Morales-Briones, D. F., & Kadereit, G. (2023). Exploring the possible role of hybridization in the evolution of photosynthetic pathways in *Flaveria* (Asteraceae), the prime model of C_4 photosynthesis evolution. *Bulletin of the Society of Systematic Biology*. https://doi.org/10.18061/bssb.v2i3.8992

Investigations

Exploring the Possible Role of Hybridization in the Evolution of Photosynthetic Pathways in *Flaveria* (Asteraceae), the Prime Model of C_4 Photosynthesis Evolution

Diego F. Morales-Briones, Ph. D[1] [a], Gudrun Kadereit, Ph. D[1]

[1] Princess Therese von Bayern chair of Systematics, Biodiversity and Evolution of Plants, Ludwig-Maximilians-Universität München, Munich, Germany.

Keywords: C3-C4 intermediates, C4 photosynthesis evolution, Hybridization, Reticulate evolution, Species network analysis

https://doi.org/10.18061/bssb.v2i3.8992

Bulletin of the Society of Systematic Biology

Abstract

Flaveria (Asteraceae) is the prime model for the study of C_4 photosynthesis evolution and seems to support a stepwise acquisition of the pathway through C_3-C_4 intermediate phenotypes, still existing in *Flaveria* today. Molecular phylogenies of *Flaveria* based on concatenated data matrices are currently used to reconstruct the complex sequence of trait shifts during C_4 evolution. To assess the possible role of hybridization in C_4 evolution in *Flaveria*, we re-analyzed transcriptome data of 17 *Flaveria* species to infer the extent of gene tree discordance and possible reticulation events. We found massive gene tree discordance as well as reticulation along the backbone and within clades containing C_3-C_4 intermediate and C_4-like species. An early hybridization event between two C_3 species might have triggered C_4 evolution in the genus. The clade containing all C_4 species plus the C_4-like species *F. vaginata* and *F. palmeri* is highly supported in our phylogenetic analyses, but it might be of hybrid origin involving *F. angustifolia* and *F. sonorensis* (both C_3-C_4 intermediate) as parental lineages. Hybridization seems to be a driver of C_4 evolution in *Flaveria* and likely promoted the fast acquisition of C_4 traits. This new insight can be used in further exploring C_4 evolution and can inform C_4 bioengineering efforts.

INTRODUCTION

The detection of gene tree discordance is common in the phylogenomic era. Discordance can be the product of multiple processes and is commonly attributed to either incomplete lineage sorting (ILS) and/or hybridization (Doyle, 1992; Galtier & Daubin, 2008; Pamilo & Nei, 1988). Hybridization is a fundamental process in the evolution of animals, plants, and fungi (Giraud et al., 2008; Payseur & Rieseberg, 2016; Schwenk et al., 2008; Soltis & Soltis, 2009), and methods to investigate hybridization in a phylogenetic context have recently been developed. These include methods that estimate phylogenetic networks while accounting for ILS and hybridization simultaneously (e.g., Solís-Lemus & Ané, 2016; Wen et al., 2018) and methods that detect hybridization based on site patterns or phylogenetic invariants (e.g., Durand et al., 2011; Green et al., 2010; Kubatko & Chifman, 2019). The current ease to produce phylogenomic data sets and the availability of new analytical methods facilitate the exploration of reticulate evolution in any clade across the Tree of Life—including those that have particular significance as model lineages, such as the flowering plant genus *Flaveria* (Asteraceae) for the study of C_4 photosynthesis evolution.

Flaveria Juss. belongs to the sunflower tribe Heliantheae (Anderberg et al., 2007). According to the most recent revision by Powell (1978), *Flaveria* includes 21 morphologically similar species distributed mainly in southern USA and northern Mexico, with few species occurring in the Caribbean and South America. The two weedy and self-compatible C_4 species, *F. trinervia* and *F. bidentis*, have been introduced almost worldwide (https://powo.science.kew.org/). Species of *Flaveria* usually show scattered occurrences in unconnected, localized populations near rivers, creeks, irrigation canals, fields, roadsides, and ponds, often on saline or gypseous soils (Powell, 1978). They are either robust shrubs or herbaceous perennials, or annuals (mainly the C_4 species). The genus stands out in Asteraceae for its reduced floral features and reduced and secondarily aggregated capitula (Anderberg et al., 2007). Reduction is most evident in *F. trinervia* (C_4), and aggregation of capitula mimicking a single capitulum in *F. anomala*

[a] **Correspondending author:** Diego F. Morales-Briones. Email: dfmoralesb@gmail.com

(C_3-C_4). *Flaveria* is consistently diploid (see Powell, 1978 and ref. therein; only exceptions are some tetraploid populations of *F. pringlei*) with a haploid chromosome number of $n = 18$. Artificial hybridization among 16 species of *Flaveria* was successful, and F1 hybrids could be obtained in most species' combinations (see Table 1 in Powell, 1978). F2 and backcross crosses also resulted in offspring in a high number of combinations, but only four of these were fertile. Powell (1978) excluded frequent natural hybridization in *Flaveria* mainly because of geographical isolation.

C_4 photosynthesis in *Flaveria* was first recognized by Smith and Turner (1975), and the presence of C_3-C_4 intermediate species was first noted by Brown (pers. comm. in Powell, 1978), first verified by Apel and Maass (1981), and then studied in detail biochemically in four species by Ku et al. (1983) and Nakamoto et al. (1983). Numerous publications characterizing the physiology and biochemistry of C_3-C_4 intermediate species of *Flaveria* followed (Ku et al., 1991 and ref. therein). At the same time, crossing experiments of C_3 and C_4 *Flaveria* species as well as backcrosses or crosses between C_3-C_4 intermediate species revealed the transfer of C_4 properties as well as the simultaneous functioning of C_3 and C_4 pathways in hybrids (see Apel et al., 1988 as an example and Kadereit et al., 2017 for review). Of all genera that contain C_3-C_4 intermediate species, *Flaveria* has the highest diversity of C_3-C_4 phenotypes, including C_2 photosynthesis (R. F. Sage et al., 2012; briefly described in Table 1), and arguably is the only lineage that allows to infer a detailed sequence of increasing C_4-ness (R. F. Sage et al., 2012). Against this background, *Flaveria* qualified as the model group for the establishment (Monson & Moore, 1989) and subsequent refinement of a model of stepwise acquisition of C_4 photosynthesis (R. F. Sage et al., 2014 and ref. therein).

One challenge of the *Flaveria* model is that C_3-C_4 intermediate phenotypes might also have resulted from reticulation during the diversification of the genus, especially when reproductive barriers are leaky among extant species as soon as they get into contact (Powell, 1978) and hybridization between C_3 and C_4 species is possible. Therefore, a phylogenetic study exploring the occurrence and location of past reticulation events is needed. Comprehensive molecular phylogenetic studies of *Flaveria* published so far were either based on few molecular markers only (McKown et al., 2005) or on concatenated data matrices and inference methods unable to reveal tree discordance, possible reticulation, or incomplete lineage sorting (Lyu et al., 2015). Using the chloroplast *trnL-F* and nuclear ITS and ETS regions, McKown et al. (2005) recovered two strongly supported clades, which included most of the C_4 species (Clade A) and C_3-C_4 intermediate (Clade B), respectively. In turn, Lyu et al. (2015) used a phylotranscriptomic approach and recovered a congruent topology to McKown et al. (2005). Irrespective of this, numerous current studies of evolutionary change during the establishment of the C_4 pathway rely on the *Flaveria* model (e.g., Lyu et al., 2021; Taniguchi et al., 2021).

The aim of this study is to use available transcriptome data of 17 species of *Flaveria* to assess the extent of reticulation during the diversification of the genus, and to evaluate these findings with respect to the evolution of C_4 photosynthesis in the genus and its suitability as general model of C_4 evolution.

METHODS

Taxon sampling

We included publicly available transcriptomes from 17 species of *Flaveria* (Table S1). In addition, we included outgroups from four genomes of Asteraceae, *Chrysanthemum seticuspe* (Maxim.) Hand.-Mazz., *Helianthus annuus L., Lactuca sativa* L., and *Stevia rebaudiana* (Bertoni) Bertoni, following Mandel et al. (2019; Table S1).

Homology and orthology inference

Raw read processing, transcriptome assembly, low-quality and chimeric transcript removal, transcript clustering into putative genes, translation, and final coding sequences (CDS) redundancy assessment were carried out following Morales-Briones et al. (2021) with minor modifications as follows. Sequencing errors in raw reads were corrected with Rcorrector (Song & Florea, 2015), and reads flagged as uncorrectable were discarded. Sequencing adapters and low-quality bases were removed with Trimmomatic v 0.39 (Bolger et al., 2014). Additionally, chloroplast and mitochondrial reads were filtered out with Bowtie2 v 2.4.4 (Langmead & Salzberg, 2012) using publicly available Asterales organelle genomes from the Organelle Genome Resources database (RefSeq; [Pruitt et al., 2007]; last accessed on June 4, 2021) as references. Read quality was assessed with FastQC v 0.11.9 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc), and overrepresented sequences were discarded. *De novo* assembly was carried out with Trinity v 2.13.2 (Grabherr et al., 2011) with default settings but without in silico normalization. Assembly quality was assessed with Transrate v 1.0.3 (Smith-Unna et al., 2016). Low quality and poorly supported transcripts were removed using individual cut-off values for three contig score components of Transrate: 1) proportion of nucleotides in a contig that agrees in identity with the aligned read, s(Cnuc) $\leqslant$ 0.25; 2) proportion of nucleotides in a contig that have one or more mapped reads, s(Ccov) $\leqslant$0.25; and 3) proportion of reads that map to the contig in correct orientation, s(Cord) $\leqslant$ 0.5. Furthermore, chimeric transcripts (*trans*-self and *trans*-multi-gene) were removed following the approach described in Yang and Smith (2013) using *Helianthus annuus* as the reference proteome and a percentage similarity and length cutoff of 30 and 100, respectively. To remove isoforms and assembly artifacts, filtered reads were remapped to filtered transcripts with Salmon v 1.5.2 (Patro et al., 2017), and putative genes were clustered with Corset v 1.09 (Davidson & Oshlack, 2014) using default settings, except that we used a minimum of five reads as threshold to remove transcripts with low coverage (-m 5). Only the longest transcript of each putative gene inferred by Corset was retained as suggested in Chen et al. (2019). Filtered transcripts were translated with TransDecoder v 5.3.0 (Haas

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

Photosynthetic type	Characteristics	Species representing this type[1]
C_3	Photorespiratory cycle operates completely within single M cells, BS cells small[2] with few organelles, veins widely spaced.	*F. cronquistii* A.M. Powell[2] *F. mcdougallii* M.E. Theroux, Pinkava & D.J. Keil*
C_3 proto-kranz	Functionally C_3, activated BS cells, greater vein density, mitochondria localized at the inner BS wall adjacent to the vasculature.	*F. pringlei* Gand. *F. robusta* Rose
C_2 Type I	Low or no GDC expression in M cells, high number of centripetally located organelles in the BS cells, CO_2 compensation point reduced in comparison to C_3, lack of any C_4 cycle.	*F. angustifolia* (Cav.) Pers. *F. chlorifolia* A. Gray *F. sonorensis* A.M. Powell
C_2 Type II	In addition to type I, modest C_4 cycle enhancement.	*F. anomala* B.L. Rob. *F. floridana* J.R. Johnst. *F. linearis* Lag.* *F. pubescens* Rydb. *F. ramosissima* Klatt
C_4-like	Strong C_4 metabolic cycle but also weak C_3 cycle in the M cells.	*F. brownii* A.M. Powell *F. palmeri* J.R. Johnst. *F. vaginata* B.L. Rob. & Greenm.
C_4	No C_3 cycle, CO_2-saturate photosynthesis below 500 ppm CO_2.	*F. bidentis* (L.) Kuntze *F. campestre* J.R. Johnst.* *F. kochiana* B.L. Turner *F. trinervia* (Spreng.) C. Mohr *F. australasica* Hook.

[1] *Flaveria oppositifolia** so far only classified as C_2 without further specification.

[2] *Flaveria cronquistii* qualifies as a C_3+ species (see R. F. Sage et al., 2018) because it has larger and photosynthetically more active BS cells (McKown & Dengler, 2007).

et al., 2013) with default settings and the proteomes of *Arabidopsis thaliana* (L.) Heynh., *Helianthus annuus,* and *Lactuca sativa* to identify open reading frames. Finally, coding sequences (CDS) from translated amino acids were further reduced with CD-HIT v 4.8.1 (-c 0.99; [Fu et al., 2012]) to remove near-identical sequences. Scripts used can be found at https://bitbucket.org/yanglab/phylogenomic*dataset*construction/src/master/ (Morales-Briones et al., 2021).

Homology inference was done with an all-by-all BLASTN search on CDS with an *E* value cutoff of 10. BLAST hits were filtered with a minimal hit coverage of 40%. Homolog groups were clustered with MCL v 14-137 (van Dongen, 2000) using a minimum minus log-transformed *E* value cutoff of 5 and an inflation value of 1.4, and only clusters with at least 17 taxa were retained. Homolog cluster sequences were aligned using the OMM_MACSE v 11.05 pipeline (Scornavacca et al., 2019). Alignments were further trimmed to remove columns with more than 90% missing data using Phyx (Brown et al., 2017). Homolog trees were inferred using RAxML v 8.2.11 (Stamatakis, 2014) with the GTRCAT model and 200 rapid bootstrap (BS) replicates. Monophyletic and paraphyletic tips of the same species were removed, keeping the tip with the highest number of characters in the trimmed alignment following Yang and Smith (2014). Spurious tips were detected and removed using TreeShrink v 1.3.9 (Mai & Mirarab, 2018) with the 'per-gene' mode, a false positive error rate threshold (α) of 0.05, and excluding the outgroups. Trees were visually inspected, and deep paralogs producing internal branch lengths longer than 0.20 were cut apart to retain subclades with at least 10 taxa to obtain final homolog trees. Orthology inference was done using the 'monophyletic outgroup' (MO) approach from Yang and Smith (2014). The MO approach filters for trees that have outgroup taxa being monophyletic and single-copy and therefore filters for single- and low-copy genes. This approach roots the gene tree by the outgroups, traverses the rooted tree from root to tip, and removes the side with less taxa when gene duplication is detected (Yang & Smith, 2014). If no taxon duplication is detected in a homolog tree, the MO approach outputs a one-to-one ortholog. We set all species of *Flaveria* as ingroups, and *Chrysanthemum, Helianthus, Lactuca,* and *Stevia* as outgroups, keeping only orthologs that included at least 10 taxa resulting in 5,981orthologs.

Tree inference and detection of gene tree conflict

Sequences from individual orthologs were aligned using the OMM_MACSE pipeline. Columns with more than 20% missing data were trimmed with Phyx, and only alignments with at least 500 characters and all 21 taxa were retained and concatenated resulting in final matrix of 2,124 orthologs. We estimated a maximum likelihood (ML) tree of the concatenated matrix with IQ-TREE v 2.1.3 (Minh et al., 2020), searching for the best partition scheme (Lanfear et al., 2012) followed by ML gene tree inference and 1000 ultrafast bootstrap replicates for clade support. To estimate a coalescent-based species tree, first, we inferred individual gene trees with IQ-TREE using extended model selection (Kalyaanamoorthy et al., 2017) and 200 non-parametric

bootstrap replicates for clade support. Gene trees were then used to infer a species tree with ASTRAL-III v 5.7.7 (Zhang et al., 2018), using local posterior probabilities (LPP; Sayyari & Mirarab, 2016) to assess clade support.

We explored gene tree discordance by calculating the number of concordant and discordant bipartitions on each node of the concatenated and ASTRAL trees using Phyparts (S. A. Smith et al., 2015). Calculations were done using only individual gene tree nodes with BS ⩾ 50%. Additionally, to distinguish conflict from poorly supported branches, we carried out a Quartet Sampling (QS; Pease et al., 2018) analysis using the concatenated matrix with a partition by gene (–genetrees), the concatenated IQ-TREE and ASTRAL trees, and 1000 replicates. To further visualize gene tree conflict, we built a cloudogram with the DensiTree function from Phangorn v 2.7.1 (Schliep, 2011) in R (R Core Team, 2021). We first time-calibrated individual ortholog gene trees, for visualization purposes only, with TreePL v 1.0 (S. A. Smith & O'Meara, 2012). The most recent common ancestor (MRCA) of *Helianthus* and *Flaveria* was fixed to 21.5 Ma, and the MRCA of *Flaveria* was fixed to 4.3 Ma based on Mandel et al. (2019). The cloudogram was plotted using 1,000 random trees from the final 2,127 orthologs that had all 21 taxa.

Testing for potential reticulation

First, to investigate if gene tree discordance can be explained by ILS alone, we performed coalescent simulations like Cloutier et al. (2019). An ultrametric tree with branch lengths in mutational units (μT) was inferred with PAUP v 4.0a (build 168; Swofford, 2002) by constraining a ML tree search to the ASTRAL tree and using the concatenated alignment, a GTRGAMMA model, and enforcing a strict molecular clock. The mutational branch lengths from the constrained tree and branch lengths in coalescent units ($\tau = T/4N_e$) from the ASTRAL tree were used to estimate the population size parameter theta ($\Theta = \mu T/\tau$; Degnan & Rosenberg, 2009) for internal branches. Terminal branches were set with $\Theta = 1$. We then used Phybase v 1.4 (Liu & Yu, 2010), that implements the formula from Rannala and Yang (2003), to simulate 10,000 gene trees using the constraint tree and the estimated theta values. Lastly, we calculated the distribution of Robinson and Foulds (1981) tree-to-tree distances between the ASTRAL tree and each original gene tree using Phangorn and compared this with the distribution of tree-to-tree distances between the ASTRAL tree and the simulated gene trees. Following the same logic from Maureira-Butler et al. (2008), if the distances between the gene trees and the species tree are larger than 95% of the distribution of tree-to-tree distances of the simulated trees and the species tree, then ILS alone is considered unlikely to explain most of the observed gene tree incongruence.

To test for potential reticulation, we inferred species networks using maximum pseudo-likelihood (Yu & Nakhleh, 2015) in PhyloNet v 3.8.2 (Than et al., 2008) with the command "InferNetworks_MPL" and using individual ML gene trees as input. We included all 17 species of *Flaveria* and *Helianthus* as an outgroup (18-taxon data set). Network searches were performed allowing for up to 12 reticulation events, collapsing gene tree nodes with BS < 50%, and ten runs per search. To find the network with optimal number of reticulations, we plotted the number of reticulations versus the pseudo-likelihood score to determine when the score stabilizes (Blair & Ané, 2019). We performed one additional round of searches by removing *F. pringlei* (C_3; 17-taxon data set) as it has been identified as a potential artificial hybrid (Lyu et al., 2015). This species was inferred to be a product of reticulation events in all ten original searches and was not involved in additional reticulation events (i.e., it is not a parental lineage of other reticulation events; see results). Network searches for the reduced data set were carried out similarly as with the original data set.

Additionally, we tested for hybridization with HyDe (Blischak et al., 2018), which uses site pattern frequencies (Kubatko & Chifman, 2019) to quantify the hybridization parameter γ between two parental lineages that form a hybrid lineage. We tested all triplet combinations using the 'run_hyde.py' script, the concatenated alignment, and a mapping file to assign species. Test significance was assessed with a Bonferroni correction (α= 0.05) for the number of hypothesis tests conducted with estimates of γ between 0 and 1 (Blischak et al., 2018). We also carried out HyDe hybridization tests using all species and without *F. pringlei* (C_3).

Assessment of whole genome duplication

To investigate potential whole genome duplication as a product of reticulation events in *Flaveria* (see results), we mapped gene duplication events onto the inferred species tree following Yang et al., 2018. First, we extracted rooted ingroup clades (orthogroup) from the final homolog trees by requiring at least 15 taxa, and only orthogroups with an average BS ⩾ 50 were used for mapping. Gene duplication events were then mapped onto the MRCA on the species tree when two or more taxa overlapped between the two daughter clades on the rooted ingroup clade. Each node on a species tree can be counted only once from each gene tree to avoid nested gene duplications inflating the number of recorded duplications (Yang et al., 2018). Orthogroup extraction and mapping were carried out using the scripts "extract_clades.py" and "map_dups_mrca.py" from https://bitbucket.org/blackrim/clustering (Yang et al., 2018).

RESULTS

Orthology inference and phylogenetic analysis

The final number of orthologs with at least ten species was 5,981 with a mean of 5,512 orthologs per species (Table S1). The number of orthologs that included all 21 species was 2,127. The concatenated matrix (⩾ 500 bp per ortholog) consisted of 2,796,180 aligned columns with a character occupancy of 94% from 2,124 orthologs. The topologies from the IQ-TREE and ASTRAL trees were similar and most nodes had maximum support (BS =100, LPP = 100; Fig. 1; Fig. S1). *Flaveria cronquistii* (C_3), *F. robusta* (C_3), *F. sonoren-*

sis (C_3-C_4), and *F. pringle*i (C_3) + *F. angustifolia* (C_3-C_4) were consecutive sisters to Clade A + Clade B (clades names followed McKown et al., 2005). The only difference between the IQ-TREE and ASTRAL topologies was the placement of *F. sonorensis*. Clade A comprised the same species and relationships as in Lyu et al. (2015). *Flaveria ramosissima* (C_3-C_4) and *F. palmeri* (C_4-like) were successive sisters to *F. kochiana* (C_4) + *F. vaginata* (C_4-like) and the C_4 clade [(*F. bidentis*, *F. trinervia* + *F australasica*); BS = 83]. Clade B included the same species as in Lyu et al. (2015). *Flaveria anomala* (C_3-C_4) as sister to the remaining species, followed by *F. pubescens* (C_3-C_4), *F. floridana* (C_3-C_4), and *F. chlorifolia* (C_3-C_4) + *F. brownii* (C_4-like).

Phylogenetic conflict

Overall, conflict analyses and cloudogram visualization revealed rampant gene tree discordance in *Flaveria* (Fig. 1; Fig S2). The cloudogram showed significant conflict along the backbone of the phylogeny as well as within clades A and B (Fig. 1). The placements of *F. cronquistii* (C_3), *F. robusta* (C_3), and *F. sonorensis* (C_3-C_4) showed moderate support of informative gene trees and moderate to high QS support with clear signal of alternative placement between these taxa. In turn, the placement of *F. pringlei* (C_3) + *F. angustifolia* (C_3-C_4) as sister of Clade A + Clade B was supported only by 193 (of 1,493) gene trees and had counter QS support (-0.11/0.45/0.41) with clear signal for an alternative topology with *F. sonorensis* (C_3-C_4), also reflected in the discordant topology between the IQ-TREE and ASTRAL topologies regarding this node (Fig. S1). The clade composed of clades A and B was supported only by 241 (of 1,217) gene trees and had low QS support (0.17/1/0.44) with no clear alternative topology.

Clade A was supported by 725 (of 1,424) gene trees and had moderate QS support (0.19/0.78/0.47) with signals of an alternative topology for *F. ramosissima* (C_3-C_4) outside the clade. The remaining species of Clade A formed a clade supported by most gene trees (1,653 of 1,882) and a strong QS score (0.75/0.89/0.77) with no signals of an alternative topology. Species within this clade, *F. palmeri* (C_4-like), *F. vaginata* (C_4-like), *Flaveria kochiana* (C_4), and *F. bidentis* (C_4), showed low gene tree and QS support with clear signals of alternative topologies among them, while *Flaveria trinervia* (C_4) + *F. australasica* (C_4) was supported by most gene trees (1,960 of 2,021) and had strong QS support. Clade B was supported by 965 (of 1,415) gene trees and moderate QS support (0.29/0.90/0.48) with no signal of an alternative topology. Placement of *F. anomala* (C_3-C_4) and *F. pubescens* (C_3-C_4) were strongly supported by gene trees and QS scores, while the placement of the remaining species, *F. floridana* (C_3-C_4), *F. chlorifolia* (C_3-C_4), and *Flaveria brownii* (C_4-like) showed low gene tree and QS support with clear signal of alternative topologies among them.

Potential widespread reticulation

The distribution of tree-to-tree distances of the empirical and simulated gene trees to the ASTRAL tree showed some overlap (Fig. S4), but there was a skew towards larger distances in the empirical trees (mean 16.68; 95% CI [16.87, 16.51]) compared to distances in the simulated trees (mean 11.16; 95% CI [11.29, 11.15]). This suggested that ILS alone cannot explain most of the observed gene tree incongruence (Maureira-Butler et al., 2008).

Species network searches were allowed up to 12 reticulations and recovered topologies with up to nine reticulation events for the 18-taxon data set (Fig. S5). The pseudo-likelihood score continually improved with the inclusion of additional reticulation (Fig. S6). The network search that was allowed 10 reticulations and recovered nine reticulation events had the best pseudo-likelihood score (Fig. 2A; Fig. S6). Although the best-scored network showed complicated and nested reticulation patterns (Fig. 2A), there are several clear patterns among most networks (Fig. S5). *Flaveria pringlei* (C_3) was consistently recovered as a hybrid in all 12 networks with parental lineages and inheritance probabilities consistent across networks (Fig. S5). Other reticulation patterns recovered across networks were the hybrid origin of clades A (mainly C_4 and C_4-like) and B (mainly C_3-C_4), which both had *F. sonorensis* (C_3-C_4; itself a hybrid in several networks) and *F. angustifolia* (C_3-C_4; or closely related to lineage) as potential parental lineages. Several reticulation events were also detected among species of Clade A. The last reticulation event recovered in all networks involved the C_3 species *F. cronquistii* and *F. robusta* which were potential parental lineages of all remaining *Flaveria* species. The analyses of the 17-taxon data set (Fig. S5) resulted in a best-scoring network with eight reticulation events (Fig. 2B). This showed patterns like the 18-taxon data set regarding the hybrid origin of Clades A and B, and the deep reticulation involving *F. cronquistii* (C_3) and *F. robusta* (C_3). Also, it recovered several reticulation events within Clade A, although they slightly differ from the 18-taxon data set network (Fig. 2).

The HyDe analysis of all possible triples using the 17 species of *Flaveria* resulted in 2,040 hybridization tests, of which 376 triples were significant (Table S2). The analyses without *F. pringlei* (C_3) resulted in 296 significant hybridization tests of 1,680 overall tests (Table S3). HyDe analyses detected 15 species as potential hybrids (Fig. 3). These species included all members of Clade A which comprises four C_4, two C_4-like species, and one C_3-C_4 species. Similarly, it included all members of Clade B (except *F. pubescens*) which comprises five C_3-C_4 and one C_4-like species. In Clade A, most species had several potential parental lineages from across *Flaveria* and γ consistent with ancient hybridization (either closer to zero or one; Fig. 3). These patterns are compatible with the hybrid origin of the clade detected with PhyloNet (Fig. 2). *Flaveria kochiana* (C_4) had notably fewer potential parental lineages, suggesting that this species is also involved in a more recent reticulation event. Hybrids detected in Clade B also had several potential parental lineages from across *Flaveria,* and γ values were either closer to zero or one (Fig. 3), consistent with the PhyloNet results regarding the hybrid origin of the clade (Fig. 2). Interestingly, *F. pubescens* (C_3-C_4) was not detected as a hybrid by HyDe. Outside Clades A and B, *F.*

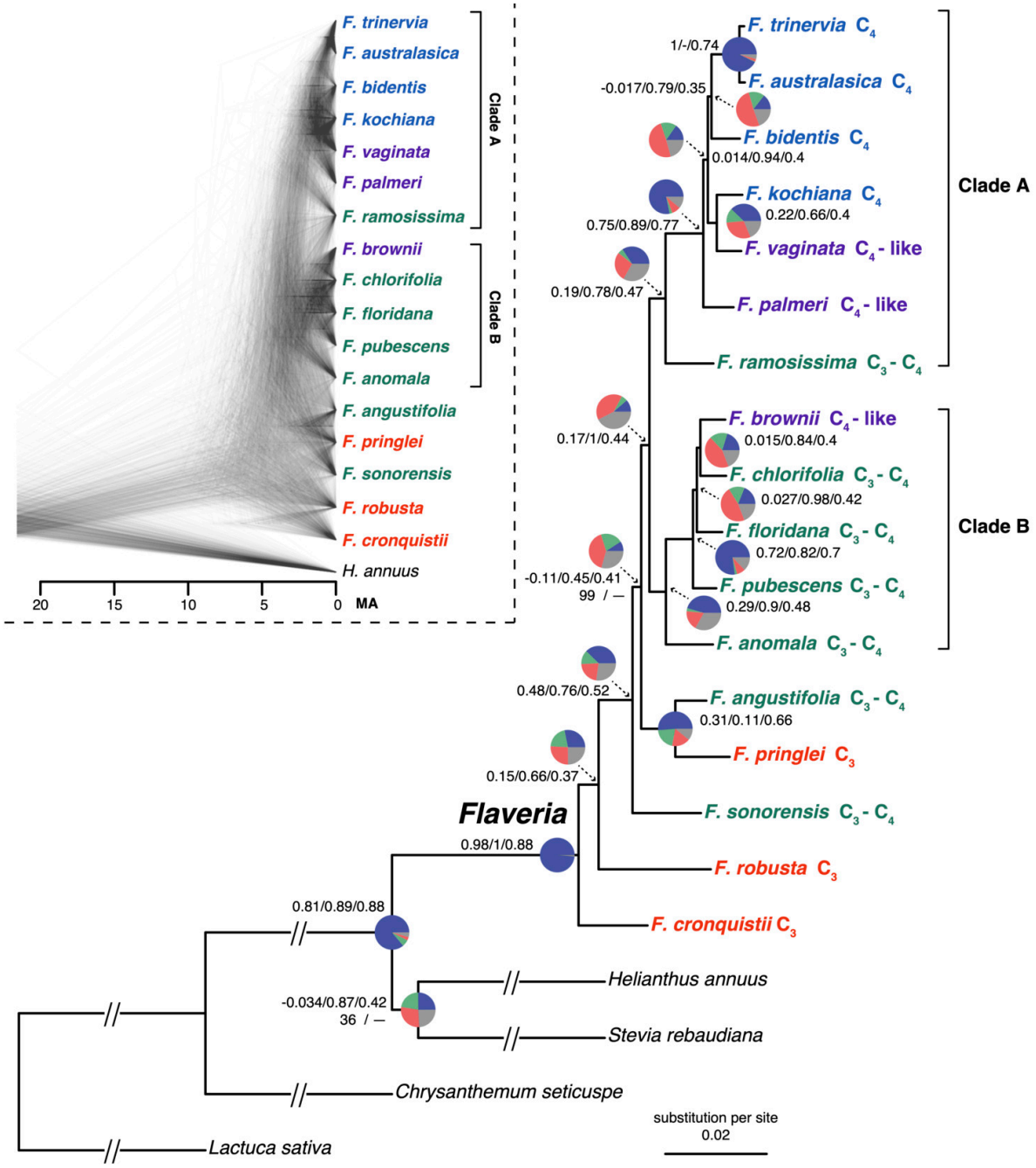

Figure 1. Maximum likelihood phylogeny of *Flaveria* inferred with IQ-TREE from the concatenated 2124-nuclear gene supermatrix.

Species names are colored by photosynthetic type. Quartet Sampling (QS) scores are shown above branches. QS scores: Quartet concordance/Quartet differential/Quartet informativeness. All nodes have full bootstrap support (BS =100) and local posterior probability (LPP =1) unless noted next to branches. Em dashes (—) denotes an alternative topology compared to the ASTRAL tree (Fig. S1). Pie charts represent the proportion of ortholog trees that support a clade (blue), the main alternative bifurcation (green), the remaining alternatives (red), and bifurcations (conflict or support) with < 50% bootstrap support (gray). Branch lengths as number of substitutions per site (scale bar). Exceptionally long branches were shortened with a broken segment (//) for illustration purposes (See Fig. S1 for original branch lengths). Inset: Cloudogram inferred from 1,000 random nuclear ortholog trees. Scale in millions of years ago (Ma).

pringlei (C_3) was detected as a hybrid as expected (Lyu et al., 2015), but the parental lineages come from across *Flaveria* with γ values suggesting ancient reticulation as also seen on the PhyloNet results. On the other hand, *F. angustifolia* and *F. sonorensis,* both C_3-C_4, also showed hybridization patterns consistent with ancient reticulations. Finally, HyDe detected *F. cronquistii* (C_3) as a potential ancient hybrid, which is consistent with some of the PhyloNet networks (Fig. S5) but not with the best one that recovered only *F. robusta* (C_3) as a hybrid. Both species had clear signals of alternative topologies in the discordance analyses (Fig. 1). Overall, HyDe detected all C_3-C_4 (except *F. pubescens*) and C_4-like species as well as *F. robusta* and *F. pringlei,* both C_3, as potential ancient hybrids with all 17 species involved in reticulation events.

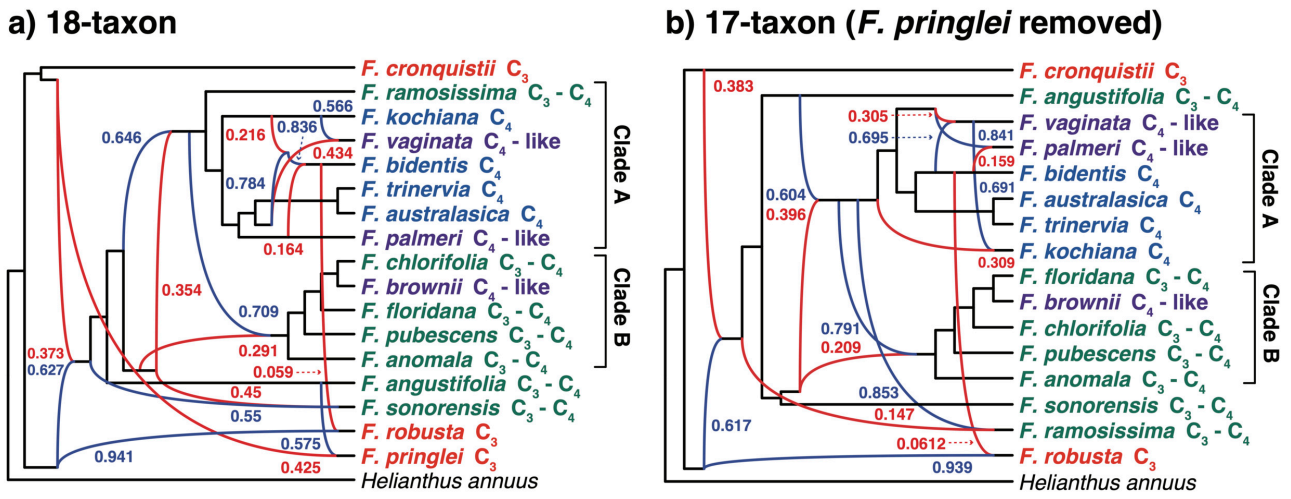

Figure 2. *Flaveria* species networks with the best maximum pseudo-likelihood scores with PhyloNet for the a) 18-taxon and b) 17-taxon data sets.

Species names are colored by photosynthetic type. Red and blue curved branches indicate the minor and major edges, respectively, of hybrid nodes. Numbers next to curved branches indicate inheritance probabilities for each hybrid node.

Lack of whole genome duplications

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

DISCUSSION

In C_4 photosynthesis, high rates of net photosynthesis and a highly competitive water and nitrogen use efficiency are achieved by spatially separated carbon fixation in the outer mesophyll cells preceding the Calvin-Benson cycle and by effectively fueling Rubisco with always high CO_2 concentration in the controlled seclusion of the Kranz cells (S. P. Long, 1999). Resulting low levels of photorespiration make C_4 plants competitive in various stressful environments where carbon deficiency poses a problem to C_3 plants (R. F. Sage et al., 2012). C_4 evolved more than 60 times in angiosperms with hotspots of C_4 origins in Poaceae and Amaranthaceae (R. F. Sage et al., 2018). Due to the anatomical and gene regulatory complexity of the C_4 pathway, it seems clear that there must have been intermediate stable phenotypes during the evolution of C_4 from a C_3 ancestor (Monson & Moore, 1989). The established generalized model of C_4 evolution tries to explain the sequence of intermediate adaptive events during the transition from the ancestral C_3 to C_4 using the naturally occurring C_3-C_4 intermediate phenotypes of *Flaveria* (and other lineages) as proxies of intermediate stages in C_4 evolution (Bräutigam & Gowik, 2016; R. F. Sage et al., 2014). However, *Flaveria* is also known for rampant interfertility of its 21 (mostly diploid) species (R. W. Long & Rhamstine, 1968; Powell, 1978) and for the transferability of photosynthetic traits to hybrid offspring in hybridization experiments (Apel et al., 1988; Kadereit et al., 2017 and ref. therein). Therefore, a clear understanding of the phylogenetic history of this model lineage of C_4 evolution, including tests for possible reticulate evolution, is fundamental as this might have strong implications for our understanding of the evolution of C_4 photosynthesis in general.

Our analyses using transcriptome data of 17 *Flaveria* species revealed rampant gene tree discordance along the backbone of the phylogeny as well as the clades containing most C_3-C_4 intermediate and C_4-like species (Clades A and B; visualized in the cloudogram in Fig. 1). Coalescence simulations showed that gene tree discordance in *Flaveria* cannot be attributed to ILS alone. Our initial species network analyses including all 17 species consistently identified *F. pringlei* (C_3) as hybrid (Fig. 2A), confirming the results of Lyu et al. (2015) that this accession is a recent hybrid which likely originated in the greenhouse. However, the exclusion of this accession did not alter the overall results (Fig. 2B). More importantly, the network analyses revealed several ancient hybridization events. An early reticulation between two C_3 lineages (*F. robusta* and *F. cronquistii*) gave rise to the ancestor of the lineage containing all C_3-C_4 intermediate species, C_4-like species, and C_4 species (Fig. 2B). This result is supported by the HyDe analysis which identified nearly all C_3-C_4 and C_4-like species as potential ancient hybrids. The ancestors of *F. robusta* and *F. cronquistii* seem to have contributed equally to the origin of this ancient hybrid lineage (Fig. 2B). Within this lineage there exist two parental lineages, *F. angustifolia* (C_3-C_4) and *F. sonorensis* (C_3-C_4), that seem to have contributed equally to a robust clade containing all C_4 species plus the two C_4-like species *F. vaginata* and *F. palmeri* (clade A without the C_3-C_4 *F. ramosissima* in Fig. 1). In contrast, Clade B, which consists of four C_3-C_4 intermediate species (*F. pubescens*, *F. chlorifolia*, *F. floridana*, *F. anomala,* and one C_4-like species, *F.*

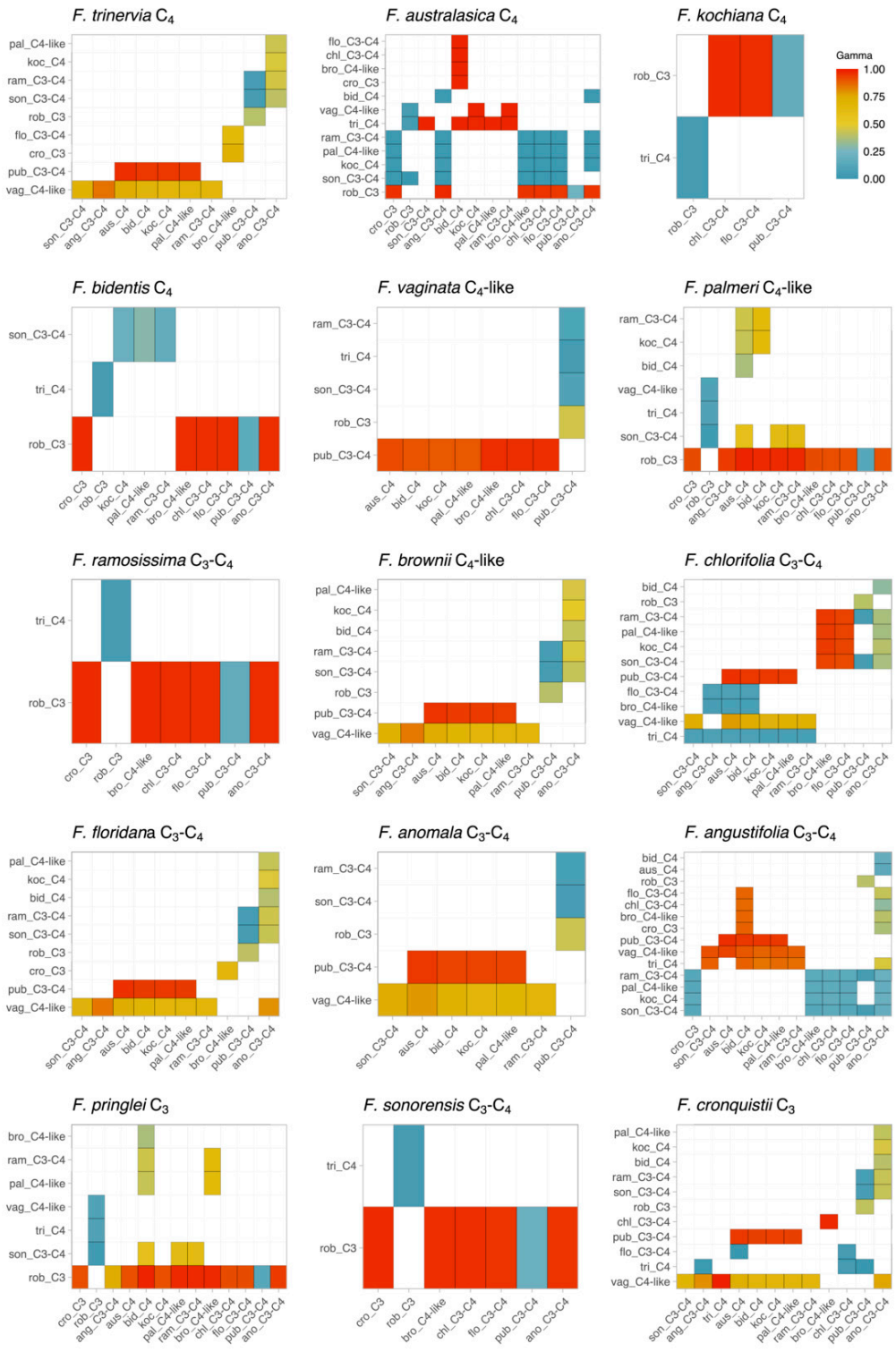

Figure 3. HyDe significant hybridization test for the 15 species of *Flaveria* identified as potential hybrids.

Species on the x-axis are parental lineage 1 (P1) and species on the y-axis are parental lineage 2 (P2). Only colored boxes denote possible combinations of P1 and P2 as parents of hybrid species. The color scale represents the value of the hybridization parameter γ for each hybridization event. Recent 50:50 hybrids would be represented by a γ ~0.5. Values of γ approaching 0 indicate a major hybrid contribution from P1, and values approaching 1 indicate a major hybrid contribution from P2, with both cases representing ancient hybridization. Parental species names as follow: ang_C3-C4 - *F. angustifolia* C_3-C_4; ano_C3-C4 - *F. anomala* C_3-C_4; aus_C4 - *F. australasica* C_4; bid_C4 - *F. bidentis* C_4; bro_C4-like - *F. brownii* C_4-like; chl_C3-C4 - *F. chlorifolia* C_3-C_4; cro_C3 - *F. cronquistii* C_3; flo_C3-C4 - *F. floridana* C_3-C_4; koc_C4 - *F. kochiana* C_4; pal_C4-like - *F. palmeri* C_4-like; pri_C3 - *F. pringlei* C_3; pub_C3-C4 - *F. pubescens* C_3-C_4; ram_C3-C4 - *F. ramosissima* C_3-C_4; rob_C3 - *F. robusta* C_3; son_C3-C4 - *F. sonorensis* C_3-C_4; tri_C4 - *F. trinervia* C_4; vag_C4-like - *F. vaginata* C_4-like.

brownii (Fig. 1), seems to originate from a hybridization event between the ancestor of Clade A (without *F. ramosissima*) and *F. sonorensis* (C_3-C_4). *Flaveria ramosissima* seems to be an ancient hybrid, too, with the ancestor of Clade A and the ancestor of the lineage containing all C_3-C_4 intermediate species, C_4-like species, and C_4 species as parental lineages.

These results have several important implications that will be discussed in the following three sections: 1) natural C_3-C_4 intermediates species in *Flaveria* do not seem to result from hybridization between a C_3 and a C_4 lineage (as similar C_3-C_4 intermediate phenotypes resulting from crossing experiments might suggest); 2) recurrent homoploid hybridization possibly played a major role in the evolution of C_4 photosynthesis in *Flaveria* with an initial hybridization event between two C_3 species (*F. robusta* and *F. cronquistii*) as a possible trigger; 3) the ancestral lineages of *F. angustifolia* and *F. sonorensis* (both C_3-C_4 intermediates) seem to be involved in the formation of the lineage that evolved full C_4 photosynthesis (Clade A).

Hybridization is an important factor in plant evolution and speciation in general (Abbott et al., 2013), and *Flaveria* is no exception to this but also shows hybridization in a more complex way than we expected. From the phenotypic outcome of the numerous $C_3 \times C_4$ hybridization experiments in *Flaveria,* it was conceivable that naturally occurring C_3-C_4 intermediate species might be the result of crosses between parental lineages with different photosynthetic types (Kadereit et al., 2017; Monson & Moore, 1989). However, our results indicate that the origin of C_3-C_4 intermediates might be more complex and resulting from several deep reticulation events challenging the reconstruction of the backbone (Fig. 1; see below). Recent reticulations involve *F. pringlei* (C_3 proto-kranz) as a hybrid between *F. cronquistii* (C_3) and *F. angustifolia* (C_3-C_4; Fig. 2a). Therefore, the proto-kranz phenotype in this species might be of different origin than that of *F. robusta* (Table 1), although HyDe results suggested a large contribution from *F. robusta* to *F. pringlei* (Fig. 3). In fact, Lyu et al. (2015) showed the *F. pringlei* sample used in their study to be of artificial hybrid origin (possibly resulting from unintended crosses in the greenhouse). Trait assessments in these two species, e.g., of leaf anatomical and ultrastructural traits (McKown & Dengler, 2007; T. L. Sage et al., 2013), might lead to re-evaluation of their assessment as evolutionary "basal" as they likely acquired the proto-kranz phenotype differently. Also, the similarities in bundle sheath and mesophyll characteristics between *F. cronquistii* and *F. pringlei* found by McKown and Dengler (2007: Table 1) and their differences to *F. robusta* can now be explained. According to our analyses, *F. robusta* and *F. cronquistii* seem to be representatives of ancient C_3 lineages in *Flaveria*, but not *F. pringlei*. The rare *Flaveria mcdougallii* (C_3, not represented in this analysis), which is morphologically and geographically distinct and might be sister to all other species of *Flaveria* (McKown et al., 2005; Powell, 1978), would be an important species to add in further studies.

The most ancient hybridization event in *Flaveria* involved two C_3 lineages, *F. robusta* and *F. cronquistii*. According to Powell (1978), *Flaveria cronquistii* most closely resembles *F. robusta*. The two are geographically separated, and the former is distributed in the Tehuacán Valley region and the latter in Colima and Michoacán (both Mexico). There are no reports of artificial hybridization between these two species. The resulting ancient hybrid lineage seems to include all C_3-C_4 intermediate, C_4-like, and C_4 species (Fig. 2; Table 1). Concerning leaf anatomy, *Flaveria robusta* differs from *F. cronquistii* by a higher vein density achieved through higher vein branching and by a higher number of organelles in the bundle sheath cells where the organelles are located in a centripetal position along the cell wall connecting bundle sheath cells and vascular tissue. Organelles are more evenly distributed in *F. cronquistii,* and its leaves are more succulent and show larger bundle sheath cells (McKown & Dengler, 2007; T. L. Sage et al., 2013). The combination of these traits in a hybrid lineage and subsequent segregation effectively leading to a higher bundle sheath to mesophyll ratio and activated large bundle sheath cells may have triggered the evolution of "pre kranz" cells in *Flaveria*. Possibly the strong leaf anatomical differences between the two parental lineages promoted the origin of novel traits which then allowed this lineage to occupy new niches from the parents. There are multiple examples for plant lineages in which homoploid hybridization resulted in potentially adaptive genotypes through transgressive segregation, eventually leading to speciation (Gross & Rieseberg, 2004; Nieto Feliner et al., 2020). An increase of the bundle sheath to mesophyll ratio has been suggested to play an initial role in the evolution of C_4 in several plant lineages (Christin et al., 2012; Griffiths et al., 2012; Lauterbach et al., 2019; Marshall et al., 2007; McKown & Dengler, 2007).

Two C_3-C_4 intermediate lineages, *F. angustifolia* and *F. sonorensis*, are involved in the formation of the lineage that eventually evolved C_4 photosynthesis. It is remarkable that only these two and not any other species (especially not the C_4-like species) contributed to the origin of this clade. In their extant distribution, the two species do not overlap. *Flaveria angustifolia* grows in sclerophyllous scrub in Puebla and Oaxaca, while *F. sonorensis* is found only in the short-tree forests of tropical Sonora (Powell, 1978). Looking more closely at these two descendants of the lineages that seem to have hybridized in the past and possibly gave rise to C_4 photosynthesis might give new insights into important preconditions for the evolution of C_4 in *Flaveria*. Both *F. angustifolia* and *F. sonorensis* were categorized as C_2 Type I species (Table 1) but with weaker C_2 photosynthesis (relatively high amounts of glycine decarboxylase in the mesophyll cells and relatively high CO_2 compensation points) than other C_2 species of *Flaveria* (R. F. Sage et al., 2018). This finding implies the evolution of a C_2 phenotype prior to C_4, supporting the 'photorespiratory bridge hypothesis' (R. F. Sage et al., 2018) in case of *Flaveria*. However, it seems that C_4 evolution in *Flaveria* was triggered by hybridization of C_2 lineages, a scenario never suggested before. For *Flaveria* this somewhat shifts the focus to C_2 photosynthesis and under which selective conditions this type of photosynthesis might have evolved. To gain

further insights into the evolution of C_2 photosynthesis, detailed studies of the origin, anatomy, and ecophysiology of *F. angustifolia* and *F. sonorensis* might be rewarding. Organelle enrichment and their flux-optimized positioning as well as glycine decarboxylase accumulation in the proto-kranz cells (Khoshravesh et al., 2016) seem to be essential for C_2 photosynthesis (in eudicots and monocots). These traits enable species to more efficiently re-cycle respired CO_2 and to longer maintain a positive assimilation rate under carbon deficient conditions. Since the majority of C_4 lineages do not have any known C_2 relatives (see R. F. Sage et al., 2018 for an overview), and C_2 lineages without close C_4 relative are known (see Lundgren, 2020 for an overview), the question remains whether there exist several evolutionary pathways to C_4 photosynthesis (Edwards, 2019) and whether C_2 should be considered an independent carbon concentrating mechanism not always intimately connected to C_4 photosynthesis (Lundgren, 2020 and ref. therein). Finally, a C_2 lineage might also be the result of ancient hybridization of a C_3 and a C_4 lineage, as has been suggested for *Salsola divaricata*, which then might have thrived through the plasticity of photosynthetic traits inherited from photosynthetically divergent parental lineages (Tefarikis et al., 2022); this, however, does not seem to be the case in *Flaveria*.

CONCLUSIONS

The young genus *Flaveria,* which includes four C_3, four C_4, three C_4-like, and around ten C_3-C_4 intermediate species, is remarkable for the high number of evolutionary reticulations creating an enormous diversity of phenotypes with different photosynthetic traits. Due to this and its strictly diploid chromosome number, the genus is a highly interesting system to study the genetic basis of the C_4 syndrome and the role of transgressive segregation in the origin of genotypes eventually leading to the evolution of C_4 photosynthesis. We suggest that homoploid hybridization of C_3 lineages might have triggered the evolution of C_2 photosynthesis and that homoploid hybridization of C_2 lineages gave rise to C_4-like or C_4 photosynthesis. In both cases the hybrid derivatives possibly surpassed the parental performance under conditions of high photorespiration and had an adaptive advantage. However, since reticulation occurred in recent as well as ancient lineages of the genus, the sequence of evolutionary events needs to be studied in the light of a carefully reconstructed phylogenetic history of the genus. Comparison of the entire genomes of parental and hybrid lineages should allow us to detect whether the origin of new photosynthetic traits is indeed the result of transgressive segregation (de los Reyes, 2019), as suggested here, or of stepwise mutational change of relevant genes.

Data Availability

Analyses and results files can be accessed at the Dryad repository https://doi.org/10.5061/dryad.msbcc2g1d. Details of publicly available transcriptomes and genomes used are provided in Table S1.

Funding

Financial support for this study came from LMU Munich and the German Science Foundation (DFG grant KA1816/9-1).

Acknowledgments

We thank Joachim W. Kadereit, Rowan Sage, and Matt Stata for valuable comments on the manuscript.

Supporting Information

Table S1. Taxon sampling and source of data

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

Table S3. HyDe significant hybridization tests for all species of *Flaveria* excluding *F. pringlei* (C_3).

Fig. S1. a) Maximum likelihood phylogeny of *Flaveria* inferred with IQ-TREE from the concatenated 2124-nuclear gene supermatrix. Numbers above branches represent bootstrap support (BS). Branch lengths as substitutions per site (scale bar on the bottom). b) ASTRAL tree of *Flaveria* inferred from the 2,127 nuclear gene trees. Local posterior probabilities (LPP) are shown next to nodes. Internal branch lengths are in coalescent units (scale bar on the bottom).

Fig. S2. a) Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the concatenated 2124-nuclear gene supermatrix. b) ASTRAL cladogram of *Flaveria* inferred from the 2,127 nuclear gene trees. Pie charts represent the proportion of gene trees that support that clade (blue), the main alternative bifurcation (green), the remaining alternatives (red), and conflict or support that have <50% bootstrap support (gray). Number above and below branches represent the number of concordant and discordant informative gene trees, respectively.

Fig. S3. a) Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the concatenated 2124-nuclear gene supermatrix. b) ASTRAL cladogram of *Flaveria* inferred from the 2,127 nuclear gene trees. Quartet Sampling (QS) scores are shown above branches. QS scores: Quartet concordance/Quartet differential/Quartet informativeness. Circles at nodes are colored by quartet concordance support.

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

Fig. S5. Maximum pseudo-likelihood species networks inferred with PhyloNet using the a) 18-taxon, b) and 17-taxon data sets and allowing up to 12 reticulation events. Red and blue curved branches indicate the minor and major edges, respectively of hybrid nodes. Numbers next to curved branches indicate inheritance probabilities for each hybrid node.

Fig. S6. Maximum pseudo-likelihood scores for species networks inferred with PhyloNet using the a) 18-taxon, b) and 17-taxon data sets. The x-axis notes the maximum number of reticulations for each of the network searches allowing up to 12 reticulation events.

Fig. S7. Maximum likelihood cladogram of *Flaveria* inferred with IQ-TREE from the concatenated 2124-nuclear gene supermatrix. Numbers above branches are gene duplication counts and numbers below branches are gene duplication percentages.

Submitted: November 17, 2022 EST, Accepted: February 09, 2023 EST

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Anderberg, A. A., Baldwin, B. G., Bayer, R. G., Breitwieser, J., Jeffrey, C., Dillon, M. O., Eldenäs, P., Funk, V., Garcia-Jacas, N., Hind, D. J. N., Karis, P. O., Lack, H. W., Nesom, G., Nordenstam, B., Oberprieler, Ch., Panero, J. L., Puttock, C., Robinson, H., Stuessy, T. F., … Watson, L. E. (2007). Compositae. *Flowering Plants · Eudicots: Asterales*, 61–588. https://doi.org/10.1007/978-3-540-31051-8_7

Apel, P., Bauwe, H., Bassüner, B., & Maass, I. (1988). Photosynthetic properties of *Flaveria cronquistii*, *F. palmeri*, and hybrids between them. *Biochemie Und Physiologie Der Pflanzen*, *183*(4), 291–299. https://doi.org/10.1016/s0015-3796(88)80021-0

Apel, P., & Maass, I. (1981). Photosynthesis in species of *Flaveria* CO_2 compensation concentration, O_2 influence on photosynthetic gas exchange and $\delta^{13}C$ values in species of *Flaveria* (Asteraceae). *Biochemie Und Physiologie Der Pflanzen*, *176*(4), 396–399. https://doi.org/10.1016/s0015-3796(81)80052-2

Blair, C., & Ané, C. (2019). Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. *Systematic Biology*, *69*(3), 593–601. https://doi.org/10.1093/sysbio/syz056

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A python package for genome-scale hybridization detection. *Systematic Biology*, *67*(5), 821–829. https://doi.org/10.1093/sysbio/syy023

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bräutigam, A., & Gowik, U. (2016). Photorespiration connects C_3 and C_4 photosynthesis. *Journal of Experimental Botany*, *67*(10), 2953–2962. https://doi.org/10.1093/jxb/erw056

Brown, J. W., Walker, J. F., & Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics*, *33*(12), 1886–1888. https://doi.org/10.1093/bioinformatics/btx063

Chen, L.-Y., Morales-Briones, D. F., Passow, C. N., & Yang, Y. (2019). Performance of gene expression analyses using *de novo* assembled transcripts in polyploid species. *Bioinformatics*. https://doi.org/10.1101/380063

Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., Garrison, L. M., Vorontsova, M. S., & Edwards, E. J. (2012). Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences*, *110*(4), 1381–1386. https://doi.org/10.1073/pnas.1216777110

Cloutier, A., Sackton, T. B., Grayson, P., Clamp, M., Baker, A. J., & Edwards, S. V. (2019). Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. *Systematic Biology*, *68*(6), 937–955. https://doi.org/10.1093/sysbio/syz019

Davidson, N. M., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biology*, *15*(7). https://doi.org/10.1186/s13059-014-0410-6

de los Reyes, B. G. (2019). Genomic and epigenomic bases of transgressive segregation – New breeding paradigm for novel plant phenotypes. *Plant Science*, *288*(110213), 110213. https://doi.org/10.1016/j.plantsci.2019.110213

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Doyle, J. J. (1992). Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany*, *17*(1), 144. https://doi.org/10.2307/2419070

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Edwards, E. J. (2019). Evolutionary trajectories, accessibility, and other metaphors: the case of C_4 and CAM photosynthesis. *New Phytologist*, *223*(4), 1742–1755. https://doi.org/10.1111/nph.15851

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Galtier, N., & Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1512), 4023–4029. https://doi.org/10.1098/rstb.2008.0144

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., & Hood, M. E. (2008). Speciation in fungi. *Fungal Genetics and Biology*, *45*(6), 791–802. https://doi.org/10.1016/j.fgb.2008.02.001

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A draft sequence of the neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Griffiths, H., Weller, G., Toy, L. F. M., & Dennis, R. J. (2012). You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell & Environment*, *36*(2), 249–261. https://doi.org/10.1111/j.1365-3040.2012.02585.x

Gross, B. L., & Rieseberg, L. H. (2004). The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, *96*(3), 241–252. https://doi.org/10.1093/jhered/esi026

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Kadereit, G., Bohley, K., Lauterbach, M., Tefarikis, D. T., & Kadereit, J. W. (2017). C_3–C_4 intermediates may be of hybrid origin – a reminder. *New Phytologist*, *215*(1), 70–76. https://doi.org/10.1111/nph.14567

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., & Sage, T. L. (2016). C_3–C_4intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C_2photosynthesis. *Journal of Experimental Botany*, *67*(10), 3065–3078. https://doi.org/10.1093/jxb/erw150

Ku, M. S. B., Monson, R. K., Littlejohn, R. O. Jr., Nakamoto, H., Fisher, D. B., & Edwards, G. E. (1983). Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species. *Plant Physiology*, *71*(4), 944–948. https://doi.org/10.1104/pp.71.4.944

Ku, M. S. B., Wu, J., Dai, Z., Scott, R. A., Chu, C., & Edwards, G. E. (1991). Photosynthetic and photorespiratory characteristics of *Flaveria* species. *Plant Physiology*, *96*(2), 518–528. https://doi.org/10.1104/pp.96.2.518

Kubatko, L. S., & Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evolutionary Biology*, *19*(1). https://doi.org/10.1186/s12862-019-1439-7

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lauterbach, M., Zimmer, R., Alexa, A. C., Adachi, S., Sage, R. F., Sage, T. L., MacFarlane, T., Ludwig, M., & Kadereit, G. (2019). Variation in leaf anatomical traits relates to the evolution of C_4 photosynthesis in Tribuloideae (Zygophyllaceae). *Perspectives in Plant Ecology, Evolution and Systematics*, *39*(125463), 125463. https://doi.org/10.1016/j.ppees.2019.125463

Liu, L., & Yu, L. (2010). Phybase: an R package for species tree analysis. *Bioinformatics*, *26*(7), 962–963. https://doi.org/10.1093/bioinformatics/btq062

Long, R. W., & Rhamstine, E. L. (1968). Evidence for the hybrid origin of *Flaveria latifolia* (Compositae). *Brittonia*, *20*(3), 238. https://doi.org/10.2307/2805449

Long, S. P. (1999). Environmental Responses. In R. F. Sage & R. K. Monson (Eds.), *C_4 Plant Biology* (pp. 215–249). Academic Press. https://doi.org/10.1016/b978-012614440-6/50008-2

Lundgren, M. R. (2020). C_2 photosynthesis: a promising route towards crop improvement? *New Phytologist*, *228*(6), 1734–1740. https://doi.org/10.1111/nph.16494

Lyu, M.-J. A., Gowik, U., Kelly, S., Covshoff, S., Hibberd, J. M., Sage, R. F., Ludwig, M., Wong, G. K.-S., Westhoff, P., & Zhu, X.-G. (2021). The coordination of major events in C_4 photosynthesis evolution in the genus *Flaveria*. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-93381-8

Lyu, M.-J. A., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., Hibberd, J. M., Stata, M., Sage, R. F., Lu, H., Wei, X., Wong, G. K.-S., & Zhu, X.-G. (2015). RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications. *BMC Evolutionary Biology*, *15*(1), 116. https://doi.org/10.1186/s12862-015-0399-9

Mai, U., & Mirarab, S. (2018). TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, *19*(S5). https://doi.org/10.1186/s12864-018-4620-2

Mandel, J. R., Dikow, R. B., Siniscalchi, C. M., Thapa, R., Watson, L. E., & Funk, V. A. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences*, *116*(28), 14083–14088. https://doi.org/10.1073/pnas.1903871116

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., Sage, R. F., & Hibberd, J. M. (2007). *Cleome*, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. *The Plant Journal*, *51*(5), 886–896. https://doi.org/10.1111/j.1365-313x.2007.03188.x

Maureira-Butler, I. J., Pfeil, B. E., Muangprom, A., Osborn, T. C., & Doyle, J. J. (2008). The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, *57*(3), 466–482. https://doi.org/10.1080/10635150802172168

McKown, A. D., & Dengler, N. G. (2007). Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany*, *94*(3), 382–399. https://doi.org/10.3732/ajb.94.3.382

McKown, A. D., Moncalvo, J.-M., & Dengler, N. G. (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *American Journal of Botany*, *92*(11), 1911–1928. https://doi.org/10.3732/ajb.92.11.1911

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Monson, R. K., & Moore, B. d. (1989). On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. *Plant, Cell and Environment*, *12*(7), 689–699. https://doi.org/10.1111/j.1365-3040.1989.tb01629.x

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., Timoneda, A., Yim, W. C., Cushman, J. C., & Yang, Y. (2021). Disentangling sources of gene tree discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology*, *70*(2), 219–235. https://doi.org/10.1093/sysbio/syaa066

Nakamoto, H., Ku, M. S. B., & Edwards, G. E. (1983). Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species II. Kinetic properties of phosphoenolpyruvate carboxylase from C_3, C_4 and C_3-C_4 intermediate species. *Plant and Cell Physiology*, *24*(8), 1387–1393. https://doi.org/10.1093/oxfordjournals.pcp.a076659

Nieto Feliner, G., Casacuberta, J., & Wendel, J. F. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics*, *11*(792). https://doi.org/10.3389/fgene.2020.00792

Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, *5*, 568–583. https://doi.org/10.1093/oxfordjournals.molbev.a040517

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., & Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany*, *105*(3), 385–403. https://doi.org/10.1002/ajb2.1016

Powell, A. M. (1978). Systematics of *Flaveria* (Flaveriinae-Asteraceae). *Annals of the Missouri Botanical Garden*, *65*(2), 590. https://doi.org/10.2307/2398862

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(Database), D61–D65. https://doi.org/10.1093/nar/gkl842

R Core Team. (2021). *R: A language and environment for statistical computing. R foundation for statistical computing*. https://www.R-project.org/

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*(4), 1645–1656. https://doi.org/10.1093/genetics/164.4.1645

Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*(1–2), 131–147. https://doi.org/10.1016/0025-5564(81)90043-2

Sage, R. F., Khoshravesh, R., & Sage, T. L. (2014). From proto-Kranz to C_4 Kranz: Building the bridge to C_4 photosynthesis. *Journal of Experimental Botany*, *65*(13), 3341–3356. https://doi.org/10.1093/jxb/eru180

Sage, R. F., Monson, R. K., Ehleringer, J. R., Adachi, S., & Pearcy, R. W. (2018). Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia*, *187*(4), 941–966. https://doi.org/10.1007/s00442-018-4191-6

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, *63*(1), 19–47. https://doi.org/10.1146/annurev-arplant-042811-105511

Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., Sultmanis, S., Rahman, B. A., Rawsthorne, S., & Sage, R. F. (2013). Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. *Plant Physiology*, *163*(3), 1266–1276. https://doi.org/10.1104/pp.113.221119

Sayyari, E., & Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, *33*(7), 1654–1668. https://doi.org/10.1093/molbev/msw079

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2805–2811. https://doi.org/10.1098/rstb.2008.0055

Scornavacca, C., Belkhir, K., Lopez, J., Dernat, R., Delsuc, F., Douzery, E. J. P., & Ranwez, V. (2019). OrthoMaM v10: Scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian Genomes. *Molecular Biology and Evolution*, *36*(4), 861–862. https://doi.org/10.1093/molbev/msz015

Smith, B. N., & Turner, B. L. (1975). Distribution of Kranz syndrome among Asteraceae. *American Journal of Botany*, *62*(5), 541–545. https://doi.org/10.1002/j.1537-2197.1975.tb14082.x

Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, *15*(1). https://doi.org/10.1186/s12862-015-0423-0

Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*, *26*(8), 1134–1144. https://doi.org/10.1101/gr.196469.115

Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genetics*, *12*(3), e1005896. https://doi.org/10.1371/journal.pgen.1005896

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, *60*(1), 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039

Song, L., & Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0089-y

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swofford, D. (2002). *PAUP*. Phylogenetic analysis using parsimony (*and other methods) version 4*. Sinauer Associates.

Taniguchi, Y. Y., Gowik, U., Kinoshita, Y., Kishizaki, R., Ono, N., Yokota, A., Westhoff, P., & Munekage, Y. N. (2021). Dynamic changes of genome sizes and gradual gain of cell-specific distribution of C_4 enzymes during C $_4$ evolution in genus *Flaveria*. *Plant Genome*, *14*(2). https://doi.org/10.1002/tpg2.20095

Tefarikis, D. T., Morales-Briones, D. F., Yang, Y., Edwards, G., & Kadereit, G. (2022). On the hybrid origin of the C_2 *Salsola divaricata* agg. (Amaranthaceae) from C_3 and C_4 parental lineages. *New Phytologist*, *234*(5), 1876–1890. https://doi.org/10.1111/nph.18098

Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*, *9*(1), 322–16. https://doi.org/10.1186/1471-2105-9-322

van Dongen, S. M. (2000). *Graph Clustering by Flow Simulation*. https://dl.acm.org/citation.cfm?id=868979

Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Systematic Biology*, *67*(4), 735–740. https://doi.org/10.1093/sysbio/syy015

Yang, Y., Moore, M. J., Brockington, S. F., Mikenas, J., Olivieri, J., Walker, J. F., & Smith, S. A. (2018). Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist*, *217*(2), 855–870. https://doi.org/10.1111/nph.14812

Yang, Y., & Smith, S. A. (2013). Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics*, *14*(1), 328. https://doi.org/10.1186/1471-2164-14-328

Yang, Y., & Smith, S. A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, *31*(11), 3081–3092. https://doi.org/10.1093/molbev/msu245

Yu, Y., & Nakhleh, L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*, *16*(S10). https://doi.org/10.1186/1471-2164-16-s10-s10

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(S6). https://doi.org/10.1186/s12859-018-2129-y