Upregulation of C_4 characteristics does not consistently improve photosynthetic performance in intraspecific hybrids of a grass

Matheus E. Bianconi*, Graciela Sotelo, Emma V. Curran, Vanja Milenkovic, Emanuela Samaritani, Luke T. Dunning, Colin P. Osborne, Pascal-Antoine Christin

Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

Abstract

C_4 photosynthesis is thought to have evolved via intermediate stages, with changes towards the C_4 phenotype gradually enhancing photosynthetic performance. This hypothesis is widely supported by modelling studies, but experimental tests are missing. Mixing of C_4 components to generate artificial intermediates can be achieved via crossing, and the grass *Alloteropsis semialata* represents an outstanding system since it includes C_4 and non-C_4 populations. Here, we analyse F1 hybrids between C_3 and C_4, and C_3+C_4 and C_4 genotypes to determine whether the acquisition of C_4 characteristics increases photosynthetic performance. The hybrids have leaf anatomical characters and C_4 gene expression profiles that are largely intermediate between those of their parents. Carbon isotope ratios are similarly intermediate, which suggests that a partial C_4 cycle coexists with C_3 carbon fixation in the hybrids. This partial C_4 phenotype is associated with C_4-like photosynthetic efficiency in C_3+C_4 x C_4, but not in C_3 x C_4 hybrids, which are overall less efficient than both parents. Our results support the hypothesis that the photosynthetic gains from the upregulation of C_4 characteristics depend on coordinated changes in anatomy and biochemistry. The order of acquisition of C_4 components is thus constrained, with C_3+C_4 species providing an essential step for C_4 evolution.

Keywords: C_3-C_4 intermediates, C_4 photosynthesis, complex traits, hybrids, leaf anatomy, Poaceae, transcriptome.

Introduction

C_4 photosynthesis is a complex adaptation that allows plants to sustain high photosynthetic rates in conditions that limit CO_2 availability in the leaf, such as hot, dry and saline environments (Sage 2004). The C_4 metabolism relies on a biochemical cycle operating within a specialized leaf anatomy, which concentrates CO_2 around Rubisco, suppressing the energetically-costly photorespiratory pathway (Hatch 1987). The C_4 trait is highly polyphyletic, but its multiple origins

are mostly clustered in a few groups (Sage et al. 2011). The increased evolutionary accessibility to a C_4 physiology in these groups has been associated with a number of ecological, anatomical and genetic features that acted as evolutionary enablers (Sage 2001; Christin et al. 2013; Moreno-Villena et al. 2018; Edwards 2019). The classic model of C_4 evolution hypothesizes that, once these enablers are in place, an intermediate metabolism that relies on a photorespiratory CO_2 pump is established (in 'type I' C_3-C_4 intermediates or 'C_2' plants), and gradually complemented by a weak C_4 cycle (in 'type II' C_3-C_4 *sensu* Edwards and Ku 1987, or 'C_3+C_4' plants *sensu* Dunning et al. 2017), which is subsequently optimized (Hylton et al. 1988; Monson and Moore 1989; Sage 2004). The transition through the intermediate stages has been predicted by mechanistic modelling studies, which support the idea that the sequential acquisition of components of the C_4 trait successively increases fitness through increases in photosynthetic output (Heckmann et al. 2013, 2016; Mallmann et al. 2014). This model however currently lacks experimental support. Efforts to engineer the C_4 trait into C_3 crops have provided the first opportunities to study the effects of C_4 components in isolation (Ishimaru et al. 1997; Taniguchi et al. 2008; Wang et al. 2017; Ermakova et al. 2021), but the recipient species do not necessarily possess the anatomical properties required for an efficient C_4 metabolism. An alternative strategy is to segregate the components of the C_4 trait via crosses between C_4 and non-C_4 plants.

Successful interspecific crosses between C_4 and non-C_4 plants have been reported particularly for the eudicot genera *Atriplex* (e.g. Björkman et al. 1969; Oakley et al. 2014) and *Flaveria* (e.g. Araus et al. 1990; Byrd et al. 1992; reviewed by Brown and Bouton 1993), with a recent interest in exploring this strategy to dissect the C_4 trait (Oakley et al. 2014; Lin et al. 2021). Most of the diversity of C_4 lineages however lies in monocots, and especially the grass family, with at least 22 independent origins of C_4 photosynthesis (GPWG II 2012). The only existing crosses between photosynthetic types in monocots concerned C_3 and C_3-C_4 parents (Bouton et al. 1986), and the grass *Alloteropsis semialata* offers a particularly suitable system to complement these studies, as it includes not only C_3 and C_4 (Ellis 1974a), but also C_3+C_4 individuals (Lundgren et al. 2016). While the different photosynthetic types of *A. semialata* can hybridize (Bianconi et al. 2020), the effects of different crosses remain to be described.

Alloteropsis semialata has a paleotropical distribution, with C_3 and C_3+C_4 populations in Southern and Central/Eastern regions of Africa, respectively, and C_4 individuals occurring across Africa, South and Southeast Asia, and Oceania (Lundgren et al. 2015; Olofsson et al. 2016; Bianconi et al. 2020; Olofsson et al. 2021). The photosynthetic types in *A. semialata* are associated with distinct genetic lineages (Lundgren et al. 2015; Olofsson et al. 2016), which initially diverged around 3 Mya (Bianconi et al. 2020). Although the ranges of C_4 and non-C_4 lineages geographically overlap in some regions of Africa, with C_3+C_4 and C_4 in Central/Eastern Africa, and C_3 and C_4 in

Southern Africa, genetic analyses have shown that natural hybrids are rare (Olofsson et al. 2016; Bianconi et al. 2020; Olofsson et al. 2021). The lack of a clear hybrid zone has been partially explained by the differences in ploidy levels among photosynthetic types where they are found in sympatry (Olofsson et al. 2021). However, the broad ranges of C_4 and C_3+C_4 diploids overlap, and they have been coexisting for at least one million years (Lundgren et al. 2015; Bianconi et al. 2020). Besides ploidy differences, the low frequency of natural hybridization might be associated with pre- and post-zygotic barriers, such as possible hybrid depression associated with pleiotropic costs of upregulating components of C_4 photosynthesis without a fully functional C_4 metabolism in place (Olofsson et al. 2021). Experimental crosses between C_4 and non-C_4 individuals of *A. semialata* provide an opportunity to test this hypothesis.

Here we generate hybrids between C_3 and C_4 (C_3 x C_4), and C_3+C_4 and C_4 (C_3+C_4 x C_4) accessions of *A. semialata* and analyse their phenotype to test the hypotheses that components of the C_4 trait are additive, and that hybrids rank between their non-C_4 and C_4 parents in terms of photosynthetic performance. Our study confirms that crosses between photosynthetic types in *A. semialata* are viable in experimental conditions, but shows that the C_4 metabolism is disrupted in the hybrids despite the significant upregulation of anatomical and biochemical components of the C_4 trait.

Material and Methods

Plant material and growth conditions

Crosses between accessions of *Alloteropsis semialata* (R. Br.) Hitchc. (Poaceae) from distinct geographical origins were generated from parental plants collected in the wild as seeds or cuttings (Table S1) and grown in greenhouse conditions at the Arthur Willis Environment Centre, University of Sheffield (UK). Initial attempts of controlled cross-pollination using pollination bags to isolate hand-pollinated inflorescences were unsuccessful. For this reason, we adopted a non-controlled pollination strategy, in which plants were allowed to receive pollen from any other plant in the greenhouse, with subsequent determination of the pollen parent via genotyping (see below). The resulting seeds were collected while still attached to the mother plant, and subsequently germinated on Petri dishes before being potted. Here, a total of 30 seedlings from seven mother plants (two C_3, two C_3+C_4 and three C_4) were obtained. Subsequent genotyping of the seedlings identified multiple F1 hybrids between parents with distinct photosynthetic types, including 14 C_3 x C_4 and six C_3+C_4 x C_4 individuals (Table S1). The remaining seedlings consisted of 10 C_4 x C_4 crosses, and an additional C_3+C_4 plant resulting from self-pollination (Table S2). Five C_3 x C_4 individuals grew to maturity and were sampled for DNA analyses but died before being phenotyped, and another two died before being sampled for RNA sequencing and leaf gas-exchange analyses (Table S1). For the

105 phenotypic analyses, we included the mother plants, the known and possible pollen parents (i.e.
106 cases in which the genotyping analyses could not distinguish between genetically similar C_4
107 accessions; see Table S1), and seven additional accessions of *A. semialata* (three C_3, two C_3+C_4 and
108 two C_4) that were growing alongside them to increase the phenotypic diversity within
109 photosynthetic types (Table S1). All plants were grown in 11-L, free-draining pots containing a 2:1
110 mix of M3 compost (Levington) and perlite (Sinclair) under well-watered conditions, and were
111 fertilized once every three months with a NK fertilizer 16-0-5 containing Iron (Evergreen Extreme
112 Green Lawn Food). Plants were grown on a 12-hour day/night cycle, with metal halide lamps
113 providing supplementary light (additional 200 µmol m^{-2} s^{-1} at bench level), 25/20°C day/night
114 temperature, ambient CO_2 concentration and relative humidity between 30% and 60%.
115

116 *Genotyping*

117 We used three genetic markers to determine the genetic lineage of the pollen parent of each plant.
118 These markers correspond to genomic regions (~500 bp long) with sufficient variation to
119 distinguish between different photosynthetic types and geographical groups, and were selected from
120 those previously assembled to determine the origin of allopolyploids in *A. semialata* (Bianconi et al.
121 2020). Custom primers were designed to PCR amplify the target genes (Table S3). Genomic DNA
122 (gDNA) of the putative crosses was isolated from fresh leaves using the DNeasy Plant Mini Kit
123 (Qiagen, Hilden, Germany). PCR reactions contained ca. 10-40 ng of gDNA template, 5 µl of 5X
124 GoTaq Flexi reaction buffer (Promega, Madison, WI, USA), 2 mM of $MgCl_2$, 0.08 mM of dNTPs,
125 0.2 µM of each primer and 0.5 unit of GoTaq polymerase (Promega) in a total volume of 25 µl. The
126 PCR mixtures were initially incubated in a thermocycler for 2 min at 94°C followed by 35 cycles
127 consisting of 30 s at 94°C (denaturation), 1 min at 48°C (annealing) and 1 min at 72°C (elongation).
128 Amplicons were cleaned using Exo-SAP-IT (Affymetrix, Santa Clara, CA, USA), and Sanger-
129 sequenced at the Core Genomic Facility at the University of Sheffield. Sequencing chromatograms
130 were individually inspected for heterozygous sites that are polymorphic among genetic groups
131 corresponding to distinct photosynthetic types, since the presence of such heterozygous sites
132 indicates a F1 hybrid between photosynthetic types. Because the maternal origin was known in
133 advance (the maternal plant was the one on which the seed was collected), the information from the
134 alternative bases was used to narrow down the pollen parent to a genetic lineage (Table S2). The
135 hybrid origin of the plants was subsequently confirmed through an additional genotype analysis
136 using RNAseq data (see below).
137

138 *Stable carbon isotopes*

139 Carbon isotope composition was determined from the central portion of fully expanded leaf blades.

140 Samples were dried in silica gel, and subsequently ground to a fine powder using a TissueLyzer II
141 (Qiagen). Carbon isotope analysis was conducted on 1-2 mg of leaf sample using an ANCA GSL
142 preparation module coupled to a Sercon 20-20 stable isotope ratio mass spectrometer (PDZ Europa,
143 Cheshire, UK). Carbon isotopic ratios ($\delta^{13}C$, in ‰) were reported relative to the standard Pee Dee
144 Belemnite (PDB). Values of $\delta^{13}C$ higher than -16‰ indicate that the plants grew using C_4
145 photosynthesis (O'Leary 1988; Stata et al. 2019). $\delta^{13}C$ values of additional accessions of *A.*
146 *semialata* that were previously phenotyped and identified as C_3, C_3+C_4 or C_4 (Lundgren et al. 2016,
147 2019; Dunning et al. 2017, 2019a) were retrieved from Lundgren et al. (2015, 2016).
148

149 *Leaf anatomy*

150 Leaf cross sections were obtained from the central portion of fully expanded leaf blades. Fresh
151 samples were initially dehydrated in an ethanol series from 70% to 100% EtOH, and resin-
152 infiltrated with Technovit 7100 (Heraeus Kulzer GmbH, Wehrheim, Germany) following the
153 manufacturer's instructions. Cross sections 7-10 µm thick were obtained using a microtome (Leica
154 RM 2245, Leica Biosystems Nussloch GmbH, Nussloch, Germany), and stained with Toluidine
155 Blue O (Sigma-Aldrich, St. Louis, MO, USA). An Olympus BX51 microscope coupled to an
156 Olympus DP71 camera (Olympus Corporation, Tokyo, Japan) was used to photograph the cross
157 sections. Leaf anatomical traits were measured using ImageJ v.1.51q (Schneider et al. 2012) in one
158 cross section per individual. Measurements were taken following Lundgren et al. (2019) within
159 segments, where the segment was defined as the area between two consecutive secondary veins (i.e.
160 veins with large metaxylem vessels; see Fig. S1). Segments that were adjacent to the mid vein or to
161 the lateral margins of the cross section were avoided to maintain consistency between samples. The
162 variables measured here included: leaf thickness (at the leftmost secondary vein in the segment), the
163 number of minor veins (4th and 5th order veins), vein density, distance between edges of consecutive
164 bundle sheaths (BSD), width of inner (IS) and outer sheath (OS), and the areas of epidermis
165 (including bulliform cells, Epd), mesophyll (M), extraxylary fibres (Fb), bundle sheath (IS and OS)
166 and vascular tissue (V; see Fig. S1 for an illustrated key for each variable). Areas were expressed
167 relative to the total area of the segment before analysis, except IS and OS, which were expressed
168 relative to the mesophyll area (IS/M and OS/M). A principal component analysis (PCA) was
169 performed on these 12 leaf anatomical variables using the function *prcomp* in R v. 3.6.3 (R Core
170 Team 2020).
171

172 *Leaf transcriptome*

173 Leaf mRNA was isolated and sequenced as previously described (Dunning et al. 2019a). In short,
174 the distal halves of fully expanded leaves were sampled in the middle of the light period, flash-

175 frozen in liquid N_2 and kept at -80°C. All samples were collected on the same day. Total RNA was
176 extracted using the RNeasy Plant Mini Kit (Qiagen, Germany) with an on-column DNA digestion
177 step (RNase-Free DNase Set; Qiagen). A total of 20 RNA-Seq libraries (one per individual) were
178 prepared with the TruSeq RNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA) using
179 0.5 µg of starting RNA and aiming at a median insert size of ~ 155 bp (standard fragmentation
180 protocol). Libraries were paired-end sequenced (read length = 100 bp) on 1/24 of a single Illumina
181 HiSeq 2500 flow cell in rapid mode (with four additional samples from an unrelated project) at the
182 Sheffield Diagnostic Genetics Service. Raw sequence reads were filtered to remove adaptor
183 contamination and low quality reads (i.e. < 80% of bases with Phred score > 20) using
184 NGSQCToolkit v. 2.3.3 (Patel and Jain 2012). Reads were further trimmed from the 3' end to
185 remove bases with Phred score < 20. The quality of the filtered data was assessed using FastQC v.
186 0.11.9 (Andrews 2010).
187 Transcript abundance was quantified by mapping the filtered reads to a reference dataset
188 consisting of coding sequences from multiple transcriptomes of *A. semialata* retrieved from
189 Dunning et al. (2017) and modified by Bianconi et al. (2018). In short, this dataset consists of 5,540
190 groups of homologous genes that are common to the Panicoideae grasses (the subfamily that
191 includes *Alloteropsis*), each group containing all paralogs detected in the transcriptomes of *A.*
192 *semialata* (total of 12,234 groups of co-orthologs). We used this dataset as it includes manually
193 curated sets of co-orthologs of 23 gene families known to have a function in the C_4 biochemistry
194 (Bianconi et al. 2018). This curated gene set increases the read mapping accuracy when paralogs
195 exist with a high sequence similarity. Furthermore, this dataset includes the sequences of paralogs
196 of core C_4 genes that are absent from the reference genome of *A. semialata*, such as four different
197 genes for the enzyme phosphoenolpyruvate carboxylase (PEPC) known to be highly expressed in
198 other *A. semialata* accessions (Dunning et al. 2017, 2019a). Filtered reads were mapped to the
199 reference dataset using Bowtie2 v. 2.3.5 (Langmead and Salzberg 2012) with default parameters.
200 Read counts are reported here in reads per million mapped reads (RPM). To visualize sample-to-
201 sample distances, a PCA on the global patterns of gene expression was conducted using the function
202 *prcomp* in R after log2-transforming the RPM values. Only genes with transcript abundance > 10
203 RPM in at least five samples were used for the PCA. To test whether the global patterns of gene
204 expression were consistent when the whole gene set of *A. semialata* was considered, we repeated
205 the analysis using the chromosome-level genome assembly of this species (Dunning et al. 2019b) as
206 reference for read mapping. Finally, to verify whether the expression level differences between
207 photosynthetic types were consistent with previous reports for *A. semialata*, we quantified transcript
208 abundance (using the co-orthologous gene set as reference) for 20 additional accessions (four C_3,
209 six C_3+C_4 and 10 C_4) retrieved from two previous RNAseq studies (Dunning et al. 2017, 2019a).

210 To confirm the paternity of the putative hybrids, we genotyped the plants using the RNAseq
211 dataset mapped to the reference genome of *A. semialata*. Here we also included individuals of *A.*
212 *semialata* from the RNAseq studies of Dunning et al. (2017, 2019a) to increase the genetic diversity
213 of the dataset. Read alignment files were filtered to remove duplicates using Picard Tools v. 1.102
214 (https://broadinstitute.github.io/picard/), and variants were called on all individuals combined using
215 BCFtools v. 1.9 (Li et al. 2009). Variant sites were filtered using VCFtools v. 0.1.17 (Li et al. 2009)
216 to remove low quality genotype calls (i.e. quality score < 30 and read depth < 3), indels, and sites
217 with more than 10% missing data, which resulted in an initial set of 521,929 single nucleotide
218 polymorphisms (SNPs). BCFtools was then used to retain only multiallelic sites in which all
219 hybrids were heterozygous and all potential parents were homozygous, which resulted in 15,406
220 SNPs that were then used for the genotype tests. First, to determine the type of cross (i.e. whether
221 C_3 x C_4 or C_3+C_4 x C_4), we identified SNPs that were common to all individuals of the same
222 photosynthetic type and exclusive to them. Sites with more than 25% missing data within each
223 photosynthetic group were removed, and a total of 114 SNPs were retained (Table S4). We then
224 tested, for each SNP, whether hybrid individuals had one C_4 allele and one allele that was either C_3
225 or C_3+C_4, according to the type of cross (Table S4). With the confirmation of the type of cross, we
226 then narrowed down the genetic lineage of the pollen parent (as the mother is already known). Only
227 one individual representing each genetic lineage was retained for this analysis (Table S5). First, we
228 used the 15,406 SNP set to identify all singletons of each potential pollen parent, which in this case
229 were defined as homozygous genotypes that were unique to a single individual among all
230 individuals with the same photosynthetic type. A total of 7,779 singletons were identified (minimum
231 of 287 per potential pollen parent; Table S5). We then verified, for each singleton, which hybrid
232 individuals carried that same allele, and counted the number of positive matches (Table S5). All
233 genotyping analyses were conducted using R (scripts are provided as Supplementary Files).
234

235 *Leaf gas-exchanges*

236 The photosynthetic response to intercellular CO_2 concentration (*A*/*Ci* curve) was measured using
237 two portable gas-exchange systems Li-6400XT (Li-Cor, Lincoln, NE, USA). *A*/*Ci* curves were
238 measured within the first six hours of the photoperiod on the widest fully expanded leaf at a block
239 temperature (T_{block}) of 25 ºC, flow rate = 300 µmol s^{-1} , photosynthetic photon flux density (*PPFD*) =
240 1500 µmol m^{-2} s^{-1}, and were started after both net CO_2 uptake (*A*) and stomatal conductance (g_S)
241 reached steady-state at reference CO_2 = 400 µmol mol^{-1}. Reference CO_2 was then changed in a
242 stepwise manner to 250, 150, 120, 100, 85, 70, 50, 35, 400, 600, 800, 1000 and 1200 µmol mol^{-1}.
243 Readings were automatically logged after 2 to 3 minutes of leaf acclimation to each CO_2 level. The
244 CO_2 compensation point (CCP) and maximum carboxylation efficiency (CE) were calculated

following Bellasio et al. (2016). In this approach, the CO_2-dependence of *A* is described by an empirical non-rectangular hyperbola, and allows for the estimation of parameters irrespective of the photosynthetic physiology, which is therefore a suitable approach for hybrids (Bellasio et al. 2016). Water-use efficiency (WUE) was calculated as the ratio between *A* and g_S at steady-state at reference CO_2 = 400 µmol mol^{-1}.

Results

Carbon isotopic ratios are intermediate between the parents

Carbon isotopic ratios ($\delta^{13}C$) of all hybrids between C_4 and non-C_4 plants were below the range of C_4 values in *A. semialata* (Fig. 1; Table S1). However, $\delta^{13}C$ was consistently higher in the hybrids than in their non-C_4 parents, with average increases of 2.8 ‰ in C_3 x C_4 and 5.4 ‰ in C_3+C_4 x C_4 hybrids relative to their C_3 and C_3+C_4 parents, respectively. The $\delta^{13}C$ ranged between -27.7 and -25.3 ‰ in C_3 x C_4, and -21.4 and -18.5 ‰ in C_3+C_4 x C_4 hybrids.

Leaf anatomy is additive in the hybrids

C_4 and non-C_4 leaves in *A. semialata* are mainly distinguished by the presence of minor veins in C_4 accessions (Lundgren et al. 2019). Here, minor veins were observed in all hybrids, except two C_3 x C_4 individuals (Table S6). The presence of chloroplasts in the inner sheath of all hybrids is suggested by the strong and consistent staining patterns in these cells, which are similar to the patterns observed in C_4 and C_3+C_4 *A. semialata*, but not in C_3 accessions, where the staining is very weak (Fig. 2A). The first component of a PCA on 12 quantitative anatomical traits explains 57.1% of the total variation and separates the accessions per photosynthetic types (Fig. 2B). C_4 individuals are associated with negative value of the first component, which indicate increased vein density (including minor veins) and increased abundance of bundle sheath tissue (IS and OS) relative to the mesophyll (Fig. 2B). C_3+C_4 and C_3 accessions partially overlap, and are associated with positive values of the first component. Hybrid individuals are intermediate between their respective parents along this first component. The second principal component explains 13.9% of the variance, and is correlated with leaf thickness (r = 0.88, $p < 0.001$), width of inner bundle sheath cells (ISW; r = 0.64, $p < 0.001$) and epidermis area (Epd; r = -0.54, $p < 0.001$), but these variables do not differentiate photosynthetic or cross types. The proportion of inner bundle sheath relative to mesophyll (IS/M) was on average doubled in C_3 x C_4 hybrids compared to C_3 accessions, and was within the range of C_3+C_4, but the values were still 62% lower than the C_4 average (Fig. 2C). The distance between consecutive bundle sheaths (BSD) in C_3 x C_4 and C_3+C_4 x C_4 hybrids was on average 49% and 55% smaller than in C_3 and C_3+C_4 plants, respectively (Fig. 2D). The mean width of IS cells was similar between C_3+C_4 and C_4 accessions, and values were on average 48% higher

280 than in C_3, but there were no significant differences among photosynthetic types or between hybrids
281 and the respective parental types (Fig. 2E). OS cells were however significantly smaller in C_4 than
282 in C_3 accessions, with hybrids having intermediate values that were not significantly different from
283 those of C_3+C_4 accessions (Fig. 2F). In the hybrids, the variation in BSD was largely explained by
284 vein density ($R^2 = 0.92$, $p < 0.001$), with no significant effects of OS or IS widths. The IS/M ratio,
285 on the other hand, was not significantly associated with vein density, but with both OS and IS width
286 ($R^2 = 0.75$, $p < 0.001$). Overall, C_3 x C_4 hybrids were not significantly different from C_3+C_4 in
287 relation to the quantitative anatomical traits analysed here, except for BSD, which was significantly
288 reduced in C_3 x C_4 as a result of the presence of minor veins in most accessions. In C_3+C_4 x C_4
289 hybrids, a quantitatively distinct phenotype was generated, with trait values ranking between C_3+C_4
290 and C_4 plants. Overall, our results show that leaf anatomical traits in *A. semialata* are mostly
291 inherited in an additive manner.
292

293 *Gene expression*

294 Global gene expression patterns were assessed via a PCA on 7,482 genes (Fig. 3A). The first
295 principal component mostly separated three sister C_3+C_4 x C_4 plants (H06, H11, and H23) from the
296 other individuals, and accounted for 24.7% of the variation. Photosynthetic types were clearly
297 separated by the second principal component, which accounted for 16.8% of the variation and
298 placed the hybrids as intermediate to the photosynthetic types of their parents. Similar clustering
299 patterns were observed when transcript abundance was quantified for the complete gene set
300 extracted from the genome of *A. semialata* (Fig. S2).
301 Transcript abundance of five genes encoding key enzymes for the C_4 biochemistry, and which
302 are known to be upregulated in C_4 plants of *A. semialata* (Dunning et al. 2019a), were analysed in
303 detail (Fig. 3B,C; Table S7). The patterns of transcript abundance of C_3, C_3+C_4 and C_4 plants are
304 highly similar to those reported by Dunning et al. (2017, 2019a), where a larger sample size was
305 used (Fig. S3). Transcript abundance of the genes encoding the C_4-specific forms of the enzymes
306 aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), NADP-malic enzyme (NADP-ME, gene
307 *nadpme-1P4*) and pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*) were four- to 17-fold
308 higher in C_4 than in C_3 plants, with the transcript abundances of the C_3+C_4 plant slightly above those
309 of C_3 plants (Fig. 3B). In each case, the transcript abundance of the hybrids was intermediate to the
310 abundance of the photosynthetic types of the parents, showing that these expression patterns are
311 mostly additive. C_3 accessions of *A. semialata* lack the gene copy encoding phosphoenolpyruvate
312 carboxykinase (PCK) that is highly expressed in C_4 and C_3+C_4 accessions (*pck-1P1_LGT*; Olofsson
313 et al. 2016); the transcript abundance of this gene reaches 20% of the mean C_4 values in C_3 x C_4
314 hybrids, and 40% in C_3+C_4 x C_4.

315 Out of the nine genes encoding the key C_4 enzyme phosphoenolpyruvate carboxylase (PEPC)
316 that have been identified in *A. semialata* accessions (Dunning et al. 2017), only five were reported
317 to be highly expressed in leaves of some C_4 accessions (Christin et al. 2012; Dunning et al. 2017,
318 2019a). One of these genes, *ppc-1P3_LGT:A* is only present in the C_4 lineage from Oceania, which
319 is represented here by an Australian accession (AUS1) that is the pollen parent of two C_3 x C_4
320 hybrids (H19 and H21; Table S5); both plants have expression levels of this gene that are similar to
321 the Australian parent (Fig. 3C). Only one of the other two laterally acquired PEPC copies is present
322 in the plants analysed here (*ppc-1P3_LGT:M*), and it is highly expressed (> 10,000 RPM) in four
323 C_3+C_4 x C_4 individuals that are offspring from the C_4 accessions from Burkina Faso or Madagascar
324 (H06, H11, H14 and H23; Table S5; Fig. 3C), which have been previously shown to be part of the
325 only C_4 lineage to carry this gene (Olofsson et al. 2016). The expression levels of *ppc-1P3* are
326 similar across C_3+C_4 and C_4 accessions, and are on average ten-fold higher than in C_3 accessions
327 (except the C_4 AUS1, which has a truncated copy of this gene; Olofsson et al. 2016); in the C_3+C_4 x
328 C_4 hybrids, however, the transcript abundance of this gene is on average 42% higher than the
329 median of C_3+C_4 and C_4 accessions (Fig. 3C). Finally, the gene *ppc-1P6* reaches high expression
330 levels in C_4 accessions (except in AUS1, where it is absent; Dunning et al. 2019b), but is
331 disproportionally upregulated in the C_4 Taiwanese lineage (TPE; Dunning et al. 2017), where it
332 reaches levels that are up to 100-fold higher than in C_3 plants, and at least twice higher than the
333 values in other C_4 and C_3+C_4 accessions. Here, the *ppc-1P6* copy reaches high levels in the C_3 x C_4
334 and C_3+C_4 x C_4 offspring of TPE, mostly ranging from 10,000 to 60,000 RPM, except for one
335 C_3+C_4 x C_4 which has a transcript abundance similar to TPE (~ 100,000 RPM).
336 Overall, the analyses of expression profiles show that the inheritance of C_4 gene expression is
337 largely additive, with some exceptions. In almost all cases, the transcript abundance of the hybrids
338 ranks between those of the parental lineages.
339

340 *Evidence of reduced photosynthetic performance in the hybrids*

341 Photosynthetic performance was assessed from CO_2 response curves (A/Ci) and steady-state
342 measurements at ambient CO_2 (Fig. S4; Tables S8 and S9). Photosynthetic rates in the hybrids
343 ranked between the parents at low Ci levels, but the differences between hybrids and parents were
344 progressively reduced as CO_2 levels increased, until they were equal to their non-C_4 parents at Ci of
345 100 in C_3 x C_4, and 250 µmol mol^{-1} in C_3+C_4 x C_4 hybrids (Fig. 4). The C_3 and C_3+C_4 accessions had
346 on average similar values of maximum carboxylation efficiency (CE), and these were 57% lower
347 than in C_4 accessions (Fig. 5A; Table S9). Despite the larger variation among individuals, C_3+C_4 x
348 C_4 hybrids had CE values that were on average 83% and 96% higher than in C_3 and C_3+C_4
349 accessions, respectively, and these were not significantly different from the average values of C_4

350 accessions. In C_3 x C_4 hybrids, however, CE was consistently lower than in C_3 accessions, and
351 reached the lowest mean values of all photosynthetic types/crosses.
352 The CO_2 compensation point (CCP) ranged between 43 and 48 µmol mol^{-1} in C_3 accessions, and
353 between 5 and 15 µmol mol^{-1} in C_3+C_4 individuals (Fig. 5B). All C_4 had CCP below 10 µmol mol^{-1}.
354 In the hybrids, CCP was below 13 µmol mol^{-1} in C_3+C_4 x C_4 individuals, and ranged between 6 and
355 40 µmol mol^{-1} in C_3 x C_4 hybrids. CCP estimates varied substantially among A/Ci curves in C_3 x C_4
356 individuals (Fig. 5B), but mean values were 52% lower than in C_3 plants. The steady-state net CO_2
357 assimilation rate at ambient CO_2 (400 µmol mol^{-1}; A_{400}) was highest in two C_4 accessions, but the
358 differences between C_3, C_3+C_4 and C_4 plants were not significant (Fig. 5C). C_3+C_4 x C_4 hybrids had
359 the lowest A_{400}, but mean values were not significantly different from C_3 x C_4, C_3 and C_3+C_4
360 accessions. The intrinsic water use efficiency (iWUE) at ambient CO_2 was on average 38% higher
361 in C_4 than in other accessions, although a few C_3, C_3+C_4 and C_3+C_4 x C_4 individuals also showed
362 C_4-like values (Fig. 5D). The lowest iWUE values were observed in C_3 and C_3 x C_4 individuals, but
363 the means were not significantly different from C_3+C_4 and C_3+C_4 x C_4 hybrids. To investigate
364 whether changes in leaf temperature (T_{leaf}) could explain the observed patterns, we inspected the
365 variation in T_{leaf} during the A/Ci curves. Although T_{leaf} values up to 3ºC above the median (26.3ºC)
366 were observed in a few accessions, 81% of all data points were collected within a 2ºC interval (Fig.
367 S5). There were no significant differences between groups when all points were considered, nor
368 when only above- or below-ambient CO_2 levels were analysed (Fig. S5). Similar patterns among
369 cross/photosynthetic types were observed in all photosynthetic parameters after removing
370 individuals for which the median T_{leaf} was 1ºC above the median of all curves (Fig. S6).
371 Overall, these results show that the physiological characters are not consistently additive in the
372 hybrids, with some individuals performing below their two parents for several traits.
373

374 **Discussion**

375 *Partial contribution of the C_4 cycle to carbon fixation in hybrids*

376 Here we show that hybrids between C_4 and non-C_4 individuals of the grass *A. semialata* do not
377 exhibit a full C_4 physiology, as indicated by their $\delta^{13}C$, which are clearly outside the range of C_4
378 values in the species (Fig. 1). This indicates that a significant fraction of CO_2 assimilation in these
379 hybrids occurs directly via Rubisco in the mesophyll, which implies that the genetic contribution of
380 the C_4 parent was insufficient to 1) completely suppress Rubisco expression and/or activity in the
381 mesophyll cells, and/or 2) create an effective C_4 cycle. Nonetheless, because $\delta^{13}C$ values are
382 significantly higher in the hybrids than in their non-C_4 parents (Fig. 1), and the CCP is substantially
383 reduced, reaching C_4-like values in C_3+C_4 x C_4 hybrids (Fig. 4B), we conclude that, in most hybrids,
384 a fraction of the CO_2 that enters the leaf is fixed by PEPC in the mesophyll and follows the C_4 route

385 to be refixed by Rubisco in the inner bundle sheath. Furthermore, given the C_4-like CCP and CE
386 values, and $\delta^{13}C$ above -20‰ (in most cases), the degree of C_4 activity is higher in C_3+C_4 x C_4 than
387 in C_3 x C_4 hybrids. Our findings therefore recapitulate what has been shown for interspecific
388 hybrids between photosynthetic types in several eudicot and monocot genera (Björkman et al. 1969;
389 Brown et al. 1985; Araus et al. 1990; Byrd et al. 1992; Brown and Bouton 1993; Oakley et al. 2014;
390 Kadereit et al. 2017). The causes for this partial contribution of the C_4 pathway to the overall carbon
391 assimilation in the hybrids are potentially the same in all these cases, and are probably related to the
392 insufficient expression of components of C_4 metabolism (Brown and Bouton 1993).
393 Two interrelated leaf anatomical traits were significantly different from the C_4 individuals in
394 both C_3 x C_4 and C_3+C_4 x C_4 hybrids, namely the proportion of inner sheath to mesophyll tissue (IS/
395 M) and the distance between consecutive bundle sheaths (BSD). These traits are mostly determined
396 by the presence of minor veins in C_4 leaves, which has been previously shown to be the major
397 anatomical difference between C_4 and non-C_4 *A. semialata* (Lundgren et al. 2019). Here, significant
398 changes towards the C_4 phenotype in IS/M and BSD were observed in the hybrids, but realized
399 values were still far from those observed in C_4 accessions (Fig. 2C,D). This is particularly clear for
400 IS/M, where values were below the mean of both parents in the two hybrid types. Since CO_2
401 refixation by Rubisco takes place in the IS cells in C_4 *A. semialata* (Ellis 1974b; Ueno and Sentoku
402 2006), the leaf anatomy of the hybrids might prevent an optimal coupling between carbon
403 assimilation and reduction reactions, therefore decreasing the efficiency of the C_4 pathway. The size
404 of the OS cells might also play a role in creating an effective C_4 cycle, as suggested by the large
405 differences between C_3 and C_4 accessions (Fig. 2E). Few other C_4 species have an extra layer of BS
406 cells outside of the BS cells containing chloroplasts, and the presence of an OS may represent an
407 increased resistance to the C_4 acid shuttle (Lundgren et al. 2014). Here, the OS cells were
408 significantly reduced in the hybrids relative to their non-C_4 parents, with most C_3+C_4 x C_4
409 individuals and a few C_3 x C_4 having OS cell sizes within the C_4 range (Fig. 2F), which suggests
410 that this character might have played a minor role, if any, in reducing the efficiency of the C_4 acid
411 shuttle in the hybrids.
412 C_4 activity could also have been limited by gene expression and protein activity. Despite
413 significant increases in the transcript abundance of core C_4 genes, particularly the genes encoding
414 PEPC (Fig 4), it is possible that the insufficient expression levels of some genes might have reduced
415 the efficiency of the C_4 cycle, for example by decreasing the rate of carboxylation and/or
416 decarboxylation reactions. Furthermore, the lack of proper compartmentation of the
417 expression/activity of these genes due to a non-tissue-specific expression of C_3 or C_3+C_4 alleles
418 could have impaired the formation of intermediate metabolite pools, therefore preventing the
419 operation of an efficient C_4 cycle (Brown and Bouton 1993; Ermakova et al. 2020). Finally, the lack

420 or insufficiency of post-transcriptional and post-translational modifications might also have affected
421 the realized enzyme activity. This might be particularly relevant in the case of hybrids, where the
422 cell environment may be substantially modified by pleiotropic effects emerging from the expression
423 of two divergent genomes. Detailed analyses of cellular localization and activity of C_4 enzymes
424 coupled to CO_2 labelling studies are necessary to identify the major determinants of the phenotype
425 observed in the hybrids reported here.
426

427 *Photosynthetic performance of plants expressing partial C_4 traits and implications*
428 *for the model of C_4 evolution*

429 The current model of C_4 evolution assumes that the sequential acquisition of certain features, such
430 as Rubisco activity in bundle sheath cells and upregulation of core C_4 enzymes, increases
431 photosynthetic output and consequently fitness in an environment that favours the C_4 physiology
432 (Monson and Moore 1989; Sage 2004; Heckmann et al. 2013; Malmmann et al. 2014). However,
433 this Mt. Fuji-like fitness landscape (Heckmann et al. 2013) is conditional on the presence of
434 anatomical enablers; when these are lacking, the C_4 phenotype is not accessible because
435 intermediate states have lower fitness than the C_3 state (Heckmann 2016). The two groups of
436 intraspecific hybrids generated in this study provide an opportunity to examine this prediction while
437 controlling for phylogenetic effects, with C_3 x C_4 and C_3+C_4 x C_4 plants being proxies for the
438 upregulation of C_4 traits at an early and intermediate stage, respectively, along C_4 evolution. Here,
439 we show that the carboxylation efficiency of C_3 x C_4 hybrids is lower than that of both parents (Figs
440 5, 6A), despite significant changes towards the C_4 phenotype in leaf anatomy and gene expression.
441 In C_3+C_4 x C_4 hybrids, however, carboxylation efficiency was increased in comparison to the C_3+C_4
442 parent, and was not significantly different from C_4 accessions. Such disparate effects of
443 upregulating C_4 components in a non-C_4 background are compatible with the initial prediction that
444 fitness gains are conditional on the presence of enabling factors, which here might have been
445 present in C_3+C_4 parents, but lacking in the C_3 plants. Such changes would have acted as a switch
446 that permitted a full C_4 pathway to operate, and therefore increase photosynthetic performance.
447 Note however that such gains in photosynthetic performance are restricted to conditions of low CO_2
448 concentration in the leaf (Fig. 4). In fact, at ambient CO_2 levels, several individuals of both hybrid
449 groups had lower photosynthetic rates than both parents, suggesting that the partial upregulation of
450 C_4 components, all at once, might have negative consequences, possibly due to pleiotropic effects
451 (e.g. C_4 enzymes competing for reducing power and ATP with other cell reactions). This in turn
452 supports the idea that the changes towards a full C_4 physiology must be built sequentially upon
453 preexisting characteristics that enable the next change to provide a fitness advantage (Heckmann
454 2016). Further characterization of the photosynthetic performance of F2 hybrids expressing C_4

455 components independently, particularly under conditions expected to favour C_4 plants, including
456 long-term exposure to low CO_2 levels, should provide further insights on whether C_4 evolution must
457 follow a particular sequence of events for it to be viable.
458

459 ## Conclusions

460 In this study, we characterize the phenotype of hybrids between different photosynthetic types in the
461 grass *Alloteropsis semialata* to investigate how the upregulation of components of the C_4 trait
462 affects photosynthetic performance. We show that the hybrids have in most cases anatomical traits
463 and gene expression patterns that are intermediate between those of the parents, and this leads to C_4
464 activity that is equally intermediate between the two parents. The physiological benefits of a partial
465 C_4 metabolism in the hybrids appear only at low CO_2 levels; at ambient CO_2, there is no evidence of
466 enhanced photosynthetic performance in the hybrids relative to their non-C_4 parents. Some hybrid
467 individuals in fact perform worse than both parents at ambient CO_2, and this possible hybrid
468 depression could explain the lack of a clear hybrid zone in regions where the distributions of C_4 and
469 non-C_4 *A. semialata* lineages overlap. Overall, our results support the hypothesis that photosynthetic
470 gains arising from the upregulation of C_4 features are conditional on coordinated changes in leaf
471 anatomy and biochemistry. Future studies with the F2 offspring, where the different C_4 components
472 can be segregated, will be able to pinpoint the key genetic and phenotypic changes that lead to
473 fitness gains, providing a unique opportunity to experimentally test long-standing hypotheses about
474 the evolution of C_4 photosynthesis.
475

476 ## Supplementary data

477 **Fig. S1.** Leaf anatomical variables measured in this study.
478 **Fig. S2.** Principal component analysis on 11,988 genes from the chromosome-level genome
479 assembly of *Alloteropsis semialata.*
480 **Fig. S3.** Transcript abundance in reads per million (RPM) of selected genes encoding core C_4
481 enzymes in *Alloteropsis semialata*: (A) aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), (B)
482 NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*), (C) phosphoenolpyruvate
483 carboxykinase (PCK, gene *pck1P1_LGT*), (D) pyruvate phosphate dikinase (PPDK, gene *ppdk-*
484 *1P2*), and (E-I) phosphoenolpyruvate carboxylase (PEPC) genes. Transcript abundance computed
485 for *A. semialata* samples extracted from Dunning et al. (2019a) are shown on the right for each
486 gene.
487 **Fig. S4.** Photosynthetic response to intercellular CO_2 (*A*/*Ci*) of F1 hybrids and the parental
488 photosynthetic types in *Alloteropsis semialata*. Panels contain all *A*/*Ci* curves collected for each
489 accession, with individual curves coloured with different shades of grey. Accession name and cross/

490 photosynthetic type are indicated on the top left. Vertical dashed lines indicate the CO_2
491 compensation point of each curve.
492 **Fig. S5.** Leaf temperature variation during *A*/*Ci* curves. Data points are individual *A*/*Ci*
493 measurements for each plant (left panels) or with individuals grouped into cross/photosynthetic
494 types (right panels). (A) All measurements, and measurements collected at reference CO_2 (B) < 100
495 µmol mol^{-1}, (C) between 100 and 400 µmol mol^{-1}, and (D) > 400 µmol mol^{-1}. Arrows indicate
496 outlier individuals that were removed for the analysis in Fig. S6.
497 **Fig. S6**. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in
498 *Alloteropsis semialata* after removing outlier accessions (i.e. with T_{leaf} 1ºC above the median). (A)
499 Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C)
500 net photosynthetic rate (A_{400}), and (D) intrinsic water use efficiency (iWUE, A/g_s) at reference CO_2
501 = 400 µmol mol^{-1}.
502
503 **Table S1.** Sample information.
504 **Table S2.** Genotyping using PCR/Sanger-sequencing.
505 **Table S3.** List of primer sequences used for genotyping.
506 **Table S4.** RNA-seq markers selected for genotyping analysis I (photosynthetic type).
507 **Table S5.** RNA-seq markers selected for genotyping analysis II (pollen parent).
508 **Table S6.** Raw leaf anatomy data.
509 **Table S7.** Normalized transcript abundance for the co-ortholog gene set.
510 **Table S8.** Raw A/Ci data.
511 **Table S9.** Estimated photosynthetic parameters and steady-state measurements.
512

513 Author contributions

514 MEB, CPO and PAC designed the study. MEB phenotyped the plants and analysed the data. EVC,
515 GS and LTD PCR-genotyped the plants. GS and VM generated transcriptome data. ES produced the
516 hybrids. MEB and PAC wrote the manuscript with the help of all co-authors.
517

518 Acknowledgements

519 This work was funded by the European Research Council (grant ERC-2014-STG-638333) and the
520 Royal Society (grant RGF\EA\181050). M.E.B is funded by a Royal Society grant (RGF\EA\
521 181050), and P.A.C. by a Royal Society University Research Fellowship (grant URF\R\180022).
522

523 Data availability

524 Newly generated RNA sequencing datasets were deposited in the NCBI SRA database under

Bioproject PRJNA752516. Scripts used for the genotyping analysis are available as Supplementary Data.

References

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. See http://www. bioinformatics.babraham.ac.uk/projects/fastqc/

Araus JL, Brown RH, Bouton JH, Serret MD. 1990. Leaf anatomical characteristics in *Flaveria trinervia* (C_4), *Flaveria brownii* (C_4-like) and their F1 hybrid. Photosynthesis Research 26, 49–57.

Bellasio C, Beerling DJ, Griffiths H. 2016. An Excel tool for deriving key photosynthetic parameters from combined gas exchange and chlorophyll fluorescence: theory and practice. Plant, Cell & Environment 39, 1180–1197.

Bianconi ME, Dunning LT, Moreno-Villena JJ, Osborne CP, Christin P-A. 2018. Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus *Alloteropsis*. Journal of Experimental Botany 69, 1967–1980.

Bianconi ME, Dunning LT, Curran E V, et al. 2020. Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. Proceedings of the Royal Society B: Biological Sciences, 287.

Björkman O, Gauhl E, Nobs MA. 1969. Comparative studies of Atriplex species with and without ß-carboxylation photosynthesis. Carnegie Institution of Washington Yearbook 68, 620–633.

Bouton JH, Brown RH, Evans PT, Jernstedt JA. 1986. Photosynthesis, leaf anatomy, and morphology of progeny from hybrids between C_3 and C_3/C_4 *Panicum* species. Plant Physiology 80, 487–492.

Brown RH, Bouton JH, Evans PT, Malter HE, Rigsby LL. 1985. Photosynthesis, morphology, leaf anatomy, and cytogenetics of hybrids between C_3 and C_3/C_4 *Panicum* species. Plant Physiology 77, 653–658.

Brown HR, Bouton JH. 1993. Physiology and genetics of interspecific hybrids between photosynthetic types. Annual Review of Plant Biology 44, 435–456.

Byrd GT, Brown RH, Bouton JH, Bassett CL, Black CC. 1992. Degree of C_4 photosynthesis in C_4 and C_3-C_4 *Flaveria* Species and their hybrids: I. CO_2 assimilation and metabolism and activities of phosphoeno/pyruvate carboxylase and NADP-malic enzyme. Plant Physiology 100, 939–946.

von Caemmerer S, Furbank RT. 2003. The C_4 pathway: an efficient CO_2 pump. Photosynthesis Research 77, 191–207.

Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J,

560 **Osborne CP.** 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene
561 transfer. Current Biology 22, 445–449.
562 **Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison**
563 **LM, Vorontsova MS, Edwards EJ.** 2013. Anatomical enablers and the evolution of C_4
564 photosynthesis in grasses. Proceedings of the National Academy of Sciences 110, 1381–1386.
565 **Danila FR, Thakur V, Chatterjee J, Bala S, Coe RA, Acebron K, Furbank RT, von**
566 **Caemmerer S, Quick WP.** 2021. Bundle sheath suberisation is required for C_4
567 photosynthesis in a *Setaria viridis* mutant. Communications Biology 4, 1–10.
568 **Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P,**
569 **Osborne CP, Christin P-A.** 2017. Introgression and repeated co-option facilitated the
570 recurrent emergence of C_4 photosynthesis among close relatives. Evolution 71, 1541–1555.
571 **Dunning LT, Moreno-Villena JJ, Lundgren MR, et al.** 2019a. Key changes in gene expression
572 identified for different stages of C_4 evolution in *Alloteropsis semialata*. Journal of
573 Experimental Botany 70, 3255–3268.
574 **Dunning LT, Olofsson JK, Parisod C, et al.** 2019b. Lateral transfers of large DNA fragments
575 spread functional genes among grasses. Proceedings of the National Academy of Sciences of
576 the United States of America 116, 4416–4425.
577 **Edwards G, Ku MB.** 1987. Biochemistry of C_3–C_4 intermediates. In: Hatch MD, Boardman NK,
578 eds. The biochemistry of plants: a comprehensive treatise, Vol 14: photosynthesis. New York:
579 Academic Press, 275–325.
580 **Edwards EJ.** 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and
581 CAM photosynthesis. New Phytologist 223, 1742–1755.
582 **Ellis RP.** 1974a. Significance of the occurrence of both Kranz and non-Kranz leaf anatomy in the
583 grass species *Alloteropsis semialata*. South African Journal of Science 70, 169-173.
584 **Ellis RP.** 1974b. Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades.
585 Bothalia 11, 273–275.
586 **Ermakova M, Arrivault S, Giuliani R, et al.** 2021. Installation of C_4 photosynthetic pathway
587 enzymes in rice using a single construct. Plant Biotechnology Journal 19, 575–588.
588 **Grass Phylogeny Working Group II (GPWG II).** 2012. New grass phylogeny resolves deep
589 evolutionary relationships and discovers C_4 origins. New Phytologist 193, 304–312.
590 **Hatch MD.** 1987. C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and
591 ultrastructure. Biochimica et Biophysica Acta 895, 81–106.
592 **Hattersley PW, Browning AJ.** 1981. Occurrence of the suberized lamella in leaves of grasses of
593 different photosynthetic types. I. In parenchymatous bundle sheaths and PCR ('Kranz')
594 sheaths. Protoplasma 109, 371–401.

595 **Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ.** 2013.
596 Predicting C_4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji
597 fitness landscape. Cell 153, 1579–1588.
598 **Heckmann D.** 2016. C_4 photosynthesis evolution: The conditional Mt. Fuji. Current Opinion in
599 Plant Biology 31, 149–154.
600 **Hylton CM, Rawsthorne S, Smith AM, Jones DA, Woolhouse HW.** 1988. Glycine decarboxylase
601 is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta 175,
602 452–9.
603 **Ishimaru K, Ichikawa H, Matsuoka M, Ohsugi R.** 1997. Analysis of a C_4 maize pyruvate,
604 orthophosphate dikinase expressed in C_3 transgenic Arabidopsis plants. Plant Science 129,
605 57–64.
606 **Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW.** 2017. C_3–C_4 intermediates
607 may be of hybrid origin – a reminder. New Phytologist 215, 70–76.
608 **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9,
609 357–359.
610 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin**
611 **R; 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map
612 format and SAMtools. Bioinformatics 25, 2078–2079.
613 **Lin M, Schlüter U, Stich B, Weber APM.** 2021. Cis-regulatory divergence underpins the
614 evolution of C_3-C_4 intermediate photosynthesis in *Moricandia*. bioRxiv [Preprint] doi:
615 10.1101/2021.05.10.443365
616 **Lundgren MR, Osborne CP, Christin PA.** 2014. Deconstructing Kranz anatomy to understand C_4
617 evolution. Journal of Experimental Botany 65, 3357–3369.
618 **Lundgren MR, Besnard G, Ripley BS, et al.** 2015. Photosynthetic innovation broadens the niche
619 within a single species. Ecology Letters 18, 1021–1029.
620 **Lundgren MR, Christin P-A, Escobar EG, Ripley BS, Besnard G, Long CM, Hattersley PW,**
621 **Ellis RP, Leegood RC, Osborne CP.** 2016. Evolutionary implications of C_3–C_4 intermediates
622 in the grass *Alloteropsis semialata*. Plant, Cell and Environment 39, 1874–1885.
623 **Lundgren MR, Dunning LT, Olofsson JK, et al.** 2019. C_4 anatomy can evolve via a single
624 developmental change. Ecology Letters 22, 302–312.
625 **Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U.**
626 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus
627 Flaveria. ELife 3.
628 **Monson RK, Moore BD.** 1989. On the significance of C_3-C_4 intermediate photosynthesis to the
629 evolution of C_4 photosynthesis. Plant, Cell & Environment 12, 689–699.

Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA. 2018. Highly expressed genes are preferentially co-opted for C_4 Photosynthesis. Molecular Biology and Evolution 35, 94–106.

O'Leary M.H. 1988. Carbon isotopes in photosynthesis. BioScience 38, 328–336.

Oakley JC, Sultmanis S, Stinson CR, Sage TL, Sage RF. 2014. Comparative studies of C_3 and C_4 *Atriplex* hybrids in the genomics era: Physiological assessments. Journal of Experimental Botany 65, 3637–3647.

Olofsson JK, Bianconi M, Besnard G, et al. 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Molecular Ecology 25, 6107–6123.

Olofsson JK, Curran EV, Nyirenda F, et al. 2021. Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. Molecular Ecology, 1–15.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLos One 7, e30619.

R Development Core Team. 2020. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sage RF. 2001. Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. Plant Biology 3, 202–213.

Sage RF. 2004. The evolution of C_4 photosynthesis. New Phytologist 161, 341–370.

Sage RF, Christin P-A, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. Journal of Experimental Botany 62, 3155–3169.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675.

Stata M, Sage TL, Sage RF. 2019. Mind the gap: the evolutionary engagement of the C_4 metabolic cycle in support of net carbon assimilation. Current Opinion in Plant Biology 49, 27–34.

Taniguchi Y, Ohkawa H, Masumoto C, et al. 2008. Overproduction of C_4 photosynthetic enzymes in transgenic rice plants: An approach to introduce the C_4-like photosynthetic pathway into rice. Journal of Experimental Botany 59, 1799–1809.

Ueno O, Sentoku N. 2006. Comparison of leaf structure and photosynthetic characteristics of C_3 and C_4 *Alloteropsis semialata* subspecies. Plant, Cell & Environment 29, 257–268.

Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA. 2017. Re-creation of a key step in the evolutionary switch from C_3 to C_4 leaf anatomy. Current Biology 27, 3278-3287.e6.

665 **Figures**

666 **Fig. 1.** Distribution of carbon isotopic ratios ($\delta^{13}C$) in *Alloteropsis semialata*. $\delta^{13}C$ values of
667 accessions of *A. semialata* that are not part of this study were retrieved from Bianconi et al. (2020).
668 Blue, green and red arrows indicate $\delta^{13}C$ values of C_3, C_3+C_4 and C_4 known or potential parents of
669 the hybrids analysed in this study.

670 **Fig. 2.** Leaf anatomy of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*
671 (A) Representative cross sections of C_3 (RSA9), C_3 x C_4 (H08), C_4 (TPE1-10), C_3+C_4 x C_4 (H11)
672 and C_3+C_4 (TAN1602-03c) accessions. Arrows indicate minor veins (i.e. fourth and fifth order
673 veins). Scale bar = 100 µm. (B) Principal component analysis of selected leaf anatomical variables
674 (C) Proportion of inner sheath to mesophyll area (IS/M). (D) Distance between consecutive bundle
675 sheaths (BSD). (E) Mean inner sheath cell width. (F) Mean outer sheath cell width. For (C-F),
676 different lower case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-
677 hoc Tukey HSD; $n \geq 4$). Bu = bulliform cells, BSD = distance between consecutive bundle sheaths,
678 Epd = proportion of epidermis area, Fb = extraxylary fibres (= sclerenchyma girder), IS = inner
679 bundle sheath (= mestome sheath), ISW = IS width, M = mesophyll, Mx = metaxylem, OS = outer
680 bundle sheath, OSW = OS width, Ph = phloem, V = proportion of vascular tissue area.

681 **Fig. 3.** Leaf gene expression of F1 hybrids and the parental photosynthetic types in *Alloteropsis*
682 *semialata*. (A) Principal component analysis on 7,482 genes. (B) Transcript abundance in reads per
683 million (RPM) of selected genes encoding core C_4 enzymes: aspartate aminotransferase (ASP-AT,
684 gene *aspat-3P4*), NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*),
685 phosphoenolpyruvate carboxykinase (PCK, gene *pck1P1_LGT*), pyruvate phosphate dikinase
686 (PPDK, gene *ppdk-1P2*). (C) Transcript abundance of selected phosphoenolpyruvate carboxylase
687 (PEPC) genes.

688 **Fig. 4.** Response of net CO_2 assimilation rate (*A*) to intercellular CO_2 (*Ci*) of F1 hybrids and the
689 parental photosynthetic types in *Alloteropsis semialata*. Data points are predicted values from an
690 empirical non-rectangular hyperbola model (Bellasio et al. 2016) based on the mean parameter
691 estimates per individual within each cross/photosynthetic type. Error bars show standard errors for
692 the predicted values ($n \geq 3$). Arrows indicate Ci values at reference CO_2 = 400 µmol mol^{-1} for each
693 cross/photosynthetic type.

694 **Fig. 5.** Photosynthetic performance of F1 hybrids and the parental photosynthetic types in
695 *Alloteropsis semialata*. (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point
696 (CCP), and steady-state (C) net photosynthetic rate (A_{400}), and (D) intrinsic water use efficiency
697 (iWUE, A/g_s) at reference CO_2 = 400 µmol mol^{-1}. Different lower case letters indicate statistical
698 differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; $n \geq 4$).

Fig. 1

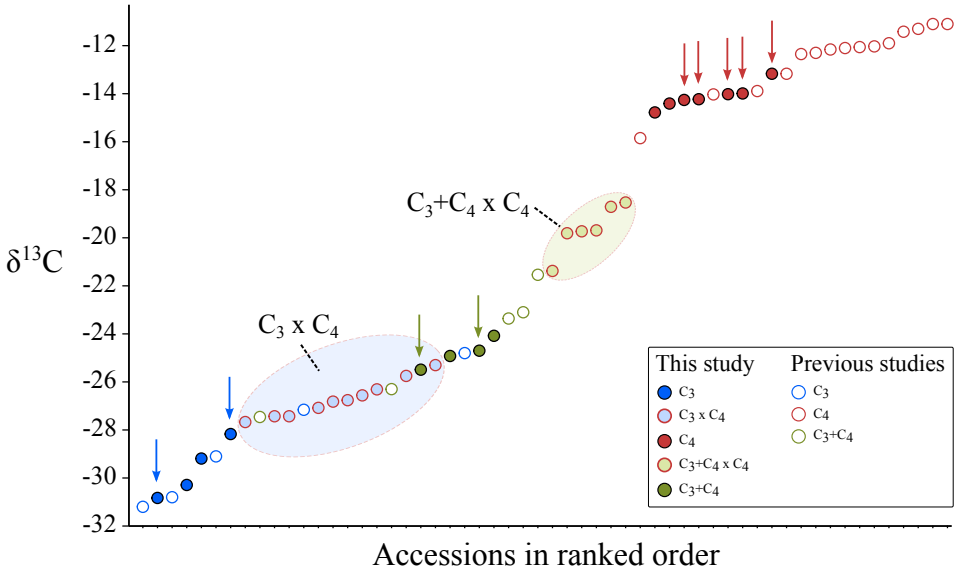

Fig. 1. Distribution of carbon isotopic ratios ($\delta^{13}C$) in *Alloteropsis semialata*. $\delta^{13}C$ values of accessions of *A. semialata* that are not part of this study were retrieved from Bianconi et al. (2020). Blue, green and red arrows indicate $\delta^{13}C$ values of C_3, C_3+C_4 and C_4 known or potential parents of the hybrids analysed in this study.

Fig. 2

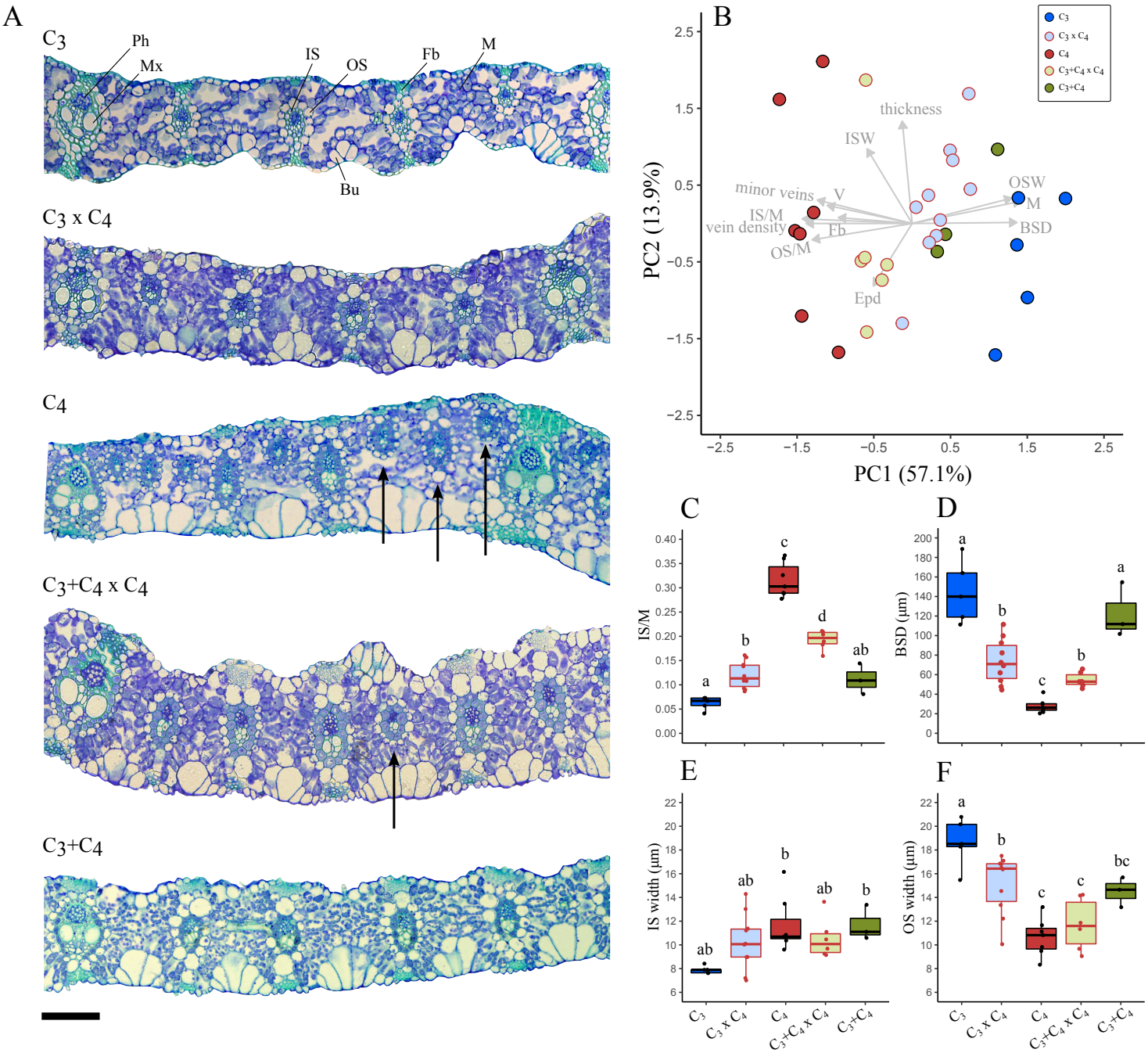

Fig. 2. Leaf anatomy of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata* (A) Representative cross sections of C_3 (RSA9), C_3 x C_4 (H08), C_4 (TPE1-10), C_3+C_4 x C_4 (H11) and C_3+C_4 (TAN1602-03c) accessions. Arrows indicate minor veins (i.e. fourth and fifth order veins). Scale bar = 100 µm. (B) Principal component analysis of selected leaf anatomical variables (C) Proportion of inner sheath to mesophyll area (IS/M). (D) Distance between consecutive bundle sheaths (BSD). (E) Mean inner sheath cell width. (F) Mean outer sheath cell width. For (C-F), different lower case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; $n \geq 4$). Bu = bulliform cells, BSD = distance between consecutive bundle sheaths, Epd = proportion of epidermis area, Fb = extraxylary fibres (= sclerenchyma girder), IS = inner bundle sheath (= mestome sheath), ISW = IS width, M = mesophyll, Mx = metaxylem, OS = outer bundle sheath, OSW = OS width, Ph = phloem, V = proportion of vascular tissue area.

Fig. 3

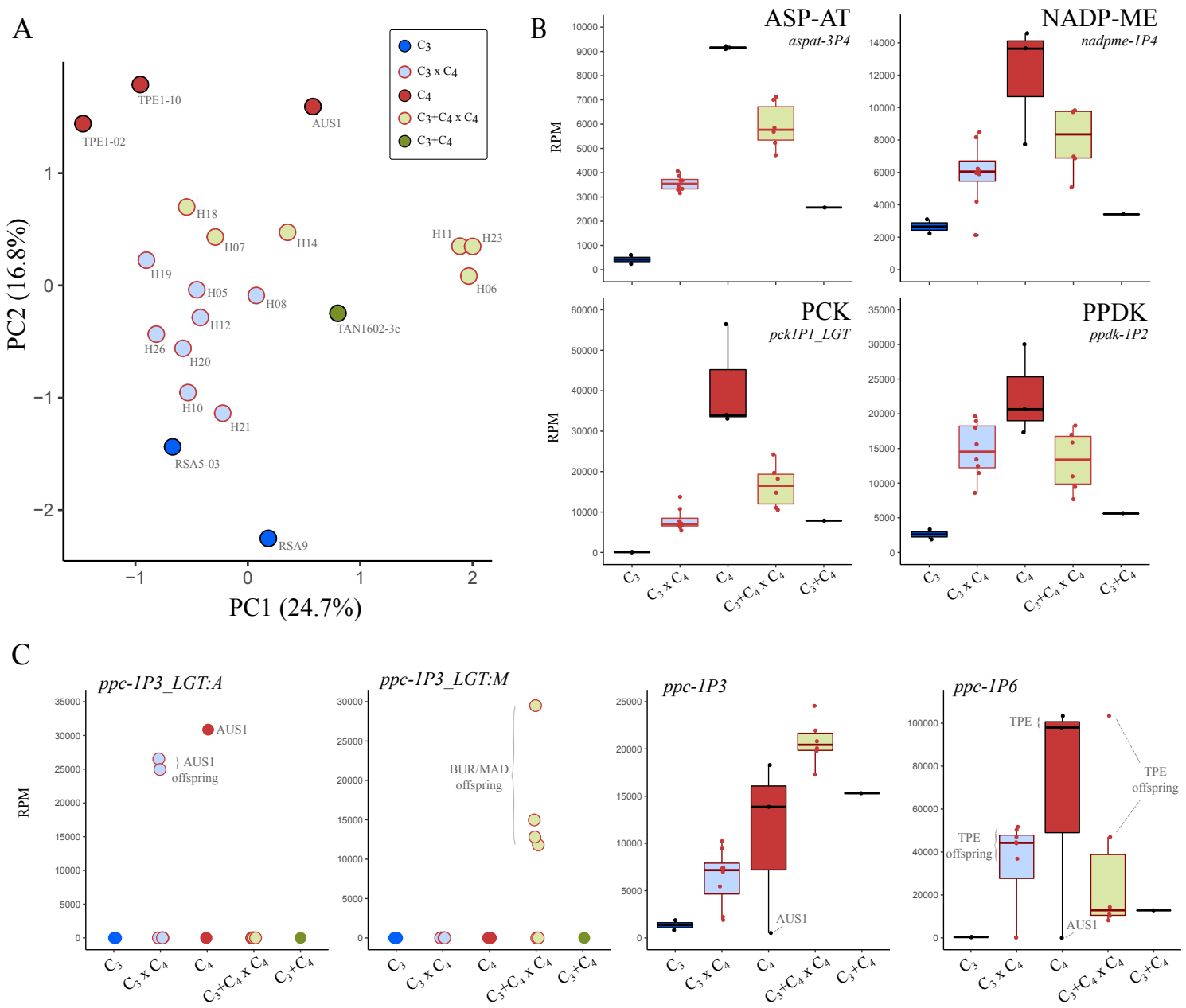

Fig. 3. Leaf gene expression of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) Principal component analysis on 7,482 genes. (B) Transcript abundance in reads per million (RPM) of selected genes encoding core C_4 enzymes: aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*), phosphoenolpyruvate carboxykinase (PCK, gene *pck1P1_LGT*), pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*). (C) Transcript abundance of selected phosphoenolpyruvate carboxylase (PEPC) genes.

Fig. 4

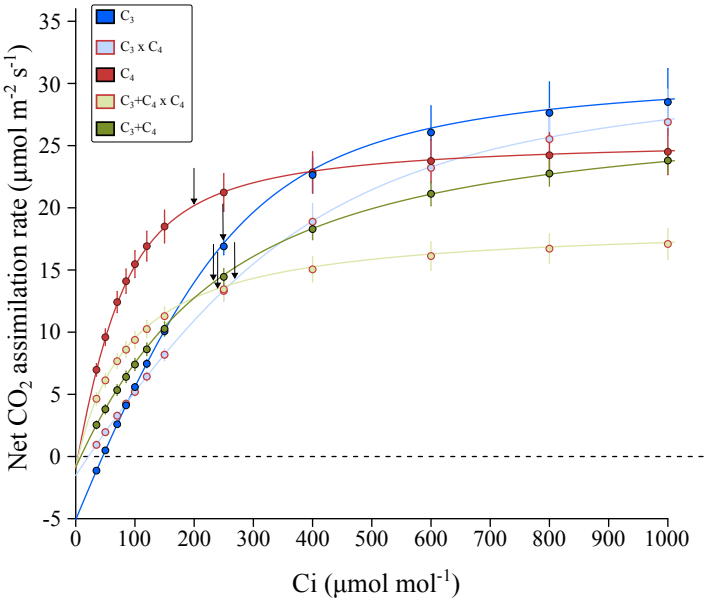

Fig. 4. Response of net CO_2 assimilation rate (*A*) to intercellular CO_2 (*Ci*) of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. Data points are predicted values from an empirical non-rectangular hyperbola model (Bellasio et al. 2016) based on the mean parameter estimates per individual within each cross/photosynthetic type. Error bars show standard errors for the predicted values ($n \geq 3$). Arrows indicate *Ci* values at reference CO_2 = 400 µmol mol^{-1} for each cross/photosynthetic type.

Fig. 5

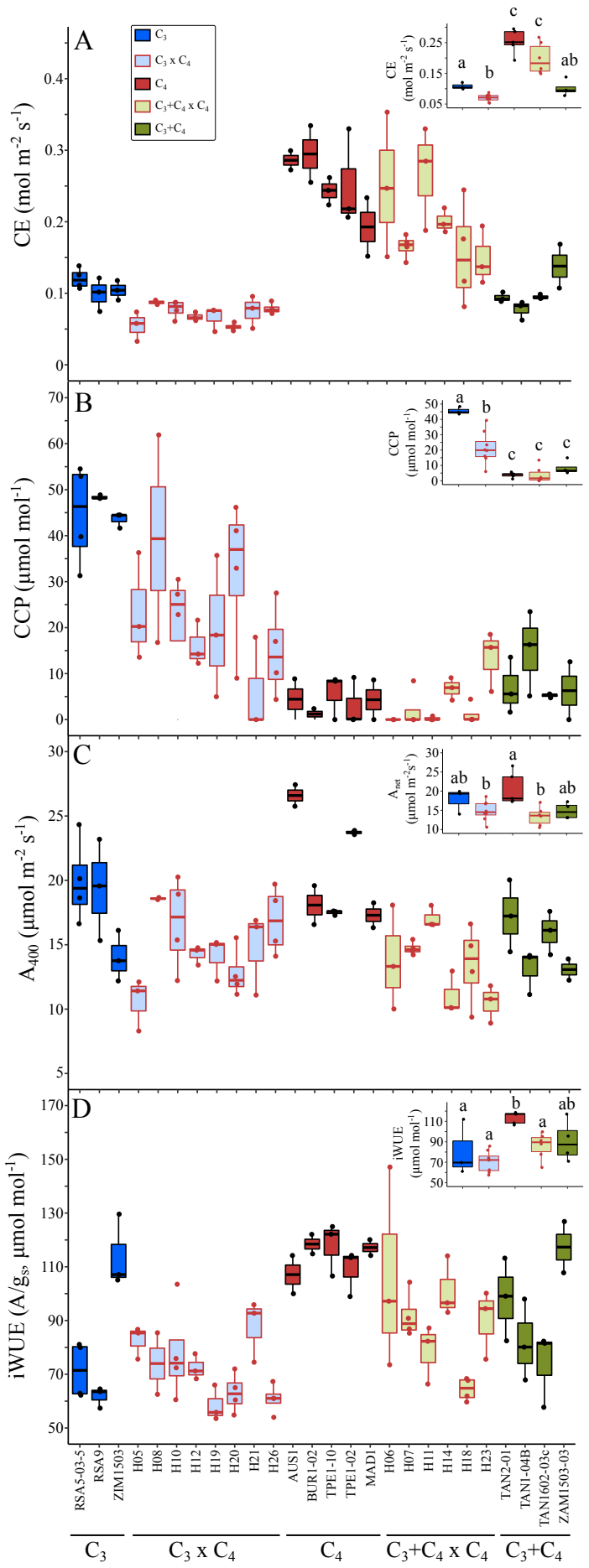

Fig. 5. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C) net photosynthetic rate (A400), and (D) intrinsic water use efficiency (iWUE, *A*/*gs*) at reference CO_2 = 400 µmol mol^{-1}. Different lower case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; n ≥ 4).