On the hybrid origin of the C_2 *Salsola divaricata* agg. (Amaranthaceae) from C_3 and C_4 parental lineages

Delphine T. Tefarikis[1]*, Diego F. Morales-Briones[2], Ya Yang[2], Gerald Edwards[3] & Gudrun Kadereit[1,4]

[1] AG Biodiversity and Evolution of Plants, Institute of Molecular Physiology, Johannes Gutenberg University Mainz, Mainz, Germany

[2] Department of Plant and Microbial Biology, University of Minnesota–Twin Cities, St. Paul, MN, USA

[3] School of Biological Sciences, Washington State University, Pullman, WA, USA

[4] Section Systematics, Biodiversity and Evolution of Plants, Ludwig Maximilians University Munich, Germany

*** Correspondence:** Delphine T. Tefarikis; tefarikisd@gmail.com

Total word count: 6,989

Summary: 200

Introduction: 835

Material & Methods: 1500

Results: 2349

Discussion: 2115

Conclusion: 190

Figures: 6 (only Figure 6 in color)

Tables: 2

Supporting Information:

Figures S1 – S8;

Tables S1– S3;

Material & Methods S1.

Abstract

- C_2 photosynthesis is characterized by recapturing photorespiratory CO_2 by RuBisCO in Kranz-like cells and is therefore physiologically intermediate between C_3 and C_4 photosynthesis. C_2 is either interpreted as an evolutionary precursor of C_4 or as the result of hybridization between a C_3 and C_4 lineage.
- We compared the expression of photosynthetic traits among populations of the *Salsola divaricata* agg. (C_2) from humid subtropical to arid habitats on the coasts of the Canary Islands and Morocco, and subjected them to salt and drought treatments. We screened for enhanced C_4-like expression of traits related to habitat or treatment. We estimated species trees with a transcriptome dataset of Salsoleae and explored patterns of gene tree discordance. With phylogenetic networks and hybridization analyses we tested for hybrid origin of the *Salsola divaricata* agg.
- We observed independent variation of photosynthetic traits within and among populations and no clear evidence for selection towards C_4-like trait expression in more stressful habitats or treatments. We found reticulation and gene tree incongruencies in the Salsoleae supporting a putative hybrid origin of the *Salsola divaricata* agg.
- C_2 photosynthesis in the *Salsola divaricata* agg. combines traits inherited from its C_3 and C_4 parental lineages and seems well adapted to a wide climatic amplitude.

Keywords: C_2 photosynthesis, carbon isotope values, CO_2 compensation point, Chenopodiaceae, phosphoenolpyruvate carboxylase, *Salsola divaricata* agg., phylogenomics, hybridization events.

Introduction

In current models of C_4 evolution, the C_3-C_4 intermediate phenotypes (including C_2 plants) are interpreted as a transitional, evolutionary link between the ancestral C_3 photosynthesis pathway and the derived C_4 pathway and showcase the complexity of the numerous structural, genetic and functional changes necessary to establish a functioning C_4 pathway (Bräutigam & Gowik, 2016; Schlüter & Weber, 2016). There are about 50 known species with intermediate photosynthetic traits, some of which do not share a common ancestor with one of the more than 60 independent C_4 lineages (Sage *et al.*, 2011), and some of which represent lineages up to 20 or even 30 million years in their crown age (Sage *et al.*, 2018). The fact that we are able to observe these intermediate phenotypes in nature repeatedly implies that they are evolutionary stable (Lundgren, 2020). Their phenotypic diversity has been classified into four photosynthetic categories of C_3-C_4 intermediacy: proto-Kranz, C_2 type I, C_2 type II and C_4-like (Sage *et al.*, 2014). In particular, C_2 photosynthesis has been interpreted as a crucial stepping stone towards C_4 photosynthesis. In C_2 species the photorespiratory enzyme glycine decarboxylase (GDC) is restricted to the bundle sheath or Kranz-like cells and mitochondria are absent or distinctly reduced in the mesophyll cells. This induces a photorespiratory glycine shuttle to the bundle sheath cells where the glycine is then processed by GDC and the photorespiratory CO_2 is recaptured by RuBisCO (Schulze *et al.*, 2016).

Bräutigam & Gowik (2016) propose that once the photorespiratory CO_2 pump is active, establishing the C_4 cycle is inevitable. According to this model, a strong selective pressure towards an increase of phosphoenolpyruvate carboxylase (PEPC) activity in C_2 species should be expected under carbon deficient conditions (Bräutigam & Gowik, 2016).

An alternative to the evolution of C_3 to intermediates to C_4 is the proposed occurrence of intermediates through hybridization (Monson *et al.*, 1984). Experimental hybrids of C_3 and C_4 species (e.g., Oakley *et al.*, 2014) reveal similar phenotypes as naturally occurring C_3-C_4 intermediate species (Brown & Bouton, 1993). Furthermore, conflicting topologies in molecular phylogenetic studies which include C_3-C_4 intermediate species indicate possible hybridization events in the evolutionary history of these lineages (reviewed in Kadereit *et al.*, 2017). To investigate the possibility of a hybrid origin of a C_2 lineage, detailed phylogenomic studies are needed; however, such a study has not been conducted so far.

88 We conducted a study of a C_2 species at the population level and looked at C_4-adaptive traits
89 to provide empirical evidence for one of the two alternative hypotheses. In the case of C_3-C_4
90 intermediates being evolutionary stepping stones and according to the model proposed by
91 Bräutigam & Gowik (2016), we would expect a pattern of populations showing shifts towards
92 more C_4-like traits under stressful conditions. In the case of C_3-C_4 intermediates resulting from
93 hybridization, we would expect independent segregation of C_4-traits as has been found in
94 several hybridization experiments (Björkman *et al*., 1969; Holaday *et al*., 1985; Cameron *et*
95 *al*., 1989; Brown *et al*., 1993; Brown & Bouton, 1993; Oakley *et al*., 2014).
96
97 The C_2 species we chose for our study is *Salsola divaricata* Moq. (Amaranthaceae, subfamily
98 Salicornioideae, tribe Salsoleae; Schüssler *et al*., 2017; Morales-Briones *et al*., 2021) and its
99 close relatives *S. verticillata* Schousb., *S. gymnomaschala* Maire and *S. deschaseauxiana*
100 Litard. & Maire (hereafter called the *S. divaricata* agg.). The distribution area of the *Salsola*
101 *divaricata* agg. on the Canary Islands and the coasts of Western Morocco, spans a considerable
102 climatic gradient with an arid climate in the East and a mesic mediterranean climate in the West
103 (García-Herrera *et al*., 2001).
104
105 Taxa of the *Salsola divaricata* agg. are salt tolerant, perennial shrubs with terete succulent
106 leaves and morphologically difficult to distinguish (Padrón Mederos, 2012). They all have a
107 Kranz-like salsoloid leaf anatomy (Voznesenskaya *et al*., 2013; Schüssler *et al*., 2017). *Salsola*
108 *divaricata* was categorized as a C_2 type I which is characterized by the photorespiratory CO_2
109 pump through GDC being confined to the bundle sheath cells while having C_3-like activity of
110 C_4 enzymes (e.g. PEPC and NADP-ME). Their CO_2 compensation points around 30 µmol mol^{-}
111 ^{1} are intermediate to C_3 and C_4 species (Schüssler et al., 2017). The *Salsola divaricata* agg.
112 belongs to Salsoleae, a tribe rich in C_4 species, but also with several C_3 species and a number
113 of C_3-C_4 species. The *Salsola divaricata* agg. forms a well-supported but unresolved subclade
114 within a clade of C_3-C_4 intermediate and C_3 species (Schüssler *et al*., 2017).
115 The aim of this study was to find empirical evidence either for the model of C_4 evolution
116 through C_2 as an intermediate pathway or for C_2 being the result of hybridization of a C_3 and
117 C_4 parental lineage. We studied C_4 adaptive traits at the population level in the C_2 lineage *S.*
118 *divaricata* agg. and searched for locally adapted more C_4-like phenotypes. We also generated
119 991 gene trees from a transcriptome data set comprising C_3, C_4 and C_2 species of the Salsoleae
120 and explored incongruences to determine if a hybridization event is plausible in the lineage of
121 the *S. divaricata* agg.

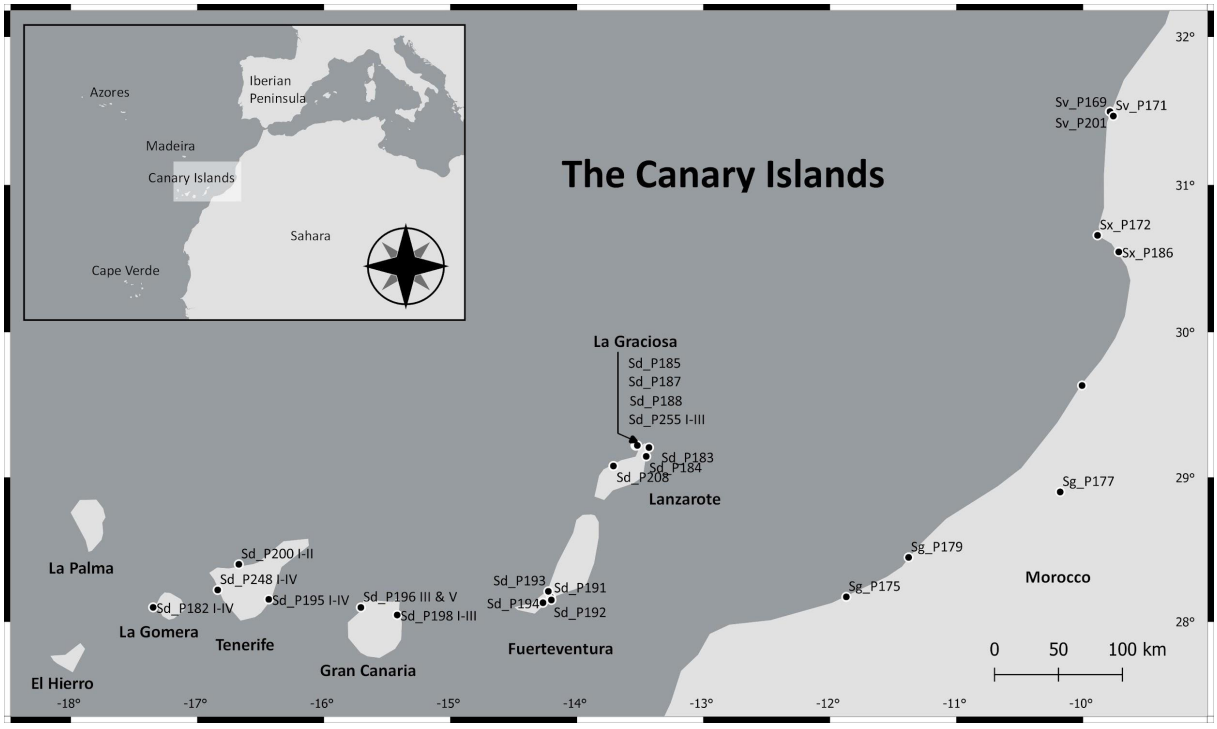

Fig. 1: Map of the collection locations of seeds and vouchers for the *S. divaricata* agg. (see also **Supporting Information Table S1** for voucher information). In total we analyzed 26 populations of the *S. divaricata* agg. on the Canary Islands and Morocco. Species name abbreviation and population number: Sd = *Salsola divaricata*, Sg = *S. gymnomaschala*, Sv = *S. verticillata*, Sx = *S. deschaseauxiana;* P*xxx* = population number. Note: Morphological studies on the species complex (Brullo, 1982; Fennane and Ibn Tattou, 1998) consider *S. verticillata* and *S. deschaseauxiana* as conspecific, while *S. gymnomaschala* seems more similar to *S. divaricata* s.s.

Materials and Methods

Plant Material

Seeds and vouchers of 26 populations of the *S. divaricata* agg. were collected in 2013 and 2014 in the Canary Islands and Morocco (**Supporting Information Table S1**; **Fig. 1**) and grown in a greenhouse at the Botanical Garden Mainz with multiple individuals per population. Plants were grown in custom mixed soil. Day/night cycles of 14h/10h were artificially maintained with natural light and supplementary light with a total light intensity of 200–400 µmol m^{-2} s^{-1}. The minimum temperature at night was 18 ºC and daytime temperatures ranged between 25 to 35 °C in summer and 20 to 25 °C in winter. All plants were watered once a week in the winter and twice a week in the summer. The individuals included in the salt treatment were watered with 2% saltwater (20 g NaCl per liter; 34.3 mM) every two weeks from June 2015 until 2021. Individuals in the dry treatment were watered every two weeks from June 2015 to March 2017. The analyses of the photosynthesis traits began in 2016. **Supporting Information**

Table S1 lists the populations included in this study as well as the collection and voucher information.

Carbon isotope measurements

C_3 plants discriminate against fixing atmospheric $^{13}CO_2$ by RuBisCO while C_4 plants do not discriminate via PEP carboxylase resulting in different isotope composition in plant biomass. Therefore, carbon isotope values ($\delta^{13}C$) can be used to detect C_4 cycle activity (Sage, 2016). We determined $\delta^{13}C$ values between and within six populations of *S. divaricata* with six individuals per treatment (control, dry and salt). We also measured 30 samples collected from wild mother plants of 24 populations and compared them with the greenhouse cultivated offspring (165 individuals in total, all populations, including 36 in dry and 36 in salt treatment). To determine carbon isotope composition, a standard procedure with PDB (Pee Dee Belemnite) limestone as the carbon isotope standard was used (Bender *et al.*, 1973). The leaf samples were dried in silica gel, and then 1–2 mg were placed in a tin capsule and combusted in an EuroVector elemental analyser (EuroVector, Italy). After separating the resulting N_2 and CO_2 gases by gas chromatography, they were fed into the IsoPrime™ isotope ratio mass spectrometer (IRMS; GV Instruments Ltd. (Micromass Ltd.), United Kingdom) for determination of $^{13}C/^{12}C$ ratios (R). $\delta^{13}C$ values were determined according to $\delta^{13}C(‰)=1000(R_{sample}/R_{standard})^{-1}$, where R is $^{13}C/^{12}C$. The measurements were taken at Washington State University and at the Geology Department at the University of Mainz.

CO_2 compensation point measurements

The CO_2 compensation point is a trait which is often used to identify C_4 species due to their very low values (usually less than 4 µmol CO_2 mol^{-1}). We measured one to two individuals (1 to 3 technical replicates for each individual) of two to four populations per island of all species in the *S. divaricata* agg. (in total 53 individuals). Measurements were made on the unstressed group and four individuals in the dry and salt treatment. C_3 *S. webbii* (2x) and C_4 *S. oppositifolia* (3x) were measured for comparison. CO_2 compensation points were determined using the portable gas exchange measurement system GFS-3000 (Heinz Walz GmbH, Germany) with the standard measuring head 3010-S equipped with a standard leaf area cuvette and a LED-Array (for details see **Supporting Information Material & Methods S1**).

179 **PEPC activity measurements**

180 Phosphoenolpyruvate carboxylase (PEPC) plays an important role in the C_4 pathway as it fixes
181 atmospheric CO_2 after its conversion to HCO_3^- by carbonic anhydrase into a four-carbon
182 compound which gave this photosynthesis pathway its name (Reyna-Llorens & Hibberd, 2017).
183 C_4 plants show high levels of PEPC activity compared to C_3 plants and an increased activity is
184 seen as an important stage in evolving from an intermediate to a full C_4 photosynthesis pathway
185 (Sage *et al.*, 2012). In total we sampled 14 populations of *S. divaricata* and six from the other
186 three species (predominantly analyzing 5 to 7 individuals per population). Treatments were
187 compared with 38 individuals in the control group, 23 in salt, and 17 in the dry treatment after
188 quality check. PEPC activity measurements were conducted using the Tecan infinite M1000
189 (Tecan Trading AG, Männedorf, Switzerland) and a Greiner UV-Star 96-well plate (Sigma-
190 Aldrich Chemie GmbH, Munich, Germany). For details of sample preparation see **Supporting**
191 **Information Material & Methods S1**.

192

193 The assay of PEPC was initiated by adding leaf extract and 3.9 mM phosphoenolpyruvate
194 (PEP). The chlorophyll content of the leaf extract was measured according to Wintermans and
195 De Mots (1965) in 96% ethanol. PEPC activity (a) was calculated on the basis of the linear part
196 of the progress curve as

197 $a = SU[\mu\text{mol ml}^{-1}\text{ min}^{-1}]V_{\text{assay}}[\text{ml}]/(1000V_{enzyme}[\text{ml}]c_{Chla+b}[\text{mg ml}^{-1}])$,

198 where SU is the substrate turnover, V_{assay} is the total assay volume, V_{enzyme} is the volume of the
199 leaf extract and c_{Chla+b} is the concentration of chlorophyll in the leaf extract. Statistical analyses
200 were conducted in R v3.1.2 (R Core Team, 2014).

201

202 **Transcriptome processing and nuclear phylogenetic analyses**

203 To evaluate the potential hybrid origin of the *S. divaricata* agg., we generated gene trees from
204 a transcriptome dataset comprising C_3, C_4 and C_2 species of Salsoleae; in total 13
205 transcriptomes representing 12 species. Three samples of these were newly sequenced on an
206 Illumina HiSeq2500 platform at the University of Minnesota Genomics Center (paired-end 125
207 bp). For RNA extraction protocol and library preparation refer to Morales-Briones *et al*. (2021).
208 Additionally, we sampled two species of Camphorosmeae and one of Caroxyloneae, and *Beta*
209 *vulgaris* L. subf. Betoideae - Amaranthaceae as outgroups. **Table 1** lists all samples with their
210 photosynthesis type, ploidy if known and SRA accession number.

211 We followed Morales-Briones *et al*. (2021) for raw read processing, transcriptome assembly,
212 transcript clustering, and homolog tree inference. We then applied the 'monophyletic outgroup'

213 (MO) approach of Yang & Smith (2014) to prune orthologs that have all 17 taxa present. We
214 used concatenation and coalescent-based methods for phylogenetic reconstruction. For the
215 concatenation approach, we prepared a supermatrix by keeping only ortholog alignments with
216 at least 300 bp and all 17 taxa. We estimated a maximum likelihood (ML) tree with RAxML v
217 8.2.11 (Stamatakis, 2014) using a partition by gene scheme. Clade support was assessed with
218 100 rapid bootstrap (BS) replicates. To estimate a species tree that is statistically consistent
219 with the multi-species coalescent (MSC), Individual gene trees were used to infer a species tree
220 using ASTRAL-III v.5.6.3 (Zhang *et al*., 2018) using local posterior probabilities (LPP; Sayyari
221 & Mirarab, 2016) to assess clade support. To examine nuclear gene tree discordance, we first
222 calculated the Internode Certainty All score (ICA; Salichos *et al*., 2014). Also, we calculated
223 the number of conflicting and concordant bipartitions on each node of the species trees. We
224 calculated both the ICA scores and the number of conflicting and concordant bipartitions with
225 Phyparts (Smith *et al*., 2015). Additionally, to distinguish strong conflict from weakly
226 supported branches, we evaluated tree conflict and branch support with Quartet Sampling (QS;
227 Pease *et al*., 2018) using 1,000 replicates (for details see **Supporting Information Material**
228 **& Methods S1**).
229
230 **Assessment of hybridization**
231 In order to detect possible hybridization, first we inferred species networks under a maximum
232 pseudo-likelihood (Yu & Nakhleh, 2015) approach using PhyloNet v.3.6.9. (Than *et al*., 2008).
233 To estimate the best number of hybridizations and test whether the species network fits our
234 gene trees better than a strictly bifurcating tree, we performed model selection using the bias-
235 corrected Akaike information criterion (Sugiura, 1978) and the Bayesian information criterion
236 (Schwarz, 1978). We also used HyDe (Blischak *et al*., 2018) to estimate the amount of
237 admixture (γ) in putative hybrid lineages. We tested all triples combinations and significance
238 was assessed with a Bonferroni correction (for details see **Supporting Information Material**
239 **& Methods S1**).
240
241 **Plastome assembly and phylogenetic analysis**
242 To investigate phylogenetic signal from plastid sequences, *de novo* assemblies were carried out
243 with the Fast-Plast v.1.2.6 pipeline (https://github.com/mrmckain/Fast-Plast) using the filtered
244 organelle reads obtained from the transcriptome raw read processing. A ML tree was inferred
245 with IQ-TREE v.1.6.1 (Nguyen *et al*., 2015) using the automated model selection
246 (Kalyaanamoorthy *et al*., 2017) and 200 standard non-parametric BS replicates for branch

support (for details see **Supporting Information Material & Methods S1**). Additionally, we used QS with 1,000 replicates, to detect potential plastome conflict in the backbone as seen in other groups of Amaranthaceae s.l. (Morales-Briones *et al*., 2021).

Analysis of photosynthetic gene trees

We inferred trees for 50 genes that are important in photosynthesis to determine if the genes in *Salsola divaricata* agg. show a tendency to group with either a C_3 or a C_4 species. We downloaded coding and protein sequences from *Arabidopsis thaliana* as baits from the TAIR website (https://www.arabidopsis.org/tools/bulk/sequences/index.jsp). These loci were determined based on Lauterbach *et al*. (2017a,b) and included two RuBisCO subunit genes, four genes coding for proteins of the glyoxylate cycle, 17 genes encoding photorespiratory proteins, and 25 genes coding for C_4-associated proteins (Lauterbach *et al*., 2017b) as well as two C_4-associated transcription factors (Lauterbach *et al*., 2017a). Gene annotation and pathway assignment can be found in **Supporting Information Table S3**. We listed whether the *S. divaricata* agg. was either in a sister relationship with C_3 species *S. montana* or the C_4 clade. When there were several gene copies or weak bootstrap support, we additionally noted whether our group of interest was in a clade with the C_3 or the C_4 species (for details see **Supporting Information Material & Methods S1**).

Results

Carbon isotope measurements

Salsola divaricata plants growing in the wild on the Canary Islands generally had higher $\delta^{13}C$ values (-21 to -28; mean -24.44) than those grown from seeds under greenhouse conditions (-22 to -34; mean -27.77). Overall, the $\delta^{13}C$ values of the *S. divaricata* agg. show a broad range but not a single individual (stressed or not stressed) came close to a C_4-like $\delta^{13}C$ value of > -16 (**Fig. 2**). We compared $\delta^{13}C$ values between control, dry and salt treatments within six populations (six individuals per treatment). The Shapiro-Wilk test was used to check for normality of distribution and the Levene's Test for homogeneity of variance. Excluding outlier 198IIr of population 198, all treatment groups of all populations were normally distributed. All populations had homogeneous variances except population 188. Since not all assumptions for an ANOVA test were met we chose the Kruskal-Wallace rank sum test to compare treatments within and between populations. In the control there was no significant difference of $\delta^{13}C$ values among populations (Kruskal-Wallis chi-squared = 5.1081, df = 5, p-value = 0.40) and the same was found for the drought treatment (Kruskal-Wallis chi-squared = 8.61, df = 5, p-

value = 0.13). In the salt treatment there was a significant difference of $\delta^{13}C$ values among populations (Kruskal-Wallis chi-squared = 15.47, df = 5, p-value = 0.009), which was due to significantly higher $\delta^{13}C$ values in salt treated individuals of populations 191 compared to 182 and 194. We tested for differences of carbon isotope values between treatments within populations of the *Salsola divaricata* agg. and found that there was a significant difference for populations 182 (Kruskal-Wallis chi-squared = 7.94, df = 2, p-value = 0.019), 191 (Kruskal-Wallis chi-squared = 10.84, df = 2, p-value = 0.004), and 198 (Kruskal-Wallis chi-squared = 8.84, df = 2, p-value = 0.012). However, there was no significant difference among the three treatments for populations 184 (Kruskal-Wallis chi-squared = 2.33, df = 2, p-value = 0.31), 188 (Kruskal-Wallis chi-squared = 0.78, df = 2, p-value = 0.68), and 194 (Kruskal-Wallis chi-squared = 2.27, df = 2, p-value = 0.32) (**Fig. 3**). After grouping $\delta^{13}C$ values according to island, significant differences were found among islands (ANOVA: df = 6, F value = 4.08, p-value = 0.001) due to differences between La Gomera and the mainland (SW Morocco), as well as between Lanzarote resp. Tenerife and the mainland. All other differences were not statistically significant. We found that the populations of *S. verticillata*, *S. deschaseauxiana* and *S. gymnomaschala* on the Moroccan mainland often had higher $\delta^{13}C$ values than the populations of *S. divaricata* on the Canary Islands (the non-stressed control groups that are regularly watered, **Supporting Information Fig. S1**). However, due to very different sample sizes further statistical tests including all populations were inconclusive.

Overall, these results show a broad plasticity of $\delta^{13}C$ values among populations associated with source population and growth conditions well within the range typical of C_3 species. The significantly higher $\delta^{13}C$ values of salt treated individuals indicate that salinity rather than drought affects the $\delta^{13}C$ in *Salsola divaricata* agg. and could be responsible for the occasional observation of relatively high $\delta^{13}C$ values of our samples collected in the wild close to the shoreline.

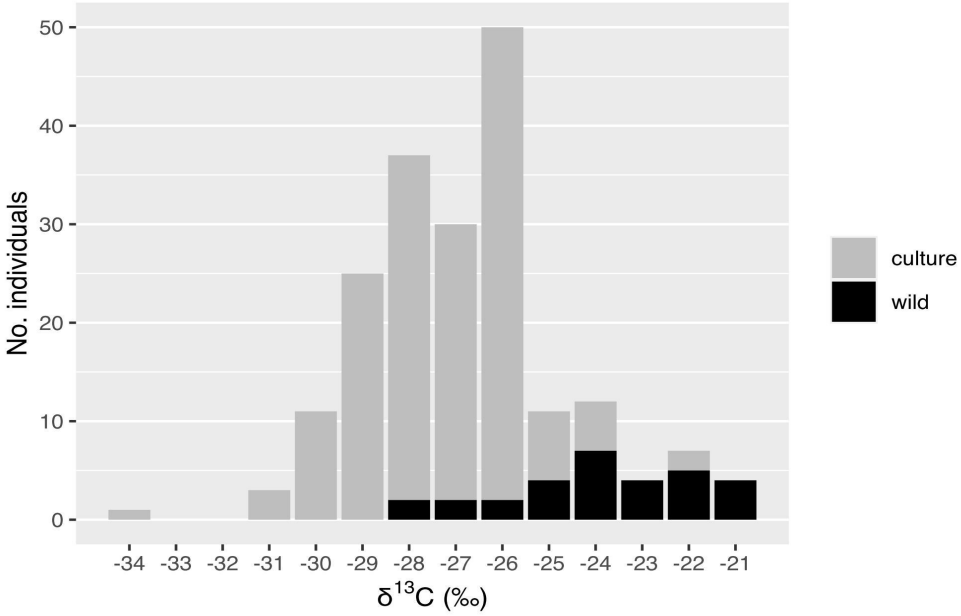

Fig. 2: Distribution of $\delta^{13}C$ (‰) of 195 samples of the *Salsola divaricata* agg., of these 30 were collected from plants growing in the wild and 165 from plants in cultivation at the Botanical Garden Mainz. This includes 36 individuals in dry treatment and 36 in salt treatment, all others are in the control group. See **Table S1** for information on populations included.

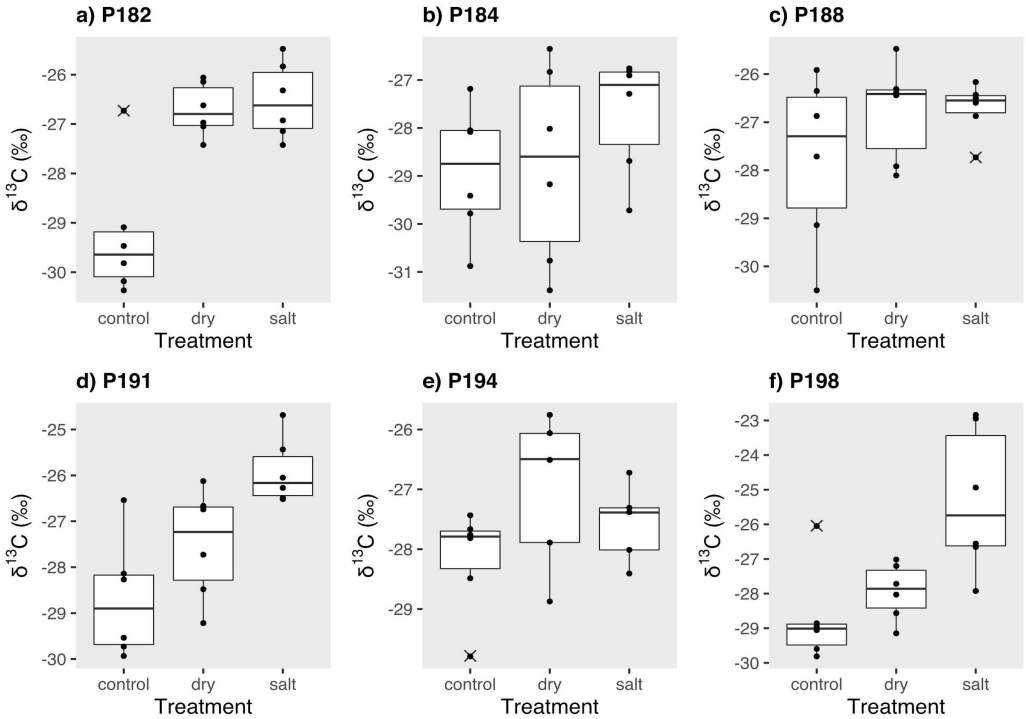

Fig. 3: Comparison of carbon isotope values ($\delta^{13}C$(‰)) between treatments within populations (P) of the *Salsola divaricata* agg. Significant differences were found for populations 182 from La Gomera (a; Kruskal-Wallis chi-squared = 7.94, df = 2, p-value = 0.02), 191 from Fuerteventura (d; Kruskal-Wallis chi-squared = 10.84, df = 2, p-value = 0.004),and 198 from Gran Canaria (f; Kruskal-Wallis chi-squared = 8.84, df = 2, p-value = 0.01). No significant differences were found between the three treatments for 184 from Lanzarote (b; Kruskal-Wallis chi-squared = 2.33, df = 2, p-value = 0.31), 188 from La Graciosa (c; Kruskal-Wallis chi-squared = 0.78, df = 2, p-value = 0.68), and 194 from from Fuerteventura (e; Kruskal-Wallis chi-squared = 2.27, df = 2, p-value = 0.32) .

CO_2 compensation point measurements

The CO_2 compensation points of the four species of the *Salsola divaricata* agg. lie between those of the C_3 (*Salsola webbii*) and C_4 species (*Salsola oppositifolia*) as expected (between 20 and 30 µmol CO_2 mol^{-1}, **Fig. 4a**). With 19.56 and 21.22 µmol CO_2 mol^{-1}, respectively, the mean values of *S. deschaseauxiana* and *S. verticillata* were slightly lower than those of *S. divaricata* and *S. gymnomaschala* (25.14 and 26.75 µmol CO_2 mol^{-1} respectively). When looking at individual measurements we again observed a broad range from 17 to 40 µmol CO_2 mol^{-1} with one outlier at 9 µmol CO_2 mol^{-1} (**Supporting Information Fig. S2**). Populations growing on the more arid Islands such as Lanzarote and Fuerteventura or in Morocco did not show lower CO_2 compensation points. Due to varying values within each population (**Supporting Information Fig. S2**) and the relatively small sample sizes (one to two individuals per population) we did not conduct further statistical analysis.

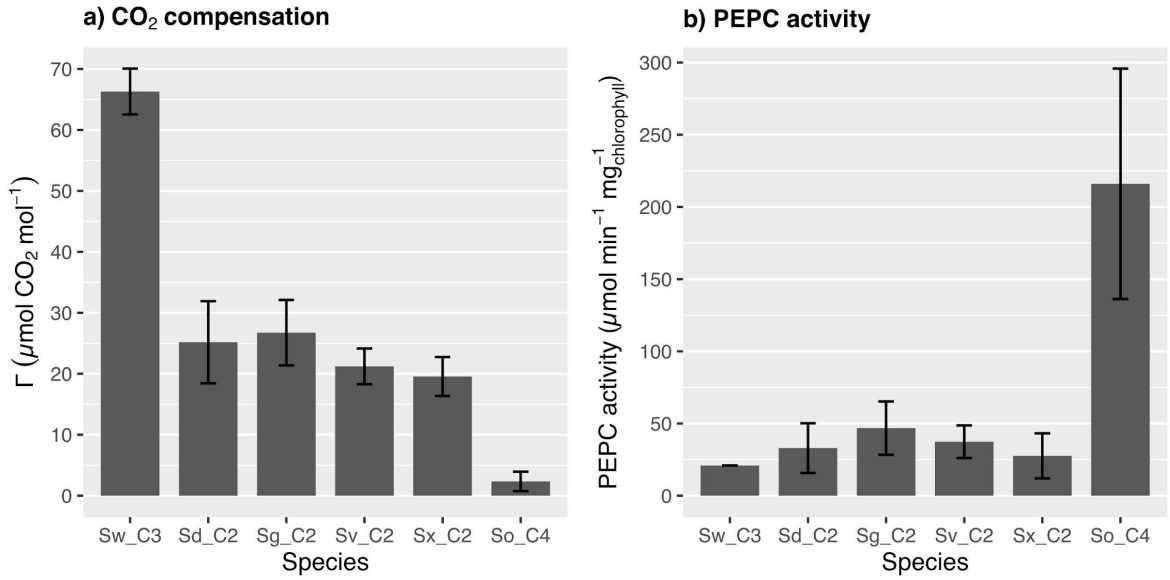

Fig. 4: a) CO_2 compensation points (Γ in µmol CO_2 mol^{-1}) of the separate species of the C_2 *Salsola divaricata* agg. (Sd_C_2 = *Salsola divaricata*, Sg_C_2 = *S. gymnomaschala*, Sv_C_2 = *S. verticillata*, Sx_C_2 = *S. deschaseauxiana*) compared to the values of C_3 species *Salsola webbii* (Sw_C_3) and C_4 species *Salsola oppositifolia* (So_C_4). **b)** PEPC activity [µmol min^{-1} $mg_{chlorophyll}^{-1}$] of the separate species of the C_2 *Salsola divaricata* agg. compared to C_4 *S. oppositifolia* and C_3 *S. webbii*.

PEPC activity measurements

First, we compared the PEPC activity of the separate species of the *Salsola divaricata* agg. to the values of the closely related C_3 species *Salsola webbii* and C_4 species *Salsola oppositifolia*. C_3 (n=1, 30 µmol min^{-1} $mg_{Chlorophyll}^{-1}$) and C_4 values (n=2, mean 216 µmol min^{-1} $mg_{Chlorophyll}^{-1}$) were as expected clearly distinguishable (**Fig. 4b**). The values of the C_2 species were closer to

those of the C_3 species and much lower than the C_4 species (*Salsola divaricata*, n=82, mean activity 33 µmol min^{-1} $mg_{Chlorophyll}^{-1}$; *S. gymnomaschala*, n=10, mean 46.8 µmol min^{-1} $mg_{Chlorophyll}^{-1}$; *S. verticillata*, n=7, mean 37.4 µmol min^{-1} $mg_{Chlorophyll}^{-1}$; *S. deschaseauxiana,* n=6, mean 27.6 µmol min^{-1} $mg_{Chlorophyll}^{-1}$; **Fig. 4b**). We used a Kruskal-Wallis test to determine if statistically significant differences exist between C_2 species (Kruskal-Wallis chi-squared = 9.94, df = 3, p-value = 0.019). This significant difference was observed due to the values of *S. divaricata* on the Canary Islands having a median of 30.05 µmol min^{-1} $mg_{Chlorophyll}^{-1}$ and *S. gymnomaschala* in Morocco having a median of 48.32 µmol min^{-1} $mg_{Chlorophyll}^{-1}$. Due to the high level of variation within species we took a closer look at 14 populations of *S. divaricata* and observed a significant difference between them (Kruskal-Wallis chi-squared = 34.64, df = 13, p-value = 0.001, see also **Fig. 5a**). When grouped according to island, we found a significant difference (Kruskal-Wallis chi-squared = 22.199, df = 5, p-value = 0.0005) which is due to the difference in values between populations on Gran Canaria and Lanzarote, as well as populations on La Graciosa and Fuerteventura and on La Graciosa and Gran Canaria (**Fig. 5b**). We did not find a significant difference between treatments in the greenhouse in *S. divaricata* (Kruskal-Wallis chi-squared = 1.21, df = 2, p-value = 0.55). Again, we observe a high variation of PEPC activity values within populations and islands.

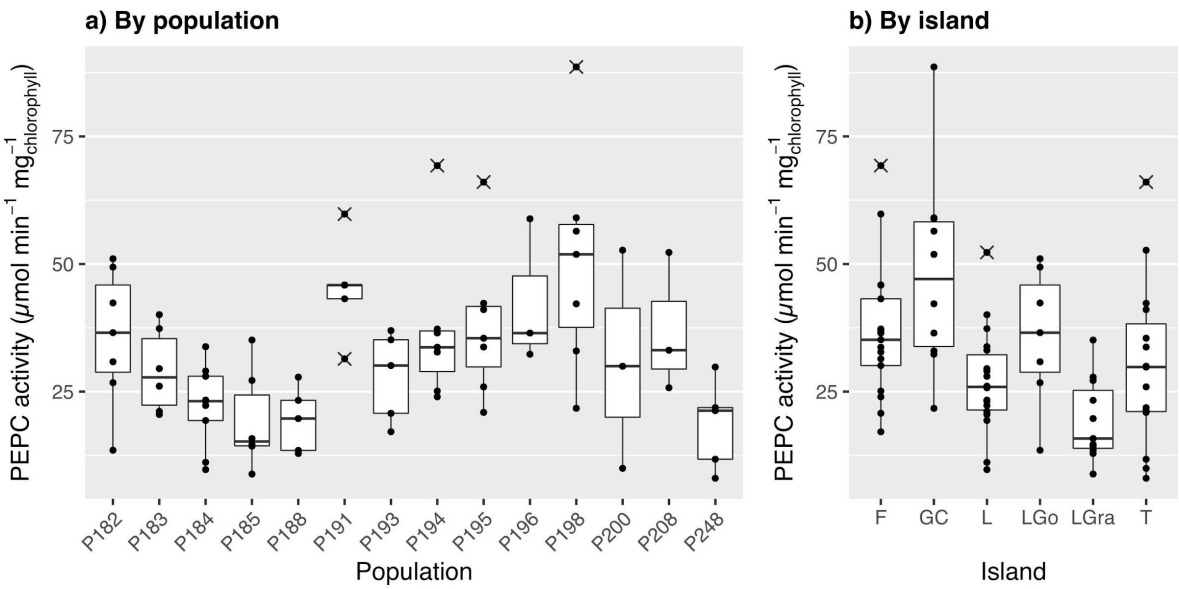

Fig. 5: Boxplots of PEPC activity [µmol min^{-1} $mg_{chlorophyll}^{-1}$] of *Salsola divaricata* Moq. **a)** Values grouped by population (P). **b)** Values grouped by island; F = Fuerteventura, GC = Gran Canaria, L = Lanzarote, LGo = La Gomera, LGra = La Graciosa, T = Tenerife.

370 **Phylogenomic analyses**

371 The raw reads for the newly generated transcriptomes are available from the NCBI Sequence
372 Read Archive (**Table 2**). We obtained 991 final MO orthologs that included all 17 taxa with
373 alignments of lengths between 315 and 6621 bp. The concatenated matrix comprises 1,427,449
374 aligned columns with overall matrix occupancy of 86% (see **Supporting Information Table**
375 **S2** alignment occupancy per taxon).

376

377 The ASTRAL species tree and the concatenated RAxML tree (**Supporting Information Fig.**
378 **S3**) had similar topologies with most nodes receiving the maximum support (BS = 100%; LPP
379 = 1). *Caroxylon vermiculatum* was recovered as sister to all other Salsoleae, followed by a
380 grade formed by *Salsola webbii*, *Salsola genistoides*, *Kali collina*, and *Salsola montana*.
381 *Salsola divaricata* and *Salsola verticillata* form a clade placed as sister of the clade composed
382 of *Anabasis articulata*, *Salsola soda, Salsola oppositifolia*, *Halogeton*, *Hammada*, and
383 *Holoxylon.* The only difference between the two topologies is that *Anabasis articulata* and
384 *Salsola soda* were sisters with maximum support in the RAxML tree but formed a grade with
385 lower support (LPP = 0.74) in the ASTRAL tree.

386

387 Concordance analyses (**Supporting Information Fig. S3** and **S4**) overall showed high support
388 for the monophyly of Salsoleae and the *Salsola divaricata* agg., while signaling conflict for the
389 remaining taxa. The placement of *Salsola genistoides* had low support with only 261 out of 794
390 informative gene trees being concordant (ICA = -0.45) and high QS score (0.22/0/0.98) but
391 also suggesting an alternative topology. The clade composed of *Salsola divaricata* + *Salsola*
392 *verticillata* had strong support with 808 out of 913 gene trees being concordant (ICA = 0.71)
393 and maximum QS support (1/–/1; i.e., all sampled quartets supported that branch). However,
394 the two samples of *Salsola divaricata* showed signals of alternative topologies with only 332
395 out of 900 gene trees supporting their sister relationship (ICA = 0.56) and QS counter support
396 (-0.25/0/1). The sister relationship of *Salsola divaricata* + *S. verticillata* and the remaining
397 Salsoleae had low support with only 178 out of 705 gene trees being concordant (ICA = 0.29)
398 and also low QS support (0.085/0/1 [ASTRAL]; (0.11/0/1) [RAxML]) suggesting an alternative
399 topology. The placement of *Anabasis articulata* and *Salsola soda,* which varied between
400 RAxML and ASTRAL trees, also had low gene tree and QS support with signals of alternative
401 topologies. The clade composed of *Salsola oppositifolia*, *Halogeton*, *Hammada*, and *Holoxylon*
402 was supported only by 280 out of 692 gene trees (ICA = 0.24), but had maximum QS support,

while within the clade all relationships have low gene tree and QS support signaling alternative topologies.

The final plastid alignment had 80,903 characters with a matrix occupancy of 70%. The plastid tree largely differed from the nuclear topologies and had most nodes with maximum support (BS = 100; **Fig. 6** and **S5**). *Caroxylon vermiculatum, Kali collina,* and *Salsola genistoides* form a well supported clade (BS = 100; QS = 0.56/0/1) that is sister to all Salsoleae. *Salsola webbii* and *Anabasis articulata* form a grade with strong support (BS =100; QS = 0.6/0/0.99 [*S. webii*]; QS = 1/–/1 [*A. articulata*]). The remaining species form an unsupported clade (BS = 0.47; QS = -0.0096/0.88/0.68) composed of two subclades. The first subclade (BS = 93; QS = 0.19/0.91/0.78) is composed of *Salsola verticillata* + *Salsola montana* (BS = 100; QS = 0.61/0/1) and the two samples of *Salsola divaricata* (BS = 100; QS = 1/–/0.77). The second subclade (BS = 100; QS = 0.59/0/0.95) has *Salsola soda* as the sister the clade (BS =100; QS = 1/–/1) composed of *Salsola oppositifolia*, *Halogeton*, *Hammada*, and *Haloxylon.*

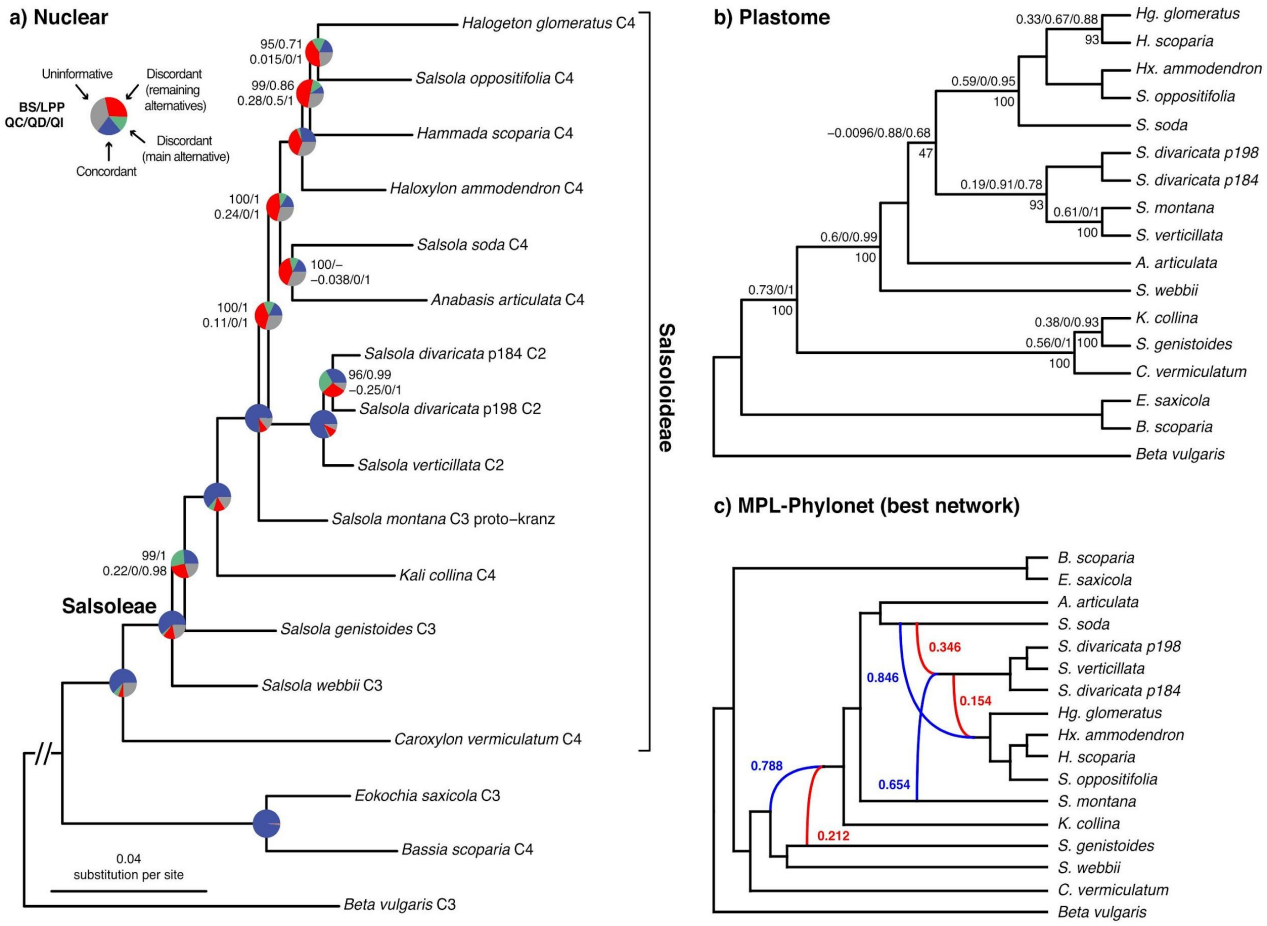

Fig. 6: a) Maximum likelihood phylogeny of Salsoloideae inferred from the RAxML analysis of the concatenated 991-nuclear gene supermatrix from the 'monophyletic outgroup' (MO) orthologs. Bootstrap support and local posterior probabilities (BS/LPP) are shown above branches. Nodes with full support (BS =

100/LPP = 1) are not shown. Em dashes (—) denote alternative topology compared to the ASTRAL tree (**Supporting Information Fig. S3**). Quartet Sampling (QS) scores are shown below branches. QS score: Quartet concordance (QC)/Quartet differential (QC)/Quartet informativeness QI). Full QS support (1/–/1) not shown. Pie charts represent the proportion of gene trees that support that clade (blue), the proportion that support the main alternative bifurcation (green), the proportion that support the remaining alternatives (red), and the proportion (conflict or support) that have < 50% bootstrap support (gray). Branch lengths are proportional to the number of substitutions per site (scale bar on the bottom). b) Cladogram of Salsoloideae inferred from IQ-TREE analysis of concatenated complete and partial plastomes. QS scores and BS support are shown above and below branches, respectively. See Supporting Information Fig. S5 for phylogram. c) Best maximum pseudo-likelihood species network inferred with PhyloNet. 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.

Assessment of hybridization

The model selection showed that a network was always a better model than a bifurcating tree (**Table 3**). A network with three hybridization events is the best model (AICc=37875.93) compared to all others, including the best bifurcating model (RAxML; AICc=39547.83). The best network had three reticulation events (**Fig. 6c**). First, the *Salsola divaricata* and *Salsola verticillata* clade was a product of a hybridization event between *Salsola soda* (C_4) (inheritance probability $p_i = 0.346$) and *Salsola montana* (C_3) ($p_i = 0.654$). The second reticulation event showed that the clade formed by *Salsola oppositifolia*, *Halogeton*, *Hammada*, and *Haloxylon* (all C_4) is the product of a hybridization event between *Salsola soda* (C_4) ($p_i = 0.846$) and the ancestor lineage of *Salsola divaricata* + *Salsola verticillata* (C_2) ($p_i = 0.154$). These two clades were also recovered as hybrids in the other networks with one, two and three reticulation events with similar parental lineages and inheritance probabilities (**Supporting Information Fig S6**). The third event is a deeper reticulation within *Salsoleae* which showed that most species of the group (excluding *Caroxylon*) were a product of an ancient hybridization event between *Salsola genistoides* ($p_i = 0.212$) and the sister lineage of *Salsola genistoides* + *Salsola webbii* ($p_i = 0.788$). The HyDe analysis of all possible triples at the individual level resulted in 1,680 hybridization tests, of which 298 triples were significant. The significant triples showed that 12 out of 13 individuals of *Salsoleae* are involved in multiple hybridization events, which is mostly in agreement with the nested reticulation events detected with Phylonet. The admixture parameter (γ) ranged from 0.013 to 0.986 (average 0.458). These results indicate frequent hybridization within Salsoleae. Additionally, Hyde also identified *Eokochia saxicola* as a hybrid (**Supporting Information Table S3**, **Fig. S7**).

Individual analysis of genes or gene families involved in photosynthesis or photorespiration

460 The number of sequences in the 50 alignments ranged from 7 to 65 due to variation in the
461 number of accessions and gene copies (**Supporting Information Table S3**). Sequence lengths
462 ranged from an average minimum of 420 (255 to 867) to an average maximum of 1,467 bp
463 (462 to 3,126) with an average length of 1,168 in all sequences. Summarizing all genes, *Salsola*
464 *divaricata* agg. appeared 22 times as sister to C_3 *S. montana* and 26 times as sister to the C_4
465 clade regardless of bootstrap support (**Supporting Information Table S3**). In three alignments
466 *S. divaricata* agg. was missing. When multiple copies per gene were recovered for *S. divaricata*
467 agg, they were often non-monophyletic, but instead separated in clades with copies from the
468 C_3 and/or C_4 species. In these cases, sequences of the *S. divaricata* agg. appeared 21 times with
469 C_3 and 26 times with C_4 species. Separated by pathway assignment, the *S. divaricata* agg. was
470 16 times sister to C_3 (11 times in clade with C_3) and 15 times sister to C_4 (17 times in clade
471 with C_4) for C_4 genes (independent from bootstrap support). When incorporating bootstrap
472 support and discarding everything below 70, the numbers shrink to 10 times sister to C_3 (8
473 times in clade with C_3) and 10 times sister to C_4 (13 times in clade with C_4) for C_4 genes. For
474 the two C_4-associated transcription factors, a sister relationship of *S. divaricata* agg. to the C_4
475 clade is supported for SHORTROOT (BS = 86). The sister relationship to the C_4 clade is not
476 well supported for BEL1-like homeodomain 7 (BS =17 sister to C_4, 57 in clade with C_4). For
477 the genes assigned to the Glyoxylate cycle *S. divaricata* agg. appears with several copies which
478 are grouped with high support with either C_3 or C_4 (BS > 80). One locus is missing in the *S.*
479 *divaricata* agg. and *S. montana*. *Salsola divaricata* agg. is also missing in the alignments for
480 the RuBisCO subunits. In the trees for photorespiration-associated genes many species are
481 represented with several copies, even though during the transcriptome processing putative
482 genes were inferred to remove isoforms and assembly artifacts (i.e. misassemblies and
483 chimeras). This is also true for *S. divaricata* agg. which is also rarely monophyletic. The sister
484 relations of *S. divaricata* agg. to the C_3 species are not well supported (5 times with BS < 70),
485 but clade memberships with the C_3 species are (9 times, BS > 70). Sister relations to the C_4
486 clade are well supported in six of the nine times this occurred, with supported shared clade
487 membership in seven of ten cases. Just like the *S. divaricata* agg. the C_4 clade is also not
488 monophyletic in most cases. Clade assignment for all 50 genes is listed in **Supporting**
489 **Information Table S3** and the individual trees are shown in the **Supporting Information Fig**
490 **S8**.
491

492 ## Discussion

493 In this study we aim to distinguish two alternative hypotheses concerning the origin of C_3-C_4
494 intermediacy of the *Salsola divaricata* agg. The model of C_4 evolution (e.g., Sage, 2004)
495 interprets C_2 as an evolutionary stepping stone to fully functional C_4 photosynthesis. *Salsola*
496 *divaricata* is currently classified as a C_2 Type I according to Sage *et al.* (2014). A strong
497 selective pressure towards an increase of PEPC activity should be expected under more
498 stressful (more carbon deficient) conditions in the more arid part of the distribution area
499 (Lanzarote, Fuerteventura, Morocco) and a transition to the C_2 Type II or even C_4-like
500 intermediate type, resulting in a reduction of photorespiration (Sage *et al*., 2014). In case of a
501 hybrid origin of C_2 in the *Salsola divaricata* agg. caused by a hybridization event at some point
502 in the evolutionary history of the Salsoleae, which is the alternative working hypothesis, we
503 should see an independent segregation of C_4-traits (Brown & Bouton, 1993; Oakley *et al*.,
504 2014); and, a phylogenetic signal indicating a hybridization event involving the *Salsola*
505 *divaricata* lineage.
506
507 **Carbon isotope measurements depict a stress reaction in the *S. divaricata* agg.; but**
508 **without occurrence of a more C_4-like photosynthesis in populations growing in drier**
509 **environments**
510 Although samples collected in the wild had slightly higher $\delta^{13}C$ values than those sampled in
511 our greenhouse (stressed or control), there is still a clear difference to typical values of C_4
512 species. None of the samples showed a $\delta^{13}C$ value typical for a C_4-like or a C_4 plant which
513 would be around -9 to -16 ‰ (Sage, 2016); instead all samples had values falling within the
514 range typical for C_3 species which would be around -22 to -32 ‰ (Sage, 2016). In C_3 plants
515 the $\delta^{13}C$ values of leaves depend on a number of factors including stomatal conductance and
516 capacity to capture CO_{-2} in photosynthesis which is dependent on light intensity, growth
517 temperature, photochemistry and capacity for carbon assimilation. The plants growing in the
518 greenhouse are under less stress for potential water loss than field conditions due to lower light
519 intensity, and temperature being more moderate than some field conditions. The results suggest
520 a greater stomatal limitation on photosynthesis in the field grown plants, which results in a low
521 intercellular CO_2 concentration. This will change the discrimination against the heavier C
522 isotope since RuBisCO will capture more $^{13}CO_2$ resulting in slightly higher $\delta^{13}C$ values (von
523 Caemmerer, 1992).
524
525 We observed a broad plasticity of $\delta^{13}C$ values. Populations on the Moroccan mainland often
526 have higher values than the populations on the Canary Islands; but there are many exceptions

527 in populations on the more arid Islands of Fuerteventura and La Graciosa as well as on the more
528 humid Islands of Gran Canaria, Tenerife and La Gomera. There does not seem to be a clear
529 pattern correlated with the general climatic conditions on the Islands. Therefore, we assume
530 that the observed variation is due to local, microclimatic differences. In any case, there are no
531 indications of populations exhibiting C_4-like photosynthesis on the basis of the $\delta^{13}C$ values.
532 *Salsola divaricata* plants from all populations grown for several years under salt stress in the
533 greenhouse showed clear morphological differences in comparison to the control plants. They
534 had more succulent leaves, did not grow to the same height as their counterparts in the control
535 group and most individuals in the salt group flowered earlier. In three of the six tested
536 populations, La Gomera (182), Fuerteventura (191), and Gran Canaria (198), this long-term
537 salt stress was reflected in a significantly higher $\delta^{13}C$ value. The three other populations from
538 Lanzarote, La Graciosa and Fuerteventura had slightly, but not significantly higher, values for
539 the salt treatment compared to the control group. This suggests *Salsola divaricata* can adapt to
540 salt stress by increasing water storage and decreasing stomatal conductance and it explains
541 some of the $\delta^{13}C$ value variation found in the wild as salinity certainly varies among sites.
542

543 **C_3-C_4 intermediate CO_2 compensation points for *S. divaricata* agg. supports a stable C_2-**
544 **state with high plasticity**

545 The CO_2 compensation points of the *S. divaricata* agg. are well within the range of other C_2
546 species (Sage *et al*., 2014 and ref. therein) and they are significantly lower than the values of
547 C_3 species *S. webbii* and significantly higher than those of C_4 species *S. oppositifolia*. There is
548 no evidence for constantly lower Γ values in populations that originated from more arid regions,
549 although individuals of *S. verticillata* and *S. deschaseauxiana* mostly show somewhat lower
550 values, many individuals from similar climatic environments (e.g. *S. gymnomaschala* or
551 individuals of *S. divaricata* from Fuerteventura or Lanzarote) show higher values which screws
552 a possible relation of climate and lower Γ values. Sage *et al*. (2014) outlined that a Γ value
553 below 10 µmol CO_2 mol^{-1} characterizes the transition from C_2 Type I to C_2 Type II and
554 represents an evolutionary progression towards C_4. An accessory C_4 metabolic cycle is
555 present in C_2 Type II plants and was so far only observed in several species of *Flaveria and*
556 *Mollugo verticillata* (Edwards & Ku, 1987). None of the Γ measurements of 53 individuals
557 of the *Salsola divaricata* agg. shows an indication of an accessory C_4 metabolic cycle and a
558 significantly higher carboxylation by PEPC.
559

560 So far the analysed traits seem to be independent between populations and islands and show a
561 high plasticity. On Lanzarote for example, populations of *S. divaricata* have comparably high
562 Γ values and medium to low carbon isotope values compared to other populations.

563 **PEPC activity does not relate to Γ values or carbon isotope values in *S. divaricata* agg.**

564 The PEPC activity of the *S. divaricata* agg. is similar to the C_3 species *S. webbii*. Even though
565 there was a high standard deviation within C_4 *S. oppositifolia*, it is evident that none of the C_2
566 individuals comes even close to a C_4-like value. The activity of PEPC for *S. verticillata* and *S.*
567 *deschaseauxiana* are slightly lower than those for *S. divaricata*; therefore, their slightly lower
568 Γ values and slightly higher carbon isotope values are not due to increased CO_2 fixation by an
569 optimized PEPC. Thus, S. *verticillata* and *S. deschaseauxiana* do not seem to be closer to a
570 more C_4-like state than *S. divaricata* or *S. gymnomaschala*.
571
572 The PEPC activity in the *S. divaricata* agg. is highly variable and shows a lack of correlation
573 with carbon isotope values and CO_2 compensation points, as discussed above for *S. verticillata*.
574 We would expect a higher PEPC activity in the salt treatment or the more arid locations e.g.
575 Lanzarote or Fuerteventura, however that was not the case. Although there are statistically
576 significant differences between islands, it is not possible to predict a value for PEPC activity
577 given the location since there is generally a large overlap of values. The differences between
578 populations could not be explained by the climatic context of the parent plant in the wild and
579 they were all treated the same in the greenhouse. We could not detect a correlation between
580 PEPC activity and the previously discussed traits, supporting the independent distribution of
581 the traits.
582
583 The independent variation in three photosynthetic traits (carbon isotope value, CO_2
584 compensation point and PEPC activity) does not support a directed selection towards C_2 Type
585 II or C_4-like photosynthesis in any of the populations studied, not even those growing in semi-
586 desert conditions. In contrast, our trait observations are more in line with the results of
587 hybridization experiments (Brown & Bouton, 1993; Oakley *et al*., 2014; Kadereit *et al*., 2017)
588 which supports the hypothesis of a hybrid origin of the *Salsola divaricata* agg. Summarizing
589 the physiology of advanced generations of interspecific C_3 and C_4 hybrids, Brown & Bouton
590 (1993) stated that the `correlation among photosynthetic traits were low´ which indicates `a
591 high degree of independence, both genetic and physiological, among C_4 traits´. Like in our
592 study, the hybridization experiment of C_3 and C_4 *Atriplex* (Oakley *et al*., 2014) showed a broad
593 range of Γ values (in nine F_2 individuals from 25-45 µmol CO_2 mol^{-1}; and one C_4-like value).

Brown & Bouton (1993) hypothesize that some of the photosynthetic traits measured can be more C_4-like due to non-disturbance of the C_3 metabolism, such as Γ, while others mirror the physiological C_3 default condition, such as enzyme activity.

First phylogenomic evidence of a hybrid origin of the *S. divaricata* agg.

The Salsoleae species tree shows a high level of conflict including the *S. divaricata* agg, despite high bootstrap support. The assessment of hybridization events using PhyloNet and HyDe showed multiple hybridization events within Salsoleae in which the *Salsola divaricata* agg. was likely involved. The incomplete sampling of Salsoleae does not allow us to identify the exact parental lineages; but a hybrid origin of the *Salsola divaricata* agg. was highly supported. The network that was recovered most often shows a possible ancient hybridization event involving an ancestor of C_4 *Salsola soda* and the C_3 proto-kranz *Salsola montana* giving rise to the *Salsola divaricata* agg. clade. This is the first genomic evidence that a C_3-C_4 intermediate species is the result of (ancient) hybridization between a C_4 and a C_3 lineage. The second reticulation event at a later time indicates a possible hybridization event involving the ancestor of *Salsola divaricata* + *Salsola verticillata* and the C_4 *Salsola soda* giving rise to the C_4 clade formed by *Salsola oppositifolia*, *Halogeton*, *Hammada*, and *Haloxylon*. This might indicate that this C_4 lineage could have inherited C_4 or antecedent traits from a parental lineage giving them an adaptive advantage which might have influenced their photosynthesis pathway, e.g. adaptations in photosynthesis or photorespiration associated enzymes or Kranz-cells. We also found hybridization events in the aforementioned two clades in the other networks with similar parental lineages, supporting the evidence that these lineages were involved in hybridization events. This could indicate either an incomplete reproductive barrier in the early lineages and recurrent backcrossing or incidences of horizontal gene transfer. We have to keep in mind though that our sampling is limited and that other lineages not sampled in the C_4 clade might also play a role (see Schüssler *et al*., 2017 for a plastid tree of the Salsoloideae).

Another explanation for the observed picture could be horizontal gene transfer (HGT). Dunning *et al*. (2019) showed that an individual of *Alloteropsis semialata* (Paniceae, Poaceae) contained at least 59 genes which were acquired via horizontal gene transfer (HGT) from approximately nine different grass donors often growing in close proximity. The mechanism is not determined as of yet though they assume allopolyploids which backcrossed with diploids, occasional interspecific cell-to-cell contact through root-to-root interaction as well as illegitimate pollination could be possibilities. Of these, HGTs functional genes included photosynthetic

genes alongside genes incorporated in disease resistance and abiotic stress response (Dunning *et al*., 2019). Key enzymes of the C_4 pathway were thus acquired by *Alloteropsis* which led to a switch to C_4 photosynthesis (Christin *et al*., 2012). There are other examples of HGT occurring in green plants which can lead to adaptive innovations (reviewed in Chen *et al*., 2021). Just like grasses, *Salsola* is wind pollinated and a similar occurrence of illegitimate pollination as described in Christin *et al*. (2012) and Roalson (2012) is not ruled out. However, the only *Salsola* currently growing in close proximity to the *Salsola divaricata* agg. is *S. oppositifolia*.

Trees of photosynthetic genes supports the hybrid origin hypothesis and suggests candidate genes for further study

A closer look at c. 50 gene trees of genes or gene families involved in photosynthesis or photorespiration revealed the same overall pattern, namely incongruence regarding the position of the *Salsola divaricata* agg. which in most cases either groups with the C_3 *S. montana* (also the case in the plastid dataset) or with the C_4 species. This pattern could be explained by hybridization of parental C_3 and C_4 lineages giving rise to an allopolyploid *S. divaricata* agg. with an intermediate expression of photosynthetic traits. The populations of the *S. divaricata* agg. studied here are all tetraploid (D. Tefarikis, unpublished flow cytometry data). The ploidy level of *S. montana* is unknown. In the C_4 clade, ploidy levels from di-, tetra- and octoploid are present (Kew C-value database and D. Tefarikis, unpublished flow cytometry data).

The kranz-like salsoloid leaf type of the *S. divaricata* agg. (Schüssler *et al.*, 2017: Fig. 7), for example, might occur as in the case of the gene SHORTROOT, a C_4 associated transcription factor which influences the development of kranz cells (Kelly *et al*., 2017, Slewinski *et al*. 2014), with expression of the parental C_4 copy. Studying transcriptome profiles of Salsoleae and Camphorosmeae leaves, respectively, Lauterbach *et al*. (2017a) and Siadjeu *et al.* (2021) revealed a high expression of SHORTROOT in C_2 species. In case of PEPC two copies were sequenced and for both the *S. divaricata* agg. group with the C_3 species supporting the C_3-like PEPC activity data found in this study.

Conclusions

The phylogenetic signal in our trees, gene trees and networks indicate hybridization events gave rise to the *Salsola divaricata* agg. Furthermore, the broad but independent segregation of three photosynthetic traits in a C_3-C_4 intermediate range without indication of selection for a more C_4-like expression supports a hybrid origin of this C_2 clade. Overall the photosynthetic traits

seem plastic and genetically fixed stress reactions are not prominent among the populations studied. The *Salsola divaricata* agg. is well adapted to a broad range of climatic conditions including mesic to arid coastal habitats and also to salinity. Part of the success of the species aggregate might be the photorespiratory pump acquired through hybridization with a C_4 lineage in the past. We therefore conclude that this C_2 species aggregate does not represent an evolutionary stepping stone towards C_4; rather the more likely explanation is a hybrid origin. However, a broader sampled phylogenomic study is needed to reveal the parental lineages. We propose that hybridization events might also have influenced the evolutionary history of other C_2 lineages, and that in particular those genera on which the model of C_4 evolution was based should be tested.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

GK designed the study. DT and GE (carbon isotope values) measured the trait data and DT analysed them. DT, DMB and YY conducted the phylogenomic analyses and hybridization networks. DMB prepared the figures. DT, GK and DMB wrote the first draft. All authors contributed to the discussion of the results and the final version of the manuscript.

Funding

Funding for this project came from the German Science Foundation (DFG grant KA1816/9-1) and the University of Minnesota.

Acknowledgements

We thank the collectors of seeds of the *Salsola divaricata* agg. for their great help to start this project, in particular R. Barone, V. Boehlke, H. Freitag, J. Gil González, F. Hernández, M. Olangua Corral, S. Scholz and E. Voznesenskaya. We are grateful to C. Wild for cultivating and maintaining the plants for this project. For conducting a part of the carbon isotope measurements we thank M. Maus (Geology Department University of Mainz). The Minnesota Supercomputing Institute provided access to computational resources.

695 ## Data Availability Statement

696 The datasets analyzed for this study can be found in the NCBI Sequence Read Archive (SRA)
697 (see **Table 2** for SRA accession numbers).
698

699 ## References

700 **Bender MM, Rouhani I, Vines HM, Black CC**. **1973**. 13C/12C Ratio Changes in
701 Crassulacean Acid Metabolism Plants. *Plant Physiology* **52**: 427–430.
702 doi:10.1104/pp.52.5.427

703 **Björkman O, Gauhl E, Nobs MA. 1969.** Comparative studies of *Atriplex* Species with and
704 without beta-Carboxylation photosynthesis and their first-generation hybrid. *Yearbook 68*
705 *of the Carnegie Institution of Washington*, 620–633.

706 **Blischak PD, Chifman J, Wolfe AD, Kubatko LS**. **2018**. HyDe: A Python Package for
707 Genome-Scale Hybridization Detection (D Posada, Ed.). *Systematic Biology* **67**: 821–829.
708 doi:10.1093/sysbio/syy023

709 **Bräutigam A, Gowik U**. **2016**. Photorespiration connects C_3 and C_4 photosynthesis. *Journal*
710 *of Experimental Botany* **67**: 2953–2962. doi:10.1093/jxb/erw056.

711 **Brown R, Bouton JH**. **1993**. Physiology and Genetics of Interspecific Hybrids Between
712 Photosynthetic Types. *Annual Review of Plant Physiology and Plant Molecular Biology*
713 **44**: 435–456. doi:10.1146/annurev.pp.44.060193.002251

714 **Brown RH, Byrd GT, Bouton JH, Bassett CL**. **1993**. Photosynthetic Characteristics of
715 Segregates from Hybrids between *Flaveria brownii* (C_4 Like) and *Flaveria linearis* (C_3-
716 C_4). *Plant Physiology* **101**: 825–831. doi: 10.1104/pp.101.3.825

717 **Brullo AS**. **1982**. Notes on the genus *Salsola* (Chenopodiaceae). 1. The *Salsola oppositifolia*
718 and *S . longifolia* groups. *Willdenowia* **12**: 241–247. http://www.jstor.org/stable/3995936

719 **Cameron RG, Bassett CL, Bouton JH, Brown RH**. **1989**. Transfer of C_4 Photosynthetic
720 Characters through Hybridization of *Flaveria* Species. *Plant Physiology* **90**: 1538–1545.
721 doi: 10.1104/pp.90.4.1538

722 **Chen R, Huangfu L, Lu Y, Fang H, Xu Y, Li P, Zhou Y, Xu C, Huang J, Yang Z**. **2021**.
723 Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances*
724 **46**: 107671. doi:10.1016/j.biotechadv.2020.107671.

725 **Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J,**
726 **Osborne CP**. **2012**. Adaptive Evolution of C_4 Photosynthesis through Recurrent Lateral
727 Gene Transfer. *Current Biology* **22**: 445–449. doi:10.1016/j.cub.2012.01.054.

728 **Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y,**
729 **Dionora J, Quick WP, Park M, Bennetzen JL, *et al.* 2019**. Lateral transfers of large
730 DNA fragments spread functional genes among grasses. *Proceedings of the National*
731 *Academy of Sciences* **116**: 4416–4425. doi:10.1073/pnas.1810031116.

732 **Edwards GE, Ku MSB. 1987.** Biochemistry of C_3-C_4 intermediates. In: Hatch MD, Boardman
733 NK, eds. *The Biochemistry of Plants*, *Volume 10 Photosynthesis*. New York, USA:
734 Academic Press, 275-325.

735 **Fennane M, Ibn Tattou M**. **1998**. Catalogue des plantes vasculaires rares, menacées ou
736 endémiques du Maroc. *Bocconea* **8**: 1–243.

737 **García-Herrera R, Gallego D, Hernández E, Gimeno L, Ribera P, Calvo N**. **2003**.
738 Precipitation trends in the Canary Islands. *International Journal of Climatology* **23**: 235–
739 241. doi:10.1175/1520-0442(2001)014<3889:IOTNAO>2.0.CO;2.

740 **Holaday AS, Talkmitt S, Doohan ME**. **1985**. Anatomical and enzymic studies of leaves of a
741 $C_3 \times C_4$ *Flaveria* F_1 hybrid exhibiting reduced photorespiration. *Plant Science* **41**: 31–39.
742 doi:10.1016/0168-9452(85)90062-7.

743 **Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW**. **2017**. C_3–C_4
744 intermediates may be of hybrid origin – a reminder. *New Phytologist* **215**: 70–76.
745 doi:10.1111/nph.14567.

746 **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS**. **2017**.
747 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*
748 **14**: 587–589. doi:10.1038/nmeth.4285.

749 **Kelly S, Covshoff S, Wanchana S, Thakur V, Quick P, Wang Y, Ludwig M, Bruskiewich**
750 **R, Fernie A, Sage RF, *et al.* 2017**. Wide sampling of natural diversity identifies novel
751 molecular signatures of C_4 photosynthesis. *bioRxiv*. doi:https://doi.org/10.1101/163097.

752 **Lauterbach M, Billakurthi K, Hankeln TM, Westhoff P, Gowik U, Kadereit G. 2017a.**
753 Organ-specific gene expression profiling of cotyledons and leaves of C_4 *Salsola*
754 *oppositifolia* and insights into regulation of C_2 and C_4 photosynthesis in Salsoleae. In:

755 Lauterbach M. *C_4 photosynthesis in Salsoleae: Ontogenetic gene expression profiling of*
756 *closely related C_3, C_3-C_4 and C_4 species using RNA-Seq*. PhD thesis, Mainz, Germany.

757 **Lauterbach M, Billakurthi K, Kadereit G, Ludwig M, Westhoff P, Gowik U. 2017**. C_3
758 cotyledons are followed by C_4 leaves: Intra-individual transcriptome analysis of *Salsola*
759 *soda* (Chenopodiaceae). *Journal of Experimental Botany* **68**: 161–176.
760 doi:10.1093/jxb/erw343.

761 **Lundgren MR. 2020**. C_2 photosynthesis: a promising route towards crop improvement? *New*
762 *Phytologist* **228**: 1734–1740. doi:10.1111/nph.16494.

763 **Monson RK, Edwards GE. 1984**. C_3-C_4 Intermediate Photosynthesis in Plants. *BioScience*
764 **34**: 563–574. doi:10.2307/1309599

765 **Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF,**
766 **Timoneda A, Yim WC, Cushman JC, Yang Y. 2021**. Disentangling Sources of Gene
767 Tree Discordance in Phylogenomic Datasets: Testing Ancient Hybridizations in
768 Amaranthaceae s.l. *Systematic Biology* **70**: 219–235. doi:10.1093/sysbio/syaa066.

769 **Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015**. IQ-TREE: A Fast and Effective
770 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular*
771 *Biology and Evolution* **32**: 268–274. doi:10.1093/molbev/msu300.

772 **Oakley JC, Sultmanis S, Stinson CR, Sage TL, Sage RF. 2014**. Comparative studies of C_3
773 and C_4 *Atriplex* hybrids in the genomics era: physiological assessments. *Journal of*
774 *Experimental Botany* **65**: 3637–3647. doi:10.1093/jxb/eru106.

775 **Padrón Mederos MA. 2012.** *Caracterización morfológica y genética de especies arbustivas*
776 *de los géneros Salsola L. y Suaeda Forssk. ex J.F. Gmel. en las Islas Canarias.* PhD thesis,
777 Universidad de La Laguna, San Cristóbal de La Laguna, Santa Cruz de Tenerife, España.

778 **Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. 2018**. Quartet Sampling
779 distinguishes lack of support from conflicting support in the green plant tree of life.
780 *American Journal of Botany* **105**: 385–403. doi:10.1002/ajb2.1016.

781 **Reyna-Llorens I, Hibberd JM. 2017**. Recruitment of pre-existing networks during the
782 evolution of C_4 photosynthesis. *Philosophical Transactions of the Royal Society B:*
783 *Biological Sciences* **372**: 20160386. http://dx.doi.org/10.1098/rstb.2016.0386

784 **Roalson EH. 2012**. C_4 Photosynthesis: Need a Gene? Borrow One! *Current Biology* **22**: R161–

785 R163. doi:10.1016/j.cub.2012.01.043.

786 **Sage RF**. **2004**. The evolution of C_4 photosynthesis. *New Phytologist* **161**: 341–370.
787 doi:10.1046/j.1469-8137.2004.00974.x.

788 **Sage RF, Christin PA, Edwards EJ**. **2011**. The C_4 plant lineages of planet Earth. *Journal of*
789 *Experimental Botany* **62**: 3155–3169. doi:10.1093/jxb/err048.

790 **Sage RF, Khoshravesh R, Sage TL**. **2014**. From proto-Kranz to C_4 Kranz: building the bridge
791 to C_4 photosynthesis. *Journal of experimental botany* **65**: 3341–3356.
792 doi:10.1093/jxb/eru180.

793 **Sage RF**. **2016**. Tracking the evolutionary rise of C_4 metabolism. *Journal of Experimental*
794 *Botany* **67**: 2919–2922. doi:10.1093/jxb/erw137

795 **Sage RF, Monson RK, Ehleringer JR, Adachi S, Pearcy RW**. **2018**. Some like it hot: the
796 physiological ecology of C_4 plant evolution. *Oecologia* **187**: 941–966.
797 doi:10.1007/s00442-018-4191-6.

798 **Sage RF, Sage TL, Kocacinar F**. **2012**. Photorespiration and the Evolution of C_4
799 Photosynthesis. *Annual Review of Plant Biology* **63**: 19–47. doi:10.1146/annurev-arplant-
800 042811-105511.

801 **Salichos L, Stamatakis A, Rokas A**. **2014**. Novel Information Theory-Based Measures for
802 Quantifying Incongruence among Phylogenetic Trees. *Molecular Biology and Evolution*
803 **31**: 1261–1271. doi:10.1093/molbev/msu061.

804 **Sayyari E, Mirarab S**. **2016**. Fast Coalescent-Based Computation of Local Branch Support
805 from Quartet Frequencies. *Molecular Biology and Evolution* **33**: 1654–1668.
806 doi:10.1093/molbev/msw079.

807 **Schlüter U, Weber APM**. **2016**. The Road to C_4 Photosynthesis: Evolution of a Complex Trait
808 via Intermediary States. *Plant and Cell Physiology* **57**: 881–889.
809 doi:10.1093/pcp/pcw009.

810 **Schulze S, Westhoff P, Gowik U**. **2016**. Glycine decarboxylase in C_3, C_4 and C_3-C_4
811 intermediate species. *Current Opinion in Plant Biology* **31**: 29–35.
812 doi:10.1016/j.pbi.2016.03.011.

813 **Schüssler C, Freitag H, Koteyeva N, Schmidt D, Edwards G, Voznesenskaya E, Kadereit**
814 **G**. **2017**. Molecular phylogeny and forms of photosynthesis in tribe Salsoleae

815 (Chenopodiaceae). *Journal of Experimental Botany* **68**: 207–223.
816 doi:10.1093/jxb/erw432.

817 **Schwarz G**. **1978**. Estimating the Dimension of a Model. *The Annals of Statistics* **6**: 461–464.
818 doi:10.1214/aos/1176344136.

819 **Siadjeu C, Lauterbach M, Kadereit G**. **2021**. Insights into regulation of C_2 and C_4
820 photosynthesis in Amaranthaceae/Chenopodiaceae using RNA-Seq. *bioRxiv*:
821 2021.09.14.460237. doi:10.1101/2021.09.14.460237.

822 **Slewinski TL, Anderson AA, Price S, Withee JR, Gallagher K, Turgeon R**. **2014**. Short-
823 Root1 Plays a Role in the Development of Vascular Tissue and Kranz Anatomy in Maize
824 Leaves. *Molecular Plant* **7**: 1388–1392. doi:10.1093/mp/ssu036.

825 **Smith SA, Moore MJ, Brown JW, Yang Y**. **2015**. Analysis of phylogenomic datasets reveals
826 conflict, concordance, and gene duplications with examples from animals and plants. *BMC*
827 *Evolutionary Biology* **15**: 150. doi:10.1186/s12862-015-0423-0.

828 **Stamatakis A**. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
829 large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033.

830 **Sugiura N**. **1978**. Further analysts of the data by akaike's information criterion and the finite
831 corrections. *Communications in Statistics - Theory and Methods* **7**: 13–26.
832 doi:10.1080/03610927808827599.

833 **Than C, Ruths D, Nakhleh L**. **2008**. PhyloNet: A software package for analyzing and
834 reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**: 1–16.
835 doi:10.1186/1471-2105-9-322.

836 **von Caemmerer S**. **1992**. Carbon isotope discrimination in C_3-C_4 intermediates. *Plant, Cell*
837 *and Environment* **15**: 1063–1072. doi:10.1111/j.1365-3040.1992.tb01656.x.

838 **Voznesenskaya E V, Koteyeva NK, Akhani H, Roalson EH, Edwards GE**. **2013**. Structural
839 and physiological analyses in Salsoleae (Chenopodiaceae) indicate multiple transitions
840 among C_3, intermediate, and C_4 photosynthesis. *Journal of Experimental Botany* **64**:
841 3583–3604. doi:10.1093/jxb/ert191.

842 **Yang Y, Smith SA**. **2014**. Orthology inference in nonmodel organisms using transcriptomes
843 and low-coverage genomes: Improving accuracy and matrix occupancy for
844 phylogenomics. *Molecular Biology and Evolution* **31**: 3081–3092.

845 doi:10.1093/molbev/msu245.

846 **Yu Y, Nakhleh L**. **2015**. A maximum pseudo-likelihood approach for phylogenetic networks.
847 *BMC Genomics* **16**: S10. doi:10.1186/1471-2164-16-S10-S10.

848 **Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018.** ASTRAL-III: polynomial time species tree
849 reconstruction from partially resolved gene trees. *BMC bioinformatics* **19**: 15–30.
850 doi:10.1186/s12859-018-2129-y.

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871 The following Supporting Information is available for this article:

872 **Fig. S1:** Carbon isotope values (δ13C(‰)) of different populations (P) of the *Salsola*
873 *divaricata* agg. on the Canary islands and Morocco

874 **Fig. S2:** CO2 compensation points (Γ in µmol CO2 mol-1) of populations (P) of the *Salsola*
875 *divaricata* agg.

876 **Fig. S3:** a) Cladogram of Salsoloideae inferred from the Maximum likelihood analyses of the
877 concatenated 991-nuclear gene supermatrix. b) Cladogram of Salsoloideae inferred from the
878 ASTRAL analyses of 991 nuclear genes trees.

879 **Fig. S4:** A) RAxML cladogram of Salsoloideae inferred from the concatenated 991-nuclear
880 gene supermatrix. B) ASTRAL Cladogram of Salsoloideae inferred from 991 nuclear genes
881 trees.

882 **Fig S5:** Maximum likelihood phylogeny of Salsoloideae s.l. inferred from the IQ-TREE
883 analysis of complete and partial plastomes.

884 **Fig. S6:** Maximum pseudo-likelihood species network inferred with PhyloNet with up to five
885 hybridization events.

886 **Fig. S7:** HyDe boxplot of the distribution of the admixture parameters (γ) from the 298
887 significant tests.

888 **Fig. S8:** Maximum likelihood cladograms of the 50 photosynthetic gene trees. Bootstrap
889 support (BS) values are shown above the branches.

890 **Table S1:** Populations (named by living collection number in the Botanical Garden Mainz;
891 see Fig. 1) included in the carbon isotope measurements, CO2 compensation point
892 measurements and PEPC activity measurements, including sampling location and collector.

893 **Table S2:** Gene and character occupancy of the concatenated matrix.

894 **Table S3:** HyDe results from the 298 significant tests.

895 **Materials and Methods S1:** More detailed description of methods of CO_2 compensation
896 point, PEPC activity measurement, Transcriptome processing and nuclear phylogenetic
897 analyses, Assessment of hybridization, Plastome assembly and phylogenetic analysis, and
898 Analysis of photosynthetic gene trees.

899 **References Materials and Methods S1**

900

901 # Tables

902 **Table 1**: List of the 17 samples in the transcriptome phylogeny; with photosynthesis type, ploidy (Kew C value
903 database 2018) and NCBI Sequence Read Archive (SRA) accession number.

Species	**Tribe**	**Ploidy**	**Photosynthesis type**	**SRA accession No.**
Anabasis articulata Choul. ex Pomel	Salsoleae	NA	C_4	SRR6435311
Halogeton glomeratus (M. Bieb.) C.A. Mey.	Salsoleae	2x	C_4	SRR1503502
Haloxylon ammodendron (C.A. Mey.) Bunge	Salsoleae	2x	C_4 (C_3 cotyledon)	SRR1697346
Hammada scoparia (Pomel) Iljin	Salsoleae	2x	C_4 (C_3 cotyledon)	ERR2060287
Kali collina Akhani & Roalson	Salsoleae	2x	C_4	SRR6435349
Salsola divaricata Moq. (Pop.184)	Salsoleae	4x	C_2	ERR2060298
Salsola divaricata Moq. (Pop.198)	Salsoleae	4x	C_2	ERR2060294
Salsola genistoides Juss. ex Poir.	Salsoleae	4x	C_3	SRR14783909
Salsola montana Litv.	Salsoleae	NA	C_3 proto-kranz	SRR14783908
Salsola oppositifolia Desf.	Salsoleae	8x	C_4	ERR2060302
Salsola soda L.	Salsoleae	2x	C_4 (C_3 cotyledon)	SRR3544552
Salsola verticillata Schousb. (Pop.171)	Salsoleae	4x	C_2	SRR14783907
Salsola webbii Moq.	Salsoleae	4x	C_3	ERR2060308
Caroxylon vermiculatum (L.) Akhani & Roalson	Caroxyloneae	NA	C_4	SRR6435345
Bassia scoparia (L.) A.J. Scott	Camphorosmeae	2x	C_4	ERR364385
Eokochia saxicola (Guss.) Freitag & G.Kadereit	Camphorosmeae	NA	C_3	SRR6435348
Beta vulgaris subsp. vulgaris L.	Beteae	2x	C_3	SRX335625

904

905

Table 2: Model testing for 17 taxa and 991 gene trees. Comparing log likelihoods of the trees and networks. The best model is the network with three hybridization events from PhyloNet (best MPL run allowing four hybridization events). k = No. branch lengths + hybridization probabilities; h = number of hybridization events.

tree/network	ln L (log likelihood)	k	h	AIC	deltaAIC	AICc	deltaAICc	BIC	deltaBIC
astral_calgtprob	-19754.54543	31	N/A	39571.091	1698.041	39573.130	1697.190	39722.951	1668.649
raxml_calgtprob	-19741.89558	31	N/A	39545.791	1672.742	39547.830	1671.891	39697.651	1643.349
iqtree_plastome_calgtprob	-22836.08408	31	N/A	45734.168	7861.119	45736.207	7860.268	45886.028	7831.726
phylonet_mpl_1hyb[1]	-19451.79848	31	1	38965.597	1092.547	38967.636	1091.696	39117.457	1063.155
phylonet_mpl_1hyb[2]	-19605.00833	33	1	39276.017	1402.967	39278.323	1402.383	39437.674	1383.372
phylonet_mpl_1hyb[3]	-19624.56237	33	1	39315.125	1442.075	39317.431	1441.491	39476.782	1422.480
phylonet_mpl_2hyb	-19435.04197	35	2	38940.084	1067.034	38942.674	1066.734	39111.539	1057.237
phylonet_mpl_3hyb	-19276.12872	37	3	38626.257	753.208	38629.147	753.208	38807.510	753.208
phylonet_mpl_4hyb	-18899.52481	37	3	37873.050	0.000	37875.940	0.000	38054.302	0.000
phylonet_mpl_5hyb	-19283.48635	41	5	38648.973	775.923	38652.512	38652.512	38849.820	38849.820