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, Lígia T. Bertolino, Colin P. Osborne, Pascal-Antoine Christin

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

* Corresponding author: m.bianconi@sheffield.ac.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 (Simpson et al. In press).

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; Sultmanis 2018; Lin et al. 2021; Simpson et al. In press). 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 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, and 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 31 seedlings from seven mother plants (two C_3,

102 two C_3+C_4 and three C_4) were obtained. Subsequent genotyping of the seedlings identified multiple
103 F1 hybrids between parents with distinct photosynthetic types, including 14 C_3 x C_4 and six C_3+C_4 x
104 C_4 individuals (Table S1). The remaining seedlings consisted of 10 C_4 x C_4 crosses, and an
105 additional C_3+C_4 plant resulting from self-pollination (Table S2). Five C_3 x C_4 individuals grew to
106 maturity and were sampled for DNA analyses but died before being phenotyped, and another two
107 died before being sampled for RNA sequencing and leaf gas-exchange analyses (Table S1). For the
108 phenotypic analyses, we included the mother plants, the known and possible pollen parents (i.e.
109 similar C_4 accessions that could not be distinguished by the genotyping analyses; see Table S1), and
110 seven additional accessions of *A. semialata* (three C_3, two C_3+C_4 and two C_4) that were growing
111 alongside them were added to increase the phenotypic diversity within photosynthetic types (Table
112 S1). All plants were grown in 11-L, free-draining pots containing a 2:1 mix of M3 compost
113 (Levington) and perlite (Sinclair) under well-watered conditions, and were fertilized once every
114 three months with a NK fertilizer 16-0-5 containing Iron (Evergreen Extreme Green Lawn Food).
115 Plants were grown on a 12-hour day/night cycle, with metal halide lamps providing supplementary
116 light (additional 200 µmol m^{-2} s^{-1} at bench level), 25/20°C day/night temperature, ambient CO_2
117 concentration and relative humidity between 30% and 60%.
118

119 *Genotyping*

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

137 alternative bases was used to narrow down the pollen parent to a genetic lineage (Table S2). The
138 hybrid origin of the plants was subsequently confirmed using RNAseq data (see below).
139

140 *Stable carbon isotopes*

141 Carbon isotope composition was determined from the central portion of fully expanded leaf blades.
142 Samples were dried in silica gel, and subsequently ground to a fine powder using a TissueLyzer II
143 (Qiagen). Carbon isotope analysis was conducted on 1-2 mg of leaf sample using an ANCA GSL
144 preparation module coupled to a Sercon 20-20 stable isotope ratio mass spectrometer (PDZ Europa,
145 Cheshire, UK). Carbon isotopic ratios ($\delta^{13}C$, in ‰) were reported relative to the standard Pee Dee
146 Belemnite (PDB). Values of $\delta^{13}C$ higher than -16‰ indicate that the plants grew using C_4
147 photosynthesis (O'Leary 1988; Stata et al. 2019). $\delta^{13}C$ values of additional accessions of *A.*
148 *semialata* that were previously phenotyped and identified as C_3, C_3+C_4 or C_4 (Lundgren et al. 2016,
149 2019; Dunning et al. 2017, 2019a) were retrieved from Lundgren et al. (2015, 2016).
150

151 *Leaf anatomy*

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

172 (R Core Team 2020).
173
174 *Leaf transcriptome*
175 Leaf mRNA was isolated and sequenced as previously described (Dunning et al. 2019a). In short,
176 the distal halves of fully expanded leaves were sampled in the middle of the light period, flash-
177 frozen in liquid N_2 and kept at -80°C. All samples were collected on the same day. Total RNA was
178 extracted using the RNeasy Plant Mini Kit (Qiagen, Germany) with an on-column DNA digestion
179 step (RNase-Free DNase Set; Qiagen). A total of 20 RNA-Seq libraries (one per individual) were
180 prepared with the TruSeq RNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA) using
181 0.5 µg of starting RNA and aiming at a median insert size of ~ 155 bp (standard fragmentation
182 protocol). Libraries were paired-end sequenced (read length = 100 bp) on 1/24 of a single lane of an
183 Illumina HiSeq 2500 flow cell in rapid mode (with four additional samples from an unrelated
184 project) at the Sheffield Diagnostic Genetics Service. Raw sequence reads were filtered to remove
185 adaptor contamination and low-quality reads (i.e. < 80% of bases with Phred score > 20) using
186 NGSQCToolkit v. 2.3.3 (Patel and Jain 2012). Reads were further trimmed from the 3' end to
187 remove bases with Phred score < 20. The quality of the filtered data was assessed using FastQC v.
188 0.11.9 (Andrews 2010).
189 Transcript abundance was quantified by mapping the filtered reads to a reference dataset
190 consisting of coding sequences from multiple transcriptomes of *A. semialata* retrieved from
191 Dunning et al. (2017) and modified by Bianconi et al. (2018). In short, this dataset consists of 5,540
192 groups of homologous genes that are common to the Panicoideae grasses (the subfamily that
193 includes *Alloteropsis*), each group containing all paralogs detected in the transcriptomes of *A.*
194 *semialata* (total of 12,234 groups of co-orthologs). We used this dataset as it includes manually
195 curated sets of co-orthologs of 23 gene families known to have a function in the C_4 biochemistry
196 (Bianconi et al. 2018). This curated gene set increases the read mapping accuracy when paralogs
197 exist with a high sequence similarity. Furthermore, this dataset includes the sequences of paralogs
198 of core C_4 genes that are absent from the reference genome of *A. semialata*, such as four different
199 genes for the enzyme phosphoenolpyruvate carboxylase (PEPC) known to be highly expressed in
200 other *A. semialata* accessions (Dunning et al. 2017, 2019a). Filtered reads were mapped to the
201 reference dataset using Bowtie2 v. 2.3.5 (Langmead and Salzberg 2012) with default parameters.
202 Read counts are reported here in reads per million mapped reads (RPM). To visualize the diversity
203 among samples, a PCA on the global patterns of gene expression was conducted using the function
204 *prcomp* in R after log2-transforming the RPM values. Only genes with transcript abundance > 10
205 RPM in at least five samples were used for the PCA. To test whether the global patterns of gene
206 expression were consistent when the whole gene set of *A. semialata* was considered, we repeated

207 the analysis using the chromosome-level genome assembly of this species (Dunning et al. 2019b) as
208 reference for read mapping. Finally, to verify whether the expression level differences between
209 photosynthetic types were consistent with previous reports for *A. semialata*, we quantified transcript
210 abundance (using the co-orthologous gene set as reference) for 20 additional accessions (four C_3,
211 six C_3+C_4 and 10 C_4) retrieved from two previous RNAseq studies (Dunning et al. 2017, 2019a).
212 To explore the global differences in gene expression between hybrids and the parental types, we
213 performed a differential expression analysis on the dataset that was generated using the complete
214 genome of *A. semialata* as reference for read mapping. Here, due to insufficient number of
215 replicates for C_3 and C_3+C_4 plants, we restricted our analyses to the comparisons between the two
216 hybrid types, and between each of these and the C_4 group. We then investigated the main metabolic
217 functions associated with the set of differentially expressed genes from each comparison (see full
218 description of the global RNAseq analyses on the Supplementary Methods).
219 Finally, to confirm the paternity of the putative hybrids, we genotyped the plants using the
220 RNAseq dataset mapped to the reference genome of *A. semialata*. Here we also included
221 individuals of *A. semialata* from the RNAseq studies of Dunning et al. (2017, 2019a) to increase the
222 genetic diversity of the dataset. Read alignment files were filtered to remove duplicates using Picard
223 Tools v. 1.102 (https://broadinstitute.github.io/picard/), and variants were called on all individuals
224 combined using BCFtools v. 1.9 (Li et al. 2009). Variant sites were filtered using VCFtools v. 0.1.17
225 (Li et al. 2009) to remove low quality genotype calls (i.e. quality score < 30 and read depth < 3),
226 indels, and sites with more than 10% missing data, which resulted in an initial set of 521,929 single
227 nucleotide polymorphisms (SNPs). BCFtools was then used to retain only multiallelic sites in which
228 all hybrids were heterozygous and all potential parents were homozygous, which resulted in 15,406
229 SNPs that were then used for the genotyping tests. First, to determine the type of cross (i.e. whether
230 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
231 photosynthetic type and exclusive to them. Sites with more than 25% missing data within each
232 photosynthetic group were removed, and a total of 114 SNPs were retained (Table S4). We then
233 tested, for each SNP, whether hybrid individuals had one C_4 allele and one allele that was either C_3
234 or C_3+C_4, according to the type of cross (Table S4). With the confirmation of the type of cross, we
235 then narrowed down the genetic lineage of the pollen parent (as the mother is already known). Only
236 one individual representing each genetic lineage was retained for this analysis (Table S5). First, we
237 used the 15,406 SNP set to identify all singletons of each potential pollen parent, which in this case
238 were defined as homozygous genotypes that were unique to a single individual among all
239 individuals with the same photosynthetic type. A total of 7,779 singletons were identified (minimum
240 of 287 per potential pollen parent; Table S5). We then verified, for each singleton, which hybrid
241 individuals carried that same allele, and counted the number of positive matches (Table S5). All

genotyping analyses were conducted using R (scripts are provided as Supplementary Files).

Leaf gas-exchanges

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

Stomatal density on the abaxial side of the leaves was quantified using leaf impressions. Dental resin (ImpressPlus Wash Light Body, Perfection Plus Ltd., Totton, UK) was applied to the central portion of fully expanded leaves, and clear nail varnish was applied to the set resin impression. Images were captured from the nail varnish impressions at 20 X using an Olympus BX51 microscope coupled to an Olympus DP71 camera, and stomatal density was quantified on 0.38 mm^2 fields using ImageJ.

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

Gene expression

Global gene expression patterns were assessed via a PCA on 7,482 genes (Fig. 3A). The first principal component mostly separated three sister C_3+C_4 x C_4 plants (H06, H11, and H23) from the

311 other individuals, and accounted for 24.7% of the variation. Photosynthetic types were clearly
312 separated by the second principal component, which accounted for 16.8% of the variation and
313 placed the hybrids as intermediate to the photosynthetic types of their parents. Similar clustering
314 patterns were observed when transcript abundance was quantified for the complete gene set
315 extracted from the genome of *A. semialata* (Fig. S2).

316 Transcript abundance of five genes encoding key enzymes for the C_4 biochemistry, and which
317 are known to be upregulated in C_4 plants of *A. semialata* (Dunning et al. 2019a), were analysed in
318 detail (Fig. 3B,C; Table S7). The patterns of transcript abundance of C_3, C_3+C_4 and C_4 plants are
319 highly similar to those reported by Dunning et al. (2017, 2019a), where a larger sample size was
320 used (Fig. S3). Transcript abundance of the genes encoding the C_4-specific forms of the enzymes
321 aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), NADP-malic enzyme (NADP-ME, gene
322 *nadpme-1P4*) and pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*) were four- to 17-fold
323 higher in C_4 than in C_3 plants, with the transcript abundances of the C_3+C_4 plant slightly above those
324 of C_3 plants (Fig. 3B). In each case, the transcript abundance of the hybrids was intermediate to the
325 abundance of the photosynthetic types of the parents, showing that these expression patterns are
326 mostly additive. C_3 accessions of *A. semialata* lack the gene copy encoding phosphoenolpyruvate
327 carboxykinase (PCK) that is highly expressed in C_4 and C_3+C_4 accessions (*pck-1P1_LGT*; Olofsson
328 et al. 2016); the transcript abundance of this gene reaches 20% of the mean C_4 values in C_3 x C_4
329 hybrids, and 40% in C_3+C_4 x C_4.

330 Out of the nine genes encoding the key C_4 enzyme phosphoenolpyruvate carboxylase (PEPC)
331 that have been identified in *A. semialata* accessions (Dunning et al. 2017), only five were reported
332 to be highly expressed in leaves of some C_4 accessions (Christin et al. 2012; Dunning et al. 2017,
333 2019a). One of these genes, *ppc-1P3_LGT:A* is only present in the C_4 lineage from Oceania, which
334 is represented here by an Australian accession (AUS1) that is the pollen parent of two C_3 x C_4
335 hybrids (H19 and H21; Table S5); both plants have expression levels of this gene that are similar to
336 the Australian parent (Fig. 3C). Only one of the other two laterally-acquired PEPC copies is present
337 in the plants analysed here (*ppc-1P3_LGT:M*), and it is highly expressed (> 10,000 RPM) in four
338 C_3+C_4 x C_4 individuals that are offspring from the C_4 accessions from Burkina Faso or Madagascar
339 (H06, H11, H14 and H23; Table S5; Fig. 3C), which have been previously shown to be part of the
340 only C_4 lineage to carry this gene (Olofsson et al. 2016). The expression levels of *ppc-1P3* are
341 similar across C_3+C_4 and C_4 accessions, and are on average ten-fold higher than in C_3 accessions
342 (except the C_4 AUS1, which has a non-functional copy of this gene; Olofsson et al. 2016); in the
343 C_3+C_4 x C_4 hybrids, however, the transcript abundance of this gene is on average 42% higher than
344 the median of C_3+C_4 and C_4 accessions (Fig. 3C). Finally, the gene *ppc-1P6* reaches high expression
345 levels in C_4 accessions (except in AUS1, where it is absent; Dunning et al. 2019b), but is

346 disproportionally upregulated in the C_4 Taiwanese lineage (TPE; Dunning et al. 2017), where it
347 reaches levels that are up to 100-fold higher than in C_3 plants, and at least twice higher than the
348 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
349 and C_3+C_4 x C_4 offspring of TPE, mostly ranging from 10,000 to 60,000 RPM, except for one
350 C_3+C_4 x C_4 which has a transcript abundance similar to TPE (~ 100,000 RPM).
351 We examined transcript abundance of selected protein-coding genes with a role in the
352 photorespiratory metabolism. We observed reduced transcript levels of the genes encoding the
353 peroxisomal enzymes flavin mononucleotide-dependent glycolate oxidase (GOX, gene *glo-1P2*)
354 and glutamate:glyoxylate aminotransferase (GGT, gene *ggat-1P6*) in C_3+C_4 x C_4 hybrids relative to
355 C_3 and C_3+C_4 accessions, with average values that are similar to or below those of C_4 accessions
356 (Fig. S4). The genes encoding the proteins P-, T-, L- and H- that compose the mitochondrial
357 multienzyme system glycine decarboxylase (GDC, genes *gldp1P*, *gcvt-1P1*, *lpd-1P2*, and *gdh-2P2*,
358 respectively) had on average 63% lower transcript levels in C_4 than in C_3 accessions; in C_3 x C_4 and
359 C_3+C_4 x C_4 hybrids, transcript levels of these four genes were on average 12% and 33% lower than
360 in C_3, respectively, suggesting additivity (Fig. S4; Table S8). Multiple genes encoding a serine
361 hydroxymethyltransferase (SHMT) are found in the genome of *A. semialata* (genes *shm-1P1*, *shm-*
362 *2P2*, *shm-3P*3 and *shm-3P4*; Dunning et al. 2019a), but none had patterns of transcript abundance
363 similar to those observed for GDC; in fact, one of the genes (*shm-3P*3) had higher average
364 transcript abundance in C_4 than in C_3 accessions. Finally, the chloroplast genes glycerate 3-kinase
365 (GLYK, gene *glyk-1P1*) and 2-phosphoglycolate (2-PG) phosphatase (PGLP, gene *pglp-2P2*) had
366 overlapping transcript abundance ranges across hybrid and photosynthetic types. Finally, a gene
367 encoding a GOLDEN2-LIKE transcription factor that has been previously demonstrated to be
368 involved in the redistribution of organelles among cell types (Wang et al. 2017) is expressed at
369 higher levels in C_4 than in C_3 and C_3+C_4 accessions (Fig. S4L; Table S8). This gene is at
370 intermediate levels in the C_3 x C_4 hybrids, but largely variable among C_3+C_4 x C_4 individuals.
371 We identified genes differentially expressed (DE) between hybrid types, and between these and
372 C_4 accessions (Table S9). We found a total of 4,104 genes differentially expressed between C_3 x C_4
373 and C_3+C_4 x C_4 hybrids (adjusted p-value < 0.05), out of which 21 were related to C_4 photosynthesis
374 and four to photorespiration (Table S9). This complete DE gene set is reduced to 206 when only
375 higher count genes (base mean > 500) and with larger differences (log2fold change > 1) are
376 considered (Table S9). Among this reduced gene set, we found seven C_4-related genes, but none
377 related to photorespiration (Fig. S5; Table S9). The individual comparisons between C_3 x C_4 or
378 C_3+C_4 x C_4 and C_4 accessions identified 1,925 and 2,727 significant DE genes, respectively. These
379 numbers are reduced to 110 and 181 when the same filters described above are applied (Table S9).
380 In this reduced set, we found 13 C_4-related genes that were differentially expressed between hybrids

381 and C_4 accessions, three of these common to both C_3 x C_4 vs C_4, and C_3+C_4 x C_4 vs C_4 comparisons,
382 namely PCK (*pck-1P1_LGT*), NADP-malate dehydrogenase (NADP-MDH, gene *nadpmdh-1P2*),
383 and soluble inorganic pyrophosphatase (PPA, gene *ppa-2P1*). While *nadpmdh-1P2* was
384 downregulated in the C_4 group, the two others had increased transcript abundance in C_4 relative to
385 both hybrid types (Fig. S5). Finally, we found four genes related to photorespiration in the DE gene
386 set between C_3 x C_4 and C_4 accessions, three of which were upregulated in C_3 x C_4, namely the
387 genes encoding the GDC proteins T- (*gcvt-1P1*), P- (*gldp1P*) and L- (*lpd-1P2;* Fig. S5). The
388 remaining gene, which encode the enzymes SHMT (*shm-3P*3) was downregulated in both C_3 x C_4
389 and C_3+C_4 x C_4 hybrids relative to C_4 accessions. Gene ontology (GO) analyses did not detect any
390 significantly enriched cellular processes for any of the full DE gene sets. However, the reduced DE
391 gene set from the comparison between hybrid types was significantly enriched with genes related to
392 the photosystems I and II, oxidoreductase activity and 1,3-beta-D-glucan synthesis (Table S10). In
393 the reduced gene set from comparisons between each of the hybrid types and the C_4 type, GO
394 analyses indicated a significant enrichment for genes associated with C_4 photosynthesis (particularly
395 malate metabolism), transmembrane transport, and photosystems I and II (Table S10).
396 Overall, the analyses of expression profiles show that the inheritance of expression levels of C_4
397 genes and those involved in the photorespiratory pathway is largely additive. In almost all cases, the
398 transcript abundance of the hybrids ranks between those of the parental lineages.
399

400 *Evidence of reduced photosynthetic performance in the hybrids*

401 Photosynthetic performance was assessed from CO_2 response curves (A/Ci) and steady-state
402 measurements at ambient CO_2 (Fig. S6; Tables S11 and S12). Photosynthetic rates in the hybrids
403 ranked between the parents at low Ci levels, but the differences between hybrids and parents were
404 progressively reduced as CO_2 levels increased, until they were equal to their non-C_4 parents at Ci of
405 100 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
406 on average similar values of maximum carboxylation efficiency (CE), and these were 57% lower
407 than in C_4 accessions (Fig. 5A; Table S11). Despite the larger variation among individuals, C_3+C_4 x
408 C_4 hybrids had CE values that were on average 83% and 96% higher than in C_3 and C_3+C_4
409 accessions, respectively, and these were not significantly different from the average values of C_4
410 accessions. In C_3 x C_4 hybrids, however, CE was consistently lower than in C_3 accessions, and
411 reached the lowest mean values of all photosynthetic types/crosses. The CO_2 compensation point
412 (CCP) ranged between 43 and 48 µmol mol^{-1} in C_3 accessions, and between 5 and 15 µmol mol^{-1} in
413 C_3+C_4 individuals (Fig. 5B). All C_4 had CCP below 10 µmol mol^{-1}. In the hybrids, CCP was below
414 13 µmol mol^{-1} in C_3+C_4 x C_4 individuals, and ranged between 6 and 40 µmol mol^{-1} in C_3 x C_4
415 hybrids. CCP estimates varied substantially among A/Ci curves in C_3 x C_4 individuals (Fig. 5B), but

416 mean values were 52% lower than in C_3 plants.

417 The steady-state net CO_2 assimilation rate at ambient CO_2 (400 µmol mol^{-1}; A_{400}) was highest in
418 two C_4 accessions, but the differences between C_3, C_3+C_4 and C_4 plants were not significant (Fig.
419 5C). C_3+C_4 x C_4 hybrids had the lowest A_{400}, but mean values were not significantly different from
420 C_3 x C_4, C_3 and C_3+C_4 accessions. Stomatal conductance (g_s) had the highest median values in C_3
421 accessions, but the differences between groups were not significant (Fig. S7A). C_3 accessions had
422 on average 50% higher stomatal density than C_3+C_4 and C_4 accessions and their hybrids (Fig. S7B).
423 We found a positive correlation between g_s and stomatal density ($R^2 = 0.31$, $p < 0.01$), but only after
424 excluding the outlier C_3 accession ZIM1503, which had the highest stomatal density, but one of the
425 lowest g_s. The correlation between stomatal density and vein density was only marginally
426 significant ($R^2 = 0.01$, $p = 0.06$). The intrinsic water use efficiency (iWUE) at ambient CO_2 was on
427 average 38% higher in C_4 than in other accessions, although a few C_3, C_3+C_4 and C_3+C_4 x C_4
428 individuals also showed C_4-like values (Fig. 5D). The lowest iWUE values were observed in C_3 and
429 C_3 x C_4 individuals, but the means were not significantly different from C_3+C_4 and C_3+C_4 x C_4
430 hybrids. To investigate whether changes in leaf temperature (T_{leaf}) could explain the observed
431 patterns, we inspected the variation in T_{leaf} during the A/Ci curves. Although T_{leaf} values up to 3ºC
432 above the median (26.3ºC) were observed in a few accessions, 81% of all data points were collected
433 within a 2ºC interval (Fig. S8). There were no significant differences between groups when all
434 points were considered, nor when only above- or below-ambient CO_2 levels were analysed (Fig.
435 S8). Similar patterns among cross/photosynthetic types were observed in all photosynthetic
436 parameters after removing individuals for which the median T_{leaf} was 1ºC above the median of all
437 curves (Fig. S9).

438 Overall, these results show that the physiological characters are not consistently additive in the
439 hybrids, with some individuals performing below their two parents for several traits.

440

441 **Discussion**

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

443 In this study, we analyse the phenotypes of hybrids between individuals of the grass *Alloteropsis*
444 *semialata* with different photosynthetic types. The C_3 x C_4 quantitatively resemble the naturally
445 occurring C_3+C_4 in most anatomical and gene expression traits (Figs 2 and 3), as previously noticed
446 (Kadereit et al. 2017). They however differ qualitatively, and minor veins and high expression of
447 some C_4-related genes are especially restricted to C_3 x C_4 individuals (Fig. 3; Table S6). These
448 findings therefore provide additional support for the hypothesis that naturally occurring C_3+C_4 in *A.*
449 *semialata* do not result from recent crosses between photosynthetic types (Lundgren et al. 2015,
450 2016), although genome analyses support a role for introgression in their distant history (Bianconi

451 et al. 2020).

452 Across all crosses, we show that hybrids between C_4 and non-C_4 individuals of the grass *A.*
453 *semialata* do not exhibit a full C_4 physiology, as indicated by their $\delta^{13}C$, which are clearly outside
454 the range of C_4 values in the species (Fig. 1). This indicates that a significant fraction of CO_2
455 assimilation in these hybrids occurs directly via Rubisco in the mesophyll, which implies that the
456 genetic contribution of the C_4 parent was insufficient to 1) completely suppress Rubisco expression
457 and/or activity in the mesophyll cells, and/or 2) create an effective C_4 cycle. However, the CCP of
458 the C_3 x C_4 hybrids is substantially reduced compared to the C_4 parents (Fig. 5B). While this change
459 could theoretically result from an increase of the photorespiratory shuttle observed in many C_3-C_4
460 plants (called 'C_2' plants; Keerberg et al. 2014; Khoshravesh et al. 2016), the expression of
461 photorespiratory is rather unchanged or decreased in our hybrids compared to the C_4 parents.
462 Changes of CCP are therefore more likely linked to an increase of the C_4 activity, which is further
463 supported by significantly higher $\delta^{13}C$ values compared to their non-C_4 parents (Fig. 1). Our results
464 therefore suggest that, in most hybrids, a fraction of the CO_2 that enters the leaf is fixed by PEPC in
465 the mesophyll and follows the C_4 route to be refixed by Rubisco in the inner bundle sheath.
466 Furthermore, given the C_4-like CCP and CE values, and $\delta^{13}C$ above -20‰ (in most cases), the
467 degree of C_4 activity is higher in C_3+C_4 x C_4 than in C_3 x C_4 hybrids. Our findings therefore
468 recapitulate what has been shown for interspecific hybrids between photosynthetic types in several
469 eudicot and monocot genera (Björkman et al. 1969; Brown et al. 1985; Araus et al. 1990; Byrd et al.
470 1992; Brown and Bouton 1993; Oakley et al. 2014; Sultmanis 2018). The causes for this only
471 partial contribution of the C_4 pathway to the overall carbon assimilation in the hybrids are
472 potentially the same in all these cases, and are probably related to the insufficient expression of
473 components of C_4 metabolism (Brown and Bouton 1993).

474 Two interrelated leaf anatomical traits were significantly different from the C_4 individuals in
475 both C_3 x C_4 and C_3+C_4 x C_4 hybrids, namely the proportion of inner sheath to mesophyll tissue
476 (IS/M) and the distance between consecutive bundle sheaths (BSD). These traits are mostly
477 determined by the presence of minor veins in C_4 leaves, which has been previously shown to be the
478 major anatomical difference between C_4 and non-C_4 *A. semialata* (Lundgren et al. 2019). Here,
479 significant changes towards the C_4 phenotype in IS/M and BSD were observed in the hybrids, but
480 realized values were still far from those observed in C_4 accessions (Fig. 2C,D). This is particularly
481 clear for IS/M, where values were below the mean of both parents in the two hybrid types. While
482 minor veins are observed in the hybrids, their number per segment remains below that observed in
483 C_4 accessions (Table S6). This indicates that the proliferation of minor veins is a continuous trait,
484 which can sustain a C_4 cycle only over a threshold in this species. Since CO_2 refixation by Rubisco
485 takes place in the IS cells in C_4 *A. semialata* (Ellis 1974b; Ueno and Sentoku 2006), the leaf

486 anatomy of the hybrids might prevent an optimal coupling between carbon assimilation and
487 reduction reactions, therefore decreasing the efficiency of the C_4 pathway. The distribution of
488 organelles among cell types was not quantified here, but a gene previously linked to the relocation
489 of organelles to BS cells (Wang et al. 2017) is more highly expressed in C_4 accessions and
490 upregulated in some hybrids compared to their non-C_4 parents, which might suggest that
491 mitochondria are partially relocated to their BS. The size of the OS cells might also play a role in
492 creating an effective C_4 cycle, as suggested by the large differences between C_3 and C_4 accessions
493 (Fig. 2E). Few other C_4 species have an extra layer of BS cells outside of the BS cells containing
494 chloroplasts, and the presence of an OS may represent an increased resistance to the C_4 acid shuttle
495 (Lundgren et al. 2014). In the grass genus *Neurachne* that presents a similar C_4 anatomy, the
496 smallest OS are also observed in C_4 individuals, with intermediate species presenting intermediate
497 OS cell sizes (Khoshravesh et al. 2020), supporting the idea that OS cells must be reduced during C_4
498 evolution. Here, the OS cells were significantly reduced in the hybrids relative to their non-C_4
499 parents, with most C_3+C_4 x C_4 individuals and a few C_3 x C_4 having OS cell sizes within the C_4
500 range (Fig. 2F), which suggests that this character might have played a minor role, if any, in
501 reducing the efficiency of the C_4 acid shuttle in the hybrids.
502 C_4 activity could also have been limited by gene expression and protein activity. Despite
503 significant increases in the transcript abundance of core C_4 genes, particularly the genes encoding
504 PEPC (Fig 4), it is possible that the insufficient expression levels of some genes might have reduced
505 the efficiency of the C_4 cycle, for example by decreasing the rate of carboxylation and/or
506 decarboxylation reactions. Furthermore, the lack of proper compartmentation of the
507 expression/activity of these genes due to a non-tissue-specific expression of C_3 or C_3+C_4 alleles
508 could have impaired the formation of intermediate metabolite pools, therefore preventing the
509 operation of an efficient C_4 cycle (Brown and Bouton 1993; Ermakova et al. 2020). Finally, the lack
510 or insufficiency of post-transcriptional and post-translational modifications might also have affected
511 the realized enzyme activity. This might be particularly relevant in the case of hybrids, where the
512 cell environment may be substantially modified by pleiotropic effects emerging from the expression
513 of two divergent genomes. Detailed analyses of cellular localization and activity of C_4 enzymes
514 coupled to CO_2 labelling studies are necessary to identify the major determinants of the phenotype
515 observed in the hybrids reported here.
516

517 *Photosynthetic performance of plants expressing partial C_4 traits and implications*
518 *for the model of C_4 evolution*

519 The current model of C_4 evolution assumes that the sequential acquisition of certain features, such
520 as Rubisco activity in bundle sheath cells and upregulation of core C_4 enzymes, increases

521 photosynthetic output and consequently fitness in an environment that favours the C_4 physiology
522 (Monson and Moore 1989; Sage 2004; Heckmann et al. 2013; Malmmann et al. 2014). However,
523 this Mt. Fuji-like fitness landscape (Heckmann et al. 2013) is conditional on the presence of
524 anatomical enablers; when these are lacking, the C_4 phenotype is not accessible because
525 intermediate states have lower fitness than the C_3 state (Heckmann 2016). The two groups of
526 intraspecific hybrids generated in this study provide an opportunity to examine this prediction while
527 controlling for phylogenetic effects, with C_3 x C_4 and C_3+C_4 x C_4 plants being proxies for the
528 upregulation of C_4 traits at an early and intermediate stage, respectively, along C_4 evolution. Here,
529 we show that the carboxylation efficiency of C_3 x C_4 hybrids is lower than that of both parents (Figs
530 5, 6A), despite significant changes towards the C_4 phenotype in leaf anatomy and gene expression.
531 In C_3+C_4 x C_4 hybrids, however, carboxylation efficiency was increased in comparison to the C_3+C_4
532 parent, and was not significantly different from C_4 accessions. Such disparate effects of
533 upregulating C_4 components in a non-C_4 background are compatible with the initial prediction that
534 fitness gains are conditional on the presence of enabling factors, which here might have been
535 present in C_3+C_4 parents, but lacking in the C_3 plants. Such changes would have acted as a switch
536 that permitted a full C_4 pathway to operate, and therefore increase photosynthetic performance.
537 Note however that such gains in photosynthetic performance are restricted to conditions of low CO_2
538 concentration in the leaf (Fig. 4). In fact, at ambient CO_2 levels, several individuals of both hybrid
539 groups had lower photosynthetic rates than both parents, suggesting that the partial upregulation of
540 C_4 components, all at once, might have negative consequences, possibly due to pleiotropic effects
541 (e.g. C_4 enzymes competing for reducing power and ATP with other cell reactions). This in turn
542 supports the idea that the changes towards a full C_4 physiology must be built sequentially upon
543 preexisting characteristics that enable the next change to provide a fitness advantage (Heckmann
544 2016). The fertility of the F1 hybrids reported here is not known yet, although evidence of rare gene
545 flow between photosynthetic types in the wild suggest that backcrossing is possible (Olofsson et al.
546 2016, 2021; Bianconi et al. 2020). If a F2 population can be produced, characterization of the
547 photosynthetic performance of such individuals expressing C_4 components independently,
548 particularly under conditions expected to favour C_4 plants, including long-term exposure to low CO_2
549 levels, should provide further insights on whether C_4 evolution must follow a particular sequence of
550 events for it to be viable.
551

552 **Conclusions**

553 In this study, we characterize the phenotype of hybrids between different photosynthetic types in the
554 grass *Alloteropsis semialata* to investigate how the upregulation of components of the C_4 trait
555 affects photosynthetic performance. We show that the hybrids have in most cases anatomical traits

556 and gene expression patterns that are intermediate between those of the parents, and this leads to C_4
557 activity that is equally intermediate between the two parents. The physiological benefits of a partial
558 C_4 metabolism in the hybrids appear only at low CO_2 levels; at ambient CO_2, there is no evidence of
559 enhanced photosynthetic performance in the hybrids relative to their non-C_4 parents. Some hybrid
560 individuals in fact perform worse than both parents at ambient CO_2, and this possible hybrid
561 depression could explain the lack of a clear hybrid zone in regions where the distributions of C_4 and
562 non-C_4 *A. semialata* lineages overlap. Overall, our results support the hypothesis that photosynthetic
563 gains arising from the upregulation of C_4 features are conditional on coordinated changes in leaf
564 anatomy and biochemistry. Future studies with the F2 offspring, where the different C_4 components
565 can be segregated, will be able to pinpoint the key genetic and phenotypic changes that lead to
566 fitness gains, providing a unique opportunity to experimentally test long-standing hypotheses about
567 the evolution of C_4 photosynthesis.
568

569 ## Supplementary data

570 **Fig. S1.** Leaf anatomical variables measured in this study.
571 **Fig. S2.** Principal component analysis on 11,988 genes from the chromosome-level genome
572 assembly of *Alloteropsis semialata.*
573 **Fig. S3.** Transcript abundance in reads per million (RPM) of selected genes encoding core C_4
574 enzymes in *Alloteropsis semialata*: (A) aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), (B)
575 NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*), (C) phosphoenolpyruvate
576 carboxykinase (PCK, gene *pck1P1_LGT*), (D) pyruvate phosphate dikinase (PPDK, gene *ppdk-*
577 *1P2*), and (E-I) phosphoenolpyruvate carboxylase (PEPC) genes. Transcript abundance computed
578 for *A. semialata* samples extracted from Dunning et al. (2019a) are shown on the right for each
579 gene.
580 **Fig. S4.** Transcript abundance of selected gene families with a role in photorespiration (A-K), and
581 the gene encoding the transcription factor GOLDEN2-like (L) in F1 hybrids and the parental
582 photosynthetic types in *Alloteropsis semialata*. (A) 2-phosphoglycolate (2-PG) phosphatase
583 (PGLP), (B-C) flavin mononucleotide (FMN)-dependent glycolate oxidase (GOX), (D)
584 glutamate:glyoxylate aminotransferase (GGT), (E-H) glycine decarboxylase (GDC) complex
585 proteins -H, -P, -T and -L, (I-J) serine hydroxymethyltransferase (SHMT), (K) glycerate 3-kinase
586 (GLYK). Transcript abundance in reads per million mapped reads (RPM).
587 **Fig. S5.** Heat map of differentially expressed genes related to C_4 photosynthesis and
588 photorespiration. Asterisks indicate significant DE genes ($p < 0.05$) for comparisons between hybrid
589 types (C_3 x C_4 vs C_3+C_4 x C_4) and between each hybrid type and the C_4 type. Only DE genes with at
590 least two-fold change and base count > 500 are shown. Count data was transformed using the VST

591 function of DESeq2 and scaled by row. See table S9 for full gene annotation

592 **Fig. S6.** Photosynthetic response to intercellular CO_2 (*A*/*Ci*) of F1 hybrids and the parental
593 photosynthetic types in *Alloteropsis semialata*. Panels contain all *A*/*Ci* curves collected for each
594 accession, with individual curves coloured with different shades of grey. Accession name and
595 cross/photosynthetic type are indicated on the top left. Vertical dashed lines indicate the CO_2
596 compensation point of each curve.

597 **Fig. S7**. Leaf stomata variables. (A) Steady-state stomatal conductance (g_s) at 400 µmol mol^{-1} (n =
598 2-4 leaves per accession). (B) Stomatal density on the abaxial side of the leaves (n = 2 leaves per
599 accession, with stomata counts averaged from 5 fields per leaf; field area = 0.38 mm^2). Data points
600 on the right are the means per accession within each cross/photosynthetic type, and different lower-
601 case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD;
602 $n \geq 3$).

603 **Fig. S8.** Leaf temperature variation during *A*/*Ci* curves. Data points are individual *A*/*Ci*
604 measurements for each plant (left panels) or with individuals grouped into cross/photosynthetic
605 types (right panels). (A) All measurements, and measurements collected at reference CO_2 (B) < 100
606 µmol mol^{-1}, (C) between 100 and 400 µmol mol^{-1}, and (D) > 400 µmol mol^{-1}. Arrows indicate
607 outlier individuals that were removed for the analysis in Fig. S9.

608 **Fig. S9**. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in
609 *Alloteropsis semialata* after removing outlier accessions (i.e. with T_{leaf} 1ºC above the median). (A)
610 Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C)
611 net photosynthetic rate (A_{400}), and (D) intrinsic water use efficiency (iWUE, A/g_s) at reference CO_2
612 = 400 µmol mol^{-1}.

613

614 **Table S1.** Sample information.

615 **Table S2.** Genotyping using PCR/Sanger-sequencing.

616 **Table S3.** List of primer sequences used for genotyping.

617 **Table S4.** RNA-seq markers selected for genotyping analysis I (photosynthetic type).

618 **Table S5.** RNA-seq markers selected for genotyping analysis II (pollen parent).

619 **Table S6.** Raw leaf anatomy data.

620 **Table S7.** Normalized transcript abundance for the co-ortholog gene set.

621 **Table S8.** Normalized transcript abundance for the *A. semialata* complete genome gene set.

622 **Table S9**. Differential expression analysis between F1 hybrids and the C_4 parental type.

623 **Table S10**. Gene ontology enrichment analyses on genes differentially expressed between F1
624 hybrids and the C_4 parental type.

625 **Table S11.** Raw A/Ci data.

Table S12. Estimated photosynthetic parameters and steady-state measurements.

Author contributions

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

Acknowledgements

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

Data availability

Newly generated RNA sequencing datasets were deposited in the NCBI SRA database under Bioproject PRJNA752516. Scripts used for the genotyping analysis are available at https://github.com/matheusbianconi/RNAgenotyping_hybrids

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.

660 **Bouton JH, Brown RH, Evans PT, Jernstedt JA.** 1986. Photosynthesis, leaf anatomy, and
661 morphology of progeny from hybrids between C_3 and C_3/C_4 *Panicum* species. Plant
662 Physiology 80, 487–492.
663 **Brown RH, Bouton JH, Evans PT, Malter HE, Rigsby LL.** 1985. Photosynthesis, morphology,
664 leaf anatomy, and cytogenetics of hybrids between C_3 and C_3/C_4 *Panicum* species. Plant
665 Physiology 77, 653–658.
666 **Brown HR, Bouton JH.** 1993. Physiology and genetics of interspecific hybrids between
667 photosynthetic types. Annual Review of Plant Biology 44, 435–456.
668 **Byrd GT, Brown RH, Bouton JH, Bassett CL, Black CC.** 1992. Degree of C_4 photosynthesis in
669 C_4 and C_3-C_4 *Flaveria* Species and their hybrids: I. CO_2 assimilation and metabolism and
670 activities of phosphoeno/pyruvate carboxylase and NADP-malic enzyme. Plant Physiology
671 100, 939–946.
672 **von Caemmerer S, Furbank RT.** 2003. The C_4 pathway: an efficient CO_2 pump. Photosynthesis
673 Research 77, 191–207.
674 **Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J,**
675 **Osborne CP.** 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene
676 transfer. Current Biology 22, 445–449.
677 **Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison**
678 **LM, Vorontsova MS, Edwards EJ.** 2013. Anatomical enablers and the evolution of C_4
679 photosynthesis in grasses. Proceedings of the National Academy of Sciences 110, 1381–1386.
680 **Danila FR, Thakur V, Chatterjee J, Bala S, Coe RA, Acebron K, Furbank RT, von**
681 **Caemmerer S, Quick WP.** 2021. Bundle sheath suberisation is required for C_4
682 photosynthesis in a *Setaria viridis* mutant. Communications Biology 4, 1–10.
683 **Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P,**
684 **Osborne CP, Christin P-A.** 2017. Introgression and repeated co-option facilitated the
685 recurrent emergence of C_4 photosynthesis among close relatives. Evolution 71, 1541–1555.
686 **Dunning LT, Moreno-Villena JJ, Lundgren MR, et al.** 2019a. Key changes in gene expression
687 identified for different stages of C_4 evolution in *Alloteropsis semialata*. Journal of
688 Experimental Botany 70, 3255–3268.
689 **Dunning LT, Olofsson JK, Parisod C, et al.** 2019b. Lateral transfers of large DNA fragments
690 spread functional genes among grasses. Proceedings of the National Academy of Sciences of
691 the United States of America 116, 4416–4425.
692 **Edwards G, Ku MB.** 1987. Biochemistry of C_3–C_4 intermediates. In: Hatch MD, Boardman NK,
693 eds. The biochemistry of plants: a comprehensive treatise, Vol 14: photosynthesis. New York:
694 Academic Press, 275–325.

695 **Edwards EJ.** 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and
696 CAM photosynthesis. New Phytologist 223, 1742–1755.
697 **Ellis RP.** 1974a. Significance of the occurrence of both Kranz and non-Kranz leaf anatomy in the
698 grass species *Alloteropsis semialata*. South African Journal of Science 70, 169-173.
699 **Ellis RP.** 1974b. Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades.
700 Bothalia 11, 273–275.
701 **Ermakova M, Arrivault S, Giuliani R, et al.** 2021. Installation of C_4 photosynthetic pathway
702 enzymes in rice using a single construct. Plant Biotechnology Journal 19, 575–588.
703 **Grass Phylogeny Working Group II (GPWG II).** 2012. New grass phylogeny resolves deep
704 evolutionary relationships and discovers C_4 origins. New Phytologist 193, 304–312.
705 **Hatch MD.** 1987. C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and
706 ultrastructure. Biochimica et Biophysica Acta 895, 81–106.
707 **Hattersley PW, Browning AJ.** 1981. Occurrence of the suberized lamella in leaves of grasses of
708 different photosynthetic types. I. In parenchymatous bundle sheaths and PCR ('Kranz')
709 sheaths. Protoplasma 109, 371–401.
710 **Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ.** 2013.
711 Predicting C_4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji
712 fitness landscape. Cell 153, 1579–1588.
713 **Heckmann D.** 2016. C_4 photosynthesis evolution: The conditional Mt. Fuji. Current Opinion in
714 Plant Biology 31, 149–154.
715 **Hylton CM, Rawsthorne S, Smith AM, Jones DA, Woolhouse HW.** 1988. Glycine decarboxylase
716 is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. Planta 175,
717 452–9.
718 **Ishimaru K, Ichikawa H, Matsuoka M, Ohsugi R.** 1997. Analysis of a C_4 maize pyruvate,
719 orthophosphate dikinase expressed in C_3 transgenic Arabidopsis plants. Plant Science 129,
720 57–64.
721 **Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW.** 2017. C_3–C_4 intermediates
722 may be of hybrid origin – a reminder. New Phytologist 215, 70–76.
723 **Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H.** 2014. C_2 photosynthesis generates
724 about 3-fold elevated leaf CO_2 levels in the C_3-C_4 intermediate species *Flaveria pubescens*.
725 Journal of Experimental Botany 65, 3649–3656.
726 **Khoshravesh R, Stinson CR, Stata M, Busch FA, Sage RF, Ludwig M, Sage TL.** 2016. C_3-C_4
727 intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase
728 expression, and the rise of C_2 photosynthesis. Journal of Experimental Botany, 67, 3065–
729 3078.

730 **Khoshravesh R, Stata M, Busch FA, Saladie M, Castelli JM, Dakin N, Hattersley PW,**
731 **Macfarlane TD, Sage RF, Ludwig M, Sage TL.** 2020. The evolutionary origin of C_4
732 photosynthesis in the grass subtribe Neurachninae. Plant Physiology 182, 566-583.
733 **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9,
734 357–359.
735 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin**
736 **R; 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map
737 format and SAMtools. Bioinformatics 25, 2078–2079.
738 **Lin M, Schlüter U, Stich B, Weber APM.** 2021. Cis-regulatory divergence underpins the
739 evolution of C_3-C_4 intermediate photosynthesis in *Moricandia*. bioRxiv [Preprint] doi:
740 10.1101/2021.05.10.443365
741 **Lundgren MR, Osborne CP, Christin PA.** 2014. Deconstructing Kranz anatomy to understand C_4
742 evolution. Journal of Experimental Botany 65, 3357–3369.
743 **Lundgren MR, Besnard G, Ripley BS, et al.** 2015. Photosynthetic innovation broadens the niche
744 within a single species. Ecology Letters 18, 1021–1029.
745 **Lundgren MR, Christin P-A, Escobar EG, Ripley BS, Besnard G, Long CM, Hattersley PW,**
746 **Ellis RP, Leegood RC, Osborne CP.** 2016. Evolutionary implications of C_3–C_4 intermediates
747 in the grass *Alloteropsis semialata*. Plant, Cell and Environment 39, 1874–1885.
748 **Lundgren MR, Dunning LT, Olofsson JK, et al.** 2019. C_4 anatomy can evolve via a single
749 developmental change. Ecology Letters 22, 302–312.
750 **Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U.**
751 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus
752 Flaveria. ELife 3.
753 **Monson RK, Moore BD.** 1989. On the significance of C_3-C_4 intermediate photosynthesis to the
754 evolution of C_4 photosynthesis. Plant, Cell & Environment 12, 689–699.
755 **Moreno-Villena JJ, Dunning LT, Osborne CP, Christin PA.** 2018. Highly expressed genes are
756 preferentially co-opted for C_4 Photosynthesis. Molecular Biology and Evolution 35, 94–106.
757 **O'Leary M.H.** 1988. Carbon isotopes in photosynthesis. BioScience 38, 328–336.
758 **Oakley JC, Sultmanis S, Stinson CR, Sage TL, Sage RF.** 2014. Comparative studies of C_3 and C_4
759 *Atriplex* hybrids in the genomics era: Physiological assessments. Journal of Experimental
760 Botany 65, 3637–3647.
761 **Olofsson JK, Bianconi M, Besnard G, et al.** 2016. Genome biogeography reveals the intraspecific
762 spread of adaptive mutations for a complex trait. Molecular Ecology 25, 6107–6123.
763 **Olofsson JK, Curran EV, Nyirenda F, et al.** 2021. Low dispersal and ploidy differences in a grass
764 maintain photosynthetic diversity despite gene flow and habitat overlap. Molecular Ecology,

765 1–15.
766 **Patel RK, Jain M.** 2012. NGS QC Toolkit: a toolkit for quality control of next generation
767 sequencing data. PLos One 7, e30619.
768 **R Development Core Team.** 2020. R: a language and environment for statistical computing.
769 Vienna, Austria: R Foundation for Statistical Computing.
770 **Sage RF.** 2001. Environmental and evolutionary preconditions for the origin and diversification of
771 the C_4 photosynthetic syndrome. Plant Biology 3, 202–213.
772 **Sage RF.** 2004. The evolution of C_4 photosynthesis. New Phytologist 161, 341–370.
773 **Sage RF, Christin P-A, Edwards EJ.** 2011. The C_4 plant lineages of planet Earth. Journal of
774 Experimental Botany 62, 3155–3169.
775 **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image
776 analysis. Nature Methods 9, 671–675.
777 **Simpson CJC, Reeves G, Tripathi A, Singh P, Hibberd JM.** Using breeding and quantitative
778 genetics to understand the C_4 pathway. Journal of Experimental Botany (in press).
779 https://doi.org/10.1093/jxb/erab486.
780 **Stata M, Sage TL, Sage RF.** 2019. Mind the gap: the evolutionary engagement of the C_4 metabolic
781 cycle in support of net carbon assimilation. Current Opinion in Plant Biology 49, 27–34.
782 **Sultmanis SD.** 2018. Developmental and expression evolution in C_3 and C_4 *Atriplex* and their
783 hybrids. PhD thesis, University of Toronto, Canada. 178p.
784 **Taniguchi Y, Ohkawa H, Masumoto C, et al.** 2008. Overproduction of C_4 photosynthetic enzymes
785 in transgenic rice plants: An approach to introduce the C_4-like photosynthetic pathway into
786 rice. Journal of Experimental Botany 59, 1799–1809.
787 **Ueno O, Sentoku N.** 2006. Comparison of leaf structure and photosynthetic characteristics of C_3
788 and C_4 *Alloteropsis semialata* subspecies. Plant, Cell & Environment 29, 257–268.
789 **Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP,**
790 **Furbank R, Sage TL, Langdale JA.** 2017. Re-creation of a key step in the evolutionary
791 switch from C_3 to C_4 leaf anatomy. Current Biology 27, 3278-3287.e6.
792

793 ## Figures

794 **Fig. 1.** Distribution of carbon isotopic ratios ($\delta^{13}C$) in *Alloteropsis semialata*. $\delta^{13}C$ values of
795 accessions of *A. semialata* that are not part of this study were retrieved from Bianconi et al. (2020).
796 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
797 the hybrids analysed in this study.
798 **Fig. 2.** Leaf anatomy of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*
799 (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)

800 and C_3+C_4 (TAN1602-03c) accessions. Arrows indicate minor veins (i.e. fourth and fifth order
801 veins). Scale bar = 100 µm. (B) Principal component analysis of selected leaf anatomical variables
802 (C) Proportion of inner sheath to mesophyll area (IS/M). (D) Distance between consecutive bundle
803 sheaths (BSD). (E) Mean inner sheath cell width. (F) Mean outer sheath cell width. For (C-F),
804 different lower-case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-
805 hoc Tukey HSD; $n \geq 4$). Bu = bulliform cells, BSD = distance between consecutive bundle sheaths,
806 Epd = proportion of epidermis area, Fb = extraxylary fibres (= sclerenchyma girder), IS = inner
807 bundle sheath (= mestome sheath), ISW = IS width, M = mesophyll, Mx = metaxylem, OS = outer
808 bundle sheath, OSW = OS width, Ph = phloem, V = proportion of vascular tissue area.
809 **Fig. 3.** Leaf gene expression of F1 hybrids and the parental photosynthetic types in *Alloteropsis*
810 *semialata*. (A) Principal component analysis on 7,482 genes. (B) Transcript abundance in reads per
811 million mapped reads (RPM) of selected genes encoding core C_4 enzymes: aspartate
812 aminotransferase (ASP-AT, gene *aspat-3P4*), NADP-dependent malic enzyme (NADP-ME, gene
813 *nadpme-1P4*), phosphoenolpyruvate carboxykinase (PCK, gene *pck1P1_LGT*), pyruvate phosphate
814 dikinase (PPDK, gene *ppdk-1P2*). (C) Transcript abundance of selected phosphoenolpyruvate
815 carboxylase (PEPC) genes.
816 **Fig. 4.** Response of net CO_2 assimilation rate (*A*) to intercellular CO_2 (*Ci*) of F1 hybrids and the
817 parental photosynthetic types in *Alloteropsis semialata*. Data points are predicted values from an
818 empirical non-rectangular hyperbola model (Bellasio et al. 2016) based on the mean parameter
819 estimates per individual within each cross/photosynthetic type. Error bars show standard errors for
820 the predicted values ($n \geq 3$). Arrows indicate Ci values at reference CO_2 = 400 µmol mol^{-1} for each
821 cross/photosynthetic type.
822 **Fig. 5.** Photosynthetic performance of F1 hybrids and the parental photosynthetic types in
823 *Alloteropsis semialata*. (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point
824 (CCP), and steady-state (C) net photosynthetic rate (A_{400}), and (D) intrinsic water use efficiency
825 (iWUE, A/g_s) at reference CO_2 = 400 µmol mol^{-1}. Different lower-case letters indicate statistical
826 differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; $n \geq 4$).

Fig. 1

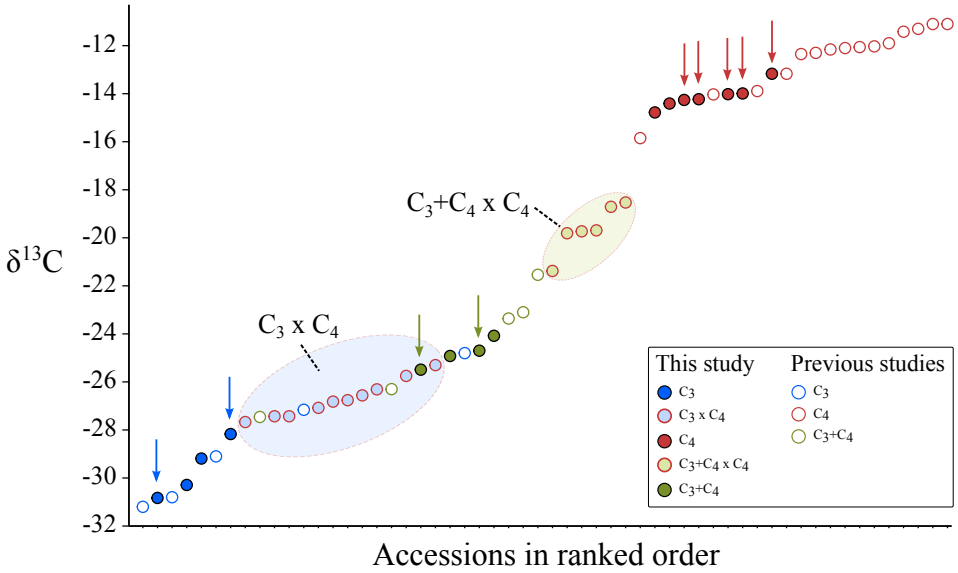

Fig. 1. Distribution of carbon isotopic ratios ($\delta^{13}C$) in *Alloteropsis semialata*. $\delta^{13}C$ values of accessions of *A. semialata* that are not part of this study were retrieved from Bianconi et al. (2020). Blue, green and red arrows indicate $\delta^{13}C$ values of C_3, C_3+C_4 and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

Fig. 2

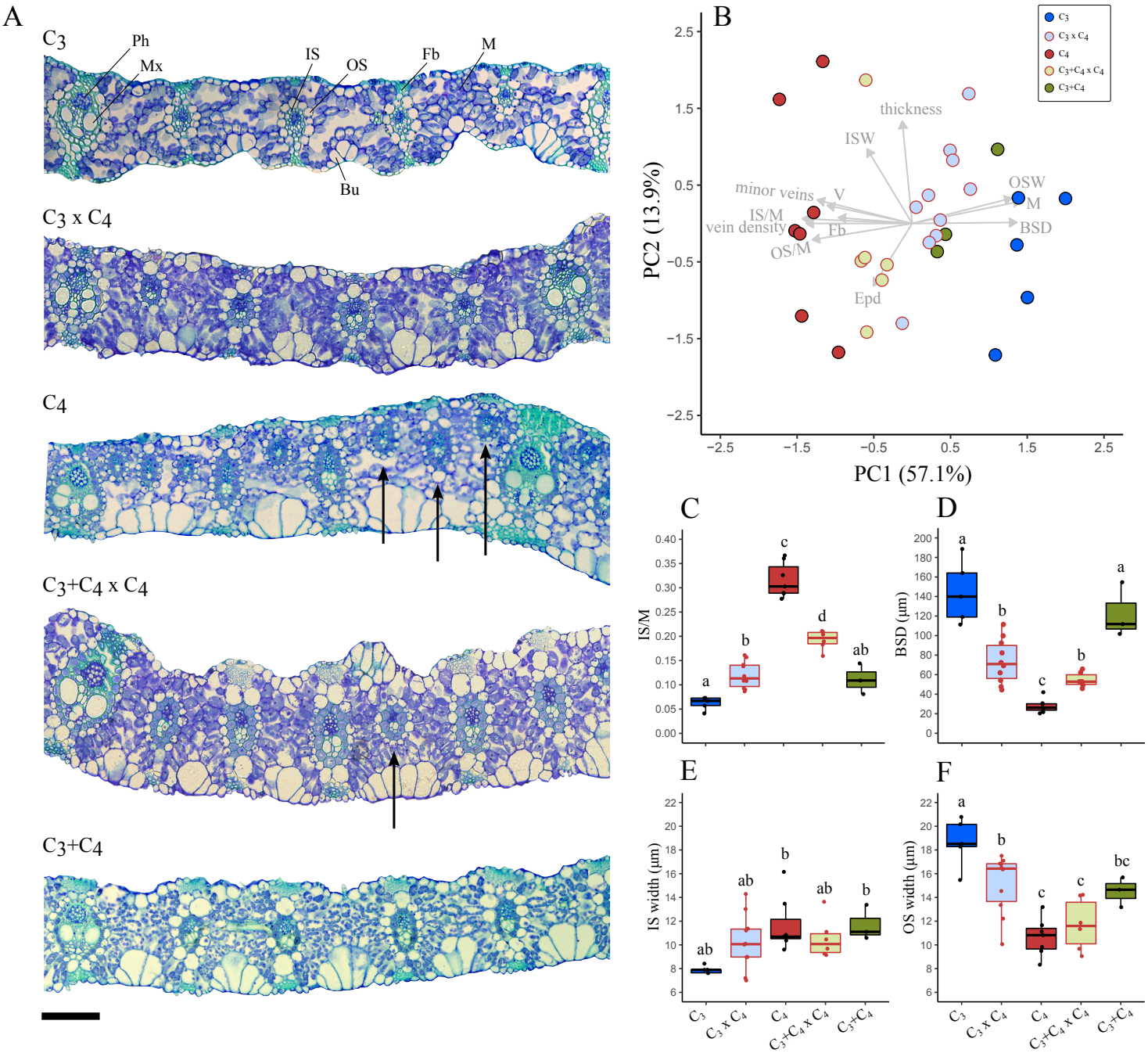

Fig. 2. Leaf anatomy of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata* (A) Representative cross sections of C_3 (RSA9), C_3 x C_4 (H08), C_4 (TPE1-10), C_3+C_4 x C_4 (H11) and C_3+C_4 (TAN1602-03c) accessions. Arrows indicate minor veins (i.e. fourth and fifth order veins). 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

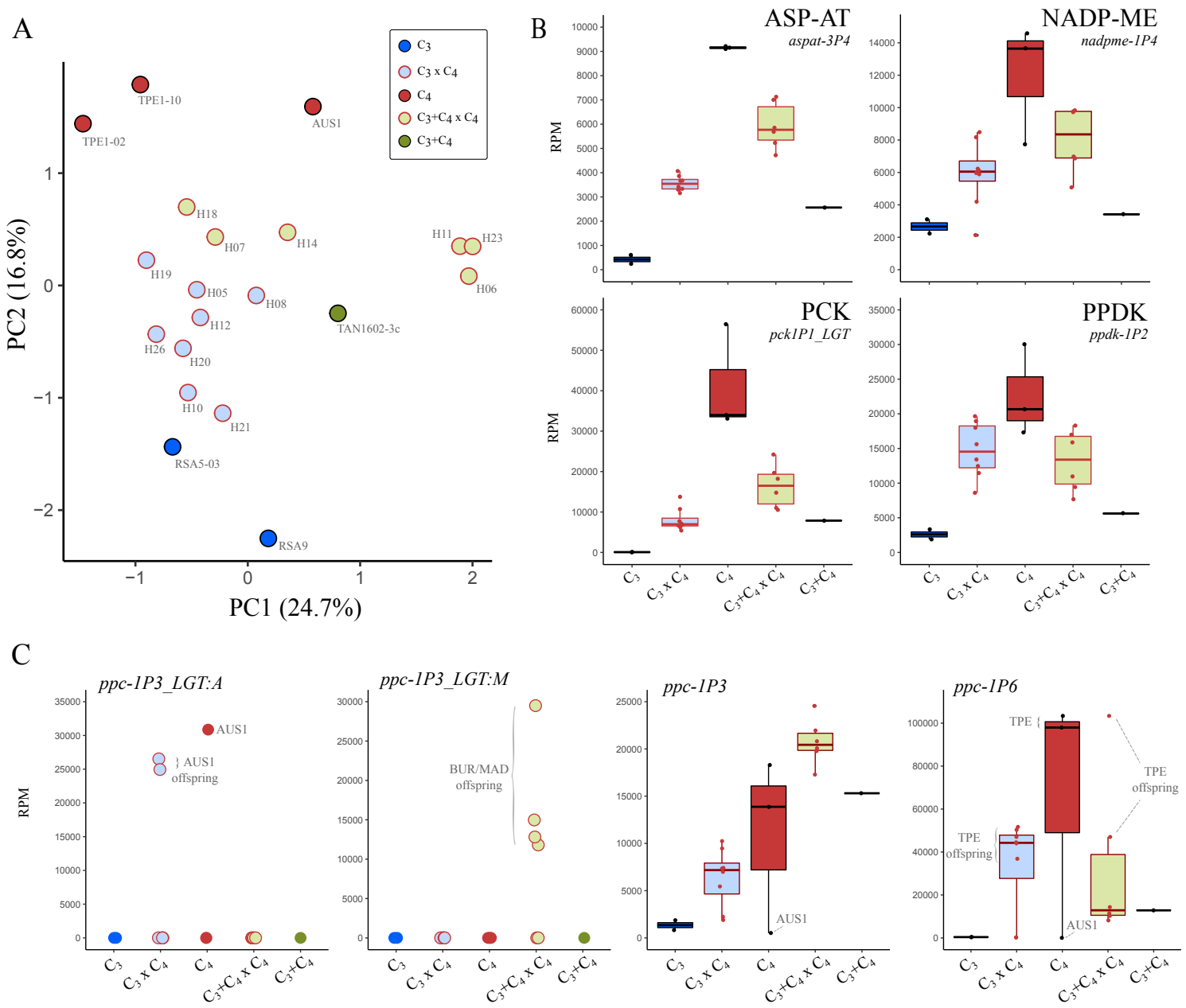

Fig. 3. Leaf gene expression of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) Principal component analysis on 7,482 genes. (B) Transcript abundance in reads per million (RPM) of selected genes encoding core C_4 enzymes: aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), 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

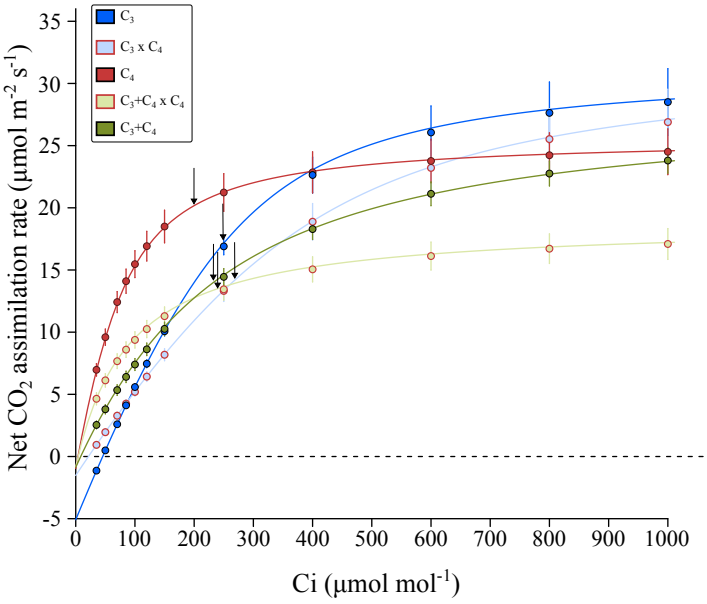

Fig. 4. Response of net CO_2 assimilation rate (*A*) to intercellular CO_2 (*Ci*) of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. Data points are predicted values from an empirical non-rectangular hyperbola model (Bellasio et al. 2016) based on the mean parameter estimates 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

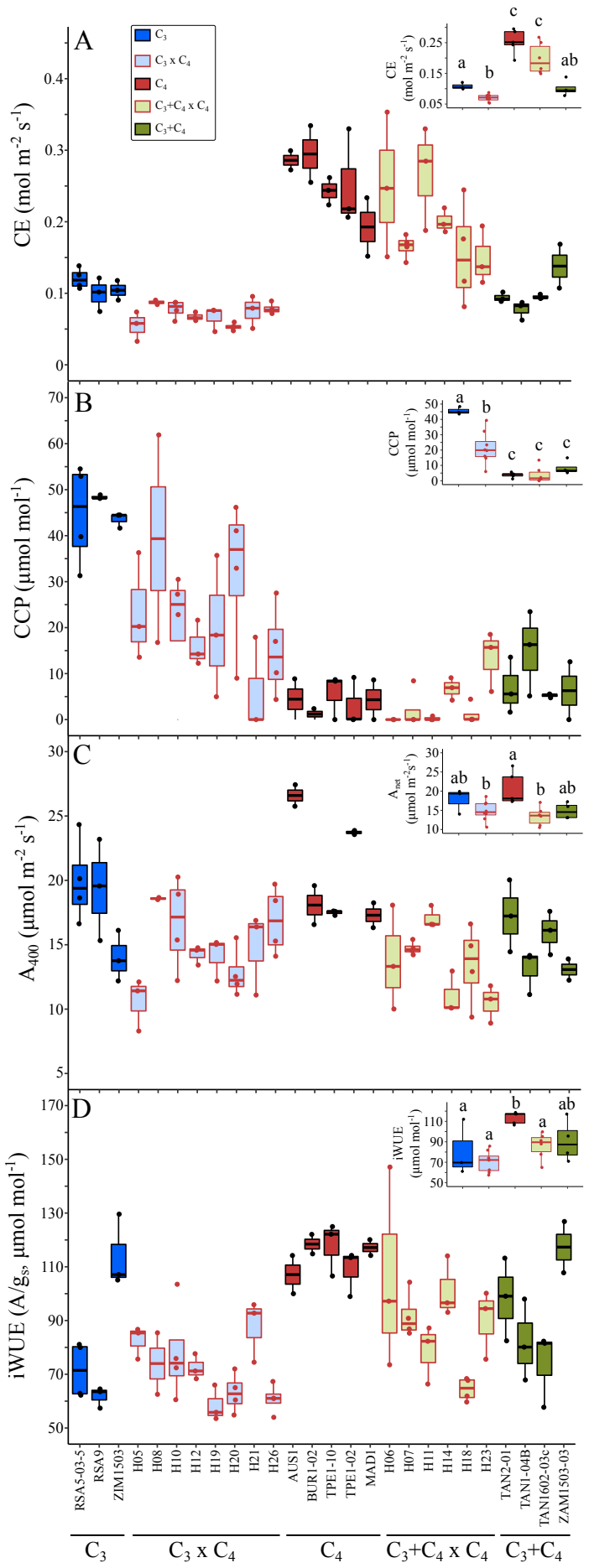

Fig. 5. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C) net photosynthetic rate (A400), and (D) intrinsic water use efficiency (iWUE, *A*/*gs*) 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).

Supplementary information for:

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, Lígia T. Bertolino, Colin P. Osborne, Pascal-Antoine Christin

This supplementary information contains 9 figures and 12 tables. Tables are provided in five separate files: Tables S1-S6, S10-S12 ("Tables_SI.xlsx") and separate XLSX files for Tables S7, S8, S9 and S10.

Fig. S1. Leaf anatomical variables measured in this study.

Fig. S2. Principal component analysis on 11,988 genes from the chromosome-level genome assembly of *Alloteropsis semialata.*

Fig. S3. Transcript abundance in reads per million (RPM) of selected genes encoding core C_4 enzymes in *Alloteropsis semialata*: (A) aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), (B) NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*), (C) phosphoenolpyruvate carboxykinase (PCK, gene *pck1P1_LGT*), (D) pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*), and (E-I) phosphoenolpyruvate carboxylase (PEPC) genes. Transcript abundance computed for *A. semialata* samples extracted from Dunning et al. (2019a) are shown on the right for each gene.

Fig. S4. Transcript abundance of selected gene families with a role in photorespiration (A-K), and the gene encoding the transcription factor GOLDEN2-like (L) in F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) 2-phosphoglycolate (2-PG) phosphatase (PGLP), (B-C) flavin mononucleotide (FMN)-dependent glycolate oxidase (GOX), (D) glutamate:glyoxylate aminotransferase (GGT), (E-H) glycine decarboxylase (GDC) complex proteins -H, -P, -T and -L, (I-J) serine hydroxymethyltransferase (SHMT), (K) glycerate 3-kinase (GLYK). Transcript abundance in reads per million mapped reads (RPM).

Fig. S5. Heat map of differentially expressed genes related to C_4 photosynthesis and photorespiration. Asterisks indicate significant DE genes ($p < 0.05$) for comparisons between hybrid types (C_3 x C_4 vs C_3+C_4 x C_4) and between each hybrid type and the C_4 type. Only DE genes with at least two-fold change and base count > 500 are shown. Count data was transformed using the VST function of DESeq2 and scaled by row. See table S9 for full gene annotation

Fig. S6. Photosynthetic response to intercellular CO_2 (*A*/*Ci*) of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. Panels contain all *A*/*Ci* curves collected for each accession, with individual curves coloured with different shades of grey. Accession name and

cross/photosynthetic type are indicated on the top left. Vertical dashed lines indicate the CO_2 compensation point of each curve.

Fig. S7. Leaf stomata variables. (A) Steady-state stomatal conductance (g_s) at 400 µmol mol^{-1} (n = 2-4 leaves per accession). (B) Stomatal density on the abaxial side of the leaves (n = 2 leaves per accession, with stomata counts averaged from 5 fields per leaf; field area = 0.38 mm^2). Data points on the right are the means per accession within each cross/photosynthetic type, and different lower-case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; $n \geq 3$).

Fig. S8. Leaf temperature variation during *A*/*Ci* curves. Data points are individual *A*/*Ci* measurements for each plant (left panels) or with individuals grouped into cross/photosynthetic types (right panels). (A) All measurements, and measurements collected at reference CO_2 (B) < 100 µmol mol^{-1}, (C) between 100 and 400 µmol mol^{-1}, and (D) > 400 µmol mol^{-1}. Arrows indicate outlier individuals that were removed for the analysis in Fig. S9.

Fig. S9. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata* after removing outlier accessions (i.e. with T_{leaf} 1ºC above the median). (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C) net photosynthetic rate (A_{400}), and (D) intrinsic water use efficiency (iWUE, A/g_s) at reference CO_2 = 400 µmol mol^{-1}.

Table S1. Sample information.

Table S2. Genotyping using PCR/Sanger-sequencing.

Table S3. List of primer sequences used for genotyping.

Table S4. RNA-seq markers selected for genotyping analysis I (photosynthetic type).

Table S5. RNA-seq markers selected for genotyping analysis II (pollen parent).

Table S6. Raw leaf anatomy data.

Table S7. Normalized transcript abundance for the co-ortholog gene set.

Table S8. Normalized transcript abundance for the *A. semialata* complete genome gene set.

Table S9. Differential expression analysis between F1 hybrids and the C_4 parental type.

Table S10. Gene ontology enrichment analyses on genes differentially expressed between F1 hybrids and the C_4 parental type.

Table S11. Raw A/Ci data.

Table S12. Estimated photosynthetic parameters and steady-state measurements.

Supplementary Methods

Global transcriptome analyses

We performed differential expression (DE) analyses using the raw counts obtained after mapping the RNA-seq datasets to the coding sequences extracted from the chromosome-level genome assembly of *A. semialata* (Dunning et al. 2019; see main text; total number of genes = 45,145). To detect genes that were differentially expressed between each of the hybrid types and the parental types, we used the R package *DESeq2* (Love et al. 2014). Due to the lack of sufficient replicates for the C_3 and C_3+C_4 parental types, we restricted our analyses to the comparisons between hybrid types (C_3 x C_4 vs C_3+C_4 x C_4), and between each of these and the C_4 type. Only genes with more than 10 counts across all accessions were retained for the DE analysis (total = 29,200 genes). For each comparison, we used a false discovery rate (FDR) of 0.05 as cut-off value. We then investigated whether each of the sets of DE genes were associated with any particular metabolic function using a gene ontology (GO) enrichment analysis with the R package *clusterProfiler* (Yu et al. 2012). We first used Orthofinder v2.5.2 (Emms and Kelly 2019) with default parameters to identify the corresponding orthologs of *A. semialata* in the closely related grass species *Setaria italica* (v2.2), *Sorghum bicolor* (v3.1.1) and *Oryza sativa* (v7.0; genomes extracted from Phytozome v13; Goodstein et al. 2012). We extracted the GO annotations for the three genomes using the Biomart tool of Phytozome 13, and transferred the annotations to the *A. semialata* gene set using the orthology information obtained from Orthofinder. We then performed a GO enrichment analysis on the DE gene set resulting from each of the three comparisons using the *enricher* function of *clusterProfiler* with a *p*-value cut-off of 0.05. The identity of C_4- and photorespiration-related genes in the genome of *A. semialata* was extracted from the annotations used in Bianconi et al. (2018) and Dunning et al. (2019).

Supplementary References

Bianconi ME, Dunning, LT, Moreno-Villena JJ, Osborne CP, Christin PA. 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.

Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, Dionora J, Paul Quick W, Park M, Bennetzen JL, et al. 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. Proceedings of the National Academy of Sciences, USA 116: 4416–4425.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology 20: 238.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. 2012. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research 40: D1178–D1186.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, **15**, 550.

Yu G, Wang L, Han Y, He Q. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, 16, 284-287.

Fig. S1

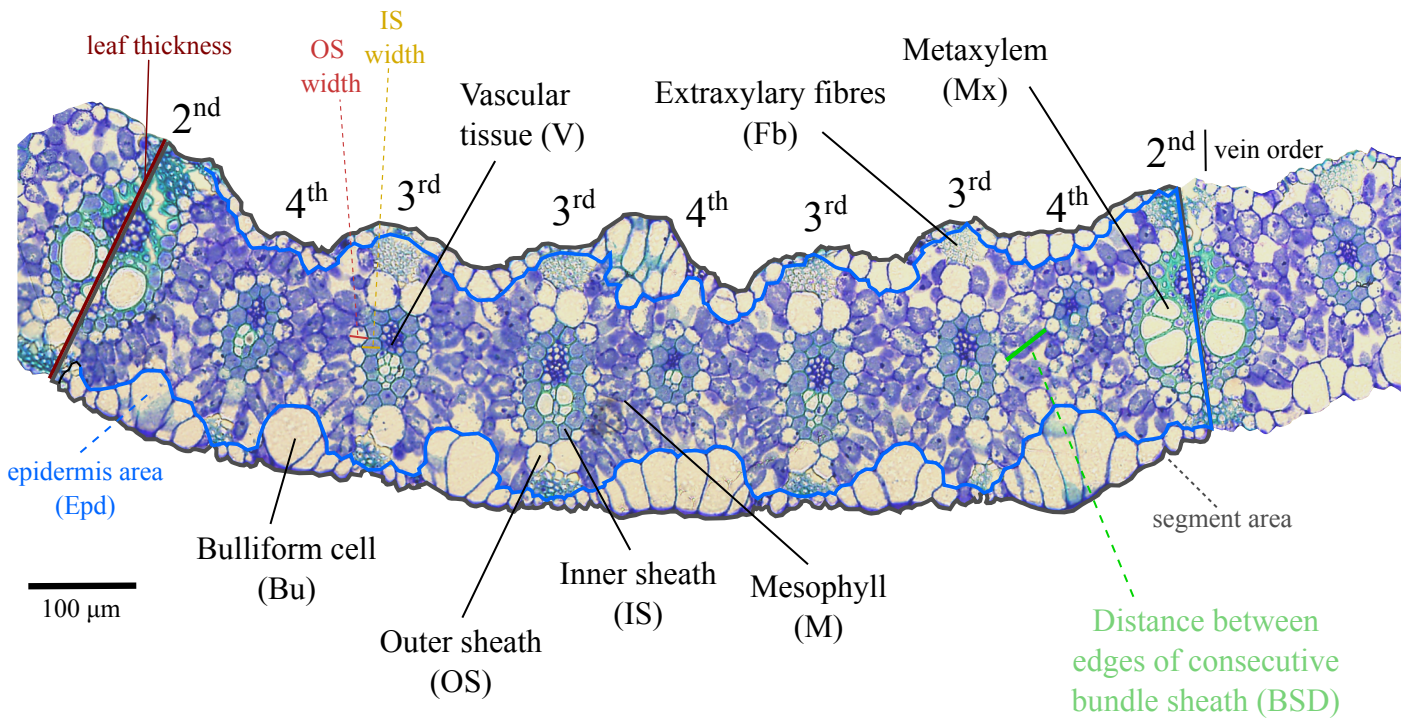

Fig. S1. Leaf anatomical variables measured in this study.

Fig. S2

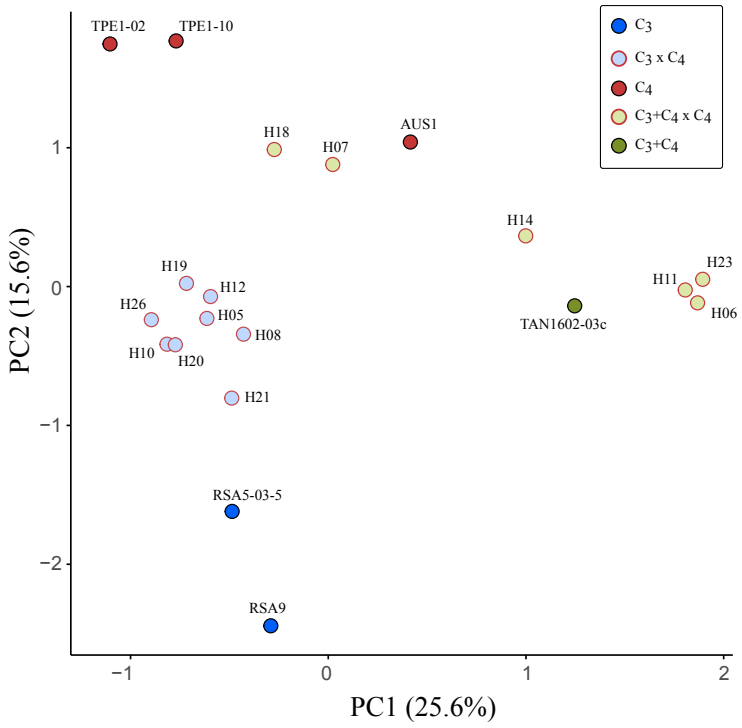

Fig. S2. Principal component analysis on 11,988 genes from the chromosome-level genome assembly of *Alloteropsis semialata*.

Fig. S3

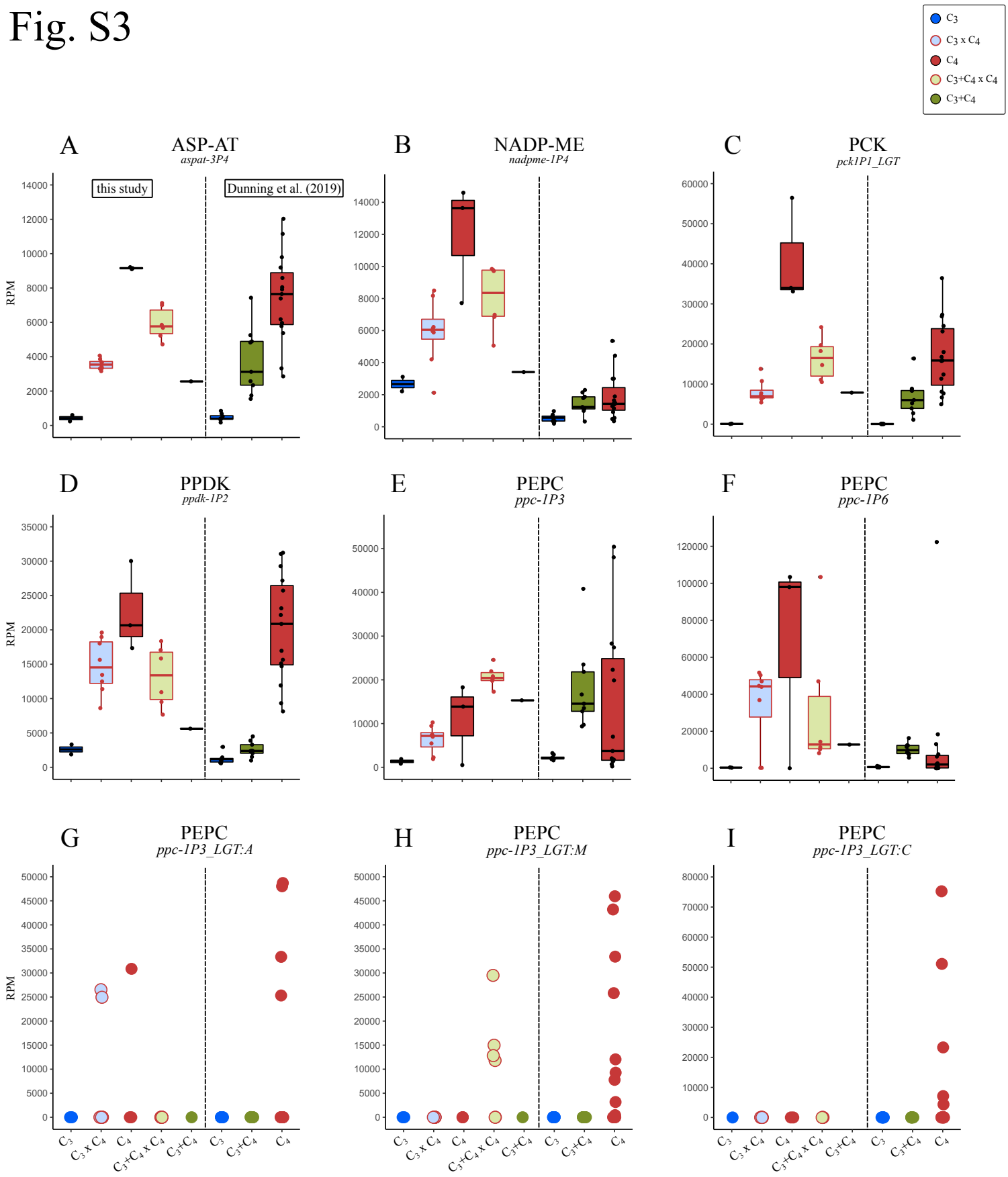

Fig. S3. Transcript abundance in reads per million (RPM) of selected genes encoding core C_4 enzymes in *Alloteropsis semialata*: (A) aspartate aminotransferase (ASP-AT, gene *aspat-3P4*), (B) NADP-dependent malic enzyme (NADP-ME, gene *nadpme-1P4*), (C) phosphoenolpyruvate carboxykinase (PCK, gene *pck1P1_LGT*), (D) pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*), and (E-I) phosphoenolpyruvate carboxylase (PEPC) genes. Transcript abundance computed for *A. semialata* samples extracted from Dunning et al. (2019a) are shown on the right for each gene.

Fig. S4

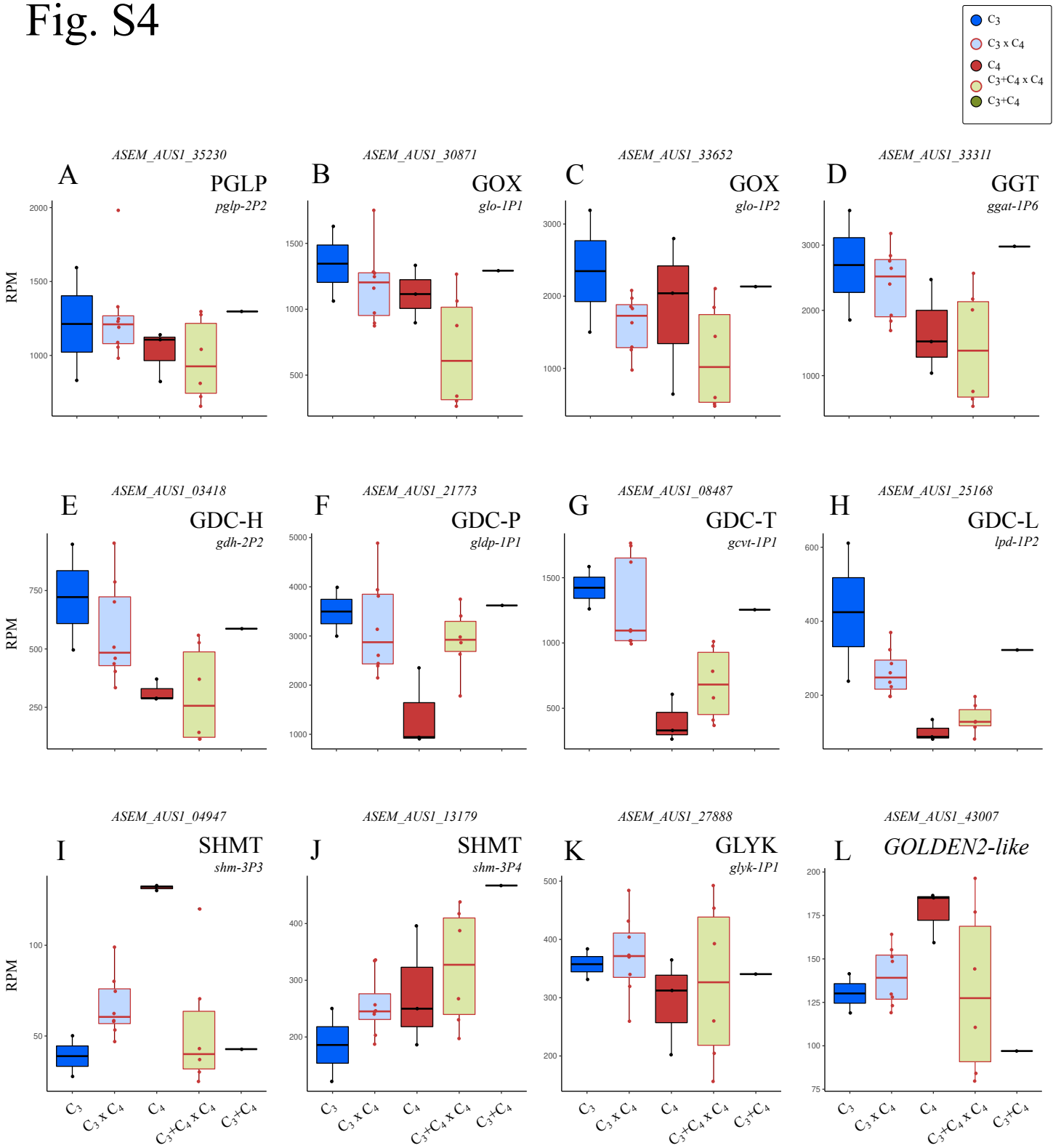

Fig. S4. Transcript abundance of selected gene families with a role in photorespiration (A-K), and the gene encoding the transcription factor *GOLDEN2*-like (L) in F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. (A) 2-phosphoglycolate (2-PG) phosphatase (PGLP), (B-C) flavin mononucleotide (FMN)-dependent glycolate oxidase (GOX), (D) glutamate:glyoxylate aminotransferase (GGT), (E-H) glycine decarboxylase (GDC) complex proteins -H, -P, -T and -L, (I-J) serine hydroxymethyltransferase (SHMT), (K) glycerate 3-kinase (GLYK). Transcript abundance in reads per million mapped reads (RPM).

Fig. S5

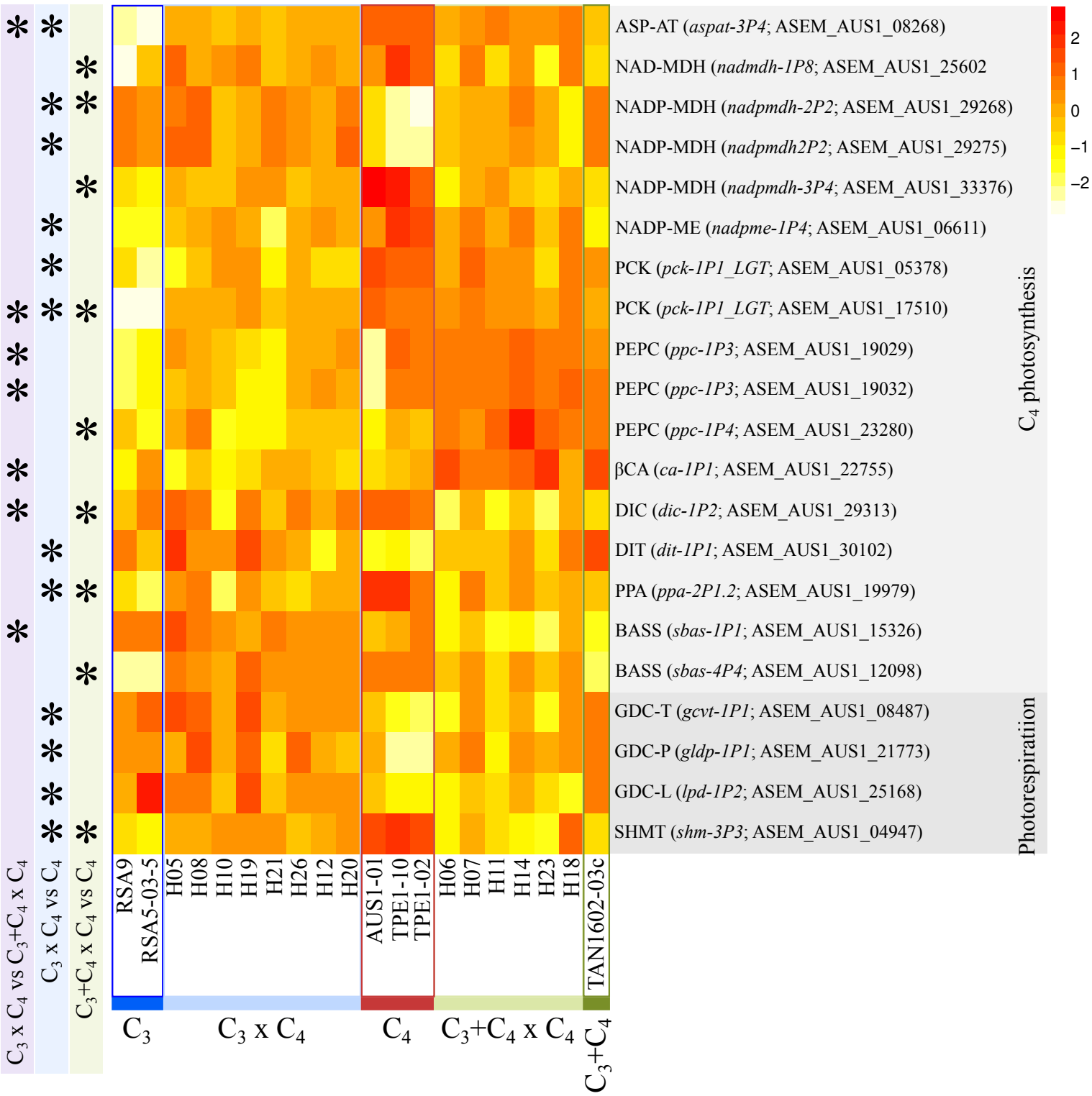

Fig. S5. Heat map of differentially expressed genes related to C_4 photosynthesis and photorespiration. Asterisks indicate significant DE genes ($p < 0.05$) for comparisons between hybrid types (C_3 x C_4 vs C_3+C_4 x C_4) and between each hybrid type and the C_4 type. Only DE genes with at least two-fold change and base count > 500 are shown. Count data was transformed using the VST function of DESeq2 and scaled by row. See table S9 for full gene annotation.

Fig. S6

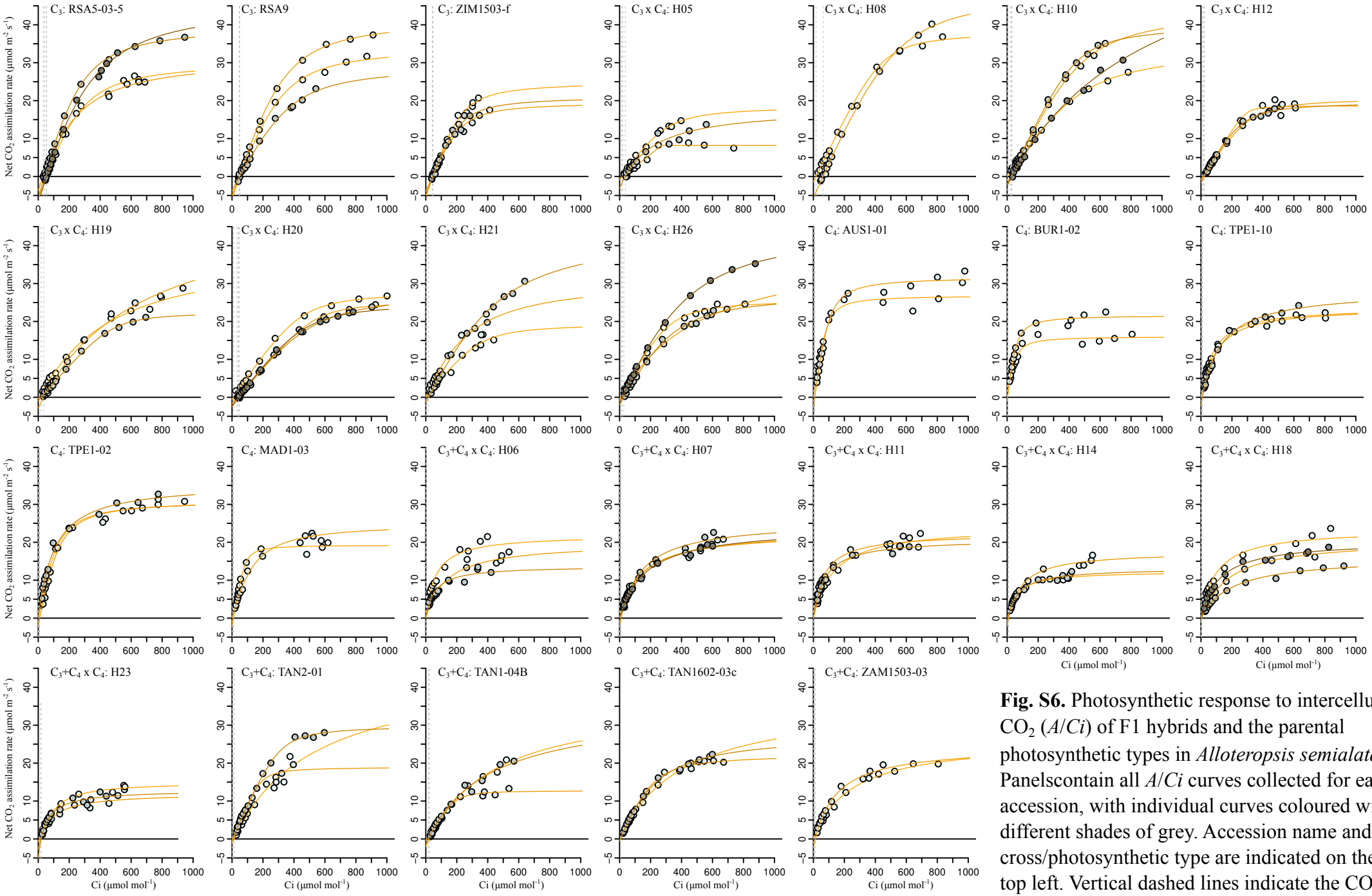

Fig. S6. Photosynthetic response to intercellular CO_2 (*A*/*Ci*) of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata*. Panelscontain all *A*/*Ci* curves collected for each accession, with individual curves coloured with different shades of grey. Accession name and cross/photosynthetic type are indicated on the top left. Vertical dashed lines indicate the CO_2 compensation point of each curve.

Fig. S7

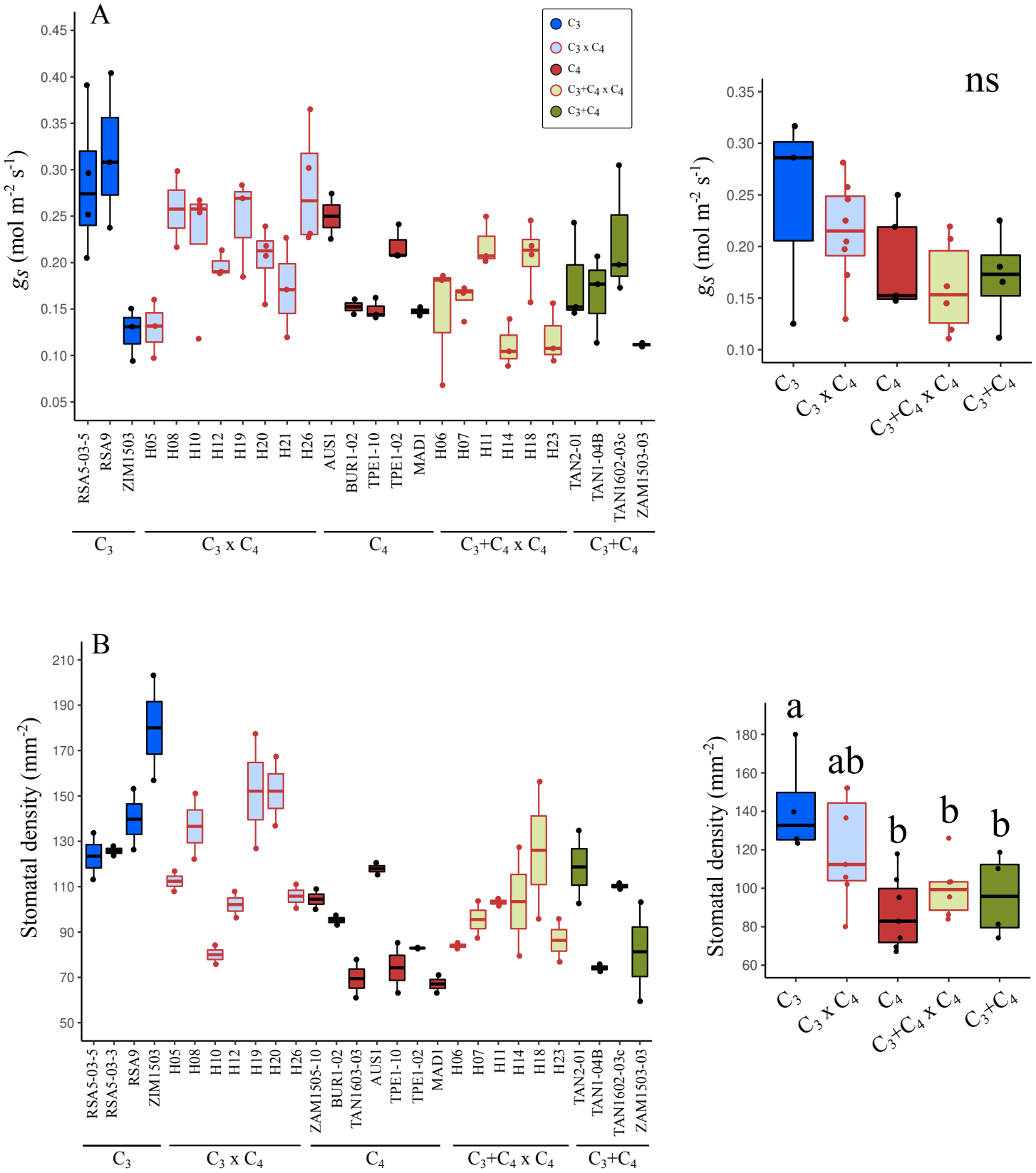

Fig. S7. Leaf stomata variables. (A) Steady-state stomatal conductance (g_s) at 400 µmol mol^{-1} (n = 2-4 leaves per accession). (B) Stomatal density on the abaxial side of the leaves (n = 2 leaves per accession, with stomata counts averaged from 5 fields per leaf; field area = 0.38 mm^2). Data points on the right are the means per accession within each cross/photosynthetic type, and different lower-case letters indicate statistical differences between groups (ANOVA, $p < 0.05$ post-hoc Tukey HSD; n ≥ 3).

Fig. S8

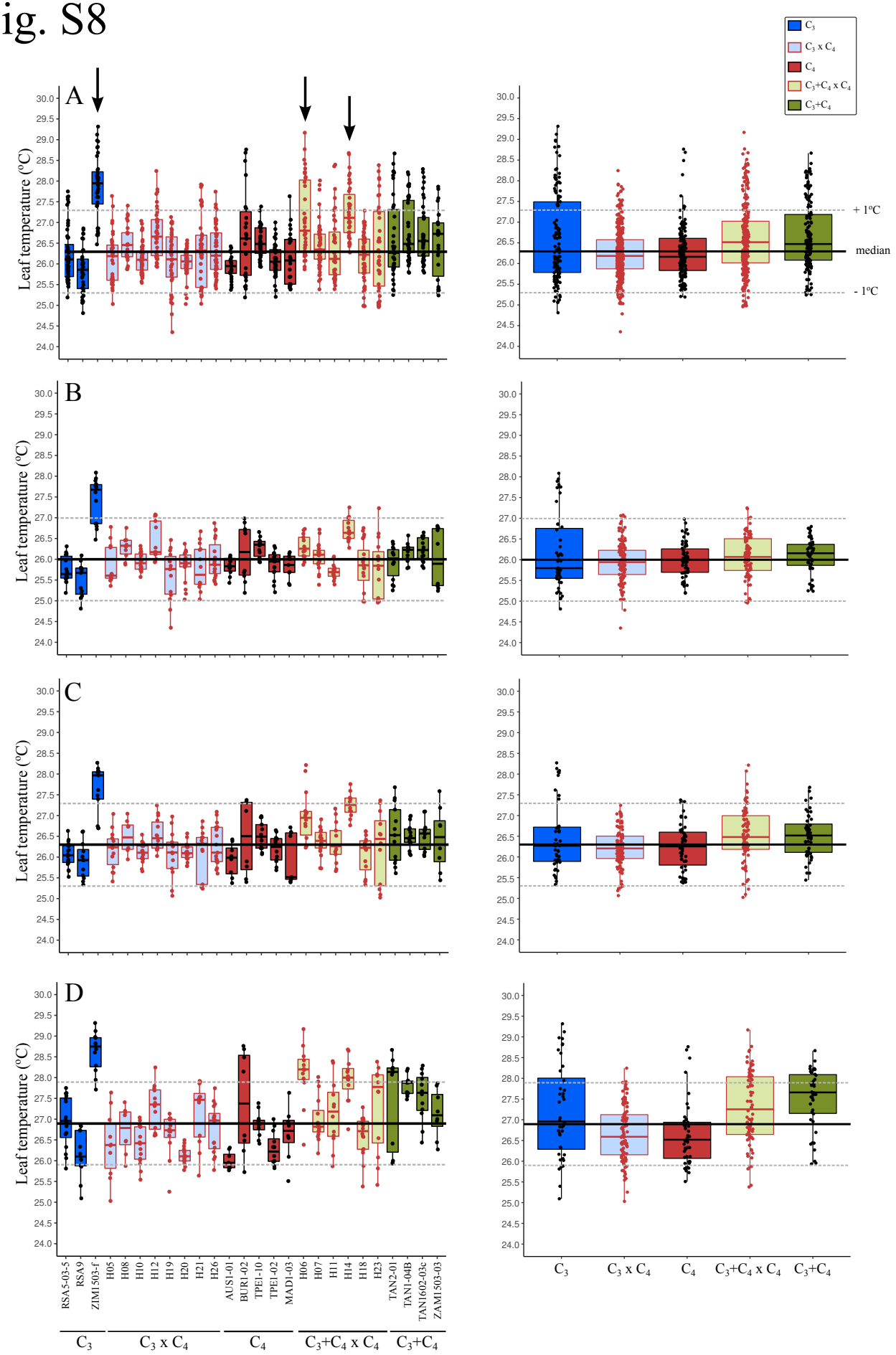

Fig. S8. Leaf temperature variation during A/Ci curves. Data points are individual A/Ci measurements for each plant (left panels) or with individuals grouped into cross/photosynthetic types (right panels). (A) All measurements, and measurements collected at reference CO_2 (B) < 100 µmol mol-1, (C) between 100 and 400 µmol mol^{-1}, and (D) > 400 µmol mol-1. Arrows indicate outlier individuals that were removed for the analysis in Fig. S9.

Fig. S9

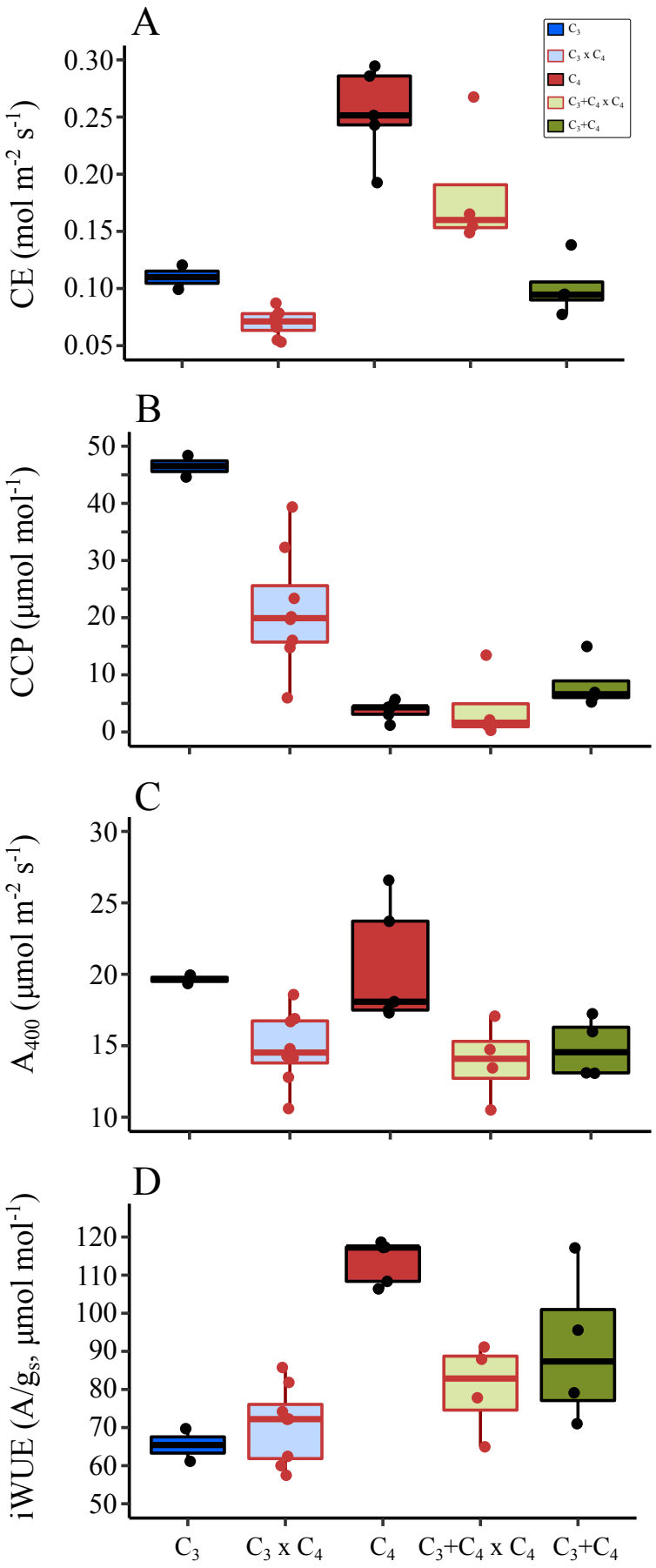

Fig. S9. Photosynthetic performance of F1 hybrids and the parental photosynthetic types in *Alloteropsis semialata* after removing outlier accessions (i.e. with T_{leaf} 1°C above the median). (A) Maximum carboxylation efficiency (CE), (B) CO_2 compensation point (CCP), and steady-state (C) net photosynthetic rate (A400), and (D) intrinsic water use efficiency (iWUE, *A*/*gs*) at reference CO_2 = 400 µmol mol^{-1}.