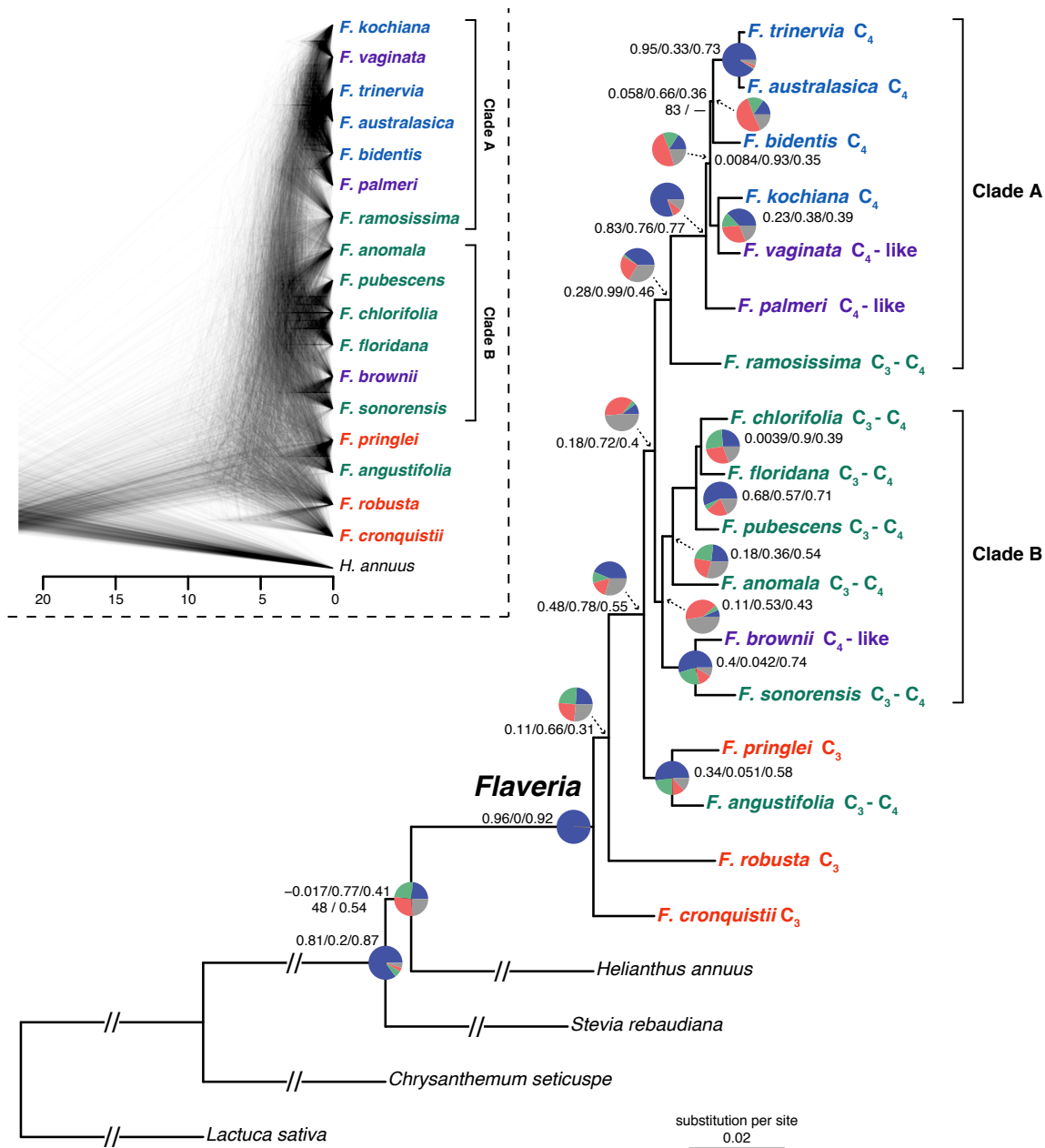
## 250 RESULTS

### 251 Orthology inference and phylogenetic analysis

252 The final number of orthologs with at least ten species was 5,374 with a mean of 4,809  
253 orthologs per species (Table S1). The number of orthologs that included all 21 species was  
254 1,295. The concatenated matrix ( $\geq 500$  bp per ortholog) consisted of 1,582,170 aligned columns  
255 with a character occupancy of 93% from 1,249 orthologs.

256 The topologies from the IQ-TREE and ASTRAL trees were similar and most nodes had  
257 maximum support (BS =100, LPP = 100; Fig. 1; Fig. S1). The two trees only differed in the  
258 placement of *Flaveria bidentis* (C<sub>4</sub>) which had lower support on both cases (Fig. 1; Fig. S1).  
259 *Flaveria cronquistii* (C<sub>3</sub>), *F. robusta* (C<sub>3</sub>) and *F. pringlei* (C<sub>3</sub>) + *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>) formed a  
260 grade which is sister to Clade A + Clade B (clades names followed McKown et al., 2005). Clade  
261 A from IQ-TREE comprised the same species and relationships as in Lyu et al. (2015). *Flaveria*  
262 *ramosissima* (C<sub>3</sub>-C<sub>4</sub>) and *F. palmeri* (C<sub>4</sub>-like) were successive sisters to *F. kochiana* (C<sub>4</sub>) + *F.*  
263 *vaginata* (C<sub>4</sub>-like) and the C<sub>4</sub> clade [(*F. bidentis*, *F. trinervia* + *F. australasica*); BS = 83]. The  
264 ASTRAL tree recovered *F. bidentis* (C<sub>4</sub>) sister to *F. kochiana* (C<sub>4</sub>) + *F. vaginata* (C<sub>4</sub>-like; LLP =  
265 0.96; Fig. S1). Clade B included the same species as in Lyu et al. (2015) but also included *F.*  
266 *sonorensis* (C<sub>3</sub>-C<sub>4</sub>). Our analyses recovered *F. brownii* (C<sub>4</sub>-like) + *F. sonorensis* (C<sub>3</sub>-C<sub>4</sub>) as sister  
267 to the remaining species in Clade B. The remaining species, all C<sub>3</sub>-C<sub>4</sub>, had similar relationships  
268 as in Lyu et al. (2015) with *F. anomala* as sister to the clade *F. pubescens*, *F. chlorifolia* + *F.*  
269 *floridana*.

270



271

272 **Fig. 1.** Maximum likelihood phylogeny of *Flaveria* inferred with IQ-TREE from the concatenated 1249-  
 273 nuclear gene supermatrix. Species names are colored by photosynthetic type. Quartet Sampling (QS)  
 274 scores are shown above branches. QS scores: Quartet concordance/Quartet differential/Quartet  
 275 informativeness. All nodes have full bootstrap support (BS =100) and local posterior probability (LLP =1)  
 276 unless noted next to branches. Em dashes (—) denotes an alternative topology compared to the ASTRAL  
 277 tree (Fig. S1). Pie charts represent the proportion of ortholog trees that support a clade (blue), the main  
 278 alternative bifurcation (green), the remaining alternatives (red), and bifurcations (conflict or support) with  
 279 <50% bootstrap support (gray). Branch lengths as number of substitutions per site (scale bar).  
 280 Exceptionally long branches were shortened with a broken segment (/) for illustration purposes (See Fig.  
 281 S1 for original branch lengths). Inset: Cloudogram inferred from 1288 nuclear ortholog trees. Scale in  
 282 millions of years ago (MA).

## 283 **Phylogenetic conflict**

284 Overall, conflict analyses and cloudogram visualization revealed rampant gene tree  
285 discordance in *Flaveria* (Fig. 1; Fig S2). The cloudogram showed significant conflict along the  
286 backbone of the phylogeny as well as within clades A and B (Fig. 1). The placement of *F.*  
287 *robusta* (C<sub>3</sub>) was supported only by 308 (of 947) informative gene trees and had low QS support  
288 (0.11/0.66/0.3) with a clear signal of an alternative topology involving *F. cronquistii*. The sister  
289 relationship of *F. pringlei* (C<sub>3</sub>) and *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>) was supported by 677 (of 1,123) gene  
290 trees and had moderate QS support (0.34/0.051/0.58) and a strong signal of an alternative  
291 topology. The placement of *F. pringlei* (C<sub>3</sub>) + *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>) as sister of Clade A +  
292 Clade B was supported by 562 (of 909) gene trees and had strong QS support (0.48/0.78/0.55)  
293 with moderate signal for an alternative topology. The clade composed of clades A and B was  
294 supported only by 131 (of 661) gene trees and had low QS support (0.18/0.72/0.4) but had no  
295 signals of an alternative topology. Clade A was supported by 505 (of 858) gene trees and had  
296 moderate QS support (0.28/0.99/0.46) but showed no signals of alternative topologies. The  
297 remaining species of Clade A [excluding *F. ramosissima* (C<sub>3</sub>-C<sub>4</sub>)] formed a clade supported by  
298 most gene trees (1,037 of 1,151) and a strong QS score (0.83/0.76/0.77) with no signals of an  
299 alternative topology. *Flaveria kochiana* (C<sub>4</sub>) + *F. vaginata* (C<sub>4</sub>-like) was supported by 479 (of  
300 1,049) gene trees and had moderate QS support (0.23/0.38/0.39) with a clear signal of an  
301 alternative topology. *Flaveria trinervia* (C<sub>4</sub>) + *F. australasica* (C<sub>4</sub>) was supported by most gene  
302 trees (1,172 of 1,220) and had strong QS support with no signals of alternative topologies.  
303 *Flaveria bidentis* (C<sub>4</sub>) as sister to *F. trinervia* (C<sub>4</sub>) + *F. australasica* (C<sub>4</sub>; IQ-TREE) was  
304 supported by 197 (of 1,054) and had very low QS support with a signal of an alternative  
305 topology. When *F. bidentis* (C<sub>4</sub>) was placed as sister to *F. kochiana* (C<sub>4</sub>) + *F. vaginata* (C<sub>4</sub>-like;

306 ASTRAL) the clade formed by these three species was supported by even fewer gene trees (170  
307 of 1,037) and QS counter-support (-0.027/0.72/0.38; Fig. S3). The clade composed of *F.*  
308 *kochiana* (C<sub>4</sub>) + *F. vaginata* (C<sub>4</sub>-like) and the C<sub>4</sub> clade *F. bidentis*, *F. trinervia* + *F. australasica*  
309 was supported by 206 (of 1,028) gene trees and had very low QS support (0.0084/0.93/0.35) with  
310 signals of alternative topologies. Clade B was supported by only 85 (of 679) gene trees and had  
311 low QS support (0.11/0.53/0.43) with signals of an alternative topology. *Flaveria brownii* (C<sub>4</sub>-  
312 like) + *F. sonorensis* (C<sub>3</sub>-C<sub>4</sub>) was supported by 708 (of 1,180) gene trees and had moderate QS  
313 support (0.4/0.042/0.74) with clear signals of an alternative topology. The four remaining species  
314 of Clade B formed a clade supported by 299 (of 910) gene trees and had low QS support  
315 (0.11/0.53/0.43) with signals of an alternative topology. *Flaveria chlorifolia* (C<sub>3</sub>-C<sub>4</sub>) + *F.*  
316 *floridana* (C<sub>3</sub>-C<sub>4</sub>) was supported by 337 (of 1,042) gene trees and had very low QS support  
317 (0.0039/0.9/0.39) with clear signals of an alternative topology. Lastly, the C<sub>3</sub>-C<sub>4</sub> clade (*F.*  
318 *pubescens*, *F. chlorifolia* + *F. floridana*) was supported by 727 (of 1,055) gene trees and had  
319 strong QS support (0.68/0.57/0.71) with a slight signal of an alternative topology.

320

### 321 **Potential widespread reticulation**

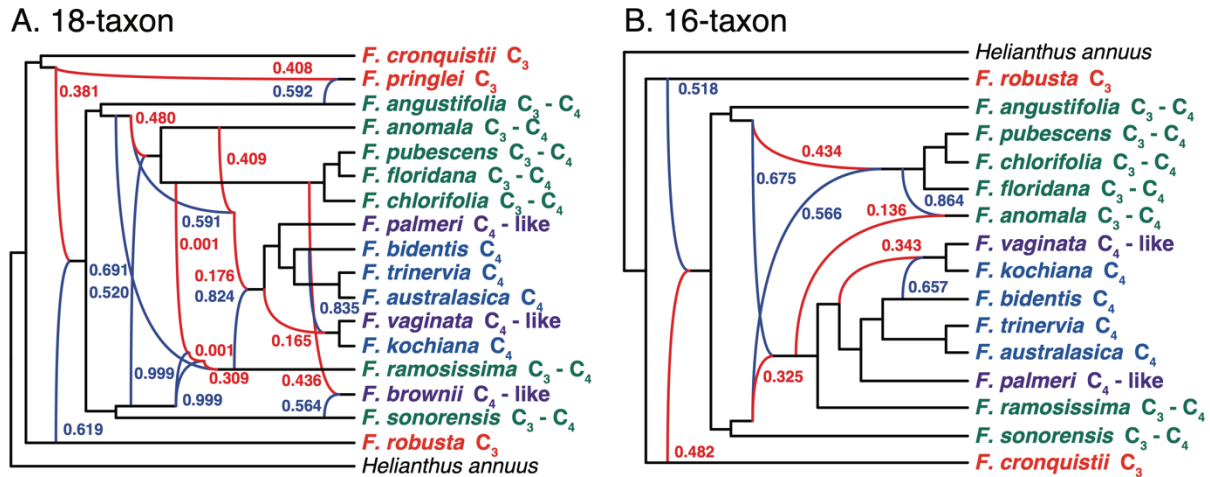
322 The distribution of tree-to-tree distances of the empirical and simulated gene trees to the  
323 ASTRAL tree showed some overlap (Fig. S4), but there was a skew towards larger distances in  
324 the empirical trees (mean 16.61) compared to distances in the simulated trees (mean 11.80). This  
325 suggested that ILS alone cannot explain most of the observed gene tree incongruence (Maureira-  
326 Butler et al., 2008).

327



328 Species network analyses recovered topologies with up to ten reticulations events for the  
329 18-taxon data set (Fig. S6). The pseudo-likelihood score (Fig. S5) continually improved with the  
330 inclusion of additional reticulation events, and the network with ten reticulations (Fig. 2A) had  
331 the best score. Although the best-scored network showed complicated and nested reticulation  
332 patterns (Fig. 2A), there are several clear patterns among most networks (Fig. S6). *Flaveria*  
333 *brownii* (C<sub>4</sub>-like) and *F. pringlei* (C<sub>3</sub>) were consistently recovered as hybrids in all ten networks  
334 (Fig. S6). In both cases, parental lineages and inheritance probabilities were consistent across  
335 networks. Other reticulation patterns recovered across networks were the hybrid origin of clades  
336 A (mainly C<sub>4</sub> and C<sub>4</sub>-like) and B (mainly C<sub>3</sub>-C<sub>4</sub>), which both had *F. sonorensis* (C<sub>3</sub>-C<sub>4</sub>; itself a  
337 hybrid in some networks) and *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>; or closely related to lineage) as potential  
338 parental lineages. The last reticulation event recovered in all networks involved the C<sub>3</sub> species *F.*  
339 *cronquistii* and *F. robusta* which were potential parental lineages of all remaining *Flaveria*  
340 species. The analyses of the 16-taxon data set (Fig. S6) resulted in a best-scoring network with  
341 five reticulation events (Fig. 2B). This showed patterns like the 18-taxon dataset regarding the  
342 hybrid origin of clades A and B, and the deep reticulation involving *F. cronquistii* (C<sub>3</sub>) and *F.*  
343 *robusta* (C<sub>3</sub>). Also, it recovered a reticulation event (consistent with the 18-taxon dataset) where  
344 *F. vaginata* (C<sub>4</sub>-like) and *F. kochiana* (C<sub>4</sub>) within Clade A had a hybrid origin.

345



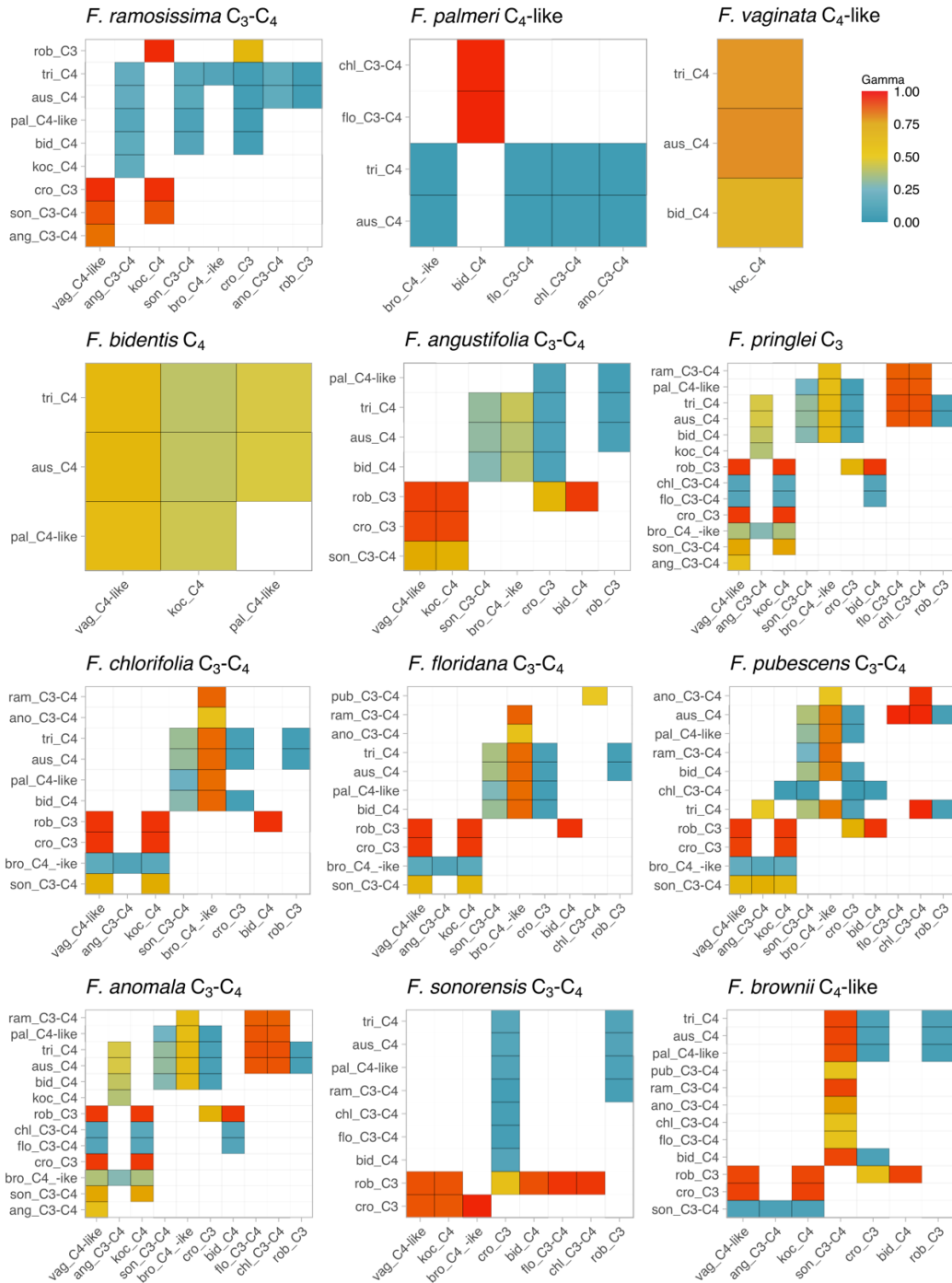
346

347 **Fig. 2.** *Flaveria* species networks with the best maximum pseudo-likelihood scores with PhyloNet for the  
348 (A) 18-taxon and (B) 16-taxon data sets. Species names are colored by photosynthetic type. Red and blue  
349 curved branches indicate the minor and major edges, respectively, of hybrid nodes. Numbers next to  
350 curved branches indicate inheritance probabilities for each hybrid node.

351

352 The HyDe analysis of all possible triples using the 17 species of *Flaveria* resulted in  
353 2,040 hybridization tests, of which 311 triples were significant (Table S2). The analyses without  
354 *F. pringlei* (C<sub>3</sub>) resulted in 247 significant hybridization tests of 1,680 overall tests (Table S3).  
355 HyDe analyses detected 12 species as potential hybrids (Fig. 3). These species included all  
356 members of Clade B which comprises five C<sub>3</sub>-C<sub>4</sub> and one C<sub>4</sub>-like species. Hybrids detected in  
357 Clade B had several potential parental lineages from across *Flaveria* and admixture ( $\gamma$ ) values  
358 were either closer to zero or one (Fig. 3). *Flaveria chlorifolia*, *F. floridana*, *F. pubescens*, *F.*  
359 *anomala* showed similar hybridization patterns consistent with a single ancient hybrid origin of  
360 these species as shown in the PhyloNet analyses. On the other hand, *F. sonorensis* and *F.*  
361 *brownii* showed hybridization patterns different from the rest of the species of the clade,  
362 suggesting their independent hybrid origins as also seen in the PhyloNet analyses. In Clade A,  
363 and unlike PhyloNet, HyDe detected only four instances of hybridization. These included the C<sub>3</sub>-  
364 C<sub>4</sub> species *F. ramosissima* and the two C<sub>4</sub>-like species, *F. vaginata* and *F. palmeri*, of the clade.

365 In these three cases  $\gamma$  was consistent with ancient hybridization. *Flaveria vaginata* had notably  
366 fewer potential parental lineages suggesting a more recent reticulation event. The fourth species,  
367 *F. bidentis*, was the only C<sub>4</sub> member of the clade which was detected as a hybrid. In this case, all  
368 potential parental lineages are members of Clade A, and  $\gamma$  values around 0.5 suggested a recent  
369 reticulation event. Outside clades A and B, *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>) and *F. pringlei* (C<sub>3</sub>) were  
370 detected as hybrids. In both cases, the potential parental lineages come from across *Flaveria* with  
371  $\gamma$  values suggesting ancient reticulation. Overall, HyDe detected all C<sub>3</sub>-C<sub>4</sub> and C<sub>4</sub>-like species as  
372 well as *F. bidentis* (C<sub>4</sub>) and *F. pringlei* (C<sub>3</sub>) as potential ancient hybrids.



373

374 **Fig. 3.** HyDe significant hybridization test for the 12 species of *Flaveria* identified as potential hybrids.  
 375 Species on the x-axis are parental lineage 1 (P1) and species on the y-axis are parental lineage 2 (P2).  
 376 Only colored boxes denote possible combinations of P1 and P2 as parents of hybrid species. The color  
 377 scale represents the value of the hybridization parameter  $\gamma$  for each hybridization event. Recent 50:50  
 378 hybrids would be represented by a  $\gamma \sim 0.5$ . Values of  $\gamma$  approaching 0 indicate a major hybrid contribution  
 379 from P1, and values approaching 1 indicate a major hybrid contribution from P2, with both cases  
 380 representing ancient hybridization.

381 **Lack of whole genome duplications**

382           The orthogroup mapping did not reveal any node in *Flaveria* with elevated levels of gene  
383 duplication (Fig S7), showing the absence of whole genome duplication in the genus. In part this  
384 is expected as *Flaveria* is consistently diploid (see Powell, 1978 and ref. therein) with a haploid  
385 chromosome number of  $n = 18$ . Furthermore, the lack of whole genome duplication in *Flaveria*  
386 suggests that reticulation events in this group are homoploid hybridizations.

387

388 **DISCUSSION**

389           In  $C_4$  photosynthesis, high rates of net photosynthesis and a highly competitive water and  
390 nitrogen use efficiency are achieved by spatially separated carbon fixation in the outer mesophyll  
391 cells preceding the Calvin-Benson cycle, and by effectively fueling Rubisco with always high  
392  $CO_2$  concentration in the controlled seclusion of the Kranz cells (Long, 1999). Resulting low  
393 levels of photorespiration make  $C_4$  plants competitive in various stressful environments where  
394 carbon deficiency poses a problem to  $C_3$  plants (Sage et al., 2012).  $C_4$  evolved more than 60  
395 times in angiosperms with hotspots of  $C_4$  origins in Poaceae and Amaranthaceae (Sage et al.,  
396 2018). Due to the anatomical and gene regulatory complexity of the  $C_4$  pathway, it seems clear  
397 that there must have been intermediate stable phenotypes during the evolution of  $C_4$  from a  $C_3$   
398 ancestor (Monson and Moore, 1989). The established generalized model of  $C_4$  evolution tries to  
399 explain the sequence of intermediate adaptive events during the transition from the ancestral  $C_3$   
400 to  $C_4$  using the naturally occurring  $C_3$ - $C_4$  intermediate phenotypes of *Flaveria* (and other  
401 lineages) as proxies of intermediate stages in  $C_4$  evolution (Sage et al., 2014; Bräutigam and  
402 Gowik, 2016). However, *Flaveria* is also known for rampant interfertility of its 21 (mostly  
403 diploid) species (Long and Rhamstine, 1968; Powell, 1978), and for the transferability of

404 photosynthetic traits to hybrid offspring in hybridization experiments (Apel et al., 1988; Kadereit  
405 et al., 2017 and ref. therein). Therefore, a clear understanding of the phylogenetic history of this  
406 model lineage of C<sub>4</sub> evolution, including tests for possible reticulate evolution, is fundamental as  
407 this might have strong implications for our understanding of the evolution of C<sub>4</sub> photosynthesis  
408 in general.

409 Our analyses using transcriptome data of 17 *Flaveria* species revealed rampant gene tree  
410 discordance along the backbone of the phylogeny as well as the clades containing most C<sub>3</sub>-C<sub>4</sub>  
411 intermediate and C<sub>4</sub>-like species (clades A and B; visualized in the cloudogram in Fig. 1).  
412 Coalescence simulations showed that gene tree discordance in *Flaveria* cannot be attributed to  
413 ILS alone. Our initial species network analyses including all 17 species sampled consistently  
414 identified *Flaveria brownii* (C<sub>4</sub>-like) and *F. pringlei* (C<sub>3</sub>) as hybrids (Figs. 2 and 3). After  
415 exclusion of these two most likely recent hybrids, deep reticulation events were recovered more  
416 clearly. The network analyses of the remaining species suggested an early reticulation event  
417 between two C<sub>3</sub> lineages (*F. robusta* and *F. cronquistii*) giving rise to the ancestor of the lineage  
418 containing all C<sub>3</sub>-C<sub>4</sub> intermediate species, C<sub>4</sub>-like species and C<sub>4</sub> species. This result is supported  
419 by the HyDe analysis which identified all C<sub>3</sub>-C<sub>4</sub> and C<sub>4</sub>-like species as potential ancient hybrids.  
420 The ancestors of *F. robusta* and *F. cronquistii* seem to have contributed equally to the origin of  
421 the hybrid lineage (Fig. 2). Within this lineage there exist two parental lineages, *F. angustifolia*  
422 (C<sub>3</sub>-C<sub>4</sub>) and *F. sonorensis* (C<sub>3</sub>-C<sub>4</sub>), that seem to have contributed to a robust clade containing all  
423 C<sub>4</sub> species plus *F. ramosissima* (C<sub>3</sub>-C<sub>4</sub>) and *F. vaginata* (C<sub>4</sub>-like; clade A in Fig. 1), and also to  
424 another robust clade of four C<sub>3</sub>-C<sub>4</sub> intermediate species (*F. pubescens*, *F. chlorifolia*, *F.*  
425 *floridana*, *F. anomala*; partial clade B in Fig. 1; supported also by the HyDe analysis). Of these  
426 four species, *F. anomala* was introgressed by the ancestor of clade A (Fig. 2).

427           These results have several important implications that will be discussed in the following  
428 three sections: 1) natural C<sub>3</sub>-C<sub>4</sub> intermediates species in *Flaveria* do not seem to result from  
429 hybridization between a C<sub>3</sub> and a C<sub>4</sub> lineage (as similar C<sub>3</sub>-C<sub>4</sub> intermediate phenotypes resulting  
430 from crossing experiments might suggest); 2) recurrent homoploid hybridization possibly played  
431 a major role in the evolution of C<sub>4</sub> photosynthesis in *Flaveria* with an initial hybridization event  
432 between two C<sub>3</sub> species (*F. robusta* and *F. cronquistii*) as a possible trigger, 3) only the ancestral  
433 lineages of *F. angustifolia* and *F. sonorensis* (both C<sub>3</sub>-C<sub>4</sub> intermediates) seem to be involved in  
434 the formation of clades A and B, of which only clade A shows C<sub>4</sub> photosynthesis.

435           Hybridization is an important factor in plant evolution and speciation in general (Abbott  
436 et al., 2013) and *Flaveria* is no exception to this but also shows hybridization in a more complex  
437 way than we expected. From the phenotypic outcome of the numerous C<sub>3</sub> × C<sub>4</sub> hybridization  
438 experiments in *Flaveria* it was conceivable that naturally occurring C<sub>3</sub>-C<sub>4</sub> intermediate species  
439 might be the result of crosses between parental lineages with different photosynthetic types  
440 (Monson and Moore, 1989; Kadereit et al., 2017). However, our results indicate that origin of  
441 C<sub>3</sub>-C<sub>4</sub> intermediates might be more complex: in *Flaveria* there are several recent hybrids  
442 expressing intermediate or deviating phenotypes compared to the parental lineages which perturb  
443 phylogenetic reconstruction (in particular, *F. pringlei* and *F. brownii*). There also were ancient  
444 hybridization events challenging the reconstruction of the backbone (Fig. 1; see below).  
445 McKown et al. (2005) and Lyu et al., 2015 interpreted *F. brownii* as representing an independent  
446 and recent origin of C<sub>4</sub>-like photosynthesis from C<sub>3</sub>-C<sub>4</sub> ancestors. According to our analyses, *F.*  
447 *brownii* is a recent hybrid between *F. sonorensis* (C<sub>3</sub>-C<sub>4</sub>) and one of the other species in clade B  
448 (all C<sub>3</sub>-C<sub>4</sub>; Fig. 1–3), probably *F. floridana* with which it forms fertile F1 and F2 hybrids  
449 (Powell, 1978). This is an important finding for the interpretation of trait evolution in *Flaveria*.

450 The C<sub>4</sub>-like photosynthesis in *F. brownii* seems to have resulted from a cross between a C<sub>2</sub> type I  
451 (*F. sonorensis*) and C<sub>2</sub> type 2 species (possibly *F. floridana*; see Table 1) resulting in stronger  
452 C<sub>4</sub>-ness of *F. brownii*. Since *F. brownii* possesses C<sub>3</sub> isoforms of the key enzymes of the C<sub>4</sub>  
453 pathway (e.g., Kubien et al., 2008; Gowik and Westhoff, 2011; Ludwig, 2011), the C<sub>4</sub>-like  
454 metabolism in *F. brownii* seems to have resulted from a highly effective gene regulatory change  
455 in comparison to its parental lineages rather than from the acquisition of C<sub>4</sub> isoforms. It clearly  
456 outcompetes its parental lineages in terms of C<sub>4</sub> physiology. Taniguchi et al. (2021), studying  
457 genome size evolution in *Flaveria*, showed dynamic alteration of genome size in the genus with  
458 *F. brownii* having the largest genome size among the 11 species investigated. Rewiring the  
459 parental genomes might have led to the C<sub>4</sub>-like phenotype found in *F. brownii* and might have  
460 facilitated adaptive divergence resulting in colonization of the coastal flats and islands of the  
461 lower Texas Gulf Coast. Similarly, our analyses revealed *F. pringlei* (C<sub>3</sub> proto-kranz) as a recent  
462 hybrid between *F. cronquistii* (C<sub>3</sub>) and *F. angustifolia* (C<sub>3</sub>-C<sub>4</sub>; Fig. 1-3). Therefore, the proto-  
463 kranz phenotype in this species might be of different origin than that of *F. robusta* (Table 1). In  
464 fact, Lyu et al. (2015) showed the *F. pringlei* sample used in their study to be of artificial hybrid  
465 origin (possibly resulting from unintended crosses in the greenhouse). Trait assessments in these  
466 two species, e.g., of leaf anatomical and ultrastructural traits (McKown and Dengler, 2007; Sage  
467 et al., 2013) might lead to re-evaluation of their assessment as evolutionary “basal” as they likely  
468 acquired the proto-kranz phenotype differently. Also, the similarities between *F. cronquistii* and  
469 *F. pringlei* found by McKown and Dengler (2007) and their differences to *F. robusta* can now be  
470 explained. According to our analyses, *F. robusta* and *F. cronquistii* seem to be representatives of  
471 ancient C<sub>3</sub> lineages in *Flaveria*, but not *F. pringlei*. The rare *Flaveria mcdougallii* (C<sub>3</sub>, not  
472 represented in this analysis), which is morphologically and geographically distinct and might be



473 sister to all other species of *Flaveria* (Powell, 1978; McKown et al., 2005), would be an  
474 important species to add in further studies.

475         The most ancient hybridization event in *Flaveria* involved two  $C_3$  lineages, *F. robusta*  
476 and *F. cronquistii*. According to Powell (1978), *Flaveria cronquistii* most closely resembles *F.*  
477 *robusta*. The two are geographically separated, and the former is distributed in the Tehuacán  
478 Valley region and the latter in Colima and Michoacán (both Mexico). There are no reports of  
479 artificial hybridization between these two species. The resulting ancient hybrid lineage seems to  
480 include all  $C_3$ - $C_4$  intermediate,  $C_4$ -like and  $C_4$  species (Fig. 2; Table 1). Concerning leaf  
481 anatomy, *Flaveria robusta* differs from *F. cronquistii* by a higher vein density achieved through  
482 higher vein branching, and by a higher number of organelles in the bundle sheath cells where the  
483 organelles are located in a centripetal position along the cell wall connecting bundle sheath cells  
484 and vascular tissue. Organelles are more evenly distributed in *F. cronquistii*, and its leaves are  
485 more succulent and show larger bundle sheath cells (McKown and Dengler, 2007; Sage et al.,  
486 2013). The combination of these traits in a hybrid lineage and subsequent segregation effectively  
487 leading to a higher bundle sheath to mesophyll ratio and activated large bundle sheath cells may  
488 have triggered the evolution of “pre kranz” cells in *Flaveria*. Possibly the strong leaf anatomical  
489 differences between the two parental lineages promoted the origin of novel traits which then  
490 allowed this lineage to occupy new niches eco-geographically separate from the parents. There  
491 are multiple examples for plant lineages in which homoploid hybridization resulted in potentially  
492 adaptive genotypes through transgressive segregation, eventually leading to speciation (Gross  
493 and Rieseberg, 2005; Nieto Feliner et al., 2020). An increase of the bundle sheath to mesophyll  
494 ratio has been suggested to play an initial role in the evolution of  $C_4$  in several plant lineages

495 (Marshall et al., 2007; McKown and Dengler, 2007; Christin et al., 2013; Griffiths et al., 2013;  
496 Lauterbach et al., 2019).

497 Two C<sub>3</sub>-C<sub>4</sub> intermediate lineages, *F. angustifolia* and *F. sonorensis*, are involved in the  
498 formation of clades A and B according to our results. It is remarkable that only these two and not  
499 any other species (especially not the C<sub>4</sub>-like species) contributed to the origin of these clades. In  
500 their extant distribution the two species do not overlap. *Flaveria angustifolia* grows in  
501 sclerophyllous scrub in Puebla and Oaxaca, while *F. sonorensis* is found only in the short-tree  
502 forests of tropical Sonora (Powell, 1978). Looking more closely at these two descendants of the  
503 lineages that seem to have hybridized in the past and possibly gave rise to C<sub>4</sub> photosynthesis  
504 might give new insights into important preconditions for the evolution of C<sub>4</sub> in *Flaveria*. Both *F.*  
505 *angustifolia* and *F. sonorensis* were categorized as C<sub>2</sub> Type I species (Table 1), but with weaker  
506 C<sub>2</sub> photosynthesis (relatively high amounts of glycine decarboxylase in the mesophyll cells and  
507 relatively high CO<sub>2</sub> compensation points) than other C<sub>2</sub> species of *Flaveria* (Sage et al., 2018).  
508 This finding implies the evolution of a C<sub>2</sub> phenotype prior to C<sub>4</sub>, supporting the ‘photorespiratory  
509 bridge hypothesis’ (Sage et al., 2018) in case of *Flaveria*. However, it seems that C<sub>4</sub> evolution in  
510 *Flaveria* was triggered by hybridization of C<sub>2</sub> lineages, a scenario never suggested before. For  
511 *Flaveria* this somewhat shifts the focus to C<sub>2</sub> photosynthesis and under which selective  
512 conditions this type of photosynthesis might have evolved. To gain further insights into the  
513 evolution of C<sub>2</sub> photosynthesis, detailed studies of the origin, anatomy, and ecophysiology of *F.*  
514 *angustifolia* and *F. sonorensis* might be rewarding. Organelle enrichment and their flux-  
515 optimized positioning as well as glycine decarboxylase accumulation in the proto-kranz cells  
516 (Khoshravesh et al., 2016) seem to be essential for C<sub>2</sub> photosynthesis (in eudicots and  
517 monocots). These traits enable species to more efficiently re-cycle respired CO<sub>2</sub> and to longer

518 maintain a positive assimilation rate under carbon deficient conditions. Since the majority of C<sub>4</sub>  
519 lineages do not have any known C<sub>2</sub> relatives (see Sage et al., 2018 for an overview), and C<sub>2</sub>  
520 lineages without close C<sub>4</sub> relative are known as well (see Lundgren, 2020 for an overview), the  
521 question remains whether there exist several evolutionary pathways to C<sub>4</sub> photosynthesis  
522 (Edwards, 2019), and whether C<sub>2</sub> should be considered an independent carbon concentrating  
523 mechanism not necessarily always intimately connected to C<sub>4</sub> photosynthesis (Lundgren, 2020  
524 and ref. therein). Finally, a C<sub>2</sub> lineage might also be the result of ancient hybridization of a C<sub>3</sub>  
525 and a C<sub>4</sub> lineage, as has been suggested for *Salsola divaricata*, which then might have thrived  
526 through the large plasticity of photosynthetic traits inherited from photosynthetically divergent  
527 parental lineages (Tefarikis et al., 2021).

528

## 529 CONCLUSIONS

530 The young genus *Flaveria* which includes four C<sub>3</sub>, four C<sub>4</sub>, three C<sub>4</sub>-like and c. ten C<sub>3</sub>-C<sub>4</sub>  
531 intermediate species is remarkable for the high number of evolutionary reticulations creating an  
532 enormous diversity of phenotypes with different photosynthetic traits. Due to this and its strictly  
533 diploid chromosome number the genus is a highly interesting system to study the genetic basis of  
534 the C<sub>4</sub> syndrome and the role of transgressive segregation in the origin of genotypes eventually  
535 leading to the evolution of C<sub>4</sub> photosynthesis. We found evidence that homoploid hybridization  
536 of C<sub>3</sub> lineages might have triggered the evolution of C<sub>2</sub> photosynthesis, and that homoploid  
537 hybridization of C<sub>2</sub> lineages gave rise to C<sub>4</sub>-like or C<sub>4</sub> photosynthesis. In both cases the hybrid  
538 derivatives possibly surpassed the parental performance under conditions of high  
539 photorespiration and had an adaptive advantage. However, since reticulation occurred in recent  
540 as well as ancient lineages of the genus, the sequence of evolutionary events needs to be studied

541 in the light of a carefully reconstructed phylogenetic history of the genus. Comparison of the  
542 entire genomes of parental and hybrid lineages should allow us to detect whether the origin of  
543 new photosynthetic traits is indeed the result of transgressive segregation (de los Reyes, 2019),  
544 as suggested here, or of stepwise mutational change of relevant genes.  
545

546 **DATA AVAILABILITY STATEMENT**

547 Analysis and results files can be accessed at the Dryad repository XXXX

548

549 **CONFLICT OF INTEREST**

550 The authors declare no conflict of interest.

551

552 **AUTHOR CONTRIBUTIONS**

553 GK and DFM-B conceived the study. DFM-B performed all analyses. GK and DFM-B evaluated

554 the results and wrote the paper.

555

556 **FUNDING**

557 Financial support for this study came from LMU Munich and the German Science Foundation

558 (DFG grant KA1816/9-1).

559

560 **ACKNOWLEDGMENTS**

561 We thank Joachim W. Kadereit (Mainz) for valuable comments on the manuscript.

562

563

564

565

566

567

568 **SUPPLEMENTARY MATERIAL**

569 **Table S1.** Taxon sampling and source of data

570

571 **Table S2.** HyDe significant hybridization tests for all species of *Flaveria*.

572

573 **Table S3.** HyDe significant hybridization tests for all species of *Flaveria* excluding *F. pringlei*

574 ( $C_3$ ).

575

576 **Fig. S1.** A. Maximum likelihood phylogeny of *Flaveria* inferred with IQ-TREE from the  
577 concatenated 1249-nuclear gene supermatrix. Numbers above branches represent bootstrap  
578 support (BS). Branch lengths as substitutions per site (scale bar on the bottom). B. ASTRAL tree  
579 of *Flaveria* inferred from the 1,295 nuclear gene trees. Local posterior probabilities (LLP) are  
580 shown next to nodes. Internal branch lengths are in coalescent units (scale bar on the bottom).

581

582 **Fig. S2.** A. Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the  
583 concatenated 1249-nuclear gene supermatrix. B. ASTRAL cladogram of *Flaveria* inferred from  
584 the 1,295 nuclear gene trees. Pie charts represent the proportion of gene trees that support that  
585 clade (blue), the main alternative bifurcation (green), the remaining alternatives (red), and  
586 conflict or support that have <50% bootstrap support (gray). Number above and below branches  
587 represent the number of concordant and discordant informative gene trees, respectively.

588

589 **Fig. S3.** A. Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the  
590 concatenated 1249-nuclear gene supermatrix. B. ASTRAL cladogram of *Flaveria* inferred from

591 the 1,295 nuclear gene trees. Quartet Sampling (QS) scores are shown above branches. QS  
592 scores: Quartet concordance/Quartet differential/Quartet informativeness. Circles at nodes are  
593 colored by quartet concordance support.

594

595 **Fig. S4.** Distribution of tree-to-tree distances between empirical gene trees and the ASTRAL  
596 tree, compared to the distribution of tree-to-tree distances between simulated trees and the  
597 ASTRAL tree.

598

599 **Fig. S5.** Maximum pseudo-likelihood scores for species networks inferred with PhyloNet using  
600 the (A) 18-taxon, (B) and 16-taxon data sets. The x-axis notes the maximum number of  
601 reticulations for each of the network searches allowing up to ten reticulation events.

602

603 **Fig. S6.** Maximum pseudo-likelihood species networks inferred with PhyloNet using the (A) 18-  
604 taxon, (B) and 16-taxon data sets and allowing up to ten reticulation events. Red and blue curved  
605 branches indicate the minor and major edges, respectively of hybrid nodes. Numbers next to  
606 curved branches indicate inheritance probabilities for each hybrid node.

607

608 **Fig. S7.** Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the  
609 concatenated 1249-nuclear gene supermatrix. Numbers above branches are gene duplication  
610 counts and numbers below branches are gene duplication percentages.

611

612 **REFERENCES**

- 613 Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., et al. (2013).  
614 Hybridization and speciation. *Journal of Evolutionary Biology* 26, 229–246.  
615 doi:10.1111/j.1420-9101.2012.02599.x.
- 616 Anderberg, A. A., Baldwin, B. G., Bayer, R. G., Breitwieser, J., Jeffrey, C., Dillon, M. O., et al.  
617 (2007). “Compositae,” in *Flowering plants · Eudicots: Asterales*, eds. J. W. Kadereit and  
618 C. Jeffrey (Berlin, Heidelberg: Springer Berlin Heidelberg), 61–588. doi:10.1007/978-3-  
619 540-31051-8\_7.
- 620 Apel, P., Bauwe, H., Bassüner, B., and Maass, I. (1988). Photosynthetic properties of *Flaveria*  
621 *cronquistii*, *F. palmeri*, and hybrids between them. *Biochemie und Physiologie der*  
622 *Pflanzen* 183, 291–299. doi:10.1016/S0015-3796(88)80021-0.
- 623 Apel, P., and Maass, I. (1981). Photosynthesis in species of *Flaveria* CO<sub>2</sub> compensation  
624 concentration, O<sub>2</sub> influence on photosynthetic gas exchange and  $\delta^{13}\text{C}$  values in species of  
625 *Flaveria* (Asteraceae). *Biochemie und Physiologie der Pflanzen* 176, 396–399.  
626 doi:10.1016/S0015-3796(81)80052-2.
- 627 Blair, C., and Ané, C. (2020). Phylogenetic trees and networks can serve as powerful and  
628 complementary approaches for analysis of genomic data. *Systematic Biology* 69, 593–  
629 601. doi:10.1093/sysbio/syz056.
- 630 Blischak, P. D., Chifman, J., Wolfe, A. D., and Kubatko, L. S. (2018). HyDe: A python package  
631 for genome-scale hybridization detection. *Systematic Biology* 67, 821–829.  
632 doi:10.1093/sysbio/syy023.
- 633 Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic - a flexible trimmer for Illumina  
634 sequence data. *Bioinformatics* 30, 2112–2120. doi:10.1093/bioinformatics/btu170/-/DC1.



- 635 Bräutigam, A., and Gowik, U. (2016). Photorespiration connects C<sub>3</sub> and C<sub>4</sub> photosynthesis.  
636 *Journal of Experimental Botany* 67, 2953–2962. doi:10.1093/jxb/erw056.
- 637 Brown, J. W., Walker, J. F., and Smith, S. A. (2017). Phyx - phylogenetic tools for unix.  
638 *Bioinformatics* 33, 1886–1888. doi:10.1093/bioinformatics/btx063/-/DC1.
- 639 Chen, L.-Y., Morales-Briones, D. F., Passow, C. N., and Yang, Y. (2019). Performance of gene  
640 expression analyses using de novo assembled transcripts in polyploid species.  
641 *Bioinformatics*. doi:10.1093/bioinformatics/btz620.
- 642 Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R.,  
643 et al. (2013). Anatomical enablers and the evolution of C<sub>4</sub> photosynthesis in grasses. *Proc*  
644 *Natl Acad Sci USA* 110, 1381. doi:10.1073/pnas.1216777110.
- 645 Cloutier, A., Sackton, T. B., Grayson, P., Clamp, M., Baker, A. J., and Edwards, S. V. (2019).  
646 Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the  
647 presence of an empirical anomaly zone. *Systematic Biology* 68, 937–955.  
648 doi:10.1093/sysbio/syz019.
- 649 Davidson, N. M., and Oshlack, A. (2014). Corset: enabling differential gene expression analysis  
650 for de novo assembled transcriptomes. *Genome Biology* 15, 57. doi:10.1186/s13059-014-  
651 0410-6.
- 652 de los Reyes, B. G. (2019). Genomic and epigenomic bases of transgressive segregation – New  
653 breeding paradigm for novel plant phenotypes. *Plant Science* 288, 110213.  
654 doi:10.1016/j.plantsci.2019.110213.
- 655 Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and  
656 the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332–340.  
657 doi:10.1016/j.tree.2009.01.009.

- 658 Doyle, J. J. (1992). Gene trees and species trees: Molecular systematics as one-character  
659 taxonomy. *Systematic Botany* 17, 144. doi:10.2307/2419070.
- 660 Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture  
661 between closely related populations. *Molecular Biology and Evolution* 28, 2239–2252.  
662 doi:10.1093/molbev/msr048.
- 663 Edwards, E. J. (2019). Evolutionary trajectories, accessibility and other metaphors: the case of C<sub>4</sub>  
664 and CAM photosynthesis. *New Phytologist* 223, 1742–1755. doi:10.1111/nph.15851.
- 665 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-  
666 generation sequencing data. *Bioinformatics* 28, 3150–3152.  
667 doi:10.1093/bioinformatics/bts565.
- 668 Galtier, N., and Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses.  
669 *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 4023–4029.  
670 doi:10.1098/rstb.2008.0144.
- 671 Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., and Hood, M. E. (2008). Speciation in  
672 fungi. *Fungal Genetics and Biology* 45, 791–802. doi:10.1016/j.fgb.2008.02.001.
- 673 Gowik, U., and Westhoff, P. (2011). “C<sub>4</sub>-Phosphoenolpyruvate Carboxylase,” in C<sub>4</sub>  
674 *Photosynthesis and Related CO<sub>2</sub> Concentrating Mechanisms*, eds. A. S. Raghavendra and  
675 R. F. Sage (Dordrecht: Springer Netherlands), 257–275. doi:10.1007/978-90-481-9407-  
676 0\_13.
- 677 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011).  
678 Full-length transcriptome assembly from RNA-Seq data without a reference genome.  
679 *Nature Biotechnology* 29, 644–652. doi:10.1038/nbt.1883.

- 680 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A  
681 draft sequence of the neandertal Genome. *Science* 328, 710–722.  
682 doi:10.1126/science.1188021.
- 683 Griffiths, H., Weller, S. G., Toy, L. F. M., and Dennis, R. J. (2013). You're so vein: Bundle  
684 sheath physiology, phylogeny and evolution in C<sub>3</sub> and C<sub>4</sub> plants. *Plant, Cell &*  
685 *Environment* 36, 249–261. doi:10.1111/j.1365-3040.2012.02585.x.
- 686 Gross, B. L., and Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid  
687 speciation. *Journal of Heredity* 96, 241–252. doi:10.1093/jhered/esi026.
- 688 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013).  
689 *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for  
690 reference generation and analysis. *Nature Protocols* 8, 1494–1512.  
691 doi:10.1038/nprot.2013.084.
- 692 Kadereit, G., Bohley, K., Lauterbach, M., Tefarikis, D. T., and Kadereit, J. W. (2017). C<sub>3</sub>-C<sub>4</sub>  
693 intermediates may be of hybrid origin – a reminder. *New Phytologist* 215, 70–76.  
694 doi:10.1111/nph.14567.
- 695 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017).  
696 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14,  
697 587–589. doi:10.1038/nmeth.4285.
- 698 Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., et al. (2016).  
699 C<sub>3</sub>-C<sub>4</sub> intermediacy in grasses: organelle enrichment and distribution, glycine  
700 decarboxylase expression, and the rise of C<sub>2</sub> photosynthesis. *Journal of Experimental*  
701 *Botany* 67, 3065–3078. doi:10.1093/jxb/erw150.

- 702 Ku, M. S. B., Monson, R. K., Littlejohn, R. O., Jr., Nakamoto, H., Fisher, D. B., and Edwards, G.  
703 E. (1983). Photosynthetic characteristics of C<sub>3</sub>-C<sub>4</sub> intermediate *Flaveria* species 1: I. Leaf  
704 anatomy, photosynthetic responses to O<sub>2</sub> and CO<sub>2</sub>, and activities of key enzymes in the  
705 C<sub>3</sub> and C<sub>4</sub> pathways. *Plant Physiology* 71, 944–948. doi:10.1104/pp.71.4.944.
- 706 Ku, M. S. B., Wu, J., Dai, Z., Scott, R. A., Chu, C., and Edwards, G. E. (1991). Photosynthetic  
707 and photorespiratory characteristics of *Flaveria* species 1. *Plant Physiology* 96, 518–528.  
708 doi:10.1104/pp.96.2.518.
- 709 Kubatko, L. S., and Chifman, J. (2019). An invariants-based method for efficient identification  
710 of hybrid species from large-scale genomic data. *BMC Evol Biol* 19, 112.  
711 doi:10.1186/s12862-019-1439-7.
- 712 Kubien, D. S., Whitney, S. M., Moore, P. V., and Jesson, L. K. (2008). The biochemistry of  
713 Rubisco in *Flaveria*. *Journal of Experimental Botany* 59, 1767–1777.  
714 doi:10.1093/jxb/erm283.
- 715 Lanfear, R., Calcott, B., Ho, S. Y. W., and Guindon, S. (2012). PartitionFinder: Combined  
716 selection of partitioning schemes and substitution models for phylogenetic analyses.  
717 *Molecular Biology and Evolution* 29, 1695–1701. doi:10.1093/molbev/mss020.
- 718 Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*  
719 *Methods* 9, 357–359. doi:10.1038/nmeth.1923.
- 720 Lauterbach, M., Zimmer, R., Alexa, A. C., Adachi, S., Sage, R., Sage, T., et al. (2019). Variation  
721 in leaf anatomical traits relates to the evolution of C<sub>4</sub> photosynthesis in Tribuloideae  
722 (Zygophyllaceae). *Perspectives in Plant Ecology, Evolution and Systematics* 39, 125463.  
723 doi:10.1016/j.ppees.2019.125463.

- 724 Liu, L., and Yu, L. (2010). Phybase: an R package for species tree analysis. *Bioinformatics* 26,  
725 962–963. doi:10.1093/bioinformatics/btq062.
- 726 Long, R. W., and Rhamstine, E. L. (1968). Evidence for the hybrid origin of *Flaveria latifolia*  
727 (Compositae). *Brittonia* 20, 238–250. doi:10.2307/2805449.
- 728 Long, S. P. (1999). “Environmental Responses,” in *C<sub>4</sub> Plant Biology*, eds. R. F. Sage and R. K.  
729 Monson (San Diego: Academic Press), 215–249. doi:10.1016/B978-012614440-6/50008-  
730 2.
- 731 Ludwig, M. (2011). The molecular evolution of  $\beta$ -carbonic anhydrase in *Flaveria*. *Journal of*  
732 *Experimental Botany* 62, 3071–3081. doi:10.1093/jxb/err071.
- 733 Lyu, M.-J. A., Gowik, U., Kelly, S., Covshoff, S., Hibberd, J. M., Sage, R. F., et al. (2021). The  
734 coordination of major events in C<sub>4</sub> photosynthesis evolution in the genus *Flaveria*.  
735 *Scientific Reports* 11, 15618. doi:10.1038/s41598-021-93381-8.
- 736 Lyu, M.-J. A., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., et al. (2015).  
737 RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria*  
738 (Asteraceae) with some modifications. *BMC Evolutionary Biology* 15, 116.  
739 doi:10.1186/s12862-015-0399-9.
- 740 Mai, U., and Mirarab, S. (2018). TreeShrink: fast and accurate detection of outlier long branches  
741 in collections of phylogenetic trees. *BMC Genomics* 19, 272. doi:10.1186/s12864-018-  
742 4620-2.
- 743 Mandel, J. R., Dikow, R. B., Siniscalchi, C. M., Thapa, R., Watson, L. E., and Funk, V. A.  
744 (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive  
745 diversifications throughout the history of Asteraceae. *Proc Natl Acad Sci USA* 116,  
746 14083–14088. doi:10.1073/pnas.1903871116.

- 747 Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., et al. (2007).  
748 *Cleome*, a genus closely related to *Arabidopsis*, contains species spanning a  
749 developmental progression from C<sub>3</sub> to C<sub>4</sub> photosynthesis. *The Plant Journal* 51, 886–896.  
750 doi:10.1111/j.1365-313X.2007.03188.x.
- 751 Maureira-Butler, I. J., Pfeil, B. E., Muangprom, A., Osborn, T. C., and Doyle, J. J. (2008). The  
752 reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57, 466–482.  
753 doi:10.1080/10635150802172168.
- 754 McKown, A. D., and Dengler, N. G. (2007). Key innovations in the evolution of Kranz anatomy  
755 and C<sub>4</sub> vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany* 94, 382–399.  
756 doi:10.3732/ajb.94.3.382.
- 757 McKown, A. D., Moncalvo, J.-M., and Dengler, N. G. (2005). Phylogeny of *Flaveria*  
758 (Asteraceae) and inference of C<sub>4</sub> photosynthesis evolution. *American Journal of Botany*  
759 92, 1911–1928. doi:10.3732/ajb.92.11.1911.
- 760 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler,  
761 A., et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic  
762 inference in the genomic era. *Molecular Biology and Evolution* 37, 1530–1534.  
763 doi:10.1093/molbev/msaa015.
- 764 Monson, R. K., and Moore, B. D. (1989). On the significance of C<sub>3</sub>-C<sub>4</sub> intermediate  
765 photosynthesis to the evolution of C<sub>4</sub> photosynthesis. *Plant, Cell & Environment* 12, 689–  
766 699. doi:10.1111/j.1365-3040.1989.tb01629.x.
- 767 Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington,  
768 S. F., et al. (2021). Disentangling sources of gene tree discordance in phylogenomic data

- 769 sets: Testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* 70, 219–  
770 235. doi:10.1093/sysbio/syaa066.
- 771 Nakamoto, H., Ku, M. S. B., and Edwards, G. E. (1983). Photosynthetic characteristics of C<sub>3</sub>-C<sub>4</sub>  
772 intermediate *Flaveria* species II. Kinetic properties of phosphoenolpyruvate carboxylase  
773 from C<sub>3</sub>, C<sub>4</sub> and C<sub>3</sub>-C<sub>4</sub> intermediate species. *Plant and Cell Physiology* 24, 1387–1393.  
774 doi:10.1093/oxfordjournals.pcp.a076659.
- 775 Nieto Feliner, G., Casacuberta, J., and Wendel, J. F. (2020). Genomics of evolutionary novelty in  
776 hybrids and polyploids. *Frontiers in Genetics* 11, 792. Available at:  
777 <https://www.frontiersin.org/article/10.3389/fgene.2020.00792>.
- 778 Pamilo, P., and Nei, M. (1988). Relationships between gene trees and species trees. *Molecular*  
779 *Biology and Evolution* 5, 568–583. doi:10.1093/oxfordjournals.molbev.a040517
- 780 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides  
781 fast and bias-aware quantification of transcript expression. *Nature Methods* 14, 417–419.  
782 doi:10.1038/nmeth.4197.
- 783 Payseur, B. A., and Rieseberg, L. H. (2016). A genomic perspective on hybridization and  
784 speciation. *Molecular Ecology* 25, 2337–2360. doi:10.1111/mec.13557.
- 785 Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet  
786 Aampling distinguishes lack of support from conflicting support in the green plant tree of  
787 life. *American Journal of Botany* 105, 385–403. doi:10.1002/ajb2.1016.
- 788 Powell, A. M. (1978). Systematics of *Flaveria* (Flaveriinae--Asteraceae). *Annals of the Missouri*  
789 *Botanical Garden* 65, 590–636. doi:10.2307/2398862.

- 790 Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a  
791 curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic*  
792 *Acids Research* 35, D61–D65. doi:10.1093/nar/gkl842.
- 793 R Core Team. 2021. R: A language and environment for statistical computing. R foundation for  
794 statistical computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 795 Rannala, B., and Yang, Z. (2003). Bayes estimation of species divergence times and ancestral  
796 population sizes using DNA sequences from multiple loci. *Genetics* 166, 1645–1656. doi:  
797 10.1093/genetics/164.4.1645.
- 798 Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical*  
799 *Biosciences* 53, 131–147. doi:10.1016/0025-5564(81)90043-2.
- 800 Sage, R. F., Khoshraves, R., and Sage, T. L. (2014). From proto-Kranz to C<sub>4</sub> Kranz: Building  
801 the bridge to C<sub>4</sub> photosynthesis. *Journal of Experimental Botany* 65, 3341–3356.  
802 doi:10.1093/jxb/eru180.
- 803 Sage, R. F., Monson, R. K., Ehleringer, J. R., Adachi, S., and Pearcy, R. W. (2018). Some like it  
804 hot: the physiological ecology of C<sub>4</sub> plant evolution. *Oecologia* 187, 941–966.  
805 doi:10.1007/s00442-018-4191-6.
- 806 Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C<sub>4</sub>  
807 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi:10.1146/annurev-arplant-042811-  
808 105511.
- 809 Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., et al. (2013).  
810 Initial events during the evolution of C<sub>4</sub> photosynthesis in C<sub>3</sub> species of *Flaveria*. *Plant*  
811 *Physiology* 163, 1266–1276. doi:10.1104/pp.113.221119.



- 812 Sayyari, E., and Mirarab, S. (2016). Fast coalescent-based computation of local branch support  
813 from quartet frequencies. *Molecular Biology and Evolution* 33, 1654–1668.  
814 doi:10.1093/molbev/msw079.
- 815 Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593.  
816 doi:10.1093/bioinformatics/btq706.
- 817 Schwenk, K., Brede, N., and Streit, B. (2008). Introduction. Extent, processes and evolutionary  
818 impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal*  
819 *Society B: Biological Sciences* 363, 2805–2811. doi:10.1098/rstb.2008.0055.
- 820 Scornavacca, C., Belkhir, K., Lopez, J., Dernas, R., Delsuc, F., Douzery, E. J. P., et al. (2019).  
821 OrthoMaM v10: Scaling-up orthologous coding sequence and exon alignments with more  
822 than one hundred mammalian Genomes. *Molecular Biology and Evolution* 36, 861–862.  
823 doi:10.1093/molbev/msz015.
- 824 Smith, B. N., and Turner, B. L. (1975). Distribution of Kranz syndrome among asteraceae.  
825 *American Journal of Botany* 62, 541–545. doi:10.2307/2441964.
- 826 Smith, S. A., Moore, M. J., Brown, J. W., and Yang, Y. (2015). Analysis of phylogenomic  
827 datasets reveals conflict, concordance, and gene duplications with examples from animals  
828 and plants. *BMC Evolutionary Biology* 15, 150. doi:10.1186/s12862-015-0423-0.
- 829 Smith, S. A., and O’Meara, B. C. (2012). treePL: divergence time estimation using penalized  
830 likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690.  
831 doi:10.1093/bioinformatics/bts492.
- 832 Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate:  
833 reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*  
834 26, 1134–1144. doi:10.1101/gr.196469.115.

- 835 Solís-Lemus, C., and Ané, C. (2016). Inferring phylogenetic networks with maximum  
836 pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* 12, e1005896-21.  
837 doi:10.1371/journal.pgen.1005896.
- 838 Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev.*  
839 *Plant Biol.* 60, 561–588. doi:10.1146/annurev.arplant.043008.092039.
- 840 Song, L., and Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina  
841 RNA-seq reads. *GigaScience* 4. doi:10.1186/s13742-015-0089-y.
- 842 Stamatakis, A. (2014). RAxML version 8 - a tool for phylogenetic analysis and post-analysis of  
843 large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033/-  
844 /DC1.
- 845 Swofford, D. (2002). PAUP\*. Phylogenetic analysis using parsimony (\*and other methods)  
846 version 4. *Sunderland, MA: Sinauer Associates*.
- 847 Taniguchi, Y. Y., Gowik, U., Kinoshita, Y., Kishizaki, R., Ono, N., Yokota, A., et al. (2021).  
848 Dynamic changes of genome sizes and gradual gain of cell-specific distribution of C<sub>4</sub>  
849 enzymes during C<sub>4</sub> evolution in genus *Flaveria*. *The Plant Genome* 14, e20095.  
850 doi:10.1002/tpg2.20095.
- 851 Tefarikis, D. T., Morales-Briones, D. F., Yang, Y., Edwards, G., and Kadereit, G. (2021). On the  
852 hybrid origin of the C<sub>2</sub> *Salsola divaricata* agg. (Amaranthaceae) from C<sub>3</sub> and C<sub>4</sub> parental  
853 lineages. *bioRxiv*, 2021.09.23.461503. doi:10.1101/2021.09.23.461503.
- 854 Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and  
855 reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9, 322–16.  
856 doi:10.1186/1471-2105-9-322.

- 857 van Dongen, S. M. (2000). Graph Clustering by Flow Simulation. Available at:  
858 <https://dl.acm.org/citation.cfm?id=868979>.
- 859 Wen, D., Yu, Y., Zhu, J., and Nakhleh, L. (2018). Inferring phylogenetic networks using  
860 PhyloNet. *Systematic Biology* 67, 735–740. doi:10.1093/sysbio/syy015.
- 861 Yang, Y., Moore, M. J., Brockington, S. F., Mikenas, J., Olivieri, J., Walker, J. F., et al. (2018).  
862 Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events  
863 in Caryophyllales, including two allopolyploidy events. *New Phytologist* 217, 855–870.  
864 doi:10.1111/nph.14812.
- 865 Yang, Y., and Smith, S. A. (2013). Optimizing *de novo* assembly of short-read RNA-seq data for  
866 phylogenomics. *BMC Genomics* 14, 328. doi:10.1186/1471-2164-14-328.
- 867 Yang, Y., and Smith, S. A. (2014). Orthology inference in nonmodel organisms using  
868 transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for  
869 phylogenomics. *Molecular Biology and Evolution* 31, 3081–3092.  
870 doi:10.1093/molbev/msu245.
- 871 Yu, Y., and Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic  
872 networks. *BMC Genomics* 16, S10. doi:10.1186/1471-2164-16-S10-S10.
- 873 Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time  
874 species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19,  
875 153. doi:10.1186/s12859-018-2129-y.