Cis-regulatory divergence underpins the evolution of C_3-C_4 intermediate photosynthesis in *Moricandia*

Meng-Ying Lin[1], Urte Schlüter[1], Benjamin Stich[2], Andreas P.M. Weber[1]

[1]Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany.

[2]Institute for Quantitative Genetics and Genomics of Plants, Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University, Universitätsstrasse 1, 40225 Düsseldorf, Germany.

Abstract

Altered transcript abundances and cell specific gene expression patterns that are caused by regulatory divergence play an important role in the evolution of C_4 photosynthesis. How these altered gene expression patterns are achieved and whether they are driven by *cis*- or *trans*-regulatory changes is mostly unknown. To address this question, we investigated the regulatory divergence between C_3 and C_3-C_4 intermediates, using allele specific gene expression (ASE) analyses of *Moricandia arvensis* (C_3-C_4), *M. moricandioides* (C_3) and their interspecific F_1 hybrids. ASE analysis on SNP-level showed similar relative proportions of regulatory effects among hybrids: 36% and 6% of SNPs were controlled by *cis*-only and *trans*-only changes, respectively. GO terms associated with metabolic processes and the positioning of chloroplast in cells were abundant in transcripts with *cis*-SNPs shared by all studied hybrids. Transcripts with *cis*-specificity expressed bias toward the allele from the C_3-C_4 intermediate genotype. Additionally, ASE evaluated on transcript-level indicated that ~27% of transcripts show signals of ASE in *Moricandia* hybrids. Promoter-GUS assays on selected genes revealed altered spatial gene expression patterns, which likely result from regulatory divergence in their promoter regions. Assessing ASE in *Moricandia* interspecific hybrids contributes to the understanding of early evolutionary steps towards C_4 photosynthesis and highlights the impact and importance of altered transcriptional regulations in this process.

Introduction

The majority of organic carbon in the biosphere is fixed by ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco) through carboxylation of ribulose-1,5-bisphosphate (RuBP) during photosynthesis. However, Rubisco has affinity not only to CO_2 but also to O_2. The oxidation of RuBP by Rubisco generates a toxic intermediate, 2-phosphoglycolate. This by-product is metabolized through the photorespiratory pathway, which is energy-consuming and leads to the release of CO_2. In C_4 plants, the oxygenation reaction is repressed by the evolution of an efficient biochemical CO_2 pump, usually functioning in concentric layers of cells known as Kranz anatomy. The enlarged bundle sheath (BS) cells with abundant organelles are located adjacent to vascular bundles and are surrounded by mesophyll (M) cells (Hatch, 1987). The CO_2 is fixed through phospho*enol*pyruvate carboxylase in M cells and the generated C_4 acid is decarboxylated in BS cells, where the released CO_2 increases the CO_2:O_2 ratio in close proximity to Rubisco, resulting in a high carboxylation rate and a low oxygenation rate (Bräutigam & Gowik, 2016; Hatch, 1987; Schlüter & Weber, 2020). The CO_2 compensation point is defined as the CO_2 concentration where photosynthetic CO_2 uptake equals CO_2 release, *i.e.*, no net gas exchange is detectable. C_4 plants have much lower CO_2 compensation points relative to C_3 plants (Krenzer et al., 1975).

The current model of C_4 evolution holds that C_4 plant species evolved from the ancestral C_3 state and C_3-C_4 intermediate species are considered as naturally occurring intermediates on the evolutionary trajectory towards C_4 photosynthesis (Blätke & Bräutigam, 2019; Mallmann et al., 2014; Sage et al., 2012). C_3-C_4 intermediates have been reported in 21 plant lineages including dicotyledonous as well as monocotyledonous species, such as *Diplotaxis*, *Flaveria*, *Moricandia*, *Neurachne*, and *Panicum* (Sage et al., 2011). C_3-C_4 intermediate species possess a photorespiratory CO_2 pump functioning in Kranz-like leaf anatomy, including BS cells with centripetally localized mitochondria and chloroplasts, and their CO_2 compensation points are in between the values of C_3 and C_4 plants (Brown & Hattersley, 1989; Holaday & Chollet, 1984; Sage et al., 2014). The photorespiratory CO_2 pump in C_3-C_4 plants efficiently recycles photorespiratory released CO_2 via the so called glycine shuttle a.k.a. C_2 photosynthesis (Kadereit et al., 2017; Sage et al., 2014; Schlüter & Weber, 2016). This system evolved via confining the expression of the gene encoding the P-subunit of glycine decarboxylase (GLDP) to the BS cells. Consequently, GLDP activity is absent from leaf M cells of C_3-C_4 plants (Monson & Rawsthorne, 2000), and hence to complete the photorespiratory pathway, glycine

61 must be shuttled to the BS cells, where CO_2 released from mitochondria can be efficiently
62 recaptured by numerous, surrounding chloroplasts.
63 A number of anatomical and biochemical adaptive steps on the evolutionary path from C_3 to
64 C_4 photosynthesis have been depicted in different models (Heckmann et al., 2013; Mallmann
65 et al., 2014; Sage et al., 2012; Williams et al., 2013). Based on studies of various naturally
66 occurring C_3-C_4 intermediate species, a stepwise model was proposed: (1) the vein density
67 increases; (2) the leaf proto-Kranz anatomy evolves; (3) a photorespiratory CO_2 pump is built
68 by a reduced M:BS ratio and confinement of mitochondrial glycine decarboxylase (GDC)
69 activity to BS cells; (4) enzymes of the C_4 metabolic cycle are established with spatial or
70 temporal expression adjustments of C_3 genes (Sage et al., 2012). The consensus trajectories of
71 the statistical (Williams et al., 2013) and mechanistic (Heckmann et al., 2013; Mallmann et al.,
72 2014) models confirmed these steps, but the order of steps was flexible and the path was smooth
73 (Heckmann, 2016; Williams et al., 2013). All C_4 evolution models unequivocally predicted
74 that the photorespiratory CO_2 pump, resulting from the confinement of GDC activity to BS
75 cells, is a crucial step.
76 In the genera *Flaveria* (*Asteraceae*) and *Moricandia* (*Brassicaceae*), the molecular
77 mechanisms by which *GLDP* expression becomes confined to BS cells during evolution of C_3-
78 C_4 intermediacy have been resolved. The genomes of C_3 *Flaveria* species encode two isoforms
79 of *GLDP*, one BS-specific isoform (*GLDPA*) and the other one ubiquitously expressed in all
80 photosynthetic tissues (*GLDPB*). *GLDPB* becomes a pseudogene in C_3-C_4 intermediate
81 *Flaverias* and thereby GDC activity is lost from M cells during C_4 photosynthesis evolution
82 (Schulze et al., 2013). A conceptually similar mechanism underpins the independent but
83 convergent evolution of C_3-C_4 intermediacy in the *Brassicaceae*. The promoter of the *GLDP1*
84 gene in this plant family carries two conserved *cis*-regulatory elements, one that drives
85 expression in the M cells (M-box), and another one that governs expression in the vasculature
86 (V-box). Deletion of the M-box from the *GLDP1* promoter led to the restriction of GDC
87 activity to BS cells (Adwy et al., 2015; Adwy, 2018). The establishment of the C_3-C_4
88 intermediate photorespiratory CO_2 pump very likely requires further metabolic adjustments
89 and anatomical modifications, probably through altered transcriptional regulation. However,
90 the genetic factors underpinning this process are still mostly unknown.
91 *Cis*-regulatory divergence has been reported to play an important role in adaptive phenotypic
92 evolution because, compared to the nonsynonymous mutation in protein sequences, it causes
93 fewer deleterious pleiotropic effects (Stern & Orgogozo, 2008; Wittkopp & Kalay, 2012;
94 Wray, 2007). Additionally, *cis*-regulatory divergences have impacts on limiting gene

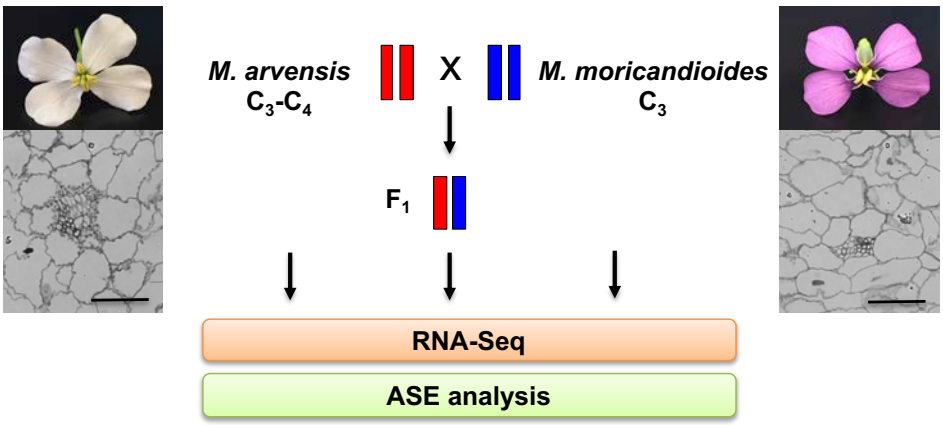
95 expression to particular tissue or cellular compartments, to specific life stages or environments
96 (Prud'homme et al., 2007). Allele specific expression (ASE) analysis on heterozygote sites in
97 diploid hybrids is considered as an effective method to identify *cis*-acting factors, as allelic
98 expressions are under the same feedback control and sharing non-*cis*-elements. Comparing the
99 allelic ratio between parental alleles and that in hybrids could distinguish the effect between
100 *cis*- and *trans*-factors (Li et al., 2017). Progress in sequencing technologies, next-generation
101 sequencing (NGS)-based approaches, such as RNA-Seq, enables analyzing ASE on a
102 transcriptome-wide scale. This strategy has been widely applied to yeast, fruit flies, and also
103 plants, including *Arabidopsis*, *Capsella*, *Atriplex*, maize, rice, millet (He et al., 2012; Lemmon
104 et al., 2014; McManus et al., 2010; Rhoné et al., 2017; Shao et al., 2019; Steige et al., 2015;
105 Sultmanis, 2018; Tirosh et al., 2009). Here we apply ASE analysis on interspecific hybrids of
106 parents displaying C_3 and C_3-C_4 intermediate photosynthesis.
107 The genus *Moricandia* provides an outstanding system to unravel the regulatory mechanisms
108 underpinning C_3-C_4 intermediacy. It comprises species with C_3 and C_3-C_4 photosynthesis
109 existing in a close phylogenetic proximity. *Moricandia* species share phylogenetic proximity
110 with species in other Brassicaceae genera, such as *Brassica* and *Diplotaxis*, as well as with
111 Arabidopsis, and were close to C_4 species in the family *Cleomaceae* (Kellogg, 1999; Beilstein
112 et al., 2010; Schlüter et al., 2017). To study the inheritance of *Moricandia* C_3-C_4 characteristics,
113 intergenic hybridizations of *Moricandia* with distant *Brassica* relatives have been reported in
114 the literature, through embryo rescue, sexual crosses and somatic hybridizations (list in
115 Warwick et al., 2009). Interspecific hybrids of *Moricandia* C_3 and C_3-C_4 species were
116 generated by *M. arvensis* as maternal and *M. moricandioides* as paternal species, of which the
117 phenotypic characterization is so far limited to records of CO_2 compensation points (Apel et
118 al., 1984). Most experimental hybrid studies were abandoned at that time, because of
119 reproductive disorders. However, genomic studies are facilitated now by the access to genetic
120 resources of closely related *M. moricandioides* (C_3) and *M. arvensis* (C_3-C_4) (Lin et al., 2021).
121 In this study, we assessed ASE on SNP- and transcript-level in *Moricandia* by means of RNA-
122 Seq on *M. arvensis* (C_3-C_4), *M. moricandioides* (C_3), and six of their interspecific F_1 hybrids.
123 Gene ontology assessments identified genes participating in chloroplast relocation and genes
124 demonstrating extreme allele imbalance. The spatial gene expression pattern of selected genes
125 flagged by ASE was validated by promoter-GUS analysis in *A. thaliana*. Our results indicate a
126 predominance of *cis*-regulatory effects on early evolutionary steps of C_4 photosynthesis.
127
128

129 **Results**

130 ***Moricandia* interspecific hybrids display phenotypes between that of parental genotypes**

131 For the analysis presented here, interspecific hybridization in *Moricandia* was performed using
132 *M. arvensis* (C_3-C_4) as maternal and *M. moricandioides* (C_3) as paternal species (Ma×Mm).
133 The reciprocal cross, Mm×Ma, had produced only one fifth of the seeds compared to Ma×Mm
134 hybridization. Additionally, the germination rate of seeds from Ma×Mm and Mm×Ma was
135 86% and 25%, respectively. In *M. arvensis* leaves, organelles are found in the BS and M cells
136 along the inner tangential walls and are abundantly accumulated toward veins in the BS cells
137 (Beebe & Evert, 1990; Schlüter et al., 2017). The same leaf anatomy was observed in this
138 study: chloroplasts are not only arranged on the inner wall of M cells, but also abundantly
139 accumulated towards veins in BS cells in *M. arvensis*. In contrast, only few chloroplasts were
140 found evenly distributed along the inner wall in BS cells, and some on the inner wall of M cells
141 in *M. moricandioides* (Figure 1). The CO_2 compensation points of *M. arvensis* and *M.*
142 *moricandioides* were 23.6±2.7 and 55.6±3.2 ppm, respectively, consistent with previous
143 studies (Apel, 1980; Bauwe & Apel, 1979; Schlüter et al., 2017). The *Moricandia* interspecific
144 hybrids displayed variation in their CO_2 compensation points, ranging from 39 to 55 ppm,
145 generally between the parental lines, but closer to that of C_3 species (Figure 2). These hybrids
146 further varied for the amount and arrangement of chloroplasts in BS cell: chloroplasts were
147 evenly distributed along the inner wall in BS cells of hybrid 1, 2, 5, 6, but only few chloroplasts
148 were found in BS cells of hybrid 3, 4 (Table 1, Supplementary file 1). The non-uniformity of
149 hybrids regarding the various phenotypic characteristics might be due to the heterozygosity
150 and chromosome mismatching of their parental species. Increased vein density compared to C_3
151 species was found in *Heliotropium* and *Flaveria* C_3-C_4 species (Muhaidat et al., 2011; Sage et
152 al., 2013), but not in *Moricandia* (Schlüter et al., 2017). The leaf venation was observed from
153 the top view of cleared leaves under the light microscope and the vein density was calculated
154 as the vein length per area. The vein density of *M. arvensis* was not significantly higher than
155 that of *M. moricandioides* but broader veins were observed in *M. arvensis* because of more
156 chloroplasts accumulating toward vascular bundles in BS cells (Supplementary file 1 and 2).
157 The vein density of hybrids showed no differences compared to parental species and the leaf
158 venation of hybrids was more similar to that of the C_3 parent with thinner veins corresponding
159 to their leaf anatomy (Supplementary file 2). They produced only very few F_2 seeds, likely
160 because of abnormal pollen produced by the F_1s, resulting in sterility of hybrids
161 (Supplementary file 3). Many interspecific or intergeneric hybrids were reported to be sterile
162 as a result of abnormal chromosome pairing or irregular meiotic division of pollen mother cells

163 (Apel et al., 1984; H. R. Brown & Bouton, 1993; Covshoff et al., 2014). Taken together, the
164 *Moricandia* interspecific hybrids demonstrated intermediate characteristics of CO_2
165 compensation points and leaf anatomy between that of C_3 and C_3-C_4 parents; however, they
166 more resembled the C_3 species. The vein width of hybrids was similar to that of C_3 plants.
167



168
169
170 **Figure 1.** Experimental design. The interspecific hybrids were obtained from the hybridization
171 of *M. arvensis* as maternal and *M. moricandioides* as paternal species. Leaf cross sections are
172 shown below the pictures of typical flowers of the parental lines. RNA-Seq was processed on
173 parents and selected hybrid lines, and further introduced to ASE analysis. Bar, 100 μm.
174

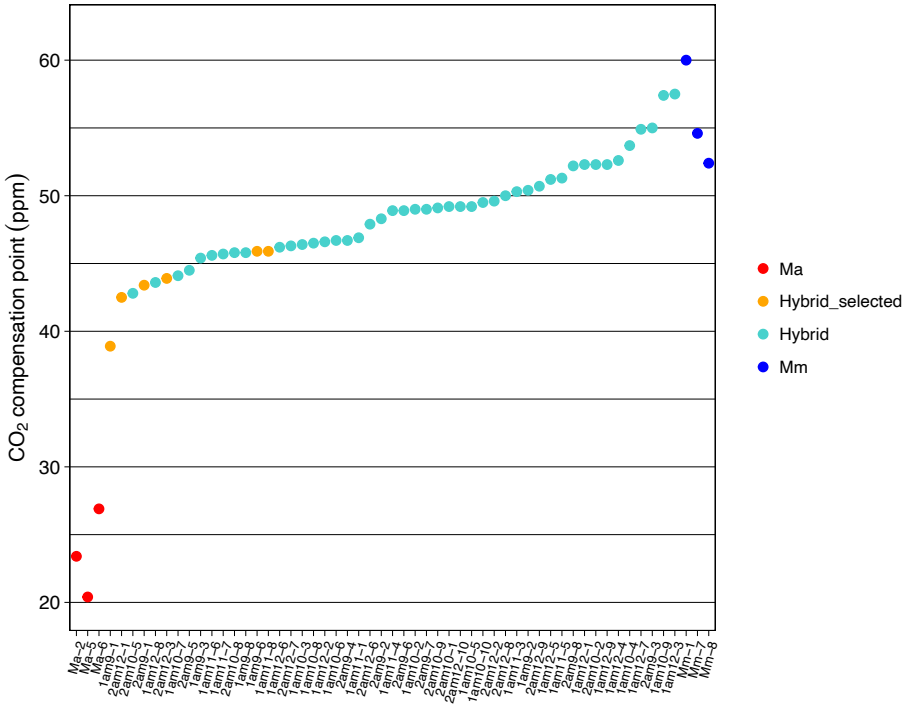


Figure 2. Distribution of CO_2 compensation points of *M. arvensis, M. moricandioides* and their interspecific hybrids. See ***Figure 2—source data 1*** for values of CO_2 compensation points. Ma, *M. arvensis*; Mm, *M. moricandioides*; Hybrid, in total 51 hybrids were tested for their CO_2 compensation point; Hybrid_selected, six interspecific hybrid lines with relative low value processed to RNA-Seq analysis.

Source data 1. CO_2 compensation points of *M. arvensis, M. moricandioides* and their interspecific hybrids.

Table 1. Phenotypic characterization of *Moricandia* parental species and their interspecific hybrids applied to RNA-Seq analysis. *M. arvensis* and *M. moricandioides* demonstrated typical C_3-C_4 and C_3 phenotypes, respectively. Interspecific hybrid lines indicated not uniform characteristics, generally intermediate between that of parents, but more resemble to C_3 parent.

Sample	Sample name	CO_2 compensation point (ppm)	Chloroplasts in BS cells
M. arvensis	Ma	24.99	Abundant, centripetally accumulated
M. moricandioides	Mm	54.91	Few, Centripetally and centrifugally accumulated
I am 9-1	Hybrid1	38.92	Centripetally and centrifugally accumulated
II am 12-1	Hybrid2	42.53	Centripetally and centrifugally accumulated
II am 9-1	Hybrid3	43.38	Nearly none
II am 12-3	Hybrid4	43.89	Nearly none
I am 9-6	Hybrid5	45.87	Centripetally and centrifugally accumulated
I am 11-8	Hybrid6	45.93	Centripetally and centrifugally accumulated

Transcript patterns showed no strong differences between C_3 and C_3-C_4 *Moricandia*, but transcripts of genes predicted to be involved in the glycine shuttle tended to be enhanced in the C_3-C_4 species

The transcriptome of *M. arvensis* (C_3-C_4) and *M. moricandioides* (C_3) was assembled using STAR v.2.5.2b with the draft genome of *M. moricandioides* serving as the reference (Lin et al., 2021). Three replicates from *M. arvensis* showed 66% mapping rate, and three replicates from *M. moricandioides* showed 94% mapping rate. Mapping rates of hybrids on *M. moricandioides* ranged from 61 to 81%. Principle component analysis (PCA) showed that the first principle component (PC1) explained 72% of the variance and clearly separated samples by species (Supplementary file 4A). PC2 underlined the separation of three replicates of *M. moricandioides* (Supplementary file 4A). The assessment of differential gene expression on 35,034 transcripts was performed with the DESeq2 tool (Love et al., 2014). Transcripts with a false discovery rate (FDR) ≤ 0.01, *P*-value adjusted with the BH procedure, were annotated as significantly differentially expressed. Using this definition, we found 3,491 transcripts that were significantly differentially expressed in *M. arvensis* and *M. moricandioides* leaves, where 2,712 transcripts were downregulated and 779 transcripts were upregulated in the C_3-C_4 species *M. arvensis*. GO terms such as metabolic process of small molecule, organic acid, and carbohydrate, transport of water and fluid, Golgi/endomembrane system organization were found in transcripts upregulated in C_3-C_4 *Moricandia* (Supplementary file 5). The downregulated transcripts encompassed significantly overrepresented GO terms, such as telomere maintenance, meiotic/nuclear chromosome segregation, chromosome organization, and regulation of organelle organization (Supplementary file 5).

The metabolic difference between C_3 and C_3-C_4 plants is predominantly caused by different intercellular arrangement of the photorespiratory process. Therefore, genes involved in pathways, such as glycine shuttle, C_4 cycle, Calvin-Benson cycle, and mitochondrial e^- transport, were screened for evidence of differential expression (Supplementary file 6). 12 out of 36 genes in the glycine shuttle were upregulated in C_3-C_4 *Moricandia*, such as *PGLP2*, *PLGG1*, *HPR1*, *SHMT2*, *GLDT*, *GLHD3*, *GS2*. 14 out of 59 genes in C_4 cycle were upregulated in C_3-C_4 *Moricandia*, such as *alpha CA1*, *gamma CA2*, *NADP-MDH*, *BASS2*, *AspAT2*, *NAD-ME1*, *PEPCK*. Out of 40 genes in the Calvin-Benson cycle, only *GAPA2* was upregulated in C_3-C_4 *Moricandia*. Out of 5 genes in mitochondrial e^- transport, *UCP1* and *UCP2* were upregulated in C_3-C_4 *Moricandia*. All in all, upregulated transcripts were overrepresented in glycine shuttle, C_4 cycles, and mitochondrial e^- transport relative to upregulated genes in all

222 pathways in C_3-C_4 intermediate species (Chi-squared test, adjusted P-value ≤ 0.01), supporting
223 the current models for glycine shuttle pathways in C_3-C_4 intermediates (Supplementary file 6).
224

225 **ASE analysis on SNP-level showed similar relative proportions of regulatory effects**
226 **among hybrids**

227 Gene expression regulation is governed by the interaction of distinct regulatory effects (*cis*-
228 and *trans*-acting factors). In hybrids, the two alleles inherited from the parental genotypes are
229 under the same cellular conditions and shared non-*cis*-elements. In this study, interspecific
230 hybrids of *M. arvensis* and *M. moricandioides* were utilized for discovering transcriptional
231 regulation through ASE analysis. Comparison of the allele ratio of characteristic SNPs between
232 hybrids and parents enabled us to distinguish between *cis*- and *trans*-regulatory effects (Li et
233 al., 2017). SNPs indicating *cis*-only regulatory effects (*cis*-SNP) are those for which the two
234 alleles are expressed unequally in hybrids and the allele ratio is the same between parents and
235 hybrids. SNPs with *trans*-only regulatory effects (*trans*-SNP) are defined by equal allele
236 expression in hybrid, but unequal in parents. To assess ASE between *M. arvensis* and *M.*
237 *moricandioides*, three replicates of each parental species and six interspecific hybrids were
238 selected for RNA-Seq (Figure 1). The ASE analysis on six hybrids was conducted individually
239 on a set of 147,883 SNPs in *Moricandia* (Figure 3—source data 1 and source data 2). The six
240 interspecific hybrids displayed similar relative proportions of regulatory effects (Figure 3). On
241 average, 36% of SNPs showed regulatory divergence by *cis*-only regulatory effect (*cis*-SNP),
242 and 7% of SNPs indicated *trans*-only effects (*trans*-SNP). Furthermore, 21% of SNPs indicated
243 mixed effects (*cis*- plus *trans*-SNP) and 37% of SNPs showed neither *cis*- nor *trans*-regulatory
244 effects (no *cis*- no *trans*-SNP). *GLDP1* of *Moricandia*, known for cell specific expression
245 between M and BS cells regulated by the M-box in the promoter region, was tagged by *cis*-
246 SNPs in all hybrids. Although relative proportions of regulatory effects were similar among
247 hybrids, most cases of ASE-SNPs were specific to individual interspecific hybrids. Around
248 8.7% of *cis*-SNPs were shared by all hybrids (3,142 transcripts harboring common *cis*-SNPs),
249 and only 1% of *trans*-SNPs (62 transcripts harboring common *trans*-SNPs) were shared by all
250 hybrids. Overall, *cis*-acting effects dominated the gene expression regulation in *Moricandia*
251 interspecific hybrids.

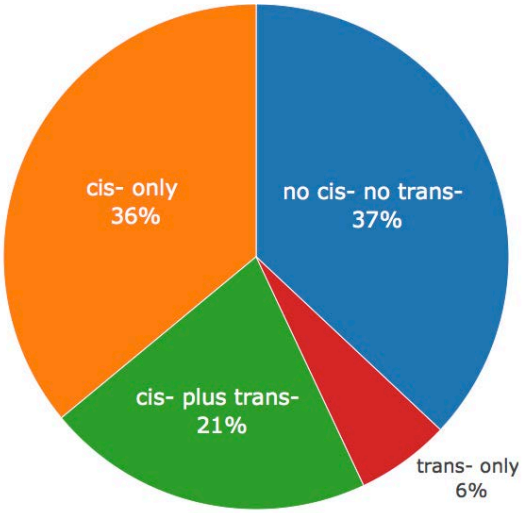


Figure 3. Average relative proportions of regulatory effects on SNP sites among six *Moricandia* interspecific hybrids. See ***source data 1*** for number of SNP sites defined with regulatory effects among six interspecific hybrids and ***source data 2*** for the raw data of ASE analysis based on SNP level.

Source data 1. Regulatory effects on SNP sites among six *Moricandia* interspecific hybrids.

Source data 2. Raw data of ASE analysis based on SNP level.

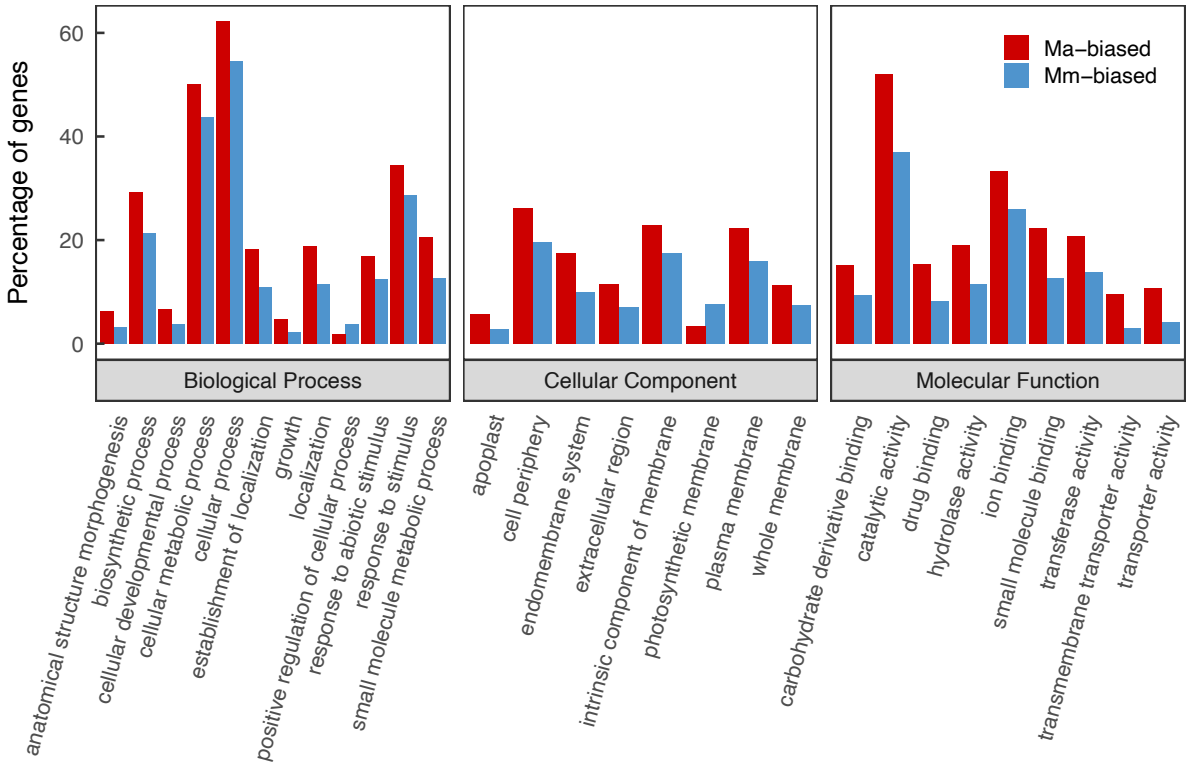
Gene ontology (GO) and pathway enrichment analysis on transcripts with common *cis*-SNPs and common *trans*-SNPs

GO and pathway enrichment analysis were used to annotate the functions of transcripts with common *cis*-SNPs and common *trans*-SNPs. A custom-mapping file was created, containing *Moricandia* transcript names and the corresponding GO terms derived from *A. thaliana* genes. The 2,236 and 45 *Moricandia* transcripts with common *cis*-SNPs and common *trans*-SNPs, respectively (Supplementary file 7), were processed with the custom mapping file by topGO R-package for gene set enrichment analysis of biological processes (Alexa et al., 2006). The top 30 most significantly enriched GO terms for transcripts with common *cis*-SNPs relate to isopentenyl diphosphate biosynthesis, carbohydrate catabolic process, oxidoreduction coenzyme metabolic process, and chloroplast relocation (Supplementary file 8). GO terms related to nucleosome assembly, RNA methylation, organophosphate biosynthetic process, and peptide metabolic process were abundant in transcripts with common *trans*-SNPs (Supplementary file 8).

To further decipher biosynthetic pathways in which transcripts with common *cis*-SNPs and common *trans*-SNPs participate, pathway enrichment analysis was conducted using the KEGG

277 Orthology Based Annotation System (KOBAS) (Xie et al., 2011). Transcripts with common
278 *cis*-SNPs were significantly ($P<0.05$) enriched in 27 pathways, including carbon metabolism,
279 protein processing in endoplasmic reticulum, carbon fixation in photosynthetic organisms,
280 porphyrin and chlorophyll metabolism, glyoxylate and dicarboxylate metabolism, and nitrogen
281 metabolism (Supplementary file 9). In contrast, transcripts with common *trans*-SNPs were
282 significantly enriched in 9 pathways, which were related to ribosome, carbon metabolism,
283 biosynthesis of amino acid and secondary metabolites, and fatty acid
284 metabolism/degradation/biosynthesis (Supplementary file 9).
285 Therefore, the GO and pathway enrichment results suggested that *Moricandia* transcripts with
286 *cis* mechanisms play a more prominent role in C_3-C_4 related functions such as major
287 photosynthetic pathways and chloroplast relocation, whereas transcripts with *trans*
288 mechanisms are involved in more general biological pathways.
289
290 **Transcripts with *cis*-specificity showed biased expression toward C_3-C_4 species in**
291 ***Moricandia***
292 *cis*-regulatory divergences generally dominate adaptive evolution because they tend to cause
293 fewer deleterious pleiotropic effects than nonsynonymous mutations in protein-coding
294 sequences (Stern & Orgogozo, 2008; Wittkopp & Kalay, 2012; Wray, 2007). Further, they
295 frequently cause altered spatiotemporal gene expression patterns (Prud'homme et al., 2007).
296 The compartmentation of CO_2 assimilatory enzymes between BS and M cells in C_4 plants
297 results from modifications in regulatory sequences (Gowik et al., 2004; Sheen, 1999). Thus,
298 genes with *cis*-specificity (*cis*-SNPs or *cis*- plus *trans*-SNPs) were candidates for selections of
299 direct targets or promotions of spatial gene expression during C_4 evolutionary trajectories. In
300 hybrid 1, we observed 4,684 *cis*-specificity SNPs (1,105 transcripts) expressed toward *M.*
301 *arvensis* (Ma-biased) and 3,871 SNPs (820 transcripts) expressed toward *M. moricandioides*
302 (Mm-biased). Similar proportions were observed in the other interspecific hybrids
303 (Supplementary file 10). In total, there were 513 Ma-biased and 326 Mm-biased transcripts
304 with *cis*-specificity shared by all hybrids.
305 To understand the gene function of biased transcripts with *cis*-specificity, the common Ma-
306 biased and common Mm-biased transcripts were further investigated by GO term classification
307 using Web Gene Ontology Annotation (WEGO) software (Ye et al., 2018). These transcripts
308 were classified into GO terms under biological process, cellular component, and molecular
309 function categories (Figure 4). There were more transcripts enriched in GO terms, such as
310 anatomical structure morphogenesis, biosynthetic process, localization, endomembrane system,

311 catalytic activity, and transmembrane transporter activity, in common Ma-biased than in
312 common Mm-biased transcripts. Additionally, under anatomical structure morphogenesis,
313 compared to common Mm-biased transcripts, more common Ma-biased transcripts were
314 enriched in leaf morphogenesis. Therefore, this directional expression bias indicated that *cis*-
315 specificity caused enhanced expression of C_3-C_4 genes involved in leaf anatomical structure
316 organization and activity of transmembrane transporters, which could be involved in shaping
317 C_3 to C_3-C_4 photosynthesis.



318

319 **Figure 4.** GO terms comparison between transcripts with *cis*-specificity biased toward *M.*
320 *arvensis* (Ma-biased) and *M. moricandioides* (Mm-biased). All GO terms shown here were
321 significantly different in gene percentages between Ma-biased and Mm-biased transcripts.

322

323 **Evaluating ASE on transcript-level in *Moricandia* interspecific hybrids**

324 Assessing ASE on SNP-level allows assigning SNPs into four categories of regulatory effects.
325 However, SNPs within a transcript might indicate different regulatory effects. For instance,
326 *GLDP1* in hybrid 1 revealed 5 *cis*-SNPs, 16 *cis*-plus *trans*-SNPs, and 2 *trans*-SNPs. Therefore,
327 studies evaluating ASE on SNP-level require either phased information of read counts at SNPs
328 from genomic data of hybrids (He et al., 2012; Rhoné et al., 2017; Skelly et al., 2011; Steige
329 et al., 2015) or an agreement across SNPs in the same transcript (Shao et al., 2019). In our
330 study, the genomic information of hybrids was not available. Therefore, we followed the

331 second strategy and applied a meta-analysis based allele-specific expression detection
332 (MBASED). This method assigns the allele with higher read counts at each SNP to the major
333 allele as a pseudo-phasing approach and therewith assumes a consistent direction of ASE
334 within a transcript. The ASE level of transcripts was estimated with the major allele frequency
335 (MAF), ranging from 0.5 to 1.0 (Mayba et al., 2014). The distribution of MAF was right-
336 skewed with the mode of 0.6, implying that most of transcripts showed mild allelic imbalance
337 (Figure 5). Transcripts with MAF ≥ 0.7 and adjusted *P*-value ≤ 0.05 were defined as ASE-
338 transcripts. Across the six interspecific hybrids, around 27% of examined transcripts showed
339 evidence of ASE, where 3% of them had extreme allelic imbalance (MAF ≥ 0.9 and adjusted
340 *P*-value ≤ 0.05) (Figure 5—source data 1). *GDLP1* in *Moricandia*, known for cell specific
341 expression regulated through *cis*-regulatory elements, showed an MAF of 0.76 on average
342 across the six hybrids (Figure 5—source data 2). Interestingly, some transcripts showed
343 extreme allele bias to one of the parental species in hybrids (Supplementary file 11). For
344 instance, *MSTRG.16015* encoding a putative chloroplast RNA binding protein showed an
345 average MAF of 0.96 toward *M. arvensis* allele, and *MSTRG.5109* encoding ATP synthase
346 subunit d revealed an average MAF of 0.96 biased to the *M. arvensis* allele (Figure 5—source
347 data 2). Overall, 27% of assayed transcripts were defined as ASE-transcript and a group of
348 transcripts had strong allelic imbalance in hybrids.
349

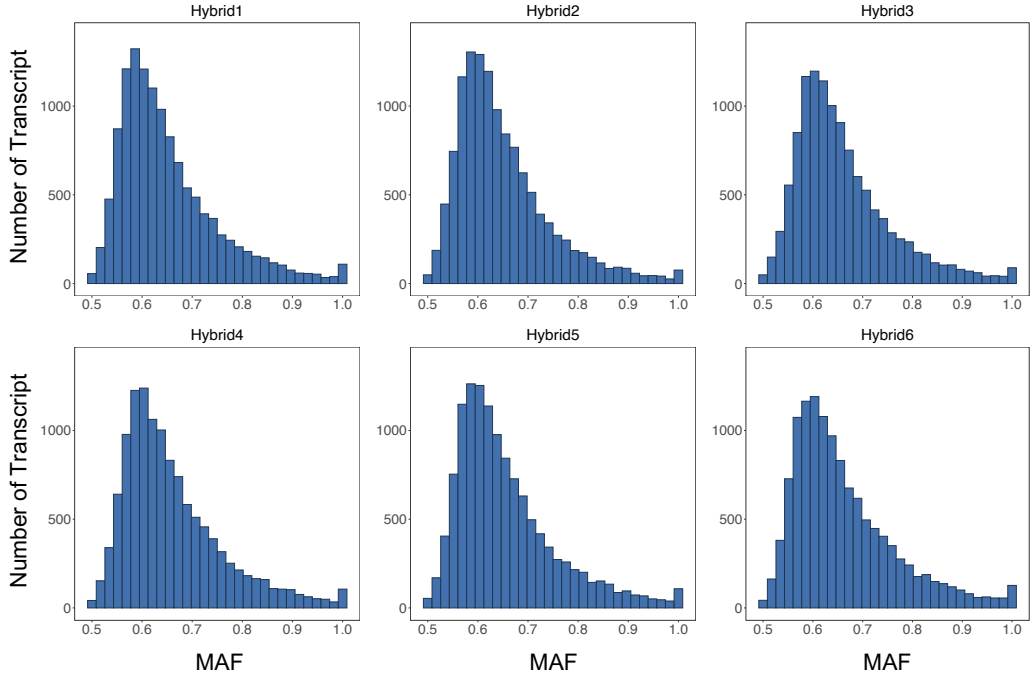


Figure 5. Distribution of allelic imbalance for all assayed transcripts among six *Moricandia* hybrids. The major allele frequency (MAF) represented the intensity of allelic imbalance in hybrids, obtained by MBASED (Mayba et al., 2014), using the allele with more read counts as major allele. See ***source data 1*** for number of assayed transcripts among six hybrids and ***source data 2*** for the list of transcripts with MAF among six hybrids and ***source data 3*** for the ***r***aw data of ASE analysis based on transcript level. Transcripts with MAF ≥ 0.7 and adjusted p-value ≤ 0.05 were defined as ASE-transcripts. Around 27% of transcripts was with MAF ≥ 0.7 and 3% of them showed MAF ≥ 0.9.

Source data 1. Allelic imbalance for all assayed transcripts among six *Moricandia* hybrids.

Source data 2. The major allele frequency of six *Moricandia* hybrids.

Source data 3. Raw data of ASE analysis based on transcript level.

Enrichment of regulatory divergences in selected pathways

Most of *Moricandia* transcripts indicated no significant differential expression in comparative transcriptome studies between the parental species using total leaf extracts. However, *cis*-acting factors, regulating spatiotemporal gene expression and transcriptional abundance, play a crucial role in adaptive phenotype evolution (Lemmon et al., 2014; Wray, 2007). Therefore, we examined the enrichment of regulatory effects on transcripts in selected pathways through detecting *cis*-SNPs and evaluating allelic imbalance in transcripts. Out of selected 140 genes, the expression of 105 genes in glycine shuttle, C_4 cycle related, and Calvin-Benson cycle

372 indicated altered *cis* regulation with the evidence of at least one *cis*-SNP or common *cis*-SNP
373 (Supplementary file 12). In addition, the ASE evaluated on transcript-level demonstrated the
374 intensity of allelic imbalance in hybrids. Transcripts with major alleles frequency ≥ 0.7 were
375 considered to possess ASE. Out of 41 genes in the C_3-C_4 glycine shuttle, 32 genes indicated
376 allele specific expression, such as *PLGG*, *GOX*, *GGAT*, *SHMT*, *GDC* complex, *DIT*, and *GS2*
377 (Figure 6A). Furthermore, C_4 cycle genes, such as *CA*, *PEPC*, *PPT*, *NADP-MDH*, *DIT*, *NADP-*
378 *ME*, *BASS2*, *PPDK*, *PPT*, *AspAT*, and *PEPCK*, and Calvin-Benson cycle genes, except *SBPase*,
379 indicated ASE (Figure 6B and 7). Taken together, most genes involving in selected pathways
380 revealed no strong differential gene expression between C_3 and C_3-C_4 species in *Moricandia*,
381 however regulatory divergences play an important role in different photosynthetic types by
382 ASE of critical genes involving in early C_4 photosynthesis evolution.
383

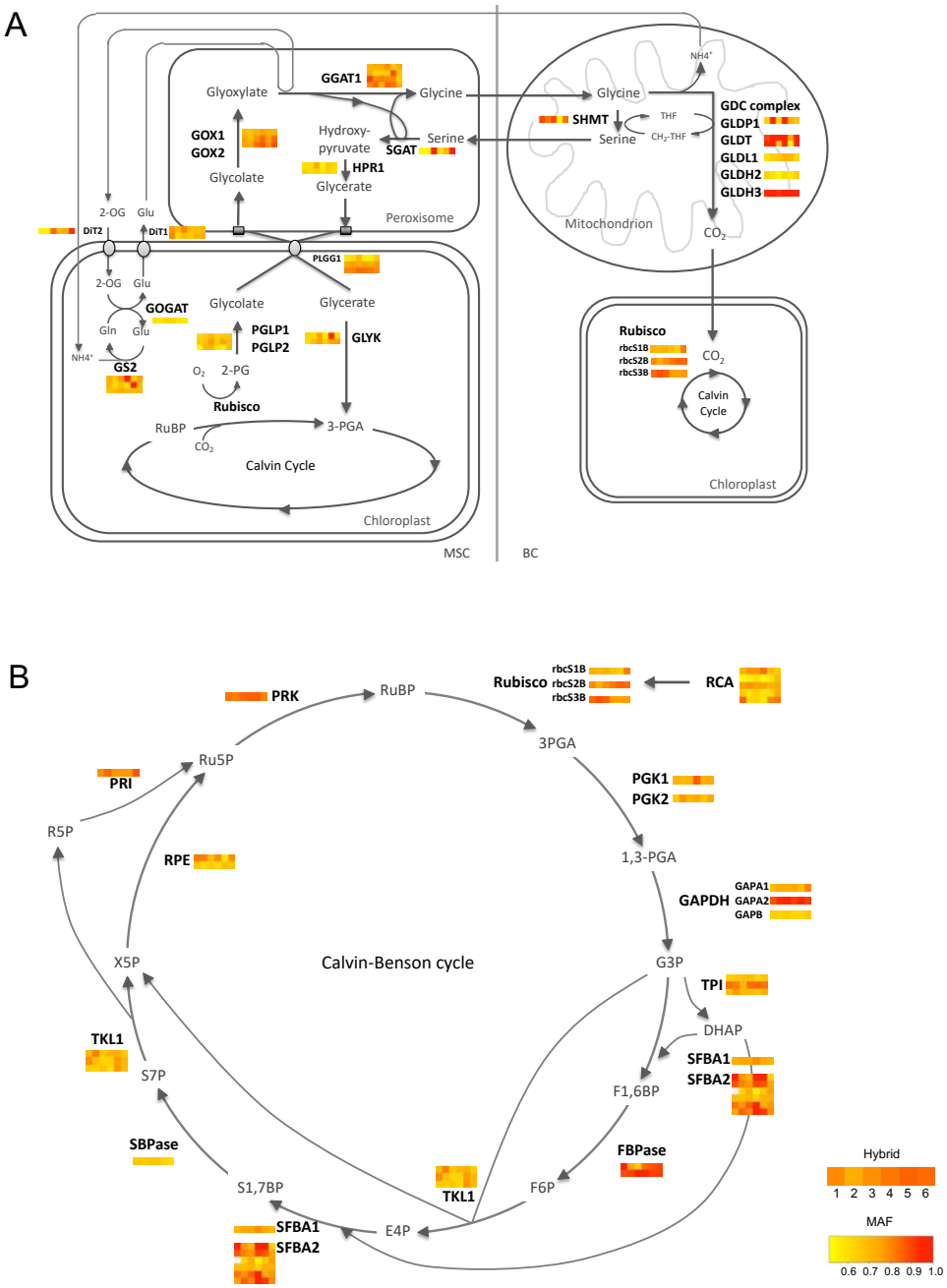


Figure 6. Overview on the allelic imbalance of genes involving in selected pathways of the C_3-C_4 glycine shuttle (A) and the Calvin-Benson cycle (B) among six *Moricandia* hybrids. The six blocks in each gene bar presented the major allele frequency (MAF) of hybrid 1 to hybrid 6 from left to right.

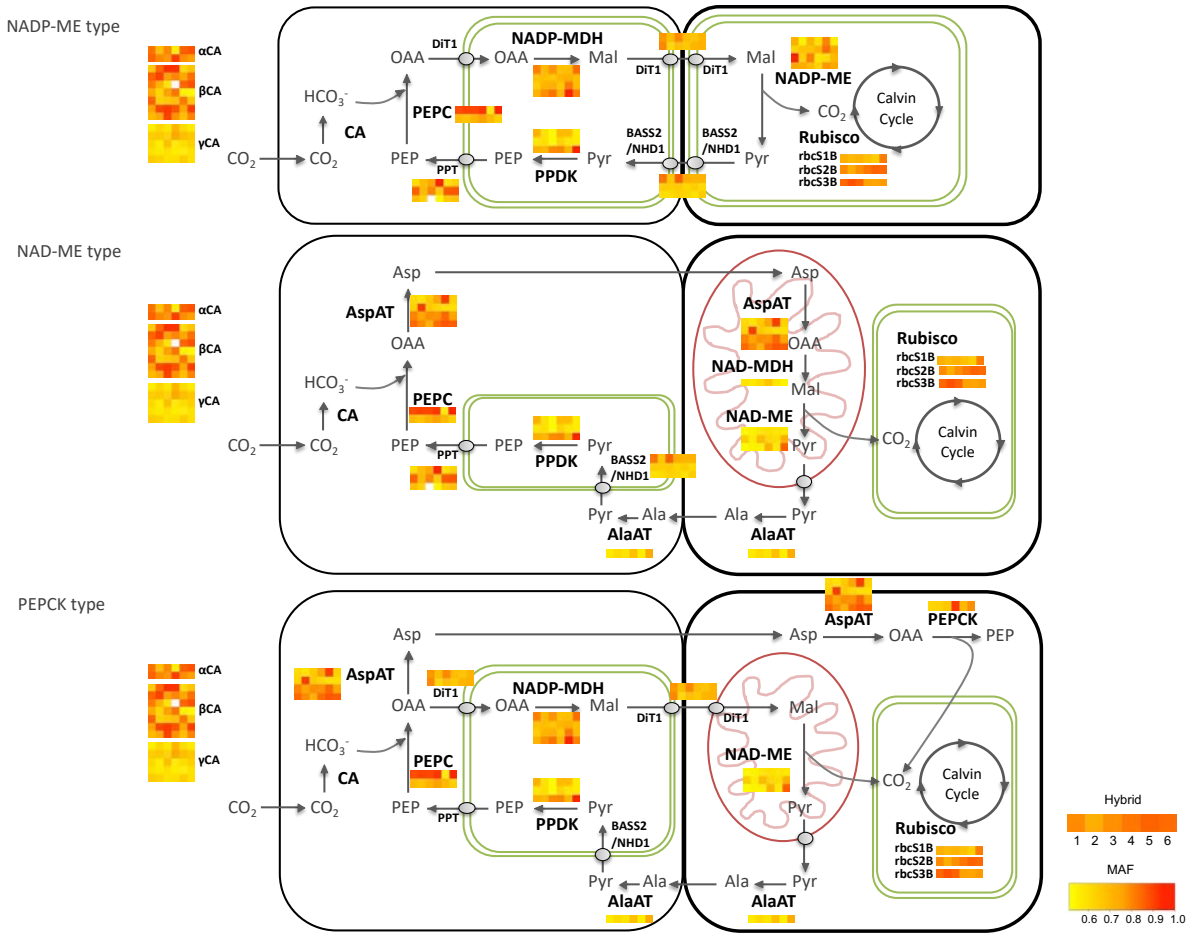


Figure 7. Overview on the allelic imbalance of C_4 cycle genes among six *Moricandia* hybrids. The six blocks in each gene bar presented the major allele frequency (MAF) of hybrid 1 to hybrid 6 from left to right.

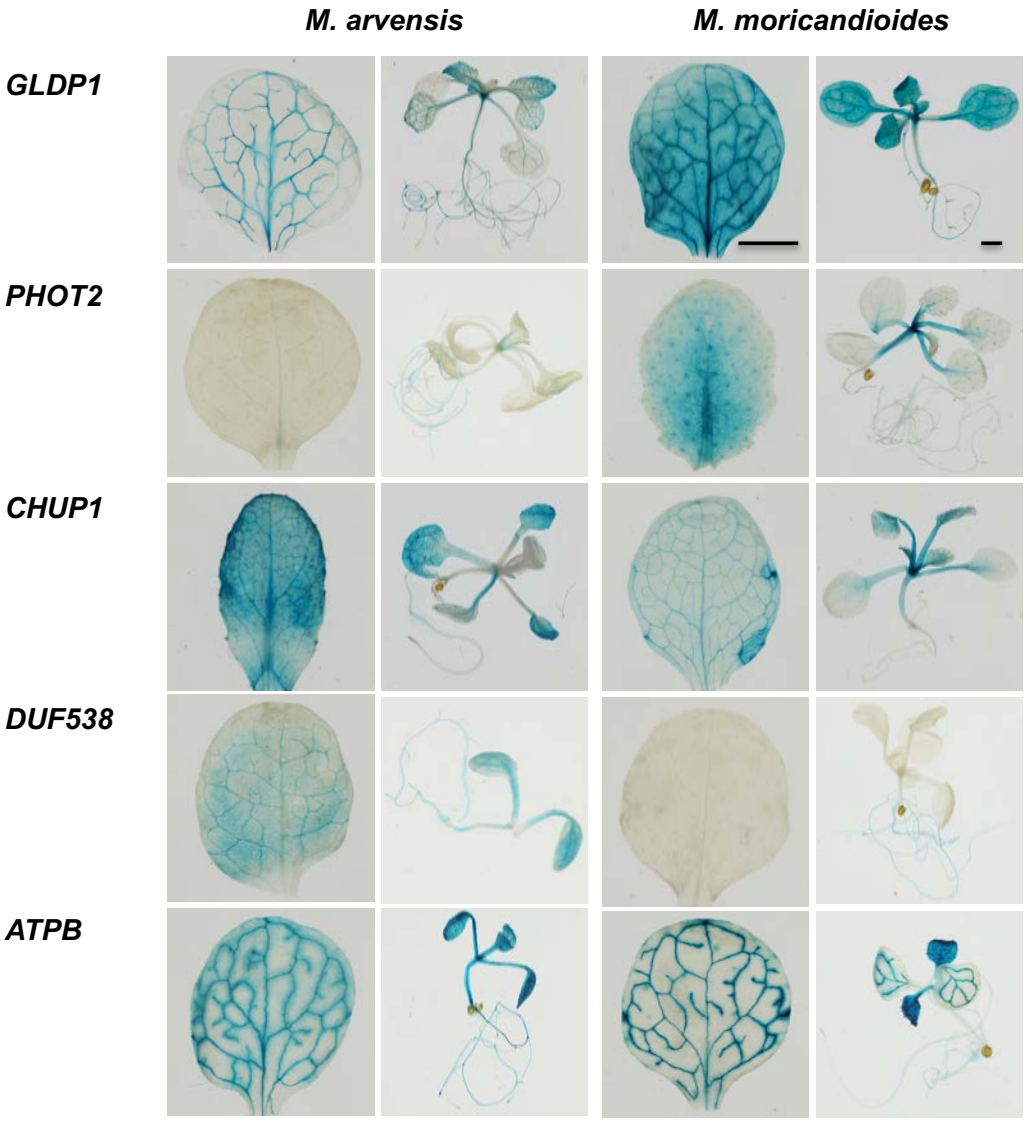
Promoter-GUS assay on selected genes confirmed the ASE result

ASE transcripts were expected to show differences in transcriptional abundance or spatiotemporal gene expression between C_3 and C_3-C_4 *Moricandia* species. To test these hypotheses, promoter-GUS assays were used. The selected genes for this assay were (1) transcripts with common *cis*-SNPs enriched in GO term chloroplast relocation (GO:0009902), and (2) ASE-transcripts with high allelic imbalance toward *M. arvensis* (MAF $\geq$ 0.8) across all six hybrids. GLDP1 localized exclusively to the BS cells of the leaf of C_3-C_4 *Moricandia* species, which is most likely regulated on transcriptional level (Adwy et al., 2019; Rawsthorne et al., 1988; Rawsthorne et al., 1988). *GLDP1* possessed *cis*-SNPs across all hybrids and an average MAF of 0.76 (Supplementary file 12; Table 2). *PHOT2* and *CHUP1* selected from the GO term chloroplast relocation had common *cis*-SNPs among six hybrids and an average MAF

410 of 0.62 and 0.72, respectively. *DUF538* and *ATPB* were selected from transcripts with high
411 allelic imbalance toward *M. arvensis*: *DUF538* encoding an unknown function protein showed
412 an average MAF of 0.83; *ATPB* was annotated as ATP synthesis coupled proton transport with
413 an average MAF of 0.96 (Supplementary file 11; Table 2).
414 Approximately 2 kb upstream promoter region of genes from *M. arvensis* and *M.*
415 *moricandioides* were amplified, fused with GUS reporter gene, and the recombinant constructs
416 were introduced into *A. thaliana* (Supplementary file 13). GUS staining demonstrated the
417 spatial gene expression difference between *M. arvensis* and *M. moricandioides* (Figure 8).
418 *GLDP1*, as control, showed different cell-specific regulation between *Moricandia* species.
419 *MaGLDP1* promoter drove GUS expression in cells surrounding veins and GUS staining of
420 that of *M. moricandioides* (*MmGLDP1*) was observed on the whole leaf. GUS expression
421 driven by *MaPHOT2* promoter was observed in roots and slightly in shoots of two-week-old
422 seedlings. However, that of *MmPHOT2* was detected predominantly surrounding the leaf mid
423 rib, trichomes in leaves, and shoots. pMaCHUP1::GUS expression was detected in the whole
424 leaf and slightly in shoots, where GUS expression driven by *MmCHUP1* promoter was stronger
425 in the veins and shoots. *MaDUF538* promoter drove GUS expression in parts of the leaf blade,
426 shoots, and roots; in contrast, GUS expression driven by *MmDUF538* promoter was only
427 detected in roots. *MaATPB* promoter resulted in vein-, root-, and shoot-preferential GUS
428 expression veins; GUS staining of p*MmATPB*::GUS transgenic plants was detected also in
429 leaves, but with different expression pattern between leaves. Thus, with the exception of *ATPB*,
430 promoter-GUS assay could associate ASE on SNP- and transcript-level with specific spatial
431 gene expression.
432
433 **Table 2** List of selected genes in *Moricandia.*

Gene	Gene model	Transcript	Regulory effect	Major allele frequency from Hybrid1 to Hybrid6	GO Biological Process
GLDP1	*AT4G33010*	*MSTRG.30006*	common *cis*-SNP	0.64; 0.96; 0.62; 0.95; 0.72; 0.66	glycine catabolic process
PHOT2	*AT5G58140*	*MSTRG.699*	common *cis*-SNP	0.63; 0.62; 0.63; 0.63; 0.60; 0.61	response to blue light, phototropism, chloroplast relocation
CHUP1	*AT3G25690*	*MSTRG.23629*	common *cis*-SNP	0.78; 0.73; 0.68; 0.74; 0.68; 0.69	chloroplast relocation
DUF538	*AT1G09310*	*MSTRG.15469*	*cis*- plus *trans*-SNP	0.86; 0.79; 0.83; 0.82; 0.82; 0.84	protein of unknown function
ATPB	*ATCG00480*	*MSTRG.7892*	*cis*- plus *trans*-SNP	0.95; 0.90; 1.00; 0.95; 0.97; 1.00	ATP synthesis coupled proton transport, response to salt stress

434



435
436
437 **Figure 8.** Promoter-GUS assay of *Moricandia* selected genes expressed in *A. thaliana*. *GLDP1*
438 served as the positive control, possessing *cis*-SNP and defined as ASE-transcript with major
439 allele frequency (MAF) ≥ 0.7; *PHOT2* and *CHUP1* had common *cis*-SNP among hybrids;
440 *DUF538* and *ATPB* demonstrated strong allelic imbalance with MAF ≥ 0.8. Bar, 1 mm.
441

442 # Discussion

443 **ASE analysis is a powerful strategy for understanding the evolution of C_3-C_4**
444 **photosynthesis**

445 The evolution of C_4 photosynthesis via intermediate forms was achieved by regulatory
446 diversions from the ancestral C_3 leaf metabolism (Heckmann, 2016; Reeves et al., 2017). The
447 regulatory mechanisms realising the C_4 specific expression patterns have however only been
448 studied for selected genes, mainly from the C_4 shuttle pathway (Schlüter und Weber 2020).
449 Our knowledge about the genetic and molecular background of C_3-C_4 intermediacy is so far
450 limited to regulatory elements in the promoter of the *GLDP* gene (Adwy et al., 2019; Schulze
451 et al., 2013).

452 By production of interspecific hybrids between C_3-C_4 *M. arvensis* and C_3 *M. moricandioides*,
453 novel insights into the regulatory divergences of photosynthetic systems in *Moricandia* are
454 now obtained by ASE on SNP and transcript level, and numerous *cis*-regulatory variations
455 between C_3-C_4 and C_3 *Moricandia* species were discovered. Transcripts harboring common
456 *cis*-SNPs dominated the major photosynthetic pathways and chloroplast relocation between
457 C_3-C_4 and C_3 *Moricandia* species.

458 Whole leaf transcriptome analyses of C_3 and C_3-C_4 *Moricandia* species identified only very
459 few differentially expressed genes. The *GLDP* gene, previously shown to display different
460 spatial expression patterns in leaves from the two photosynthesis types was not flagged in
461 whole-leave differential gene expression analysis (Schlüter et al., 2017). To the contrary, the
462 ASE analysis reported here clearly identified *cis*-derived effects for *GLDP1* in all hybrids.
463 Additional genes connected to the glycine shuttle (*GOX*, *GGAT*, other *GDC* complex subunits,
464 *SHMT*, *GS2*, *DiT*, *AspAT*) contained also at least one *cis*-SNP among hybrids and demonstrated
465 strong ASE (Supplemental Table 11; Figure 6A). Of the roughly 12,400 *Moricandia*
466 transcripts, 27% showed evidence of ASE and 3% demonstrated extreme allelic imbalance.
467 The results support our current hypothesis that the installation of the glycine shuttle requires
468 regulatory adjustments in the whole photorespiratory pathway, but also adjacent pathways such
469 as the nitrogen metabolism, Calvin Benson cycle, redox metabolism and transport processes
470 between cellular organs as well as mesophyll and bundle sheath (Mallmann et al., 2014;
471 Schlüter et al., 2017). Candidate genes from all these pathways showed evidence of *cis*
472 regulation in the investigated hybrids (e.g. *CA*, *PEPC*, *PPT*, *NADP-MDH*, *DIT*, *NADP-ME*,
473 *BASS2*, *PPDK*, *PPT*, *AspAT*, *PEPCK* and many Calvin Benson cycle genes).

474 Beside leaf biochemistry, the installation of an efficient glycine shuttle is based on general
475 activation of BS metabolism, as well as BS-specific changes in organelle number and
476 localization (Lundgren, 2020). ASE analysis identified numerous genes predicted to be
477 involved in chloroplast movement and localization. Consistent with this observation, promoter
478 GUS studies with two candidate genes involved in chloroplast positioning within the cell
479 (*PHOT2* and *CHUP1*) revealed differences in their expression pattern in C_3 and C_3-C_4 leaves.
480 Chloroplast movements are important for adjustment of photosynthesis to different light
481 conditions, the differential regulation of the two proteins in C_3 and C_3-C_4 *Moricandia* species
482 suggest they also play a role in BS specific chloroplast positioning.
483 The lack of knowledge about the mechanisms underpinning the C_3-C_4 and also C_4 specific
484 anatomical features represents a serious bottleneck for the engineering of more efficient
485 photosynthetic pathways into C_3 crop plants. ASE analysis of interspecific hybrids could help
486 to identify key players of these anatomical changes, especially when expanded to different
487 stages of leaf development.
488
489 ***Moricandia* interspecific hybrids showed C_3 parental-like phenotypes**
490 C_3-C_4 × C_3 *Moricandia* hybrids were successfully created, but they were sterile most likely
491 because of chromosome mismatching and irregular meiotic division of pollen mother cells.
492 Similar problems were observed in trials using species in *Panicium* and *Flaveria* (Bouton et
493 al., 1986; H. R. Brown & Bouton, 1993). However, interspecific hybrids of *A. prostrata* (C_3)
494 and *A. rosea* (C_4) were fertile (Oakley et al., 2014). The six hybrids in this study demonstrated
495 intermediate, non-uniform phenotypic characteristics between those of the parents. The CO_2
496 compensation points and leaf anatomy were thereby more similar to that of the C_3 parent
497 (Figure 2; Table 1; Supplementary file 1 and 2). Additionally, a PCA showed that the gene
498 expressions of *Moricandia* interspecific hybrids were also closer to that of C_3 parents compared
499 to that of the C_3-C_4 parent (Supplementary file 4B). Our finding was in accordance with
500 intermediate phenotypic characteristics of CO_2 compensation points and organelle quantities
501 in BS cells observed in interspecific hybrids of C_3-C_4 and C_3 *Panicum* species and intergeneric
502 hybrids of *M. nitens* (C_3-C_4) × *Brassica napus* (C_3) (Brown et al., 1985; Rawsthorne et al.,
503 1998). The leaf anatomy of hybrids from *M. nitens* (C_3-C_4) × *B. napus* (C_3) resembled that of
504 the C_3 parent (Rawsthorne et al., 1988). The increase in C_3-C_4 proportion also increased the
505 intermediate phenotype. The dominance of C_3-C_4 phenotypes (CO_2 exchange and confined
506 GDC in BS cells) increased with the C_3-C_4 genome constitution in *D. tenuifolia* (C_3-C_4) × *R.*
507 *sativus* (C_3) hybrids and *M. arvensis* (C_3-C_4) × *B. oleracea* (C_3) and their reciprocal crosses

508 (Ueno et al., 2003, 2007), and the same phenomenon was found in backcrosses of *B. alboglabra*
509 (C_3) × *M. arvensis* (C_3-C_4) to the C_3-C_4 parent (Apel et al., 1984). These findings, together with
510 our observation, suggest that genes regulating C_3-C_4 characteristics showed additive effects in
511 C_3-C_4 × C_3 hybrids. For the *GLDP* gene, the C_3 copy for instance could be responsible for
512 almost C_3-like expression in the M. Thereby, *Moricandia* C_3-C_4 × C_3 hybrids more resembled
513 C_3 parent, because C_3-C_4 species had also the C_3 genetic background.
514
515 ***Cis* mechanisms play a major role in evolution of C_3-C_4 intermediacy**
516 *Cis*-regulatory effects presented larger impacts than *trans*-acting divergences on *Moricandia*
517 interspecific hybrids: 36% and 6% of assayed SNPs were discovered as *cis*-SNP and *trans*-
518 SNP, respectively. Our results are in accordance with previous findings showing that *cis*-
519 regulatory changes are more prevalent in interspecific hybrids (long evolutionary time-scales),
520 whereas *trans*-regulatory divergence is more frequently observed in intraspecific hybrids (short
521 evolutionary time-scales) (McManus et al., 2010; Rhoné et al., 2017; Stern & Orgogozo, 2008).
522 The dominance of *cis*-regulatory divergence was also shown in ASE studies on interspecific
523 hybrids, such as in *Atriplex* and poplar (Sultmanis, 2018; Zhuang & Adams, 2007). The relative
524 proportions of regulatory effects were similar in all six Ma×Mm hybrids, but many of them
525 were sample-specific. Only 8.7% of *cis*-SNPs and 1% of *trans*-SNPs were found consistently
526 across all interspecific hybrids. It was also described in the literature that ASE genes are often
527 unique in different hybrids (Lemmon et al., 2014; Steige et al., 2015). This finding might
528 explain that the hybrids didn't show uniform phenotypic characteristics, and together with the
529 sterility of hybrids, it supports the problems with different chromosome arrangement between
530 the parents.
531 *cis*- and *trans*-regulatory divergences have different impacts on the inheritance and evolution
532 of gene expression, and hence also on different biological processes (McManus et al., 2010;
533 Meiklejohn et al., 2014). This is related to the genetic nature of transcriptional regulations: *cis*-
534 regulatory sequences, located in promoter regions, UTRs, and introns, modulate the binding of
535 *trans*-acting factors to DNA, therefore affecting the transcription of nearby genes, whereas
536 *trans*-element, such as transcription factors and long noncoding RNA, are able to affect the
537 expression of many genes (Wray, 2007). In our study, ASE would detect regulatory differences
538 in all affected pathways. GO and pathway enrichment analysis on transcripts with common
539 *trans*-SNPs indicated very general biological pathways. Transcripts with common *cis*-SNPs on
540 the other hand showed enrichment in major photosynthetic pathways and chloroplast
541 relocation, thus indicating that the differences in the parental photosynthesis types were *cis*-

542 regulated. For instance, transcripts with common *cis*-SNP were overrepresented in isopentenyl
543 diphosphate biosynthesis. In higher plants, the formation of isopentenyl diphosphate, the
544 central intermediate of all isoprenoids, was compartmentalized, such as sterols in cytosol or
545 carotenoid, phytol, and chlorophyll in plastids (Lange and Croteau, 1999), which is likely
546 regulated by *cis*-acting divergences. In the present study, *cis*-specificity accumulated in the
547 gene category of chloroplast relocation and anatomical structure morphogenesis in
548 *Moricandia*. The anatomical adjustment, such as centripetal accumulation of mitochondria and
549 chloroplasts in BS cells, was a prominent characteristic for reduced compensation points in C_3-
550 C_4, resulting from the glycine shuttle (Schlüter et al., 2017). Regulatory divergences were for
551 instance found for two genes responsible for chloroplast localization within the cells. *PHOT2*
552 mediates blue light responses and is involved in phototropism, chloroplast movement, stomatal
553 opening, leaf development and photosynthetic efficiency (Hart et al., 2019). *CHUP1* is
554 involved in chloroplast positioning within the cell and interaction with the actin cytoskeleton
555 (Oikawa et al., 2008). The expression of both genes was tested by promoter-GUS assay in
556 Arabidopsis and different patterns could be observed for the Ma and Mm derived promoters.
557 Of course, elements in untranslated gene regions (5'UTR and 3'UTR), coding regions, and
558 introns have been reported to affect steady-state transcript amounts could be active in addition
559 to the promoter region (Barrett et al., 2012; Hernandez-Garcia and Finer, 2014).
560 *cis*-regulatory divergences are additive in heterozygotes, which were visible and preferentially
561 accumulated during evolution, therefore playing a crucial role in evolution of adaptive traits
562 (Lemos et al., 2008; Meiklejohn et al., 2014; Wittkopp et al., 2008). In our study, transcripts
563 with *cis*-specificity showed preferential expression toward C_3-C_4 *M. arvensis* alleles in hybrids
564 (Supplementary file 10), although the phenotype of hybrids more resembled that of the C_3
565 parent. During evolution of the selfing syndrome in *Capsella*, alleles from the self-compatible
566 species *C. rubella* were expressed at higher level than the outcrossing *C. rubella* alleles in their
567 hybrids (Steige et al., 2015). *cis*-regulatory divergences dominated the positive selection and
568 the adaptive improvement during maize domestication from teosinte, and genes with *cis*-
569 regulatory effect demonstrated a directional expression bias toward maize (Lemmon et al.,
570 2014). An expression preference was however not found in $C_3 \times C_4$ *Atriplex* hybrids
571 (Sultmanis, 2018). Transcripts with *cis*-specificity showing a directional expression bias
572 toward *M. arvensis* were more abundant in GO terms, involving in anatomical structure
573 morphogenesis, transmembrane transporter activity and localization, compared to transcripts
574 expressed biased to *M. moricandioides*. The installation of the glycine shuttle could therefore

575 be associated with *cis*-mediated upregulation of genes involved in leaf anatomy and transport
576 activities.
577 Our study suggests that many *cis*-regulatory effects, favored in adaptive phenotypic traits
578 during evolution, were additive in C_3-C_4 × C_3 *Moricandia* hybrids. This is consistent with the
579 previously predicted long and smooth path of C_4 photosynthesis evolution (Heckmann et al.,
580 2013; Williams et al., 2013).
581

582 ## Conclusion

583 Interspecific hybrids between C_3 and C_3-C_4 *Moricandia* species possessed phenotypic
584 characteristics of CO_2 compensation point and chloroplast accumulation in BS cells between
585 those of the parental species, however more resembling those of the C_3 parent. We showed that
586 *cis*-regulatory divergences have a large impact on *Moricandia* interspecific hybrids, and the
587 corresponding transcripts were found to be enriched in major photosynthetic pathways and
588 chloroplast relocation. We further observed that *cis*-specificity caused enhanced expression of
589 C_3-C_4 alleles in categories such as anatomical structure morphogenesis and transmembrane
590 transporter activity. *cis* mechanisms contributed to the installation of the glycine shuttle,
591 playing an important role in the early evolutionary steps of C_4 photosynthesis. With the genetic
592 information of parental species, the RNA-Seq dataset and ASE approaches, we investigated
593 *cis*- and *trans*-acting divergences on a transcriptome-wide scale, which helps us to understand
594 the role of transcriptional regulations during evolution of C_4 photosynthesis. It also provides
595 possible targets for engineering C_3-C_4 characteristics into C_3 plants.
596

597 ## Materials and Methods

598 **Plant materials**

599 Seeds from *Moricandia* were surface-sterilized using chloride gas and germinated on half MS
600 medium for one week. The seedlings were then transferred individually to pots with soil and
601 grown in the growth chamber under 12h/12h light/dark conditions with 23°C/20°C day/night.
602 The anthers of *M. arvensis* (IPK Gatersleben: MOR1, a C_3-C_4 intermediate, as maternal plant)
603 were removed and their stigma was bagged one day before the artificial cross-pollination. The
604 pollen from *M. moricandioides* (Botanical Garden Osnabrück: 04-0393-10-00, a C_3 species, as
605 paternal plant) were collected and transferred to the receptive stigma of *M. arvensis*. The
606 reciprocal crosses were done in the same way. The seeds from parents and hybrids germinated
607 and grew following the same procedure described before. Two-week-old leaves of parents and

608 hybrids were used for DNA extraction for genotyping and for promoter region amplification.
609 The two youngest leaves from four-week-old plants were collected as materials for RNA-Seq.
610 Moreover, the mature rosette leaves were taken for leaf anatomy and gas exchange analysis.
611 Seeds from *Arabidopsis thaliana* wild-type plants ecotype Col-0 and the transgenic lines were
612 surface sterilized by vapor phase seed sterilization, further germinated on half MS medium
613 with cold treatment for two days in the dark, and then transferred to the growth chamber under
614 10h/14h light/dark conditions with 22°C/20°C day/night for 10 days. The seedlings were later
615 transferred individually to pots with soil and grown in the growth chamber. The two-week-old
616 Arabidopsis transgenic plants and the wild-type plants were collected for further GUS staining
617 analysis.
618
619 **Leaf anatomy**
620 The 2 mm^2 leaf sections were taken near the midrib of the top third of mature rosette leaves for
621 the leaf ultrastructural analysis. For araldite embedding, leaf sections were fixed with fixation
622 buffer (2% paraformaldehyde, 2% glutaraldehyde), dehydrated by an acetone dilution series,
623 and embedded with an araldite series according to the protocol (Fineran & Bullock, 1972) with
624 modifications. The sections were transferred to the mold filled with fresh araldite and
625 polymerized at 65ºC for two days. Semi-thin sections in 2.5 μm thickness obtained by cutting
626 with a glass knife were mounted on slides, stained with 1% toluidine blue for 2 min and washed
627 by distilled water. The leaf ultrastructure was examined under the light microscope, Zeiss
628 Axiophot microscope (Carl Zeiss Microscopy GmbH, Göttingen, Germany).
629
630 **Photosynthetic gas exchange**
631 The mature rosette leaves were chosen to measure gas exchange characteristics using a LI-
632 6400XT Portable Photosynthesis System (LI-COR Biosciences, Lincoln, USA) with the
633 settings according to the manufacturer's instructions with modifications: the flow of 300 μmol
634 s−1, the light source of 1500 μmol m−2 s−1, the leaf temperature of 25°C, and the vapor
635 pressure deficit based on leaf temp less than 1.5 kPa. The CO_2 response curve, the so called A-
636 Ci curve, was captured by detecting net CO_2 assimilation rates under different intercellular CO_2
637 concentrations. A partial A-Ci curve obtained with measurements at 400, 100, 80, 65, 45, 25,
638 15, and 400 ppm CO_2 was used to calculate the CO_2 compensation points of parental genotypes
639 and hybrids.
640

641 **Sample preparation and RNA sequencing**

642 We selected 12 plants including three replicates of *M. arvensis*, three plants of *M.*
643 *moricandioides* and six F_1 interspecific hybrids (Ma×Mm). The six hybrids demonstrated
644 relative lower CO_2 compensation points among hybrids and differed in leaf anatomy patterns
645 (Figure 2; Table 1). Compared to the parents, hybrids 1, 2, 5, 6 had fewer organelles in BS
646 cells, whereas hybrid 3 and 4 showed very little organelle in BS cells (Supplementary file 1).
647 Total RNA of parental species and interspecific F_1s was extracted using the RNeasy Plant Mini
648 Kit (Qiagen, Hilden, Germany). Then, 17 µl total RNA (100 ng/µl) was added with 2 µl buffer
649 and 0.5 µl RNase-free DNaseI enzyme (New England Biolabs GmbH, Frankfurt am Main,
650 Germany) incubating on ice for 30 s. The treatment was stopped by adding 2 µl 50 mM EDTA
651 and incubated at 65°C for 10 min. The quality of RNA and DNaseI treated RNA was assessed
652 on a Bioanalyzer 2100 (Agilent, Santa Clara, USA) with an RNA Integrity Number (RIN)
653 value ≥ 8. Subsequently, cDNA libraries were prepared using 1 µg of total RNA with the
654 TruSeq RNA Sample Preparation Kit (Illumina, San Diego, USA). The cDNA library was
655 qualified on the Agilent Technologies 2100 Bioanalyzer to check the library quality and
656 fragment size of the sample. RNA-Seq was performed on an Illumina HiSeq 3000 platform at
657 the BMFZ (Biologisch-Medizinisches Forschungszentrum) of the Heinrich-Heine University
658 (Düsseldorf, Germany) to gain 150 bp paired-end reads. In total, we obtained 54.97 Gb of
659 RNA-Seq data, with an average of 4.58 Gb per sample. On average, 27 million reads per library
660 were obtained. The sequencing quality was examined using FastQC v.0.11.5. Quality scores
661 across all bases were generally good but showed lower quality at the end of reads observed in
662 few samples.

663

664 **Read mapping and variant calling**

665 In our study, the C_3 *M. moricandioides* genome assembly was chosen as the draft reference
666 genome, because it possesses a higher quality over C_3-C_4 *M. arvensis* by having more reliable
667 number of repetitive elements. Mapping all the RNA-Seq reads to *M. moricandioides* might
668 lead to underestimation of the transcript level of *M. arvensis* allele, however it has no impact
669 on comparing the allele ratio between parents and hybrids as well as transcripts dominated by
670 *M. arvensis* allele. The RNA-Seq reads were mapped on a draft reference genome of *M.*
671 *moricandioides* (Lin et al., 2021) using STAR v.2.5.2b (Dobin et al., 2013). After duplication
672 marking, base quality recalibration, we used a simulated set of SNPs as known variants for
673 preparing analysis-ready RNA-Seq reads. The variant calling was conducted according to the
674 Genome Analysis Toolkit (GATK) best practice (DePristo et al., 2011). Variant discovery was

675 performed jointly the three *M. arvensis* replicates using the UnifiedGenotyper with *M.*
676 *moricandioides* as the reference, which output a raw vcf file containing 1,748,436 SNP
677 callings. The variant calling vcf file and aligned RNA-Seq reads were further input into
678 ASEReadCounter from GATK to obtain read counts at each SNP site. Only SNP sites with
679 more than 20 total read counts of parental species and less than 3 counts at the other parental
680 species were processed for further allele specific expression (ASE) analysis.
681
682 **Differential gene expression analysis**
683 Transcriptome comparison between species was performed with the DESeq2 tool (Love et al.,
684 2014) in R (www.R-project.org) using the Benjamini–Hochberg (BH) adjusted false discovery
685 rate ≤ 0.01 as the cut-off for significant differential expression (Benjamini & Hochberg, 1995).
686 The Chi-square test was applied to test the overrepresentation of upregulated transcripts in
687 selected pathways, such as glycine shuttle, C_4 cycles, Calvin-Benson cycle, and mitochondrial
688 e^- transport. The principle component analysis (PCA) analysis was performed on the rlog-
689 transformed gene expression data to explore and visualize distances between samples.
690
691 **Allele specific expression (ASE) analysis**
692 In hybrids, parental alleles were expressed under the same genetic background which made it
693 possible to distinguish between *cis*- or *trans*-regulatory effects by calculating and comparing
694 the allele ratio of the parents (A: PA1/PA2) and that of hybrids (B: F1A1/F1A2). A binomial
695 test was applied to test if F1A1 is equal to F1A2 using a *P*-value adjusted by the BH procedure.
696 On the other hand, Fischer's exact test was used to assess the significant difference between
697 the ratio of parental alleles (PA1/PA2) and the allele ratio of hybrids (F1A1/F1A2) where the
698 *P*-value was adjusted by the BH method. Four regulatory effects were defined according to the
699 following conditions: *cis*- only, $B \neq 1$ and A=B; *trans*- only, B=1 and $A \neq B$; *cis*- plus *trans*-,
700 $B \neq 1$ and $A \neq B$; no *cis*- no *trans*-, B=1 and A=B. ASE analysis on SNP-level was conducted
701 on six hybrids individually on a set of 120,200 SNPs, demonstrating polymorphisms on 14,004
702 transcripts.
703 To integrate SNP information into a gene-specific measure, R package, meta-analysis-based
704 allele-specific expression detection (MBASED) was applied to measure the allelic imbalance
705 (Mayba et al., 2014). MBASED applied the principles of meta-analysis on combining the
706 information of every SNP site within a single transcript (a single unit of expression) in the
707 absence of the prior information of phased data, the genetic information of hybrids. The ASE

708 was evaluated based on the transcripts with at least one SNP site. The pseudo-phasing based
709 "major" haplotype of genes took the allele with higher counts as the major allele, resulting in
710 the higher estimates of the major allele frequency (MAF, ranging from 0.5 to 1.0). At least 10^6
711 simulations were carried out to obtain appropriate assessment of the statistical significance.
712
713 **ASE verification by qPCR**
714 A quantitative real-time PCR (qPCR) assay using SNP-specific primers for four selected genes
715 was used to validate the RNA-Seq and ASE results (Supplementary file 14). The DNaseI
716 treated RNA was preceded to cDNA synthesis according to the manufacturer's instructions
717 with modifications. First, a total of 1 µg RNA was mixed with 1 µl oligo-dT primer, 10 mM
718 dNTP-Mix, 4 µl 5X Firstrand-Buffer, and 2 µl 0.1 M DTT and incubated at 42°C for 2 min.
719 The mix was combined with 1 µl Invitrogen SuperScript™ II Reverse Transcriptase and
720 incubated at 42°C for 50 min for cDNA synthesis. The heat inactivation of reverse transcripts
721 was conducted with incubation for 15 min at 70°C. SNPs on *Moricandia* orthologs of *GLDP1*,
722 *ASP3*, *γCA2*, *PPA2* were chosen for designing SNP-specific qPCR primers (Supplementary
723 file 15). The *Moricandia* ortholog (*MSTRG.23175*) of Arabidopsis housekeeping gene
724 Helicase (*AT1G58050*) was tested and selected as reference housekeeping gene. The qPCR
725 amplification was carried out in a total reaction of 20 µL containing 0.5 µL forward primer (10
726 ng/µL), 0.5 µL forward primer (10 ng/µL), 5 µL 5 ng/µL cDNA template, 4 µL ddH_2O, and 10
727 µL SYBR® Green qPCR SuperMix (Thermo Fisher Scientific, Schwerte, Germany). The
728 qPCR reaction was carried out on StepOne™ Real-Time PCR System (Applied Biosystems™,
729 Waltham, USA) by following program: initial denaturation at 95 ºC for 60 s, 40 cycles of
730 denaturation at 95 ºC for 15 s and annealing at 60 ºC for 30 s. The delta CT value was calculated
731 by normalized sample's CT value with that of the housekeeping gene.
732
733 **Transcriptome annotation**
734 After comparing the *M. moricandioides* predicted protein from TransDecoder (Haas et al.,
735 2013) to UniProtKB (both Swiss-Prot and TrEMBL,ed on April 3, 2019) (Camacho et al.,
736 2009) using BLASTP (UniProt, 2019) with e-value < 1e-5, we summarized the functional
737 annotation in the form of "Human Readable Description" by the AHRD pipeline
738 (https://github.com/groupschoof/AHRD). Then, to determine the phylogenetic relationships
739 among *M. arvensis* and *M. moricandioides*, the predicted protein sequences of them together
740 with *A. thaliana* were applied to OrthoFinder v.2.3.3 (Emms & Kelly, 2019).
741

742 **Biased transcript with *cis*-specificity**

743 The transcripts with *cis*-specificity (*cis*-SNPs or *cis plus trans*-SNPs) were classified to two
744 categories based on the gene expression direction in hybrids: transcripts with biased expression
745 toward *M. arvensis* (Ma-biased) and toward *M. moricandioides* (Mm-biased). The biased
746 transcripts were annotated with corresponding GO terms derived from *A. thaliana*. Afterward,
747 the gene ontology comparison between Ma-biased and Mm-biased transcripts were conducted
748 on WEGO 2.0 website (Ye J et al., 2018) and visualized using R.

749

750 **Gene ontology term and pathway enrichment analysis**

751 To recognize the function of transcripts with *cis*-SNPs (common *cis*-SNPs) and with *trans*-
752 SNPs (common *trans*-SNPs) shared by all studied hybrids, a custom mapping file was created,
753 containing *Moricandia* transcript names and the corresponding GO terms derived from *A.*
754 *thaliana* genes. The 2,236 and 45 *Moricandia* transcripts with common *cis*-SNPs and common
755 *trans*-SNPs, respectively, were processed with the custom mapping file by topGO R-package
756 for gene set enrichment analysis of biological processes (Alexa et al., 2006).

757 Pathway enrichment analysis was conducted on the mRNA sequences of 2,236 and 45
758 *Moricandia* transcripts with common *cis*-SNPs and *trans*-SNPs, respectively, with the KEGG
759 Orthology Based Annotation System (KOBAS) (Xie et al., 2011). A BH adjusted false
760 discovery rate of 0.05 served as the threshold to define the significantly enriched pathways
761 (Benjamini & Hochberg, 1995).

762

763 **Promoter-GUS assay and plant transformation**

764 The 5' upstream regions of the *GLDP1*, *CHUP1*, *DUF538*, *ATPB* genes of *M. arvensis* and *M.*
765 *moricandioides* were fused to the GUS reporter gene in the binary plant vector pCambia1381.
766 The primers for amplifying the promoter region were included a BamHI site at the 5' border
767 and a NcoI site at the 3' end of the DNA fragment. The DNA fragment was inserted into
768 pCambia1381 by homologous recombination using the Gibson Assembly Cloning kit (New
769 England Biolabs, catalog number: E5510S). The predicted promoter region of the *PHOT2* gene
770 of *M. arvensis* and *M. moricandioides* was cloned to a Gateway donor vector pDONR207, and
771 then further cloned to a Gateway destination vector pGWB3, which was for C-terminal GUS
772 fusions. The primers for amplifying the promoter region of *PHOT2* gene were included an
773 attB1 sequence at the 5' border and an attB2 sequence at the 3' end of the DNA fragment. The
774 +1 positions of the candidate genes were defined in different ways, shown in Supplemental
775 Table 13. All generated constructs were verified by colony-PCR and DNA sequencing.

776 The promoter-GUS constructs were transformed into *Agrobacterium tumefaciens* strain
777 GV310::pMP90 (Koncz and Schell, 1986) by electroporation. All constructs were verified
778 again by colony-PCR and DNA sequencing. The *Agrobacterium* introduced with the promoter-
779 GUS constructs were transformed in four to six-week-old *A. thaliana* (col-0) by floral-dip
780 method (Clough and Bent, 1998). The transformed T_1 seeds were collected in four to six weeks
781 after transformation, and then selected on Hygromycin B contained half MS plates for two
782 weeks. The survival T_1 lines were further transferred to pots with soil and verified the insertion
783 of T-DNA by PCR.
784 Primers used in promoter region amplification and colony-PCR were shown in Supplementary
785 file 16.
786
787 **GUS staining**
788 Two to four-week-old T_1 leaves were stained with GUS staining solution (100 mM Na_2HPO_4,
789 100 mM NaH_2PO_4, 1 mM Potassium-Ferricyanide $K_4[Fe(N_6)]$, 1 mM Potassium-Ferrocyanide
790 $K_3[Fe(N_6)]$, 0.2% Triton X-100, 2mM X-Gluc) and incubated at 37ºC in the dark for 2 to 72
791 hours. The GUS stained leaves were further fixed by the fixation solution (50% Ethanol, 5%
792 Glacial acetic acid, 3.7% Formaldehyde) at 65ºC for 10 min. Then, leaves were incubated in
793 80% Ethanol at room temperature in order to remove the chlorophyll.
794
795 **Accession numbers**
796 Sequencing read data have been deposited at the European Nucleotide Archive under the
797 project number PRJEB39765 (https://www.ebi.ac.uk/ena/submit/sra/#studies).
798
799 ## Acknowledgements
800 We thank Dr. Otho Mantegazza and Nils Koppers for their bioinformatics support, and
801 Samantha Flachbart for cDNA library preparation. Many thanks to Prof. Dr. Miltos Tsiantis
802 for constructive suggestions and to Prof. Juliette de Meaux for encouraging us to pursue an
803 ASE approach. We thank the gardeners for taking care of the plants and Steffen Köhler for
804 photography support.
805

Additional information

Funding

This work was funded by grants of the Deutsche Forschungsgemeinschaft to APMW under Germany's Excellence Strategy EXC-2048/1, Project ID 390686111, and by ERA-CAPS project C4BREED (WE 2231/20-1), and by a graduate fellowship of the International Max Planck Research School on "Understanding Complex Plant Traits using Computational and Evolutionary Approaches" to MY.L.

Author contributions

MY.L. designed and performed all the experimental works and data analysis and wrote the manuscript.

B. S. participated in drafting the manuscript.

U.S. and **A.P.M.W.** conceptualized the study, supervised the experimental design and participated in drafting the manuscript.

Additional files

Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and identifier
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis_plant2_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424638	Publicly available at ENA, ERS5424638
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis_plant5_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424639	Publicly available at ENA, ERS5424639
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis_plant6_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424640	Publicly available at ENA, ERS5424640
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. moricandioides_ plant1_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424641	Publicly available at ENA, ERS5424641
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. moricandioides_ plant7_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424642	Publicly available at ENA, ERS5424642
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. moricandioides_ plant8_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5424643	Publicly available at ENA, ERS5424643
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid1_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455027	Publicly available at ENA, ERS5455027
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid2_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455028	Publicly available at ENA, ERS5455028
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid3_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455029	Publicly available at ENA, ERS5455029
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid4_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455030	Publicly available at ENA, ERS5455030
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid5_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455031	Publicly available at ENA, ERS5455031
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	M. arvensis x M. moricandioides _Hybrid6_RNA-Seq reads	https://www.ebi.ac.uk/ena/browser/view/ERS5455032	Publicly available at ENA, ERS5455032
Lin M-Y, Schlüter U, Stich B, Weber AP	2021	Transcriptome assembly of M. moricandioides	https://www.ebi.ac.uk/ena/browser/view/ERS5472842	Publicly available at ENA, ERS5472842

825 **Supplementary files**

826 **Supplementary file 1.** Leaf micrographs of transverse sections of *M. arvensis*, *M.*
827 *moricandioides* and their interspecific hybrids a, *M. arvensis*; b, *M. moricandioides*; c-h,
828 hybrids 1-6. Arrow, chloroplasts. Bar, 100 μm.

829 **Supplementary file 2.** Leaf venations of *M. arvensis*, *M. moricandioides* and their
830 interspecific hybrids a, *M. arvensis*; b, *M. moricandioides*; c-h, hybrid 1-6. Bar, 500 μm.

831 **Supplementary file 3.** The pollen activity test of *M. arvensis*, *M. moricandioides* and their
832 interspecific hybrids dyed by Alexander staining method (Alexander, 1969) Pollens from
833 hybrid lines were stained red, but demonstrated abnormal shapes compared to parents' round
834 pollens. Aborted pollen grains are stained blue-green, and non-aborted pollen grains are stained
835 magenta-red. Bar, 20 μm.

836 **Supplementary file 4.** Principal component analysis of rlog-transformed gene expression data
837 A) parental species, *M. arvensis* (Ma) and *M. moricandioides* (Mm) and B) parental species
838 and their interspecific hybrids.

839 **Supplementary file 5.** GO analysis on Ma-upregulated transcripts and Ma-downregulated
840 transcripts using topGO.

841 **Supplementary file 6.** Transcriptional changes in selected pathways. The heatmap indicated
842 the log2-fold changes in transcript level of C_3–C_4 species *M. arvensis* compared to the C_3
843 species *M. moricandioides*. Blue and red indicates reduced and enhanced transcript abundance
844 in C_3–C_4, respectively. *, adjusted P-value < 0.05; **, adjusted P-value < 0.01. Bold, transcripts
845 with the highest expression among isoforms.

846 **Supplementary file 7.** List of common SNPs and transcripts harboring common SNP.

847 **Supplementary file 8.** GO analysis on common *cis*-SNPs and common *trans*-SNPs using
848 topGO.

849 **Supplementary file 9.** Significantly enriched pathways identified in transcripts with common
850 *cis*-SNPs and common *trans*-SNPs using KOBAS database.

851 **Supplementary file 10.** Number of biased transcripts with *cis*-specificity among six hybrids.

852 **Supplementary file 11.** List of transcripts showed extreme allelic imbalance with major allele
853 frequency ≥ 0.9 in all hybrids.

854 **Supplementary file 12.** Enrichment of regulatory effects in selected pathways. 0, no *cis*-SNP;
855 1, at least one *cis*-SNP found in hybrid line; 2, common *cis*-SNP among hybrids.

856 **Supplementary file 13.** Selected gene list for promoter-GUS assay.

857 **Supplementary file 14.** Confirmation of RNA-Seq data by allele-specific RT-PCR of *M.*
858 *arvensis* × *M. moricandioides* hybrid.

859 **Supplementary file 15.** qPCR primer list for ASE verification.

860 **Supplementary file 16.** Primer list for promoter-GUS assay.

861

862 **Supplementary method**

863 **Pollen activity assay**

864 The pollen viability of *M. arvensis*, *M. moricandioides*, and their hybrids was observed
865 followed modified Alexander's staining method (Alexander, 1969). The primary
866 inflorescences with mature pollens were collected one day after flowering and then incubated
867 in 1:50 staining solution for 5 min. The stock solution was comprised of 10 ml 96% ethanol, 1
868 ml 1% malachite green (w/v, in 96% ethanol), 25 ml glycerol, 5 ml 1% acid fuchsin (w/v, in
869 dH_2O), 4 ml glacial acetic acid, and 100 ml dH_2O. The phase contrast images of dyed pollens
870 were obtained under inverted microscopy (Eclipse Ti, Nikon).

871

References

Adwy, W., Schlüter, U., Papenbrock, J., Peterhansel, C., & Offermann, S. (2019). Loss of the M-box from the glycine decarboxylase P-subunit promoter in C2 Moricandia species. *Plant Gene*, *18*, 100176. https://doi.org/10.1016/j.plgene.2019.100176

Alexander, M. P. (1969). Differential Staining of Aborted and Nonaborted Pollen. *Stain Technology*, *44*(3), 117–122. https://doi.org/10.3109/10520296909063335

Apel, P. (1980). CO2 compensation concentration and its O2 dependence in Moricandia spinosa and Moricandia moricandioides (Cruciferae). *Biochemie Und Physiologie Der Pflanzen*, *175*(4), 386–388. https://doi.org/10.1016/S0015-3796(80)80079-5

Apel, P., Bauwe, H., & Ohle, H. (1984). Hybrids between Brassica alboglabra and Moricandia arvensis and their Photosynthetic Properties. *Biochemie Und Physiologie Der Pflanzen*, *179*(9), 793–797. https://doi.org/10.1016/S0015-3796(84)80008-6

Bauwe, H., & Apel, P. (1979). Biochemical Characterization of Moricandia arvensis (L.) DC., a Species with Features Intermediate between C3 and C4 Photosynthesis, in Comparison with the C3 Species Moricandia foetida Bourg. *Biochemie Und Physiologie Der Pflanzen*, *174*(3), 251–254. https://doi.org/10.1016/S0015-3796(17)30581-4

Beebe, D. U., & Evert, R. F. (1990). The Morphology and Anatomy of the Leaf of Moricandia arvensis (L.) DC. (Brassicaceae). *Botanical Gazette*, *151*(2), 184–203.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. JSTOR.

Blätke, M.-A., & Bräutigam, A. (2019). Evolution of C4 photosynthesis predicted by constraint-based modelling. *ELife*, *8*, e49305. https://doi.org/10.7554/eLife.49305

Bouton, J. H., Brown, R. H., Evans, P. T., & Jernstedt, J. A. (1986). Photosynthesis, Leaf Anatomy, and Morphology of Progeny from Hybrids between C3 and C3/C4Panicum Species 1. *Plant Physiology*, *80*(2), 487–492.

Bräutigam, A., & Gowik, U. (2016). Photorespiration connects C3 and C4 photosynthesis. *Journal of Experimental Botany*, *67*(10), 2953–2962. https://doi.org/10.1093/jxb/erw056

Brown, H. R., & Bouton, J. H. (1993). Physiology and Genetics of Interspecific Hybrids Between Photosynthetic Types. *Annual Review of Plant Physiology and Plant Molecular Biology*, *44*(1), 435–456. https://doi.org/10.1146/annurev.pp.44.060193.002251

Brown, R. H., Bouton, J. H., Evans, P. T., Malter, H. E., & Rigsby, L. L. (1985). Photosynthesis, Morphology, Leaf Anatomy, and Cytogenetics of Hybrids between C3 and C3/C4Panicum Species. *Plant Physiology*, *77*(3), 653–658.

Brown, R. H., & Hattersley, P. W. (1989). Leaf Anatomy of C3-C4 Species as Related to Evolution of C4 Photosynthesis. *Plant Physiology*, *19*, 1543–1550.

Covshoff, S., Burgess, S. J., Kneřová, J., & Kümpers, B. M. C. (2014). Getting the most out of natural variation in C4 photosynthesis. *Photosynthesis Research*, *119*(1), 157–167. https://doi.org/10.1007/s11120-013-9872-8

DePristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *BioRxiv*, 466201. https://doi.org/10.1101/466201

Fineran, B. A., & Bullock, S. (1972). A Procedure for Embedding Plant Material in Araldite for Electron Microscopy. *Annals of Botany*, *36*(1), 83–86. https://doi.org/10.1093/oxfordjournals.aob.a084579

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., & Westhoff, P. (2011). Evolution of C4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C4? *The Plant Cell*, *23*(6), 2087–2105. https://doi.org/10.1105/tpc.111.086264

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., & Westhoff, P. (2004). Cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene. *The Plant Cell*, *16*(5), 1077–1090. https://doi.org/10.1105/tpc.019729

Hart, J. E., Sullivan, S., Hermanowicz, P., Petersen, J., Diaz-Ramos, L. A., Hoey, D. J., Łabuz, J., & Christie, J. M. (2019). Engineering the phototropin photocycle improves photoreceptor performance and plant biomass production. *Proceedings of the National Academy of Sciences*, *116*(25), 12550–12557. https://doi.org/10.1073/pnas.1902915116

Hatch, M. D. (1987). C4 photosynthesis: A unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, *895*(2), 81–106. https://doi.org/10.1016/S0304-4173(87)80009-5

He, F., Zhang, X., Hu, J., Turck, F., Dong, X., Goebel, U., Borevitz, J., & de Meaux, J. (2012). Genome-wide Analysis of Cis-regulatory Divergence between Species in the Arabidopsis Genus. *Molecular Biology and Evolution*, *29*(11), 3385–3395. https://doi.org/10.1093/molbev/mss146

Heckmann, D. (2016). C4 photosynthesis evolution: The conditional Mt. Fuji. *Current Opinion in Plant Biology*, *31*, 149–154. https://doi.org/10.1016/j.pbi.2016.04.008

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., & Lercher, M. J. (2013). Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell*, *153*(7), 1579–1588.

https://doi.org/10.1016/j.cell.2013.04.058

Holaday, A. S., & Chollet, R. (1984). Photosynthetic/photorespiratory characteristics of C3−C4 intermediate species. *Photosynthesis Research*, *5*(4), 307–323. https://doi.org/10.1007/BF00034976

Kadereit, G., Bohley, K., Lauterbach, M., Tefarikis, D. T., & Kadereit, J. W. (2017). C3–C4 intermediates may be of hybrid origin – a reminder. *New Phytologist*, *215*(1), 70–76. https://doi.org/10.1111/nph.14567

Krenzer, E. G., Moss, D. N., & Crookston, R. K. (1975). Carbon dioxide compensation points of flowering plants. *Plant Physiology*, *56*(2), 194–206. https://doi.org/10.1104/pp.56.2.194

Lange, B. M., & Croteau, R. (1999). Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13714–13719.

Lemmon, Z. H., Bukowski, R., Sun, Q., & Doebley, J. F. (2014). The Role of cis Regulatory Evolution in Maize Domestication. *PLOS Genetics*, *10*(11), e1004745. https://doi.org/10.1371/journal.pgen.1004745

Lemos, B., Araripe, L. O., Fontanillas, P., & Hartl, D. L. (2008). Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14471–14476. https://doi.org/10.1073/pnas.0805160105

Li, Y., Chen, A., & Liu, Z. (2017). *Analysis of Allele-Specific Expression—Bioinformatics in Aquaculture—Wiley Online Library*. https://onlinelibrary.wiley.com/doi/10.1002/9781118782392.ch14

Lichtenthaler, H. K. (1997). *EVOLUTION OF CAROTENOID AND ISOPRENOID BIOSYNTHESIS IN PHOTOSYNTHETIC AND NON-PHOTOSYNTHETIC ORGANISMS*. 14.

Lin, M.-Y., Koppers, N., Denton, A., Schlüter, U., & Weber, A. P. M. (2021). Whole genome sequencing and assembly data of Moricandia moricandioides and M. arvensis. *Data in Brief*, 106922. https://doi.org/10.1016/j.dib.2021.106922

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., & Gowik, U. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. *ELife*, *3*, e02478. https://doi.org/10.7554/eLife.02478

Mayba, O., Gilbert, H. N., Liu, J., Haverty, P. M., Jhunjhunwala, S., Jiang, Z., Watanabe, C., & Zhang, Z. (2014). MBASED: Allele-specific expression detection in cancer tissues and cell lines. *Genome Biology*, *15*(8). https://doi.org/10.1186/s13059-014-0405-3

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

Meiklejohn, C. D., Coolon, J. D., Hartl, D. L., & Wittkopp, P. J. (2014). The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Research*, *24*(1), 84–95. https://doi.org/10.1101/gr.156414.113

Monson, R. K., & Rawsthorne, S. (2000). CO2 Assimilation in C3-C4 Intermediate Plants. In R. C. Leegood, T. D. Sharkey, & S. von Caemmerer (Eds.), *Photosynthesis: Physiology and Metabolism* (pp. 533–550). Springer Netherlands. https://doi.org/10.1007/0-306-48137-5_22

Oakley, J. C., Sultmanis, S., Stinson, C. R., Sage, T. L., & Sage, R. F. (2014). Comparative studies of C3 and C4 Atriplex hybrids in the genomics era: Physiological assessments. *Journal of Experimental Botany*, *65*(13), 3637–3647. https://doi.org/10.1093/jxb/eru106

Oikawa, K., Yamasato, A., Kong, S.-G., Kasahara, M., Nakai, M., Takahashi, F., Ogura, Y., Kagawa, T., & Wada, M. (2008). Chloroplast outer envelope protein CHUP1 is essential for chloroplast anchorage to the plasma membrane and chloroplast movement. *Plant Physiology*, *148*(2), 829–842. https://doi.org/10.1104/pp.108.123075

Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences*, *104*(suppl 1), 8605–8612. https://doi.org/10.1073/pnas.0700488104

Rawsthorne, S., Hylton, C. M., Smith, A. M., & Woolhouse, H. W. (1988). Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C3-C 4 intermediate species Moricandia arvensis (L.) DC. *Planta*, *176*(4), 527–532. https://doi.org/10.1007/BF00397660

Rawsthorne, S., Morgan, C. L., O'Neill, C. M., Hylton, C. M., Jones, D. A., & Frean, M. L. (1998). Cellular expression pattern of the glycine decarboxylase P protein in leaves of an intergeneric hybrid between the C3-C4 intermediate species Moricandia nitens and the C3 species Brassica napus. *Theoretical and Applied Genetics*, *96*(6), 922–927. https://doi.org/10.1007/s001220050821

Rawsthorne, Stephen, Hylton, C. M., Smith, A. M., & Woolhouse, H. W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C3 and C3-C4 intermediate species of Moricandia. *Planta*, *173*(3), 298–308. https://doi.org/10.1007/BF00401016

Reeves, G., Grangé-Guermente, M. J., & Hibberd, J. M. (2017). Regulatory gateways for cell-specific gene expression in C4 leaves with Kranz anatomy. *Journal of Experimental Botany*, *68*(2), 107–116. https://doi.org/10.1093/jxb/erw438

Rhoné, B., Mariac, C., Couderc, M., Berthouly-Salazar, C., Ousseini, I. S., & Vigouroux, Y. (2017). No Excess of Cis-Regulatory Variation Associated with Intraspecific Selection in Wild Pearl Millet (Cenchrus americanus). *Genome Biology and Evolution*, *9*(2), 388–397. https://doi.org/10.1093/gbe/evx004

Sage, R. F., Khoshravesh, R., & Sage, T. L. (2014). From proto-Kranz to C4 Kranz: Building the bridge to C4 photosynthesis. *Journal of Experimental Botany*, *65*(13), 3341–3356. https://doi.org/10.1093/jxb/eru180

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the Evolution of C4 Photosynthesis. *Annual Review of Plant Biology*, *63*(1), 19–47. https://doi.org/10.1146/annurev-arplant-042811-105511

Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P.-A., Kurz, S., Mettler-Altmann, T., & Weber, A. P. (2017). Photosynthesis in C3–C4 intermediate Moricandia species. *Journal of Experimental Botany*, *68*(2), 191–206. https://doi.org/10.1093/jxb/erw391

Schlüter, U., & Weber, A. P. M. (2016). The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant & Cell Physiology*, *57*(5), 881–889. https://doi.org/10.1093/pcp/pcw009

Schlüter, U., & Weber, A. P. M. (2020). Regulation and Evolution of C4 Photosynthesis. *Annual Review of Plant Biology*, *71*(1), 183–215. https://doi.org/10.1146/annurev-arplant-042916-040915

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., Gowik, U., & Westhoff, P. (2013). *Evolution of C 4 Photosynthesis in the Genus Flaveria: Establishment of a Photorespiratory CO 2 Pump*.

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., Song, J., Li, X., Xiao, J., Chen, L.-L., Ouyang, Y., & Zhang, Q. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proceedings of the National Academy of Sciences*, *116*(12), 5653–5658. https://doi.org/10.1073/pnas.1820513116

Sheen, J. (1999). C4 Gene Expression. *Annual Review of Plant Physiology and Plant Molecular Biology*, *50*(1), 187–217. https://doi.org/10.1146/annurev.arplant.50.1.187

Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J., & Akey, J. M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research*, *21*(10), 1728–1737. https://doi.org/10.1101/gr.119784.110

Steige, K. A., Reimegård, J., Koenig, D., Scofield, D. G., & Slotte, T. (2015). Cis-Regulatory Changes Associated with a Recent Mating System Shift and Floral Adaptation in Capsella. *Molecular Biology and Evolution*, *32*(10), 2501–2514. https://doi.org/10.1093/molbev/msv169

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: How predictable is genetic evolution? *Evolution; International Journal of Organic Evolution*, *62*(9), 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

Sultmanis, S. D. (2018). *Developmental and Expression Evolution in C3 and C4 Atriplex and their Hybrids*.

Tirosh, I., Reikhav, S., Levy, A. A., & Barkai, N. (2009). A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science*, *324*(5927), 659–662.

https://doi.org/10.1126/science.1169766

Ueno, O., Bang, S. W., Wada, Y., Kondo, A., Ishihara, K., Kaneko, Y., & Matsuzawa, Y. (2003). Structural and Biochemical Dissection of Photorespiration in Hybrids Differing in Genome Constitution between Diplotaxis tenuifolia (C3-C4) and Radish (C3). *Plant Physiology*, *132*(3), 1550–1559. https://doi.org/10.1104/pp.103.021329

Ueno, O., Woo Bang, S., Wada, Y., Kobayashi, N., Kaneko, R., Kaneko, Y., & Matsuzawa, Y. (2007). Inheritance of C3-C4 Intermediate Photosynthesis in Reciprocal Hybrids Between Moricandia arvensis (C3-C4) and Brassica oleracea (C3) That Differ in Their Genome Constitution. *Plant Production Science*, *10*(1), 68–79. https://doi.org/10.1626/pps.10.68

Williams, B. P., Johnston, I. G., Covshoff, S., & Hibberd, J. M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *ELife*, *2*, e00961. https://doi.org/10.7554/eLife.00961

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genetics*, *40*(3), 346–350. https://doi.org/10.1038/ng.77

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59–69. https://doi.org/10.1038/nrg3095

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, *8*(3), 206–216. https://doi.org/10.1038/nrg2063

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.-Y., & Wei, L. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, *39*(Web Server issue), W316–W322. https://doi.org/10.1093/nar/gkr483

Zhuang, Y., & Adams, K. L. (2007). Extensive Allelic Variation in Gene Expression in Populus F1 Hybrids. *Genetics*, *177*(4), 1987–1996. https://doi.org/10.1534/genetics.107.080325