

**Studies into the genetic architecture of C_3 - C_4
characteristics in *Moricandia***

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Meng-Ying Lin

from Taipei, Taiwan

Düsseldorf, January 2020

From the institute of Plant Biochemistry
at the Heinrich Heine University Düsseldorf

Published by the permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Andreas P.M. Weber

Co-supervisor: Prof. Dr. Miltos Tsiantis

Date of the oral examination: 05.05.2020

Declaration of the Doctoral Dissertation

I herewith declare under oath that this dissertation was the result of my own work without any unauthorized help in compliance with the “Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf”. This dissertation has never been submitted in this or similar format to any other institutions. I have not previously failed a doctoral examination procedure.

Düsseldorf, 23.12.2019

Meng-Ying Lin

Table of Contents

I. Summary	1
II. Introduction	3
1. Motivation.....	3
2. An overview of evolutionary trajectory from C₃, C₃-C₄, toward C₄ photosynthesis.....	4
2.1 The photosynthetic pathway and problems resulting from photorespiration	4
2.2 The CO ₂ concentrating mechanism implemented in C ₃ -C ₄ intermediate species.....	4
2.3 C ₄ plants evolve an efficient solution to repress photorespiration.....	6
2.4 The evolutionary trajectory from C ₃ to C ₄ photosynthesis through C ₃ -C ₄ intermediates	8
3. Known genetic control of Kranz anatomy and chloroplast development.....	9
3.1 Regulation of Kranz anatomy development	9
3.2 Genes related to chloroplast development, division, and movement	11
4. Old stories and new expects of experimental hybrids in C₃-C₄/C₄ gene discovery	15
4.1 Efforts have been done in the past.....	15
4.2 A new experimental hybrid studying system in the Brassicaceae: <i>Moricandia</i>	15
4.3 Obstacles and new opportunities of the application of experimental hybrids	16
5. Strategies for genetic discovery of C₃-C₄ characteristics.....	17
5.1 Exploring regulatory divergences between different photosynthesis types	17
5.2 Comparative transcriptomics of closely related photosynthesis species	18
III. Aim of the Thesis	21
IV. References.....	22
V. Manuscripts	33
1. Manuscript I: Interspecific Hybridization between <i>Moricandia</i> C₃ and C₃-C₄ Species	34
2. Manuscript II: Transcriptional Regulation Associated with the CO₂ Concentrating Mechanism in <i>Moricandia</i>	61
3. Manuscript III: Comparative Transcriptome Analysis during Leaf Development on Closely Related C₃ and C₃-C₄ <i>Moricandia</i> Species.....	115
VI. Acknowledgements	161

I. Summary

C₃-C₄ intermediate species are of great value to unravel molecular mechanisms of the early evolutionary steps of C₄ photosynthesis, owing to the fact that they are predicted to be natural occurring intermediates on the convergent evolution from C₃ to C₄ photosynthesis. C₃-C₄ plants possess Kranz-like anatomy, bundle sheath cells with numerous organelles arranged centripetally around the veins, which serves as a requisite for the photorespiratory glycine shuttle. The P-subunit of the photorespiratory enzyme glycine decarboxylase is absent in C₃-C₄ mesophyll cells. In order to digest the toxic product 2-phosphoglycolate, the photorespiratory glycine is shuttled from mesophyll to bundle sheath cells, where the released CO₂ can be efficiently recaptured by numerous, adjacent chloroplasts. Transcriptional regulation of spatial expression pattern of the P-protein of glycine decarboxylase between C₃ and C₃-C₄ species is well known in *Flaveria* and *Moricandia*. However, the genetic architecture of C₃-C₄ characteristics remains largely elusive.

In this thesis, I investigated cytogenetic characteristics of *Moricandia* C₃ and C₃-C₄ species and the potential of interspecific hybridization between C₃ and C₃-C₄ plant, serving as fundamental knowledge for further transcriptional regulation analysis and comparative transcriptomics. Different interspecific hybridizations between C₃ and C₃-C₄ species in *Moricandia* were performed, where the hybridized pair of C₃-C₄ *M. arvensis* as maternal and C₃ *M. moricandioides* as paternal species was the most prolific combination. To investigate C₃-C₄ intermediate specific gene expression, allele specific expression analysis was applied to the interspecific hybrids of *M. arvensis* and *M. moricandioides*. It demonstrated that *cis*-mechanisms play an important role in shaping C₃ to C₃-C₄ photosynthesis type and the corresponding transcripts are enriched in major photosynthetic pathways and chloroplast relocation. Candidates with strong allele specific expression are of special interest for studying the genetic architecture regarding C₃-C₄ characteristic traits.

Comparison of anatomical changes during leaf ontology between C₃ and C₃-C₄ *Moricandia* species revealed the early establishment of Kranz-like anatomy in C₃-C₄ species. Through the comparative transcriptome and cluster analysis during leaf development of closely related C₃ and C₃-C₄ *Moricandia* species, we found transcription factors as potential candidates for genetic control of C₃-C₄ characteristics, such as auxin response, sigma-like, and growth-related factor, and a cohort of genes regulating vein initiation, procambium formation, xylem

I. Summary

formation, *SCR/SHR* pathway, and vein patterning associated with the early leaf development in C₃-C₄ *M. arvensis*. The early organelle accumulation in C₃-C₄ bundle sheath cells was more relevant to plastid division than formation and biosynthesis.

II. Introduction

1. Motivation

The green revolution in the late 1960s resulted in a large increase of rice and wheat production through introduction of the dwarfing genes (Khush, 1999; Furbank et al., 2015; Nagai et al., 2018). The dwarfing genes are responsible for increasing yield by the reduction of lodging and for increasing of the harvest index in rice and wheat (Parry et al., 2011). The approach could also be implemented in *Brassica* crops (Muangprom et al., 2005). However, the increase of global production through improving harvest index is limited. The new generation of green revolution has focused on the increase of the biomass by improving the photosynthetic performance (Parry et al., 2011). C_4 photosynthetic plant species outperform the ancestral C_3 plants through considerable reduction of the energy-demanding photorespiration, which leads to a net loss of CO_2 and decreases the efficiency of photosynthesis by 40% under ambient CO_2 concentration and under conditions of water deficits and high temperatures (Gowik and Westhoff, 2011). Although the C_4 photosynthetic pathway is found in only 3% of terrestrial plants, it accounts for around 25% of primary productivity (Still et al., 2003; Sage, 2004; Edwards et al., 2010). It has been predicted that the crop yield could be boosted by 50% with an increase of 10% the photosynthetic efficiency through transferring the C_4 photosynthesis pathway into C_3 crops (Langdale, 2011; Furbank et al., 2015). Many efforts have been done in producing C_4 rice (von Caemmerer et al., 2012). However, the genetic control of C_4 characters remains largely unknown, especially the molecular mechanisms regulating Kranz anatomy development. C_3 - C_4 species are considered as naturally occurring intermediates on the evolutionary trajectory from C_3 to C_4 photosynthesis, possessing Kranz-like leaf anatomy and photorespiratory glycine shuttle, a CO_2 concentrating mechanism. A recent model suggested that introduction of C_3 - C_4 characters into C_3 rice might be an advantageous alternative (Bellasio and Farquhar, 2019). Therefore, dissecting the genetic control of C_3 - C_4 characters is not only prominent for understanding the early evolution of the C_4 photosynthesis, but also for improving the breeding of highly efficient C_3 crops through introducing C_3 - C_4 characters.

II. Introduction

2. An overview of evolutionary trajectory from C₃, C₃-C₄, toward C₄ photosynthesis

Photosynthesis is the major determinant for the conversion from H₂O and CO₂ through solar energy to the production of carbohydrates (Johnson, 2016). It functions in the two-step process, the light reaction and the dark reaction. In the light reaction, solar energy drives the transfer of electron and proton through electron transport chain between the two photosystems (PS), PSI and PSII, in the thylakoid membranes of chloroplasts. This process generates NADPH and ATP, serving as energy source for the CO₂ assimilation through Calvin-Benson cycle to generate carbohydrates in the dark reaction.

2.1 The photosynthetic pathway and problems resulting from photorespiration

C₃ plants perform the ancestral CO₂ assimilation pathway. CO₂ is fixed to the 3-carbon compound, 3-phosphoglycerate (3-PGA), through carboxylation reaction with Ribulose-1,5-bisphosphate (RuBP) carboxylase-oxygenase (RuBisCO). In addition to CO₂, RuBisCO also has affinity to oxygen, leading to the production of 2-phosphoglycolate (2-PG), which is toxic and needs to be recycled to RuBP through photorespiration (Douce et al., 2001; Bauwe et al., 2010). In chloroplasts, the toxic 2-PG is converted to glycolate by 2-PG phosphatase, and then glycolate is shuttled to peroxisome, where it is catalyzed to glyoxylate by glyoxylate oxidase (GOX). Glyoxylate is then transaminated to glycine by glutamate:glyoxylate aminotransferase (GGAT) in the mitochondria, where two molecules of glycine are converted to serine, NH₃, and CO₂ by glycine decarboxylase complex (GDC) together with serine hydroxymethyltransferase (SHMT). Serine is shuttled to the peroxisome, where it is catalyzed by serine:glyoxylate aminotransferase (SGAT) and the produced 3-hydroxypyruvate is converted to glycerate by 3-hydroxypyruvate reductase (HPR). Glycerate is then shuttled back to the chloroplast, where it is phosphorylated to 3-PGA by D-glycerate 3-kinase (GLYK), which reenters the Calvin-Benson cycle. NH₃ is reassimilated to glutamate through glutamate synthase (GS) cycle in the chloroplast. Photorespiration is energy-consuming and leads to the release of CO₂, therefore decreasing the photosynthetic efficiency. To solve this problem, land plants evolved CO₂ concentrating mechanisms, for instance the photorespiratory glycine shuttle or the C₄ cycle.

2.2 The CO₂ concentrating mechanism implemented in C₃-C₄ intermediate species

C₃-C₄ intermediate plant species are considered as naturally existing intermediates on the evolutionary path from C₃ to C₄ photosynthesis. They possess Kranz-like leaf anatomy: the

II. Introduction

bundle sheath (BS) cells are with abundant chloroplasts accumulated along the cell wall toward the vein (V) (Brown and Hattersley, 1989; Sage et al., 2014). The CO_2 concentrating mechanism implemented in $\text{C}_3\text{-C}_4$ intermediate is fulfilled through the confinement of GDC activity in the BS cells (Rawsthorne et al., 1988; Morgan et al., 1993). The GDC complex is comprised of four subunits, the H-, P-, T-, and L-proteins. H-proteins play the pivotal role, which serve as mobile substrates commuting between the other subunits, undergoing a cycle of reductive methylation (P-protein), methylamine transfer (T-protein), and electron transfer (L-protein) (Figure 1). The P-subunit of GDC (GLDP), responsible for GDC activity, is absent in $\text{C}_3\text{-C}_4$ mesophyll (M) cells. Thus, to complete the photorespiratory pathway, two molecules of photorespiratory glycine are shuttled to the BS cells (Figure 2). The first molecule of glycine binds to the GLDP and then is decarboxylated to release CO_2 , which can be efficiently recaptured by numerous, adjacent chloroplasts, reentering the Calvin-Benson cycle. With the expense of the second molecule of glycine, the conversion between THF and $\text{CH}_2\text{-THF}$ is regulated through the interaction with T-protein and SHMT, which releases NH_3 and serine, then shuttled back to chloroplasts and peroxisomes, in the M cells, respectively. This process implemented in $\text{C}_3\text{-C}_4$ intermediates increases the CO_2 concentration in the vicinity of RuBisCO in the BS cells, where the decarboxylation rate is higher relative to C_3 plants, which is called photorespiratory glycine shuttle or C_2 photosynthesis (Sage et al., 2014; Schlüter and Weber, 2016; Kadereit et al., 2017).

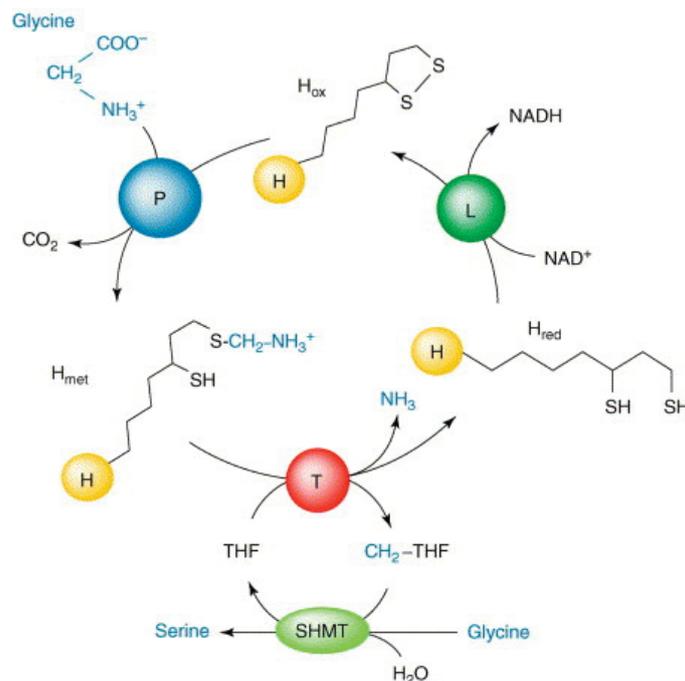


Figure 1. The GDC complex system with SHMT in mitochondrion, adapted from Hagemann and Bauwe, 2016.

II. Introduction

cytosol of M cells and then fixed to produce four-carbon compounds (OAA) by phosphoenolpyruvate carboxylase (PEPC) in M cells. The produced OAA converts to either malate or aspartate, which diffuses to BS cells through plasmodesmata. In the BS cells, the malate or aspartate-converted malate is decarboxylated resulting in elevated CO_2 concentration around RuBisCO allowing efficient CO_2 fixation. Based on the major decarboxylation enzymes, C_4 plants are categorized into three subtypes, NADP-dependent malic enzyme (NADP-ME), NAD-dependent malic enzyme (NAD-ME) or PEP carboxykinase (PEPCK) (Figure 3; Hatch, 1987), but many species are reported to implement a mixture of biochemical pathway with different degrees of engagement of subtypes (Sommer et al., 2012; Muhaidat and McKown, 2013; Wang et al., 2014).

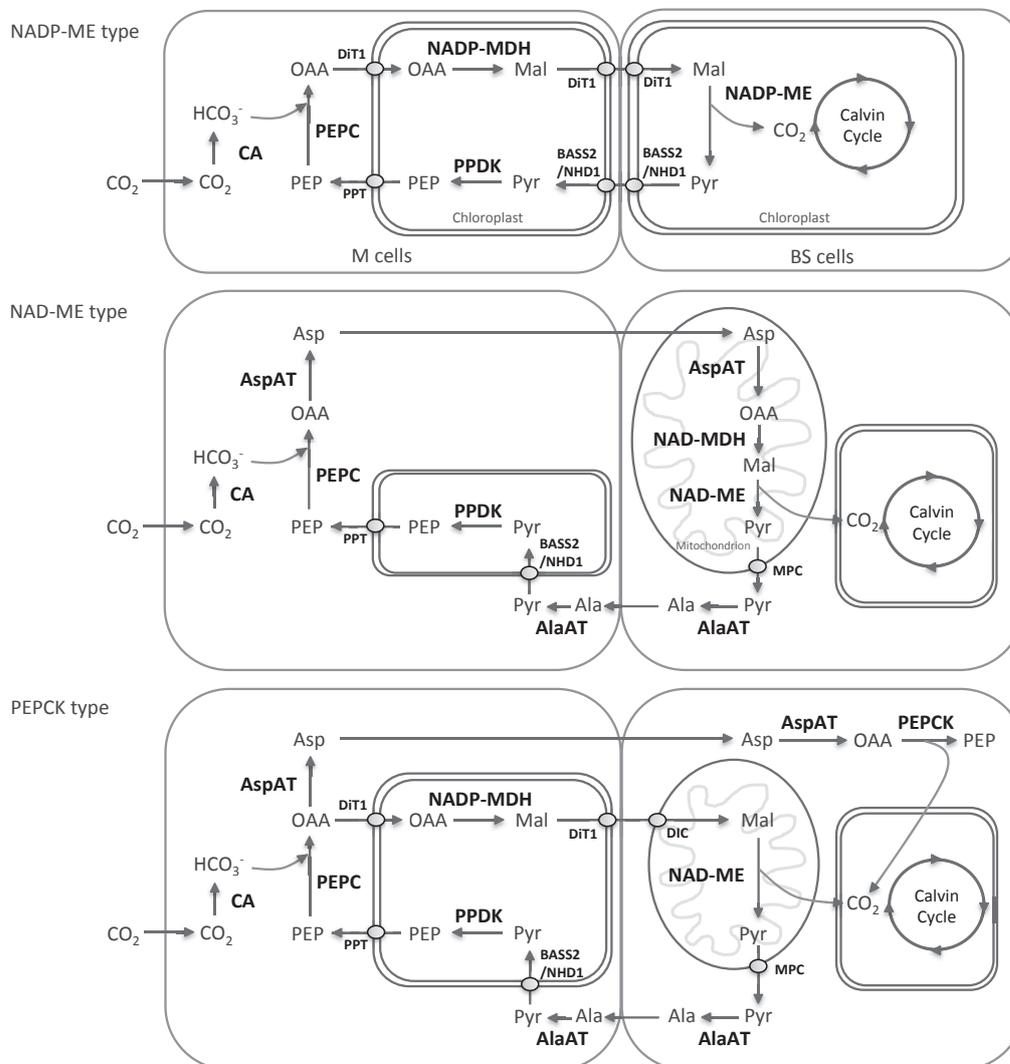


Figure 3. C_4 cycles, including NADP-ME, NAD-ME, and PEPCK type.

CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; DiT1, dicarboxylate transporter 1 (oxaloacetate/malate transporter); PPT, phosphoenolpyruvate/phosphate

II. Introduction

translocator; MDH, Malate dehydrogenase; PPDK, pyruvate, orthophosphate dikinase; BASS2/NHD1, plasma membrane pyruvate transport; DIC, mitochondrial dicarboxylate transporter; MPC, mitochondrial pyruvate carrier; NADP-ME, NADP-malic enzyme; NAD-ME, NAD-malic enzyme; AspAT, aspartate; PEPCK, PEP carboxykinase; PEP, phosphoenolpyruvate; OAA, oxaloacetate; Mal, Malate; Pyr, pyruvate; Asp, Aspartate; Ala, Alanine.

2.4 The evolutionary trajectory from C₃ to C₄ photosynthesis through C₃-C₄ intermediates

In spite of the biochemical and anatomical complexity, C₄ photosynthesis has evolved at least 66 times in independent C₃ lineages including monocots and eudicots (Sage et al., 2012). This implies that it must be a convergent evolution triggering independent C₃ lineages evolving toward C₄ photosynthesis. The evolutionary trajectory from C₃ to C₄ photosynthesis has been depicted in different models, including a number of anatomical and biochemical adaptive steps (Sage et al., 2012; Heckmann et al., 2013; Williams et al., 2013; Mallmann et al., 2014). The conceptual nature based model suggested that the vein density first increases, followed by evolution of the leaf proto-Kranz anatomy, and then a photorespiratory CO₂ pump is built through reducing M:BS cells ratio and confining GDC activity in BS cells. Later, C₄ cycle enzymes are functionalized through the compartmentalization between M and BS cells (Sage et al., 2012). The consensus trajectories of the statistical (Williams et al., 2013) and mechanistic (Heckmann et al., 2013; Mallmann et al., 2014) models confirmed these steps, but the order of steps was flexible and the path was smooth (Williams et al., 2013; Heckmann, 2016). Consistent in all model approaches, the photorespiratory glycine shuttle is considered as an earlier step prior to the compartmentalized primary PEPC and secondary RuBisCO CO₂ fixation during C₄ photosynthesis evolution. Moreover, the photorespiratory glycine shuttle causes the imbalance of nitrogen between BS and M cells, which might enforce the final transition to C₄ metabolism (Mallmann et al., 2014).

II. Introduction

3. Known genetic control of Kranz anatomy and chloroplast development

In most C_3 plants, carboxylation-decarboxylation cycles happen in M cells. The major role of C_3 BS cells is transporting metabolites between M cells and the vasculature (Leegood, 2008). C_4 plants possess enlarged BS cells with abundant chloroplasts, reflecting the expanded physiological function relative to C_3 BS cells. In C_4 plants, CO_2 fixation and decarboxylation are performed in M and in BS cells, respectively, which are arranged as the Kranz anatomy (V-BS-M-M-BS-V pattern), efficient for shuffling intercellular metabolites. C_3 - C_4 intermediates shared common features with C_4 plants, such as evolution of Kranz-like anatomy, BS cells with abundant organelles. Here, I summarize known genetic mechanisms of Kranz anatomy and chloroplast development. However, much more genetic control of Kranz or Kranz-like leaf anatomy development and the chloroplast positioning in BS cells are still undiscovered.

3.1 Regulation of Kranz anatomy development

The pathway of Kranz development with known genetic control could be categorized into three steps: (1) procambium initiation, (2) cell-types patterning in Kranz, and (3) functionalization of BS cells (chloroplast development) for C_4 cycle (Figure 4; Sedelnikova et al., 2018).

3.1.1 Initiation of procambium (Vein formation)

The leaf procambium is a meristematic tissue, which differentiates into all vascular cells. The initiation of vein-forming procambial cells, induced from ground meristem cells, is reported to be regulated through the feedback loop of auxin signal transduction in *Arabidopsis thaliana* leaves (Donner et al., 2010). The auxin induced transcription factor *MONOPTEROS/AUXIN RESPONSE FACTOR5 (MP/ARF5)* upregulates the auxin efflux carrier *PIN-FORMED1 (PIN1)*, which facilitates directional auxin flow and governs the formation of auxin maxima, where procambial strands form (Scarpella et al., 2006; Wenzel et al., 2007). *MP/ARF5* also activates the *HD-ZIP III* transcription factor *ARABIDOPSIS THALIANA HOMEBOX 8 (ATHB8)*, which identifies preprocambial cell, regulates procambium development, and stabilizes *PIN1* localization and procambial cell fate (Donner et al., 2010). Upregulation of auxin biosynthesis genes, higher auxin content and elevated levels of auxin transport are reported to be associated with higher vein density in developing C_4 leaves relative to C_3 plants

II. Introduction

in *Flaveria*, *Cleome*, and Maize (Wang et al., 2013b; Huang et al., 2017; Billakurthi et al., 2018).

3.1.2 BS and M cell-type patterning

The radial patterning of BS and M cells in Kranz anatomy is reported to be analogous to that of endodermis and cortex in roots and of starch sheath and cortex in shoots, which are regulated through GRAS family transcription factors *Short-Root/SCARECROW (SHR/SCR)* pathway (Nelson, 2011; Slewinski et al., 2014). *SHR* genes are expressed in veins and their proteins move to the adjacent cell layer where they activate *SCR* gene transcription, resulting in BS cell fate specification, and *SCR* also blocks the movement of *SHR* (Cui et al., 2014). The *SHR/SCR* signaling mechanism has been reported to govern the development of Kranz anatomy in maize and *Arabidopsis thaliana* leaves (Cui et al., 2014; Slewinski et al., 2014). Both mutants of *ZmSCR* and *ZmSHR1* showed perturbed vein patterning and vein anatomy, such as abnormal layer number of BS cells or disruption of V-BS-M-M-BS-V structure (Slewinski et al., 2012; Slewinski et al., 2014).

3.1.3 BS chloroplast functionalization

The chloroplasts of land plants developed from the proplastids, which locate in meristematic cells (Waters and Langdale, 2009). The chloroplast differentiation is dimorphic in BS and M cells in C₄ species maize: chloroplasts are agranal and contain large starch granules in BS cells, whereas those in M cells are granal and without starch granules (Langdale, 2011). Additionally, the chloroplast biogenesis differs from M and BS cells in C₄ species leaves, since they function in CO₂ capture and CO₂ reduction, respectively. The regulatory network of chloroplast development and division is reported with groups of transcriptional factors in maize and rice. The maize GARP super family *GOLDEN2-LIKE* transcriptional factors (GLK), *ZmG2* and *ZmGLK1*, are known to regulate the transition from proplastids to chloroplasts in BS and M cells, respectively (Rossini et al., 2001). The same GLK cell-specific mechanism was also found in the C₄ species *Sorghum bicolor* (Wang et al., 2013). The GLK orthologs of *ZmG2* and *ZmGLK1* are also present in C₃ plants, rice (*OsGLK2* and *OsGLK1*) and *Arabidopsis* (*AtGLK1* and *AtGLK2*), which are however functionally redundant (Rossini et al., 2001; Fitter et al., 2002). The *glk1glk2* double mutants of *Arabidopsis* showed increased amount of chlorophyll precursors and deficient photosynthetic apparatus formation (reduction in grana and thylakoids) (Waters et al., 2009). The constitutive expression of *ZmG2* or *ZmGLK1* in rice resulted in an increase of organelle size with enhanced amount of

II. Introduction

photosynthetic enzymes, mimicking the proto-Kranz leaf anatomy (Wang et al., 2017). The primary targets of GLK genes in Arabidopsis are reported to be nuclear-encoded genes responsible for chlorophyll biosynthesis, electron transport component, and light harvesting (Waters et al., 2009).

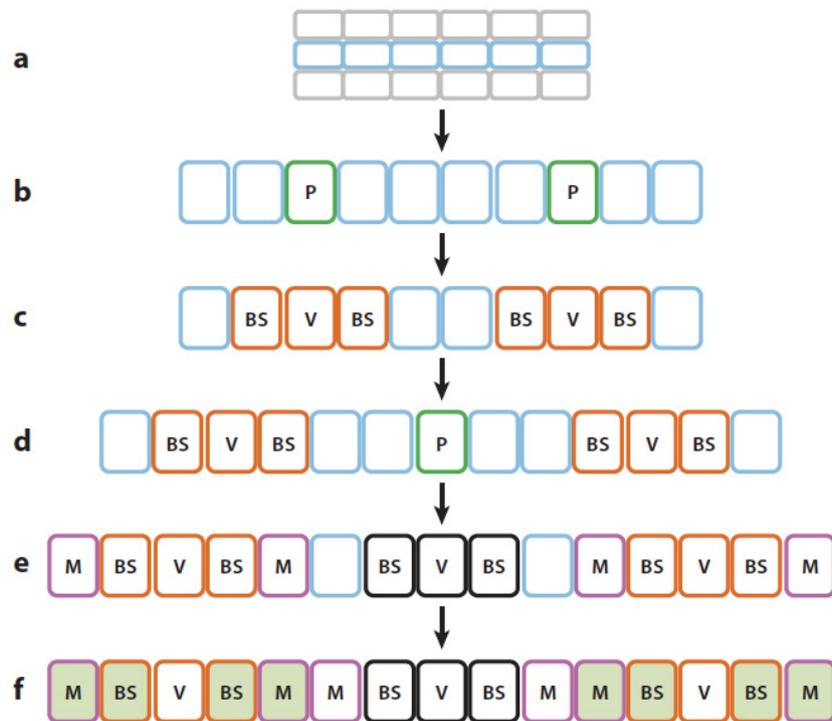


Figure 4. Kranz leaf development pathway, adapted from Sedelnikova et al., 2018.

(a)-(b) Initiation of rank-1 procambium from the cell layer (pale blue) in the innermost leaf. (c)-(e) Rank-2 procambium initiation and BS and M cell-type patterning. (f) Photosynthetic functionalization of BS and M cells.

3.2 Genes related to chloroplast development, division, and movement

3.2.1 Chloroplast development and division

In addition to the GLK family, the GATA transcription factors of Arabidopsis, *GATA NITRATE-INDUCIBLE CARBON-METABOLISM-INVOLVED* (*GNC*) and *CYTOKININ-RESPONSIVE GATA1* (*CGAI*), are expressed in chloroplast-contained tissues, act downstream of cytokinin, and stimulate chloroplast development and division (Figure 5; Chiang et al., 2012). *CGAI* is coexpressed with important chloroplast-localized genes

II. Introduction

involved in photosynthesis, chloroplast and carbon metabolism (Hudson et al., 2013). In rice and Arabidopsis, the mutants with altered expression of *CGAI* differ in chloroplast number, chlorophyll amount, and starch content (Hudson et al., 2013). Another GATA transcription factor *GROWTH REGULATING FACTOR5* (*GRF5*) of Arabidopsis is known to prolong the cell proliferation duration and stimulate chloroplast division, with evidence that the overexpression of *GRF5* caused increased chloroplast number per cell (Vercruyssen et al., 2015).

A complex regulatory system evolved in photosynthetic leaf cells to produce a group of chloroplasts with uniform size and shape through dividing in the middle of chloroplasts (Osteryoung and Pyke, 2014). Chloroplast fission is driven by the inner FtsZ division ring and outer division ring at plastid midpoint. *FtsZ1* and *FtsZ2*, known to encode the bacterial cell division component, regulate the formation of the inner FtsZ ring. The Min system, composed of *MCD1*, *MinC*, and *MinE*, controls the placement of the FtsZ ring. The outer plastid division ring is under control of *ACCUMULATION AND REPLICATION OF CHLOROPLAST 5/DYNAMIN RELATED PROTEIN 5B* (*ARC5/DRP5B*), which is recruited to the division site by *PLASTID DIVISION1* (*PDV1*) and *PDV2*. The cellular levels of *PDV2* are influenced by the CYTOKININ RESPONSE FACTOR 2 (*CRF2*) (Okazaki et al., 2009). The inner envelope-located *ARC6* and *PARALOG OF ARC 6* (*PARC6*) regulate the positioning of the *PDV1* and *PDV2* to outer envelope (Zhang et al., 2016).

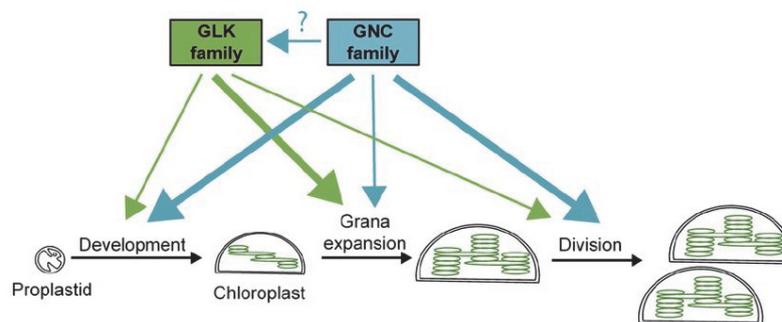


Figure 5. Functions of GLK and GNC family in chloroplast development, adapted from Zubo, 2018.

II. Introduction

3.2.2 Chloroplast movement

In C_3 plants, chloroplasts are distributed along the cell wall of M and BS cells, particularly towards the intercellular air spaces. This distribution is also found in C_3 - C_4 and C_4 M cells. In C_3 - C_4 BS cells, the enlarged chloroplasts accumulated centripetally towards veins. However, chloroplast-positioning patterns in BS cells are differently placed in C_4 plant species: in NADP-ME monocot species, chloroplasts are centrifugally located; in NADP-ME dicot and NAD-ME species, chloroplasts are in centripetal position towards veins. The chloroplast positioning in C_4 BS cells is fixed, but that in C_3 - C_4 , C_4 M cells as well as C_3 M and BS cells is analogously regulated through light-directed chloroplast movements (Luesse et al., 2006; Maai et al., 2011a). Under low light condition, chloroplasts showed accumulation response, chloroplasts move to the periclinal walls to perceive light efficiently, whereas high-intensity light induces the avoidance response that chloroplasts relocate to the anticlinal walls to protect themselves from photodamage (Figure 6A). In C_3 plants Arabidopsis, the blue-light-induced chloroplast photorelocation movement has been well studied, which is mediated by the chloroplast-actin (cp-actin) polymerization and maintenance (Wada and Kong, 2018).

Two blue-light receptors, PHOTOTROPIN 1 (PHOT1) and PHOT2, regulate the blue-light-induced chloroplast photorelocation movement through chloroplast-actin (cp-actin) filaments in most green plants (Kadota et al., 2009; Ichikawa et al., 2011). PHOT1 and PHOT2, localized on the plasma membrane, function redundantly in the accumulation response, whereas PHOT2, localized on the chloroplast outer envelope, mediates the avoidance response. The chloroplast outer membrane located CHLOROPLAST UNUSUAL POSITIONING 1 (CHUP1) was identified serving as the link between chloroplasts and plasma membrane with actin-binding motifs and showed the in vitro interaction with F-actin, G-actin and profilin (Oikawa et al., 2003; Oikawa et al., 2008; Schmidt von Braun and Schleiff, 2008). CHUP1 and the KINESIN-LIKE PROTEIN FOR ACTIN-BASED CHLOROPLAST MOVEMENT (KAC1 and KAC2) are reported to be involved in the cp-actin polymerization (Kadota et al., 2009; Suetsugu et al., 2010). THRUMIN1 (THRUM1) localizes on the plasma membrane and serves as a link between phototropin activity and actin-based chloroplast photorelocation, which is responsible for cp-actin maintenance and chloroplast movement efficiency (Whippo et al., 2011). *J-DOMAIN PROTEIN REQUIRED FOR CHLOROPLAST ACCUMULATION RESPONSE1* (JAC1) mediates the accumulation response (Suetsugu et al., 2005). Under strong light conditions, the *WEB1* (*WEAK CHLOROPLAST MOVEMENT UNDER BLUE LIGHT 1*) and *PMI2* (*PLASTID MOVEMENT IMPAIRED 2*) complex regulates the chloroplast movement via cp-actin filament

II. Introduction

reorganization through suppressing *JAC1* (Kodama et al., 2010). *PMII* might be responsible for cp-actin maintenance, not polymerization (Suetsugu et al., 2015).

In monocot C_4 M cells, the avoidance and accumulation responses of chloroplasts are regulated by blue-light and ABA in *Eleusine coracana* (finger millet) and maize (Maai et al., 2011b). Compared to C_3 plants, the avoidance movement of C_4 M chloroplasts is triggered with stronger light stimuli ($3,000\text{--}4,000 \mu\text{mol m}^{-2} \text{s}^{-1}$) in longer exposure time and also responds to environmental stresses, such as salinity and drought (Yamada et al., 2009). In C_4 plants, BS chloroplasts are located either centrifugally (close the M cells, such as maize) or centripetally (close the vein, as finger millet), whereas $C_3\text{--}C_4$ BS chloroplasts accumulate toward the vascular tissue. In addition, BS chloroplasts have no response to light irradiation. It has been shown that a mechanism in C_4 plants keeps chloroplasts in the home position in BS cells, which is associated with the actomyosin system and cytosolic protein synthesis instead of tubulin or light (Kobayashi et al., 2009). However, the molecular mechanism of chloroplast positioning in $C_3\text{--}C_4$ or C_4 BS specific system remains elusive.

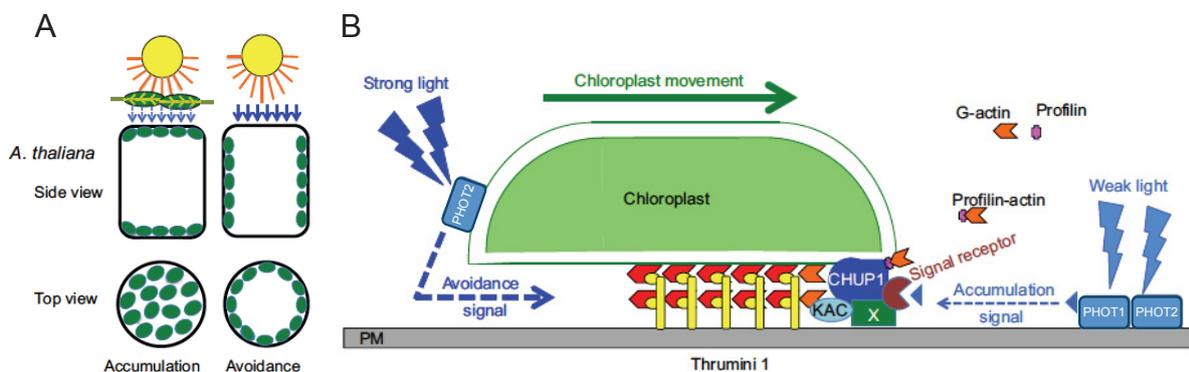


Figure 6. (A) Chloroplast movements in C_3 species, *Arabidopsis thaliana*. (B) Model of actin-mediated movement of chloroplasts.

Both were adapted from Wada and Kong, 2018. PHOT, PHOTOTROPIN; CHUP1, CHLOROPLAST UNUSUAL POSITIONING 1; KAC, KINESIN-LIKE PROTEIN FOR ACTIN-BASED CHLOROPLAST MOVEMENT; PM, plasma membrane; X, an unknown membrane protein.

II. Introduction

4. Old stories and new expects of experimental hybrids in C₃-C₄/C₄ gene discovery

Forward and reverse genetic approaches could facilitate the efficiency of gene discovery, which are widely applied on rice and Arabidopsis. In the case of genetic discovery of C₃-C₄ characters, model species are needed but remain limited. One of promising strategies is to generate experimental hybrids between closely related C₃, C₃-C₄, and C₄ species, which are beneficial for crop breeding programs introducing new genetic variations into cultivated C₃ crops as well as for discovering the genetic differences between photosynthetic types using the genetic mapping strategy.

4.1 Efforts have been done in the past

A number of interspecific and intergeneric hybridizations have been generated in *Atriplex*, *Flaveria*, *Panicum*, *Moricandia*, *Brassica*, *Diploaxis*, and *Raphanus* (Brown and Bouton, 1993; Kadereit et al., 2017). In breeding programs, the direction of hybridization is critical and matters for the hybrid fertility. In the following text, the hybridization will be shown as maternal species × paternal species. Most of experimental hybridizations between photosynthetic types were produced from the 1960s to 1990s. The first experimental C₄ × C₃ hybrids were made of *Atriplex* species (Björkman et al., 1969). In *Flaveria*, hybrids among C₃, C₃-C₄, C₄ like, C₄ species were created (Brown et al., 1986; Apel et al., 1988; Brown and Bouton, 1993). Hybrids of C₃-C₄ × C₃ species were generated in *Panicum* (Brown et al., 1986). Intergeneric hybrids were made of *D. tenuifolia* (C₃-C₄) × *R. sativus* (C₃) (Ueno et al., 2003). Hybrids between different photosynthetic types usually demonstrated intermediate phenotypes of CO₂ compensation points and leaf anatomy. Hybrids between C₃ and C₃-C₄ resembled closer to the C₃ parental species, whereas hybrids of C₄ and C₃-C₄ inherited more C₄ characteristics, such as increased amount of C₄ enzymes. Genetic controls of C₃-C₄ species were examined to be additive, with the evidence of hybrids with increased genetic ratio of C₃-C₄:C₃ genome, demonstrating increased C₃-C₄ characters, such as decreased CO₂ compensation point and increased GDC activity in BS cells (Apel et al., 1984; Ueno et al., 2003).

4.2 A new experimental hybrid studying system in the Brassicaceae: *Moricandia*

The genus *Moricandia*, belonging to the family Brassicaceae, is found in Mediterranean and Saharo-Sindian areas, adapted to arid and drought conditions (Eduardo, 1997; Tahir and Watts, 2010). It provides an outstanding system to study early evolutionary steps of C₄

II. Introduction

photosynthesis, because it comprises species with C₃ and C₃-C₄ species existing in a close phylogenetic proximity: three C₃ species (*M. moricandioides*, *M. foetida*, and *M. foleyi*) and five C₃-C₄ intermediates (*M. arvensis*, *M. suffruticosa*, *M. nitens*, *M. spinosa*, and *M. sinaica*). *Moricandia* species shared phylogenetic proximity with species in other Brassicaceae genera, such as *Brassica* and *Diplotaxis*, as well as with *Arabidopsis thaliana*, which possesses a well annotated genome, and were close to C₄ species in the family Cleome (Kellogg, 1999; Beilstein et al., 2010; Schlüter et al., 2017). Wide intergeneric hybridization of *Moricandia* with species in *Brassica* has been conducted for studying the inheritance of *Moricandia* C₃-C₄ characteristics. The first interspecific C₃-C₄ × C₃ hybrids in *Moricandia* were of *M. arvensis* × *M. moricandioides* (Apel, 1984). Of interest to plant breeders is introducing *Moricandia* C₃-C₄ characteristics into *Brassica* cultivars, improving drought tolerance as well as high photosynthetic efficiency. Subsequent efforts generated intergeneric hybrids between *Brassica* C₃ species and their wild relatives, *Moricandia* C₃-C₄ species, such as *M. arvensis* × *B. alboglabra* (Chinese kale), *M. nitens* × *B. napus* (Rapeseed) and *M. arvensis* × *B. oleracea* (such as cabbage, broccoli, cauliflower) (Apel et al., 1984; Rawsthorne et al., 1998; Ueno et al., 2007).

4.3 Obstacles and new opportunities of the application of experimental hybrids

In the most cases, the interspecific or intergeneric hybrids showed reproductive disorder, caused by alternation in pollen fertility. In very few cases of obtaining F₂ hybrids in *Atriplex* and *Flaveria*, the subsequent construction of mapping populations has failed because of chromosome mispairing and abnormalities (Osmond et al., 1980; Covshoff et al., 2014). Most experimental hybrid studies were abandoned at that time. However, with the access to genetic resources obtained from high-throughput sequencing and bioinformatics, we could focus on applying comparative transcriptomics approaches on two closely related C₃ and C₄ species, or on F₂ populations, as well as conducting allele specific expression analysis on F₁ hybrids through RNA-Seq technologies. For instance, the *Atriplex* C₄ × C₃ hybrids were regenerated and were further applied to discover the transcriptional regulation between photosynthetic types (Oakley et al., 2014; Sultmanis, 2018).

5. Strategies for genetic discovery of C₃-C₄ characteristics

5.1 Exploring regulatory divergences between different photosynthesis types

5.1.1 Transcriptional regulatory mechanism of gene expression

Divergent expression of specific genes is responsible for phenotypic differences within species or between closely related species. A complex network of *cis*-regulatory DNA sequences, *trans*-acting elements, epigenetic variations and post-transcriptional mechanisms governs the regulation of gene expression. The *cis*- and *trans*-regulatory mechanisms have different impacts on the inheritance and evolution of gene expressions. *cis*-regulatory sequences locate in promoter regions, UTRs, and introns, which modulate the binding of *trans*-acting factors to DNA, affecting the transcription of nearby genes. *trans*-acting elements, such as transcription factors and long noncoding RNA, regulate the expression of many genes (Wray, 2007). In addition, *cis*-regulatory divergences are able to cause spatiotemporal expression, activating genes in specific tissue or cellular compartments or limiting gene expression to specific life stages or environments (Prud'homme et al., 2007). They play an important role in many adaptive traits, for instance, a predicted late step of C₄ photosynthesis evolution is establishing C₄ cycle enzymes with spatial or temporal expression adjustments of C₃ genes (Sage et al., 2012).

The *cis*-mechanism dominates interspecific hybrids, representing the long evolutionary time-scales, whereas the *trans*-system is responsible for intraspecific hybrid events, the short evolutionary time-scales (Stern and Orgogozo, 2008; McManus et al., 2010; Rhoné et al., 2017). *cis*-regulatory effects are considered to play a more critical role in the adaptive phenotypic evolution relative to the nonsynonymous mutation in protein sequences, which might result in deleterious pleiotropic effects (Apel et al., 1984; Rawsthorne et al., 1998; Ueno et al., 2007). It has been discovered that *cis*-regulatory divergences dominated the positive selection and the adaptive improvement during maize domestication from teosinte (Lemmon et al., 2014). Thus, *cis*-acting regulations have been reported to play an important role in adaptive phenotypic evolution, especially on the long evolutionary time-scales.

5.1.2 Allele specific expression (ASE) analysis

ASE analysis on heterozygote sites in diploid hybrids is considered as an effective method to distinguish the regulatory effects between *cis*- and *trans*-elements. In hybrids, the allelic expressions are under the same genetic background, sharing non-*cis*-elements. Comparing the allelic ratio of parental alleles (A: PA1/PA2) and that of hybrids (B: F1A1/F1A2) could define the transcriptional regulation on the heterozygote sites to four categories: (1) *cis*-only

II. Introduction

effect ($B \neq 1$ and $A = B$), alleles are differentially expressed and the allele ratio is the same between parental species and hybrids; (2) *trans*-only effect ($B = 1$ and $A \neq B$), two alleles are equally expressed in the hybrid, but not between parental species; (3) *cis*- plus *trans*- effect ($B \neq 1$ and $A \neq B$), two alleles are differently expressed in the hybrid and the allele ratio is different between parental species and hybrids; (4) no *cis*- no *trans*-effect ($A = B = 1$), no allelic effect is observed (Figure 7; Li et al., 2017). With the advent of sequencing technologies, new generation sequencing based approaches, such as RNA-Seq, enables analyzing ASE on a genome-scale. This strategy has been widely applied to yeast, fruit flies, and plants, including *Arabidopsis*, *Capsella*, *Atriplex*, maize, rice, millet (Tirosch et al., 2009; McManus et al., 2010; He et al., 2012; Lemmon et al., 2014; Steige et al., 2015; Rhoné et al., 2017; Sultmanis, 2018; Shao et al., 2019).

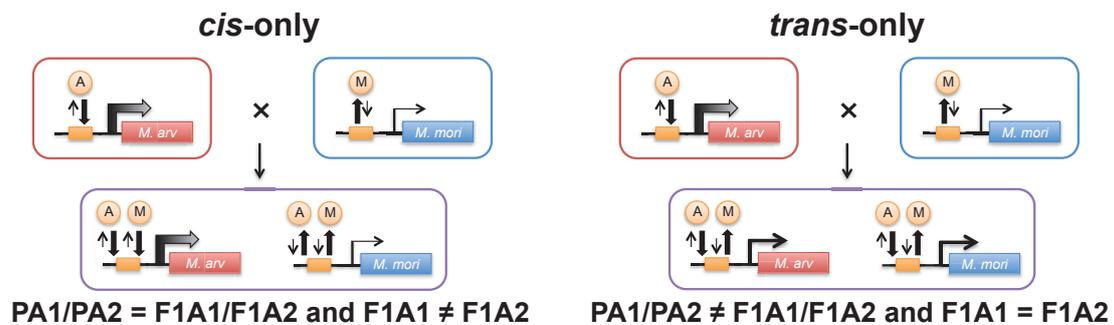


Figure 7. Comparison of the allele ratio between hybrids and parents enables distinguishing effects between *cis*- and *trans*-regulatory changes.

5.2 Comparative transcriptomics of closely related photosynthesis species

With the advent of high-throughput sequencing and the bioinformatics method, the approach applying RNA-Seq on species without complete reference genome information is feasible and has a great impact on genetic researches using closely related C_3 and C_4 species. Additionally, the strategy of mapping to cross-species reference on comparative transcriptome studies, as well as the identification orthologs utilizing the phylogenetic proximity of experimental species and the model plant *Arabidopsis*, have been established and widely applied (Bräutigam et al., 2011; Gowik et al., 2011).

II. Introduction

5.2.1 Comparative transcriptome of mature leaves

The comparative transcriptome analysis of mature leaves of *Cleome* species revealed that 603 transcripts are differentially regulated between closely related C₃ and C₄ plants, including 17 transcription factors, putative transport protein, and genes involved in chloroplast positioning (*CHUP1* and *ACTIN11*), plasmadesmatal connectivity, and cell wall modification (Bräutigam et al., 2011). In *Flaveria*, a comparative transcriptome analysis conducted on mature leaves of C₃, C₃-C₄ intermediate, and C₄ species defines thousands of alternations, revealing that genes in C₄ cycle are upregulated in C₄ and some C₃-C₄ species, many C₄-related transporters are upregulated in C₄ species, such as bile acid sodium symporter (BASS2 and BASS4), and DiT1 (chloroplast dicarboxylate transporter 1), and photorespiration genes are downregulated in C₄, whereas upregulated in C₃-C₄ species in *Flaveria* (Gowik et al., 2011). Similar results were discovered through differential gene expression analysis between C₃, C₃-C₄ intermediate, and C₄ species in Salsoleae. C₄ cycle genes were upregulated in C₄ and C₃-C₄ species, the transcriptional abundance of photorespiration genes decreased in C₄ species and increased in C₃-C₄ species relative to C₃ plants (Lauterbach et al., 2017). In *Moricandia*, the comparison of mature leaf transcriptomes between C₃-C₄ intermediates and C₃ species showed that no large changes were observed for genes involving in photorespiratory pathway as well as in Calvin-Benson cycle and few C₄ cycle genes were upregulated in C₃-C₄ intermediates, such as *AspAT*, *PEPCK*, *PPDK* (Schlüter et al., 2017). These studies significantly extended our knowledge of C₄-related genes and contribute to current models for different C₄ cycles, especially with the analysis of closely related C₃ and C₄ species pairs in different genera and families. However, these studies lacked of the genetic information of early leaf developmental stages when leaf anatomy is established.

5.2.2 Study of C₄ related anatomy via comparative transcriptome analysis during leaf ontogeny

Leaf sections from base to tip of maize could present the developmental dynamics. Comparison of leaf sections of maize revealed that genes for primary cell wall and basic cellular metabolism and for secondary cell wall biosynthesis dominated in the leaf base, whereas genes for C₄ photosynthesis development were abundant in the leaf tip (fully-developed Kranz anatomy) (Li et al., 2010). A comprehensive system biology analysis combining metabolome, transcriptome, and enzyme activity revealed that no intermediate phase was found during the ontogeny of C₄ maize leaf sections (Pick et al., 2011). The leaf primordia genes associated with vein patterning was discovered through comparing different

II. Introduction

ontogeny of Maize foliar (Kranz) and husk (non-Kranz) leaves from leaf primordia to mature leaves (Wang et al., 2013b). Comparison of two independent C₄ lineages, *Gynandropsis gynandra* (formerly known as *Cleome gynandra*) and *Zea mays*, a set of 18 homologous transcription factors was consistently expressed with C₄-related genes during the leaf development (Aubry et al., 2014).

Comparative transcriptomics on the ontogeny of leaf sections or series of developmental leaves between closely related C₃ and C₄ species were investigated in *Cleomaceae* and *Flaveria*. To underpin the shared characteristics of C₄ leaf maturation, transcriptomes of leaf sections of mature leaves of two C₃ and two C₄ *Flaveria* species were implemented and identified candidate genes for vein density, cross sections between M and BS cells, chloroplast ultrastructure, and abundance of C₄ gene transcripts (Kümpers et al., 2017). Comparative transcriptome analysis on series of developmental leaves between *Gynandropsis gynandra* (C₄) and *Tarenaya hassleriana* (C₃) discovered a delay in MS differentiation in C₄ species (Külahoglu et al., 2014). Comparison of transcriptome patterns of series of developmental leaves of C₃ and C₄ *Flaveria* species revealed that auxin metabolism plays an important role for high vein density in C₄ leaves (Billakurthi et al., 2018).

III. Aim of the Thesis

C₃-C₄ species possess the photorespiratory glycine shuttle, functioning in Kranz-like anatomy, enhances the ratio of CO₂ to O₂ in the vicinity of the bundle sheath Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). This mechanism suppresses the oxygenase function of RuBisCO, thereby reducing photorespiration. Some efforts have been made utilizing C₃-C₄ intermediates to unravel the genetic control of C₃-C₄ evolution in *Flaveria*, *Moricandia*, and *Salsola* through comparative transcriptome analysis on mature leaves (Gowik et al., 2011; Lauterbach et al., 2017; Schlüter et al., 2017). However, except the confinement of glycine decarboxylase activity to the bundle sheath cells and the shuttle of metabolites with anatomical changes, transcriptional comparison of mature leaves didn't give much more information. Genes and elements involved in regulation of these C₃-C₄ features remain poorly understood.

The aim of this thesis was to improve our understanding of genetic mechanisms underpinning C₃-C₄ characteristics using *Moricandia* species through insights into transcriptional regulations and transcriptomic comparison with the advantage of available preliminary genome and transcriptome data. As pre-experiments for further genetic investigation, we examined cytogenetic characteristics of *Moricandia* species and the potential of interspecific hybridizations between C₃ and C₃-C₄ photosynthesis type, verified with a reliable molecular marker system (Manuscript I). In Manuscript II, we focused on allele specific expression analysis using interspecific hybrids of C₃-C₄ *M. arvensis* and C₃ *M. moricandioides* based on RNA-Seq of mature leaves. Promoter-GUS assays were applied for examination of altered spatial gene expression of candidates. To dissect the genetic architecture of establishment of C₃-C₄ Kranz-like anatomy, we captured the transcriptome dynamic during leaf development of C₃ and C₃-C₄ *Moricandia* species through comparative transcriptomics and cluster analysis (Manuscript III).

IV. References

- Apel P** (1984) Photosynthetic properties in the genus *Moricandia*. In *Kinetics of Photosynthetic Carbon Metabolism in C3- Plants*, ed J 1: 141–45
- Apel P, Bauwe H, Bassüner B, Maass I** (1988) Photosynthetic Properties of *Flaveria cronquistii*, *F. palmeri*, and Hybrids between Them. *Biochemie und Physiologie der Pflanzen* **183**: 291–299
- Apel P, Bauwe H, Ohle H** (1984) Hybrids between *Brassica alboglabra* and *Moricandia arvensis* and their Photosynthetic Properties. *Biochemie und Physiologie der Pflanzen* **179**: 793–797
- Aubry S, Kelly S, Kümpers BMC, Smith-Unna RD, Hibberd JM** (2014) Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis. *PLOS Genetics* **10**: e1004365
- Bauwe H, Hagemann M, Fernie AR** (2010) Photorespiration: players, partners and origin. *Trends in Plant Science* **15**: 330–336
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S** (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **107**: 18724–18728
- Bellasio C, Farquhar GD** (2019) A leaf-level biochemical model simulating the introduction of C2 and C4 photosynthesis in C3 rice: gains, losses and metabolite fluxes. *New Phytologist* **223**: 150–166
- Billakurthi K, Wrobel TJ, Bräutigam A, Weber APM, Westhoff P, Gowik U** (2018) Transcriptome dynamics in developing leaves from C3 and C4 *Flaveria* species reveal determinants of Kranz anatomy. *bioRxiv* 473181
- Björkman O, Gauth E, Nobs M** (1969) Comparative studies of *Atriplex* species with and without β -carboxylation photosynthesis. *Carnegie Institution of Washington Yearbook* **68**: 620–633

IV. References

- Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Maß J, Lercher MJ, et al** (2011) An mRNA Blueprint for C4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiology* **155**: 142–156
- Brown HR, Bouton JH** (1993) Physiology and Genetics of Interspecific Hybrids Between Photosynthetic Types. *Annual Review of Plant Physiology and Plant Molecular Biology* **44**: 435–456
- Brown RH, Bassett CL, Cameron RG, Evans PT, Bouton JH, Black CC, Sternberg LO, Deniro MJ** (1986) Photosynthesis of F1 Hybrids between C4 and C3-C4 Species of Flaveria1. *Plant Physiol* **82**: 211–217
- Brown RH, Hattersley PW** (1989) Leaf Anatomy of C3-C4 Species as Related to Evolution of C4 Photosynthesis. *Plant Physiol* **19**: 1543–1550
- von Caemmerer S, Quick WP, Furbank RT** (2012) The Development of C4 Rice: Current Progress and Future Challenges. *Science* **336**: 1671–1672
- Chiang Y-H, Zubo YO, Tapken W, Kim HJ, Lavanway AM, Howard L, Pilon M, Kieber JJ, Schaller GE** (2012) Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant physiology* **160**: 332–348
- Covshoff S, Burgess SJ, Kneřová J, Kümpers BMC** (2014) Getting the most out of natural variation in C4 photosynthesis. *Photosynth Res* **119**: 157–167
- Cui H, Kong D, Liu X, Hao Y** (2014) SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in Arabidopsis thaliana. *The Plant Journal* **78**: 319–327
- Donner TJ, Sherr I, Scarpella E** (2010) Auxin signal transduction in Arabidopsis vein formation. *Plant Signal Behav* **5**: 70–72
- Douce R, Bourguignon J, Neuburger M, Rébeillé F** (2001) The glycine decarboxylase system: a fascinating complex. *Trends in Plant Science* **6**: 167–176
- Eduardo S** (1997) Interfertility in the genus *Moricandia* DC. *Lagascalia* **19**: 839–844

IV. References

- Edwards EJ, Osborne CP, Stromberg CAE, Smith SA, C4 Grasses Consortium, Bond WJ, Christin PA, Cousins AB, Duvall MR, Fox DL, et al** (2010) The Origins of C4 Grasslands: Integrating Evolutionary and Ecosystem Science. *Science* **328**: 587–591
- Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA** (2002) GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J* **31**: 713–727
- Furbank RT, Quick WP, Sirault XR** (2015) Improving photosynthesis and yield potential in cereal crops by targeted genetic manipulation: Prospects, progress and challenges. *Field Crops Research* **182**: 19–29
- Gowik U, Bräutigam A, Weber KL, Weber APM, 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**: 2087–2105
- Gowik U, Westhoff P** (2011) The Path from C3 to C4 Photosynthesis. *Plant Physiology* **155**: 56–63
- Hatch MD** (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **895**: 81–106
- 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. *Mol Biol Evol* **29**: 3385–3395
- Heckmann D** (2016) C4 photosynthesis evolution: the conditional Mt. Fuji. *Current Opinion in Plant Biology* **31**: 149–154
- Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ** (2013) Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588
- Huang C-F, Yu C-P, Wu Y-H, Lu M-YJ, Tu S-L, Wu S-H, Shiu S-H, Ku MSB, Li W-H** (2017) Elevated auxin biosynthesis and transport underlie high vein density in C4 leaves. *Proc Natl Acad Sci USA* **114**: E6884–E6891

IV. References

- Hudson D, Guevara DR, Hand AJ, Xu Z, Hao L, Chen X, Zhu T, Bi Y-M, Rothstein SJ** (2013) Rice Cytokinin GATA Transcription Factor1 Regulates Chloroplast Development and Plant Architecture. *Plant Physiology* **162**: 132–144
- Ichikawa S, Yamada N, Suetsugu N, Wada M, Kadota A** (2011) Red light, Phot1 and JAC1 modulate Phot2-dependent reorganization of chloroplast actin filaments and chloroplast avoidance movement. *Plant Cell Physiol* **52**: 1422–1432
- Johnson MP** (2016) Photosynthesis. *Essays In Biochemistry* **60**: 255–273
- Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW** (2017) C3–C4 intermediates may be of hybrid origin – a reminder. *New Phytologist* **215**: 70–76
- Kadota A, Yamada N, Suetsugu N, Hirose M, Saito C, Shoda K, Ichikawa S, Kagawa T, Nakano A, Wada M** (2009) Short actin-based mechanism for light-directed chloroplast movement in Arabidopsis. *Proc Natl Acad Sci U S A* **106**: 13106–13111
- Kellogg EA** (1999) 12 - Phylogenetic Aspects of the Evolution of C4 Photosynthesis. *In* RF Sage, RK Monson, eds, *C4 Plant Biology*. Academic Press, San Diego, pp 411–444
- Khush GS** (1999) Green revolution: preparing for the 21st century. *Genome* **42**: 646–655
- Kobayashi H, Yamada M, Taniguchi M, Kawasaki M, Sugiyama T, Miyake H** (2009) Differential Positioning of C4 Mesophyll and Bundle Sheath Chloroplasts: Recovery of Chloroplast Positioning Requires the Actomyosin System. *Plant Cell Physiol* **50**: 129–140
- Külahoglu C, Denton AK, Sommer M, Maß J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Buell CR, Simon R, et al** (2014) Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C3 and C4 Plant Species. *The Plant Cell* **26**: 3243–3260
- Kümpers BMC, Burgess SJ, Reyna-Llorens I, Smith-Unna R, Bournnell C, Hibberd JM** (2017) Shared characteristics underpinning C4 leaf maturation derived from analysis of multiple C3 and C4 species of Flaveria. *J Exp Bot* **68**: 177–189
- Langdale JA** (2011) C4 cycles: past, present, and future research on C4 photosynthesis. *Plant Cell* **23**: 3879–3892

IV. References

- Lauterbach M, Schmidt H, Billakurthi K, Hankeln T, Westhoff P, Gowik U, Kadereit G** (2017) De novo Transcriptome Assembly and Comparison of C3, C3-C4, and C4 Species of Tribe Salsoleae (Chenopodiaceae). *Front Plant Sci.* doi: 10.3389/fpls.2017.01939
- Leegood RC** (2008) Roles of the bundle sheath cells in leaves of C3 plants. *J Exp Bot* **59**: 1663–1673
- Lemmon ZH, Bukowski R, Sun Q, Doebley JF** (2014) The Role of cis Regulatory Evolution in Maize Domestication. *PLOS Genetics* **10**: e1004745
- Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al** (2010) The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42**: 1060–1067
- 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>
- Luesse DR, DeBlasio SL, Hangarter RP** (2006) Plastid Movement Impaired 2, a New Gene Involved in Normal Blue-Light-Induced Chloroplast Movements in Arabidopsis. *Plant Physiology* **141**: 1328–1337
- Maai E, Miyake H, Taniguchi M** (2011a) Differential positioning of chloroplasts in C4 mesophyll and bundle sheath cells. *Plant Signal Behav* **6**: 1111–1113
- Maai E, Shimada S, Yamada M, Sugiyama T, Miyake H, Taniguchi M** (2011b) The avoidance and aggregative movements of mesophyll chloroplasts in C4 monocots in response to blue light and abscisic acid. *J Exp Bot* **62**: 3213–3221
- Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U** (2014) The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *eLife* **3**: e02478
- McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ** (2010) Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res* **20**: 816–825

IV. References

- Morgan CL, Turner SR, Rawsthorne S** (1993) Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and of serine hydroxymethyltransferase in leaves of C3-C4 intermediate species from different genera. *Planta* **190**: 468–473
- Muangprom A, Thomas SG, Sun T, Osborn TC** (2005) A Novel Dwarfing Mutation in a Green Revolution Gene from *Brassica rapa*. *Plant Physiol* **137**: 931–938
- Muhaidat R, McKown AD** (2013) Significant involvement of PEP-CK in carbon assimilation of C4 eudicots. *Ann Bot* **111**: 577–589
- Nagai K, Hirano K, Angeles-Shim RB, Ashikari M** (2018) Breeding Applications and Molecular Basis of Semi-dwarfism in Rice. *In* T Sasaki, M Ashikari, eds, *Rice Genomics, Genetics and Breeding*. Springer, Singapore, pp 155–176
- Nelson T** (2011) The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C4 leaves. *J Exp Bot* **62**: 3039–3048
- Oakley JC, Sultmanis S, Stinson CR, Sage TL, Sage RF** (2014) Comparative studies of C3 and C4 Atriplex hybrids in the genomics era: physiological assessments. *J Exp Bot* **65**: 3637–3647
- Oikawa K, Kasahara M, Kiyosue T, Kagawa T, Suetsugu N, Takahashi F, Kanegae T, Niwa Y, Kadota A, Wada M** (2003) CHLOROPLAST UNUSUAL POSITIONING1 Is Essential for Proper Chloroplast Positioning. *Plant Cell* **15**: 2805–2815
- 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 Physiol* **148**: 829–842
- Okazaki K, Kabeya Y, Suzuki K, Mori T, Ichikawa T, Matsui M, Nakanishi H, Miyagishima S** (2009) The PLASTID DIVISION1 and 2 Components of the Chloroplast Division Machinery Determine the Rate of Chloroplast Division in Land Plant Cell Differentiation. *The Plant Cell* **21**: 1769–1780
- Osmond CB, Björkman O, Anderson DJ** (1980) *Physiological Processes in Plant Ecology: Toward a Synthesis with Atriplex*. Springer-Verlag, Berlin Heidelberg

IV. References

- Osteryoung KW, Pyke KA** (2014) Division and Dynamic Morphology of Plastids. Annual Review of Plant Biology **65**: 443–472
- Parry MAJ, Reynolds M, Salvucci ME, Raines C, Andralojc PJ, Zhu X-G, Price GD, Condon AG, Furbank RT** (2011) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. J Exp Bot **62**: 453–467
- Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW** (1988) Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C3-C4 intermediate species *Moricandia arvensis* (L.) DC. Planta **176**: 527–532
- Rawsthorne S, Morgan CL, O'Neill CM, Hylton CM, Jones DA, Freen ML** (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*. Theor Appl Genet **96**: 922–927
- Rhoné B, Mariac C, Couderc M, Berthouly-Salazar C, Ousseini IS, Vigouroux Y** (2017) No Excess of Cis-Regulatory Variation Associated with Intraspecific Selection in Wild Pearl Millet (*Cenchrus americanus*). Genome Biol Evol **9**: 388–397
- Rossini L, Cribb L, Martin DJ, Langdale JA** (2001) The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. The Plant Cell **13**: 1231–1244
- Sage RF** (2004) The evolution of C4 photosynthesis. New Phytologist **161**: 341–370
- Sage RF, Khoshravesh R, Sage TL** (2014) From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. J Exp Bot **65**: 3341–3356
- Sage RF, Sage TL, Kocacinar F** (2012) Photorespiration and the Evolution of C4 Photosynthesis. Annu Rev Plant Biol **63**: 19–47
- Scarpella E, Marcos D, Friml J, Berleth T** (2006) Control of leaf vascular patterning by polar auxin transport. Genes Dev **20**: 1015–1027
- Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin P-A, Kurz S, Mettler-Altmann T, Weber AP** (2017) Photosynthesis in C3–C4 intermediate *Moricandia* species. J Exp Bot **68**: 191–206

IV. References

- Schlüter U, Weber APM** (2016) The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol* **57**: 881–889
- Schmidt von Braun S, Schleiff E** (2008) The chloroplast outer membrane protein CHUP1 interacts with actin and profilin. *Planta* **227**: 1151–1159
- Shao L, Xing F, Xu C, Zhang Q, Che J, Wang X, Song J, Li X, Xiao J, Chen L-L, et al** (2019) Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *PNAS* **116**: 5653–5658
- Slewinski TL, Anderson AA, Price S, Withee JR, Gallagher K, Turgeon R** (2014) Short-Root1 Plays a Role in the Development of Vascular Tissue and Kranz Anatomy in Maize Leaves. *Molecular Plant* **7**: 1388–1392
- Slewinski TL, Anderson AA, Zhang C, Turgeon R** (2012) Scarecrow Plays a Role in Establishing Kranz Anatomy in Maize Leaves. *Plant Cell Physiol* **53**: 2030–2037
- Sommer M, Bräutigam A, Weber APM** (2012) The dicotyledonous NAD malic enzyme C4 plant *Cleome gynandra* displays age-dependent plasticity of C4 decarboxylation biochemistry. *Plant Biology* **14**:
- Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T** (2015) Cis-Regulatory Changes Associated with a Recent Mating System Shift and Floral Adaptation in *Capsella*. *Mol Biol Evol* **32**: 2501–2514
- Stern DL, Orgogozo V** (2008) The loci of evolution: how predictable is genetic evolution? *Evolution* **62**: 2155–2177
- Still CJ, Berry JA, Collatz GJ, DeFries RS** (2003) Global distribution of C3 and C4 vegetation: Carbon cycle implications: C4 PLANTS AND CARBON CYCLE. *Global Biogeochemical Cycles* **17**: 6-1-6–14
- Suetsugu N, Dolja VV, Wada M** (2010) Why have chloroplasts developed a unique motility system? *Plant Signal Behav* **5**: 1190–1196
- Suetsugu N, Kagawa T, Wada M** (2005) An Auxilin-Like J-Domain Protein, JAC1, Regulates Phototropin-Mediated Chloroplast Movement in *Arabidopsis*. *Plant Physiology* **139**: 151–162

IV. References

- Sultmanis SD** (2018) Developmental and Expression Evolution in C3 and C4 Atriplex and their Hybrids.
- Tahir M, Watts R** (2010) Moricandia. *In* C Kole, ed, Wild crop relatives: genomic and breeding resources. Springer Verlag, Heideberg, Germany, pp 191–198
- Tirosh I, Reikhav S, Levy AA, Barkai N** (2009) A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* **324**: 659–662
- Ueno O, Bang SW, 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**: 1550–1559
- 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**: 68–79
- Vercruyssen L, Tognetti VB, Gonzalez N, Van Dingenen J, De Milde L, Bielach A, De Rycke R, Van Breusegem F, Inzé D** (2015) GROWTH REGULATING FACTOR5 stimulates *Arabidopsis* chloroplast division, photosynthesis, and leaf longevity. *Plant Physiol* **167**: 817–832
- Voznesenskaya EV, Franceschi VR, Kiirats O, Artyusheva EG, Freitag H, Edwards GE** (2002) Proof of C4 photosynthesis without Kranz anatomy in *Bienertia cycloptera* (Chenopodiaceae). *Plant J* **31**: 649–662
- Voznesenskaya EV, Franceschi VR, Kiirats O, Freitag H, Edwards GE** (2001) Kranz anatomy is not essential for terrestrial C4 plant photosynthesis. *Nature* **414**: 543–546
- Wada M, Kong S-G** (2018) Actin-mediated movement of chloroplasts. *J Cell Sci* **131**: jcs210310
- Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al** (2013a) Evolution of GOLDEN2-LIKE gene function in C3 and C4 plants. *Planta* **237**: 481–495

IV. References

- Wang P, Kelly S, Fouracre JP, Langdale JA** (2013b) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *The Plant Journal* **75**: 656–670
- Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA** (2017) Re-creation of a Key Step in the Evolutionary Switch from C3 to C4 Leaf Anatomy. *Curr Biol* **27**: 3278-3287.e6
- Wang Y, Bräutigam A, Weber APM, Zhu X-G** (2014) Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *J Exp Bot* **65**: 3567–3578
- Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA** (2009) GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. *The Plant Cell* **21**: 1109–1128
- Wenzel CL, Schuetz M, Yu Q, Mattsson J** (2007) Dynamics of MONOPTEROS and PINFORMED1 expression during leaf vein pattern formation in Arabidopsis thaliana. *The Plant Journal* **49**: 387–398
- Whippo CW, Khurana P, Davis PA, DeBlasio SL, DeSloover D, Staiger CJ, Hangarter RP** (2011) THRUMIN1 Is a Light-Regulated Actin-Bundling Protein Involved in Chloroplast Motility. *Current Biology* **21**: 59–64
- Williams BP, Johnston IG, Covshoff S, Hibberd JM** (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife* **2**: e00961
- Wittkopp PJ, Kalay G** (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**: 59–69
- Wray GA** (2007) The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216
- Yamada M, Kawasaki M, Sugiyama T, Miyake H, Taniguchi M** (2009) Differential Positioning of C4 Mesophyll and Bundle Sheath Chloroplasts: Aggregative Movement of C4 Mesophyll Chloroplasts in Response to Environmental Stresses. *Plant Cell Physiol* **50**: 1736–1749

IV. References

Zhang M, Chen C, Froehlich JE, TerBush AD, Osteryoung KW (2016) Roles of Arabidopsis PARC6 in Coordination of the Chloroplast Division Complex and Negative Regulation of FtsZ Assembly. *Plant Physiology* **170**: 250–262

V. Manuscripts

1. Meng-Ying Lin, Nils Koppers, Urte Schlüter, Andreas P.M. Weber. **Interspecific Hybridization between *Moricandia* C₃ and C₃-C₄ Species.** Unpublished work.
2. Meng-Ying Lin, Urte Schlüter, Benjamin Stich, Andreas P.M. Weber. **Transcriptional Regulation Associated with the CO₂ Concentrating Mechanism in *Moricandia*.** In preparation for publication.
3. Meng-Ying Lin, Alisandra K Denton, Urte Schlüter, Andreas P.M. Weber. **Comparative Transcriptome Analysis during Leaf Development on Closely Related C₃ and C₃-C₄ *Moricandia* Species.** In preparation for publication.

1. Manuscript I

Interspecific Hybridization between *Moricandia* C₃ and C₃-C₄ Species

Meng-Ying Lin¹, Nils Koppers¹, Urte Schlüter¹, Andreas P.M. Weber¹

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

Abstract

C₄ plants evolved more than 66 times from independent C₃ lineages through C₃-C₄ intermediates. To better understand the C₄ genetic mechanisms, we need to know more about the intermediate steps on the evolutionary path. However, plant models for discovering the genetic control of C₃-C₄ photosynthetic characters remain limited. The genus *Moricandia* provides a promising system for such purpose because C₃ and C₃-C₄ species exist in a close phylogenetic proximity. Additionally, *Moricandia* species shared phylogenetic proximity with *Brassica* economically important crops, which would be beneficial for improving photosynthetic efficiency in *Brassica* breeding program. In this study, we focused on cytogenetic characteristics of *Moricandia* C₃ and C₃-C₄ species and the potential of interspecific hybrids between C₃ and C₃-C₄ photosynthesis types. The ISSR and SSR molecular marker system was implemented for the validation of *Moricandia* interspecific hybrids. *M. arvensis* and *M. moricandioides* were both diploid with basic 14 chromosomes ($2n=2x=28$), and *M. arvensis* had larger genome size than *M. moricandioides*. The percentage of pod set depended on the parental species and the direction of hybridizations. The crossing of *M. arvensis* as maternal and *M. moricandioides* as paternal species produced the most viable seeds, serving as the potential hybridization system. This study, together with the available genome and transcriptome information, will contribute to the following studies: the allele specific analysis for investigating the gene regulatory network and comparative transcriptomes during leaf development between C₃ and C₃-C₄ *Moricandia* species for understanding the establishment of C₃-C₄ leaf anatomy.

Introduction

Many plants operate an adaptive photosynthetic trait, known as C_4 photosynthesis, which is more efficient than C_3 plants under drought and heat conditions. C_4 plants evolved an efficient CO_2 concentrating mechanism to reduce the carbon and energy loss of photorespiration, usually involving two distinct cell types, mesophyll (M) and bundle sheath (BS) cells. They were organized in a specific leaf anatomy, called Kranz anatomy, the enlarged BS cells with centripetally or centrifugally accumulated organelles located surrounding vascular bundles, which are separated by two layers of M cells (Hatch, 1987). The CO_2 is initially fixed by phosphoenolpyruvate carboxylase (PEPC) in M cells, and the generated C_4 acid is decarboxylated in BS cells, where the released CO_2 is refixed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) (Hatch, 1987; Bräutigam and Gowik, 2016). C_3 - C_4 plants are considered as natural occurring intermediates on the evolutionary trajectory from C_3 to C_4 photosynthesis. They possess a Kranz-like anatomy, BS cells with abundant organelles locating towards the vein, and a CO_2 concentrating mechanism, the so called photorespiratory glycine shuttle, which is formed through the confinement of glycine decarboxylase activity to BS cells. Both C_3 - C_4 and C_4 CO_2 concentrating mechanisms increase the ratio of $CO_2:O_2$ in the vicinity of RuBisCO in BS cells, resulting in repression of oxygenation of RuBisCO.

Despite the complexity of leaf anatomy and biochemistry, C_4 photosynthesis evolved more than 66 times from C_3 progenitors through C_3 - C_4 intermediates (Sage et al., 2012). Studies have been carried out to predict the evolutionary steps of C_4 photosynthesis using modeling approaches with either photosynthesis parameters (Heckmann et al., 2013) or phenotypic landscape including anatomical and biochemical traits (Williams et al., 2013). These studies supported that the photorespiratory glycine shuttle and the confinement of GDC activity to BS cells occurred in early stages, which are considered as important stepping-stones of metabolism during evolutionary transition from C_3 to C_4 (Monson and Rawsthorne, 2000). On the contrary, compartmentalized C_4 pathway and C_4 cycle activation were predicted to have evolved in later stages. The basic biochemistry of C_3 - C_4 and C_4 photosynthesis is well understood. In the evolution process of C_4 photosynthesis, hundreds of genes have been altered in transcriptional abundances between closely related C_3 and C_4 species in *Cleome* and *Flaveria* (Bräutigam et al., 2011; Gowik et al., 2011). However, we still have very limited understanding of genes responsible for transports across organelles, controlling altered cell biology, and C_3 - C_4 / C_4 associated leaf morphologies.

Many efforts have been made to introduce C_4 characters into C_3 rice, aiming at improving the photosynthetic efficiency and increase water and nitrogen use efficiency, the so called C_4 rice

project (Hibberd et al., 2008; von Caemmerer et al., 2012). In addition, it has been suggested that introduction of C₃-C₄ characters into C₃ rice might be an advantageous alternative (Bellasio and Farquhar, 2019). Therefore, identification of genes involved in C₃-C₄ evolution will facilitate the understanding of not only C₄ photosynthesis mechanisms, but also the specific genetic regulation of C₃-C₄ characters, which will contribute to engineering C₃-C₄ characteristics into C₃ crops.

Forward and reverse genetic approaches could facilitate the efficiency of gene discovery. However, in the case of C₃-C₄/C₄ photosynthesis, model species are needed but have not been well developed. One promising strategy is to generate experimental hybrids between closely related C₃ and C₄ species. Interspecific hybridizations between different photosynthesis types (C₃, C₃-C₄, C₄-like, C₄ species) have been conducted in *Atriplex*, *Flaveria*, *Panicum*, and *Moricandia*, and intergeneric hybrids were generated between *Moricandia* and *Brassica*, *Diplotaxis* and *Raphanus* (Brown and Bouton, 1993; Kadereit et al., 2017). The first interspecific hybrids were produced using C₃ and C₄ species in *Atriplex* (Björkman et al., 1969). In *Flaveria*, interspecific hybrids of C₃ × C₄, C₄-like/C₄ × C₃, and C₄ × C₃-C₄ were generated (Brown et al., 1986; Apel et al., 1988; Cameron et al., 1989). Crosses of C₃-C₄ × C₃ were carried out in *Panicum* and *Moricandia*, such as *P. milioides* × *P. laxum* and *M. arvensis* × *M. moricandioides* (Apel, 1984; Brown et al., 1986). Intergeneric hybrids were made of *Diplotaxis tenuifolia* (C₃-C₄) × *Raphanus sativus* (C₃) (Ueno et al., 2007). Moreover, *Moricandia* C₃-C₄ species have also been applied to the hybridization with C₃ *Brassica* species, such as *M. arvensis* × *B. alboglabra*, *M. nitens* × *B. napus*, and *M. arvensis* × *B. oleracea* (Apel et al., 1984; Rawsthorne et al., 1998; Ueno et al., 2007). However, the literature on interspecific hybrids between C₃ and C₃-C₄ species in *Moricandia* is not very extensive, only *M. arvensis* (C₃-C₄) × *M. moricandioides* (C₃) was reported (Apel, 1984).

Above interspecific or intergeneric hybridization studies mainly focused on understanding the inheritance of C₃-C₄/C₄ anatomical and physiological characteristics in the experimental hybrids, such as leaf ultrastructure, photosynthetic features, and C₄ enzyme activity (Brown and Bouton, 1993; Kadereit et al., 2017). Experimental hybrids between C₃ and C₃-C₄ showed characteristics between those of parental species, but closer to the C₃ species. For instance, hybrids of *M. nitens* × *B. napus* and *D. tenuifolia* × *R. sativus* displayed CO₂ compensation points between those of the parents, but the values were closer to that of C₃ parent (Rawsthorne et al., 1998; Ueno et al., 2003). On the other hand, crossing between C₃-C₄ and C₄ demonstrated more apparent C₄ traits, such as high C₄ enzyme activity of *F. trinervia* × *F. linearis* (Brown et al., 1986). These studies aimed at discovery of inheritance of C₃-C₄/C₄

traits through hybridization and would like to apply to breeding programs. However, in most cases, interspecific hybrids were sterile. Even those generated through embryo rescue, ovary culture, *in vitro* fertilization, and somatic fusion eventually failed to construct a mapping population, resulting from failure of chromosome pairing and abnormalities (Osmond et al., 1980; Covshoff et al., 2014). Advances in sequencing technologies have opened a novel era of understanding genetic architecture of interest with available genomic, transcriptomics information and bioinformatics tools (Bansal and Boucher, 2019). These tools provide new insights into studies on sterile hybrids, such as the allele specific expression (ASE) analysis on genome-wide scale could benefit from quantifications of transcriptional abundance in hybrids and parental species through RNA-Sequencing (RNA-Seq). ASE analysis has been applied on hybrids of *Atriplex* C₃ and C₄ species, revealing that 80% of expression changes resulted from *cis*-regulatory divergences, which played an important role in driving C₄ evolution (Sultmanis, 2018). It would be of special interest what transcriptional regulations dominate in C₃-C₄ species, which could be investigated by ASE analysis on hybrids of C₃ and C₃-C₄ species through RNA-Seq technology.

The genus *Moricandia*, belonging to the family *Brassicaceae*, represents an appropriate model for investigating the genetic mechanism of photorespiratory glycine shuttle and Kranz-like leaf anatomy, because it's photosynthetically diversified with C₃ and C₃-C₄ intermediate species. C₃-C₄ intermediate species possess traits, which do not exist in their ancestral C₃ plants, could contribute to understanding biochemical and anatomical aspects of C₄ evolution. In addition, *Moricandia* C₃-C₄ characters are beneficial for improving photosynthetic efficiency of *Brassica* species. In our study, the genome size, the ploidy level and the chromosome number were detected in *Moricandia* species. Crossings between C₃ and C₃-C₄ plants were conducted to test the cross compatibility among *Moricandia* species. Additionally, a molecular marker system was established for validating hybrids. These results gained insights into the genetic nature of *Moricandia* species and their hybrids, together with genome and transcriptome information, which could be applied for further ASE analysis and comparative transcriptome during leaf ontology.

Materials and Methods

Plant materials

Seeds of *Moricandia arvensis* (IPK: MOR1), *M. suffruticosa* (the seed bank at the Royal Kew Gardens: 0105433), *M. moricandioides* (Botanical Garden Osnabrück: 04-0393-10-00) and interspecific hybrids were surface-sterilized using chloride gas and germinated on half MS medium for one week. Then, the seedlings were transferred individually to pots with soil and grown in a growth chamber under 12 h light/12 h dark conditions with 23 °C day/20 °C night temperatures.

Leaf anatomy

The 2 mm² leaf sections were taken near the midrib of the top third of mature rosette leaves for the leaf ultrastructural analysis. The sections were fixed with fixation buffer (2% paraformaldehyde, 2% glutaraldehyde), dehydrated by an acetone series, and embedded with an araldite series. The sections were transferred to the mold filled with fresh araldite and polymerized at 65°C for two days. Semi-thin sections in 2.5 µm thickness obtained by cutting with a glass knife were mounted on slides, stained with 1% toluidine blue for 2 min and washed by distilled water. The leaf ultrastructure was examined under the light microscope, Zeiss Axiophot microscope (Carl Zeiss Microscopy GmbH, Göttingen, Germany).

Estimation of relative nuclear DNA content by flow cytometry

Around 20 mg fresh one-week old leaves were chopped in 1 mL ice cold Otto I solution (0.1 M citric acid, 0.5 % v/v Tween 20) with razor blade in the Petri dish. The standard species young leaves were also chopped together with the sample. The homogenate was filtered through a 42 µm nylon mesh and centrifuged for 5 min at 150 g at RT. The supernatant was carefully removed and left around 100 µl to suspend the pellet. Then, the suspension was added 100 µl Otto I solution, 800 µl Otto II solution (0.4 M Na₂HPO₄·12H₂O), and 50 µg ml⁻¹ propidium iodide and measured at Core Flow Cytometry Facility of the Düsseldorf University Hospital (Düsseldorf, Germany).

Cytogenetic tests for chromosome counting and ploidy level analyses

The meristematic tissues of root tip were collected and incubated in ice-cold water for 12 to 24 hours as the pretreatment. Then, the tissue was fixed in cold freshly prepared Carnoy's solution composed of ethanol and glacial acetic acid (3:1) at RT for 2 to 4 hours. The fixated material was rinse with H₂O for 10 min, followed by acid hydrolysis (1 N HCl) at 60 °C for 6

to 10 min, and washed again with H₂O. The material was stained with Schiff's reagent (3952016 Sigma-Aldrich, St. Louis, USA) at RT for 1 to 2 hours, and then washed with H₂O. The stained material was placed on the microscope slide together with a drop of 45% acetic acid, covered with the slip, and squashed for microscope examination.

Interspecific hybridization

Hybridizations of C₃-C₄ intermediates (*Moricandia arvensis* and *M. suffruticosa*) and C₃ species (*M. moricandioides*) have been conducted twice in June and August 2015 in the green house of Heinrich Heine University, Düsseldorf. The procedure of hybridization was shown in Supplemental Figure 3. The flowers of the maternal parent were emasculated and immediately bagged in the afternoon one day before self-pollination. Then, in the next morning, the stigma of emasculated flower was pollinated with paternal parent' pollens and tagged with the hybridization combination and the date. Therefore, there were four different combinations of hybridizations: *M. arvensis* × *M. moricandioides* (Ma×Mm), *M. moricandioides* × *M. arvensis* (Mm×Ma), *M. suffruticosa* × *M. moricandioides* (Ms×Mm), and *M. moricandioides* × *M. suffruticosa* (Mm×Ms).

DNA extraction

DNA of *M. arvensis*, *M. suffruticosa*, *M. moricandioides* and Ma×Mm hybrids was extracted from leaf tissue followed modified CTAB method. The one-week old leaf tissue was ground in liquid N₂ using Retsch Mill (Retsch, Haan, Germany). The ground sample was suspended in 700 µL of CTAB extraction buffer (100 mM Tris-HCl, pH 8.0, 20 mM EDTA, pH 8.0, 1.4 M NaCl, 1% w/v polyvinylpyrrolidone, 2% w/v CTAB), mixed by vortexing and incubated at 65 °C in a water bath for 30 min (inverting every 10 min). After centrifugation for 10 min at 13,000 rpm at RT, the supernatant was mixed with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and incubated on ice for 10 min. The supernatant was taken after centrifugation for 10 min at 13,000 rpm at RT, then mixed with an equal volume of ice-cold isopropanol and incubated on ice for 30 min. The DNA pellet was collected with centrifugation at 13,000 rpm at RT for 10 min and washed by 70% ethanol 2 to 3 times. The dried pellet was suspended in 100 µL TE buffer. Then, the isolated DNA was qualified and quantified by Nanodrop 2000 spectrophotometer (Thermo Scientific™, Waltham, USA), and then used for genotyping.

Molecular markers and genotyping by PCR

A total of 15 inter simple sequence repeat (ISSR) markers were examined for specific bands and polymorphism within and between three species (Supplemental Table 1). In addition to ISSR markers, SSR mining was performed based on the whole genome sequencing data of *M. moricandioides*. From the initial assembled scaffold covering 766.2 Mb sequence data, 94,443 SSRs were identified by MISA scripting language. The SSR primers were designed using BatchPrimer3 interface modules. Two sets of primers from each SSR type were selected randomly for validation (Supplemental Table 2).

The PCR amplification of ISSR primers was performed in a total reaction of 20 μ L containing 4 μ L 5X buffer, 0.4 μ L 10 mM dNTP, 2.5 μ L 10 mM ISSR primer, 3 μ L 20 ng/ μ L DNA template, 10 μ L ddH₂O, and 0.1 μ L G2 GoTaq DNA polymerase (Promega, Madison, USA). The reaction was carried out on a PCR cycler by following program: initial denaturation at 95 °C for 2 min, 30-35 cycles of denaturation at 95 °C for 30 s, annealing at 45-50 °C for 30 s, extension at 72 °C for 1 min and then final extension at 72 °C for 5 min. The electrophoresis of the PCR product was performed on a 2% agarose gel (my-budget standard Agarose, Bio-Budget Technologies GmbH, Krefeld, Germany) at 80 V for 1 to 2 hours.

The PCR amplification of SSR primers was performed in a total reaction of 25 μ L containing 5 μ L 5X buffer, 0.4 μ L 10 mM dNTP, 2.5 μ L SSR primer forward, 2.5 μ L SSR primer reverse, 3 μ L 20 ng/ μ L DNA template, 13 μ L ddH₂O, and 0.5 μ L HomemadeTag with the same PCR program for ISSR markers. The electrophoresis of the PCR product was performed on a 3 % Agarose super fine resolution gel (VWR, Pennsylvania, USA) at 80 V for 1 to 3 hours.

Results

Genome size, chromosome number and ploidy level of *M. moricandioides* and *M. arvensis*

To facilitate further genetic analysis, the genome size, the chromosome number, the ploidy level of *Moricandia* species were observed through flow cytometry and cytogenetic tests. The estimated genome size of *M. arvensis* (C₃-C₄) and *M. moricandioides* (C₃) by whole genome sequencing differed from each other, with 1.1 Gb and 766 Mb, respectively (personal communication with Nils Koppers). To confirm the genomic sequencing result, the relative nuclear DNA content of *M. arvensis* and *M. moricandioides* was observed using flow cytometry technique. However, no proper internal control of plant species with known genome sizes was found to calculate the absolute genome size, because of the small scale of differences between these two *Moricandia* species. Under the same voltage setting, the observation of 2C peaks, representing G1 phase DNA content, of *M. arvensis* and *M. moricandioides* at around 72 and 66 (propidium iodide fluorescent intensity), respectively, showed that genome of *M. arvensis* was larger than that of *M. moricandioides*, corresponding to the estimated genome sizes from whole genome sequencing (Figure 1).

Previous cytological studies of the basic information on chromosome number and ploidy level indicated that *M. arvensis* is diploid with basic chromosome number of 14 ($2n = 2x = 28$), however the information of *M. moricandioides* genome remains limited (Warwich and Al-Shehbaz, 2006). We counted the chromosome number of *M. arvensis* and *M. moricandioides* using the cytogenetic method based on light microscopic examination. Fresh root tips were collected and processed following the protocol. Both of the species were counted to have 2n number of 28 in a single root tip cell (Figure 2). All in all, *M. arvensis* and *M. moricandioides* are both diploid species with basic chromosome number of 14, whereas the genome size of *M. arvensis* is larger than that of *M. moricandioides*.

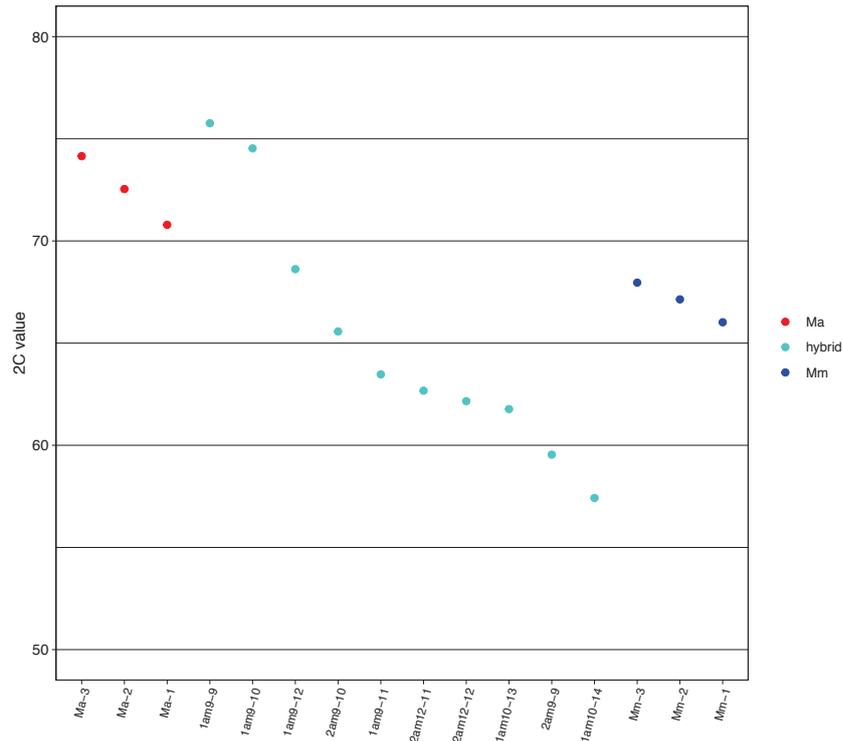


Figure 1. Comparison of genome size between *M. arvensis* (red, Ma), *M. moricandioides* (blue, Mm) and their interspecific hybrids (mediumturquoise).

The 2C value obtained from flow cytometry were based on propidium iodide fluorescent intensity, which represented the relative genome size.

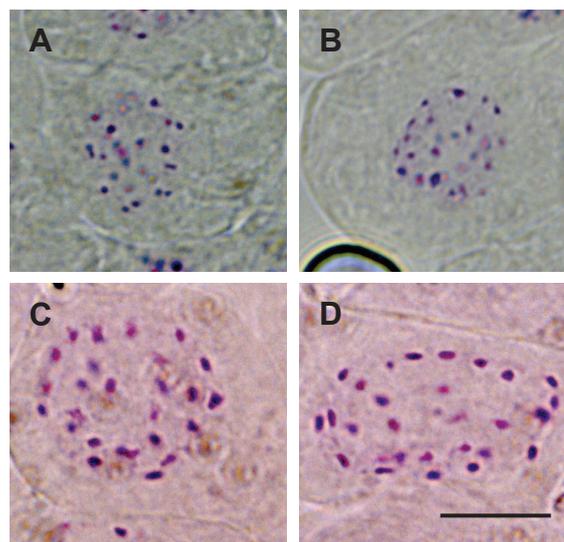


Figure 2. Chromosome counting of *M. arvensis* (A, B) and *M. moricandioides* (C, D) under light microscope.

Root tip cells were stained in Feulgen solution. Bar, 10 μ m. All images are at the same magnification.

Interspecific hybridizations of *Moricandia* species

To investigate the genetic nature of different photosynthesis types in *Moricandia* and to produce genetic materials for plant breeding and gene discovery, hybridizations of C₃-C₄ plants (*M. arvensis* and *M. suffruticosa*) and C₃ species (*M. moricandioides*) have been conducted twice in June and August of 2015, respectively. Generally, *M. arvensis* needed fewer days from seed germination to flowering and showed higher germination rate, followed by *M. suffruticosa* and *M. moricandioides* (Table 1). We designed four different combinations of hybridizations: *M. arvensis* × *M. moricandioides* (Ma×Mm), *M. moricandioides* × *M. arvensis* (Mm×Ma), *M. suffruticosa* × *M. moricandioides* (Ms×Mm), and *M. moricandioides* × *M. suffruticosa* (Mm×Ms) (Figure 3). The successful rate of hand hybridization of Ma×Mm, Mm×Ma, Ms×Mm, and Ms×Mm was 87%, 67%, 23%, and 53%, respectively (Table 2). At the end, we received numerous mature pods from hybridization lines: the pod amount of each line was 165, 24, 15, 21 and 150, 30, 50, 26 in two planting times, respectively. The germination rate of the hybrids from Ma×Mm, Mm×Ma, Ms×Mm, and Ms×Mm was 86%, 25%, 32% and 33%, respectively. In addition, we found very little seeds from mature pods of Mm×Ma, Ms×Mm, and Ms×Mm, and even if few seeds from them successfully germinated, the plant development and growth were abnormal (dwarf and unable to flower). The observation of F₁ hybrids from Ma×Mm and Ms×Mm revealed that leaf shapes and growth habits of hybrids were neither always intermediate shape between those of their parents, nor uniform between each other (Supplemental Figure 1). To sum up, the C₃ and C₃-C₄ species in *Moricandia* could be hybridized, however the seed viability of interspecific hybrids depended on the combination of parental species as well as the direction for hybridization.

Table 1. Comparison of days from germination to flowering between two planting seasons and the germination rate of *Moricandia* species.

Species	Days from gemination to flowering		Germination rate (%)
	Season I (06.2015)	Season II (08.2015)	
<i>M. arvensis</i>	39	34	100
<i>M. moricandioides</i>	55	49	63
<i>M. suffruticosa</i>	49	41	83

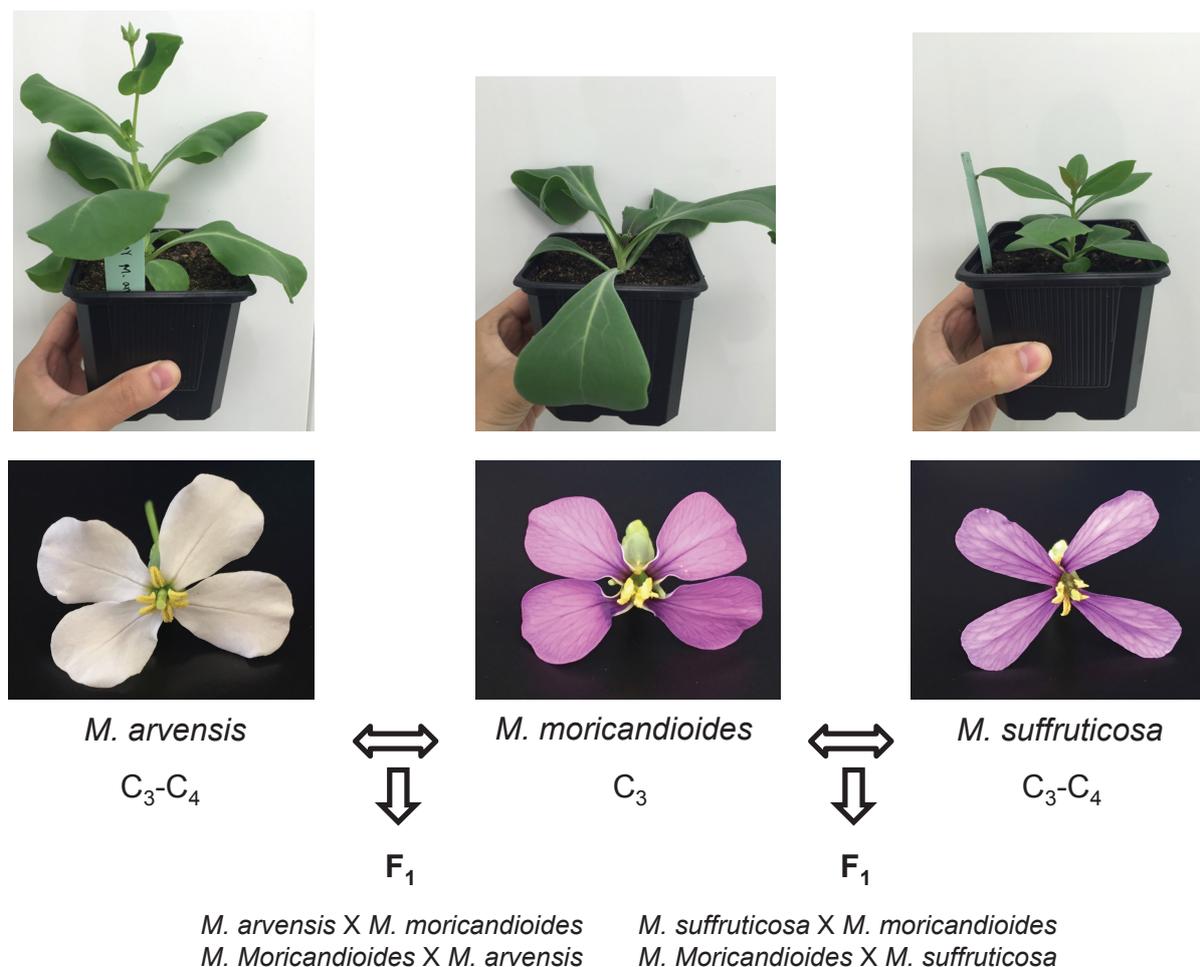


Figure 3. Four combinations of interspecific hybridization between C₃-C₄ (*M. arvensis* and *M. suffruticosa*) and C₃ (*M. moricandioides*) species.

Table 2. An overview of interspecific hybridizations of *Moricandia* species on successful rate of hybridization, total number of mature pods, viable seeds number and germination rate.

The hybridization rate was calculated by successful hybridized pods to all hand hybridization events.

Hybridization	Abbreviation	Successful rate of hybridization (%)	Number of mature pods from Jun/Aug 2015	Number of seeds per mature pods	Germination rate (%)
<i>M. arvensis</i> × <i>M. moricandioides</i>	Ma×Mm	87	165/150	8-45	85.9
<i>M. moricandioides</i> × <i>M. arvensis</i>	Mm×Ma	67	24/30	0	25.0
<i>M. suffruticosa</i> × <i>M. moricandioides</i>	Ms×Mm	23	15/50	1-22	31.8
<i>M. moricandioides</i> × <i>M. suffruticosa</i>	Mm×Ms	53	21/26	0	33.0

Genetic Validation of Hybrids through ISSR and SSR markers

Seedlings of *Moricandia* hybrids were not easy to identify as hybrids by the appearance, although leaf shapes of Ma×Mm and Ms×Mm were a bit more similar to the paternal species, *M. moricandioides* (Supplemental Figure 1). The inter simple sequence repeat (ISSR) system is independent from genomic information, because they were designed to be complimentary to single sequence repeats found in the eukaryotic genomes, which is ideal for species without complete reference genome. Therefore, to discover the molecular markers for genotypic verification of hybrids, we first screened with ISSR markers for the polymorphism between *Moricandia* species, because the genome sequencing data of *Moricandia* species was still missing at the time of the experiments. A total of 15 ISSR markers were examined for specific bands and polymorphism between C₃-C₄ (*M. arvensis*, *M. suffruticosa*) and C₃ species (*M. moricandioides*) (Supplemental Table 1). ISSR001 ((TC)₈G), ISSR004 ((GA)₈YA), ISSR008 ((GA)₈C), ISSR009 ((TC)₈C), and ISSR012 ((GA)₈YT) were confirmed to serve as genetic markers for validating the hybrids (Supplemental Figure 2). ISSR001 could be used in identifying hybrids of Ma×Mm, Ms×Mm, and Mm×Ms; ISSR004, ISSR008, ISSR009, and ISSR012 were markers for all four combinations. Thus, there were at least four markers available for each hybridized combination (Table 3).

ISSR markers are convenient and cost-effective, which could be applied on different species, however they are dominant markers, amplifying multiple locus, which are not always stable (Sarwat, 2012). Simple sequence repeat (SSR) markers are designed by the DNA region anchoring simple sequence repeats, which are high reproducible and has codominant nature. Therefore, SSR mining was performed as the preliminary genome sequencing data of *M. moricandioides* was available. From the initial assembled scaffold covering 766.2 Mb *M. moricandioides* sequence data, 94,443 SSRs were identified by MISA scripting language. In total, we discovered five types of SSR from dinucleotide to hexanucleotide, and dinucleotide SSRs were the most abundant type with 81.3% in all SSRs (Table 4). The SSR primers were designed using BatchPrimer3 interface modules. Two sets of primers from each SSR type were randomly selected for PCR validation (Supplemental Table 2). SSR02, SSR04, SSR08, and SSR10 showed polymorphisms between *M. arvensis* and *M. moricandioides*. The SSR02, anchoring dinucleotide repeat of (TC)₁₀, amplified a product with about 143 bp, showing the potential as genetic marker for Ma×Mm hybridization. Taken together, the combination of ISSR and SSR marker could confirm *Moricandia* interspecific hybrids. For instance, six Ma×Mm hybrids were verified using ISSR004 and SSR02 (Figure 4).

Table 3. The inter simple sequence repeat (ISSR) markers utilized for hybrids verification.

Primer	Nucleotide sequence (5'-3')	Fragment size range (bp)	Ma×Mm	Mm×Ma	Ms×Mm	Mm×Ms
ISSR001	TCTCTCTCTCTCTCG	380-2000	V		V	V
ISSR004	GAGAGAGAGAGAGAYA	200-1350	V	V	V	V
ISSR008	GAGAGAGAGAGAGAC	220-1500	V	V	V	V
ISSR009	TCTCTCTCTCTCTCC	250-1700	V	V	V	V
ISSR012	GAGAGAGAGAGAGAYT	220-1600	V	V	V	V

Table 4. The simple sequence repeats (SSR) identified in *M. moricandioides*.

Type of SSR	Number	Proportion in all SSR (%)
Dinucleotide	76,758	81.3
Trinucleotide	15,76	16.7
Tetranucleotide	992	1.0
Pentanucleotide	333	0.4
Hexanucleotide	599	0.6

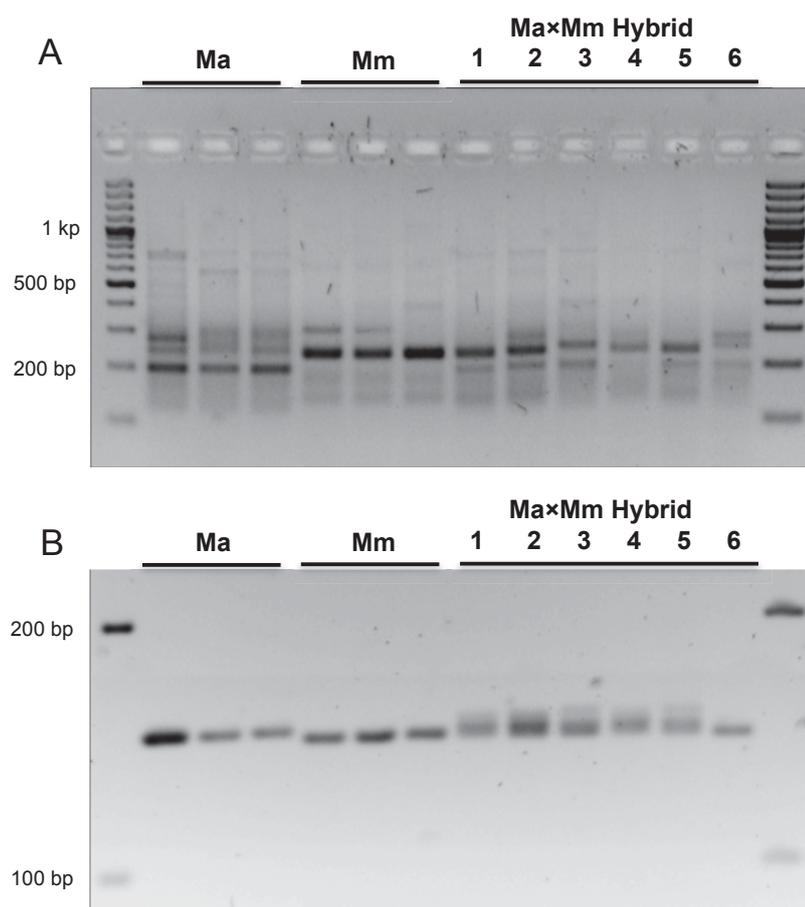


Figure 4. Genotyping of Ma×Mm Hybrid using (A) ISSR (ISSR004) and (B) SSR (SSR02) marker. Ma, *M. arvensis*; Mm, *M. moricandioides*; Ma×Mm Hybrid1-6 represented I Ma×Mm 9-1, II Ma×Mm 12-1, II Ma×Mm 9-1, II Ma×Mm 12-3, I Ma×Mm 9-6, and I Ma×Mm 11-8, respectively. 100 bp DNA Ladder was shown as the reference.

Interspecific hybridization of Ma×Mm and generation of F₂ population

We chose the most prolific cross of maternal species *M. arvensis* and paternal species *M. moricandioides* (Ma×Mm) for further genetic investigation, since its higher germination rate (85.9%) and higher capability to bloom compared to the other three hybridized combinations. Moreover, we also examined the genome size and the ploidy level of Ma×Mm hybrids through detecting the propidium iodide fluorescent intensity at 2C peaks derived from flow cytometry experiments. The genome size of Ma×Mm hybrids was located generally between those of parents, whereas some of them showed larger genome than *M. arvensis* genome, and some showed smaller genome than that of *M. moricandioides* (Figure 1). This indicated that hybrids did not diversify from parents at ploidy level, thus all of them were diploid, with one genome set from each parental species, but they possessed different genome sizes.

Hybrid lines of Ma×Mm were preceded to self-pollinate for producing the F₂ population. Most of F₂ pods contained no viable seeds (Figure 5A). Only two F₁ lines (II Ma×Mm 12-1 and II Ma×Mm 9-6) produced few mature F₂ pods, but they often possessed wrinkled seeds and had difficulties to germinate (Figure 5B). Additionally, the rarely successfully germinated F₂ seedlings were dwarfs and failed to flower. The low successful rate of producing F₂ population might result from incompatible sterility of interspecific hybrids. To test whether pollen from hybrid plants was active, pollen grains were taken from mature anthers and dyed by Alexander staining method. Under light microscope, pollen grains from hybrids were stained magenta-red, meaning that pollens were active, but showed abnormal shapes, which looked like the development was in the midway of pollen synthesis (Lin Manuscript II).

To conclude, Ma×Mm hybrids were diploid containing one set genome of C₃ and the other set of C₃-C₄ species, however the genome size of them were various from parents and from each other. The F₁ hybrids were sterile because of malformed pollen grains, causing very few viable F₂ seeds.

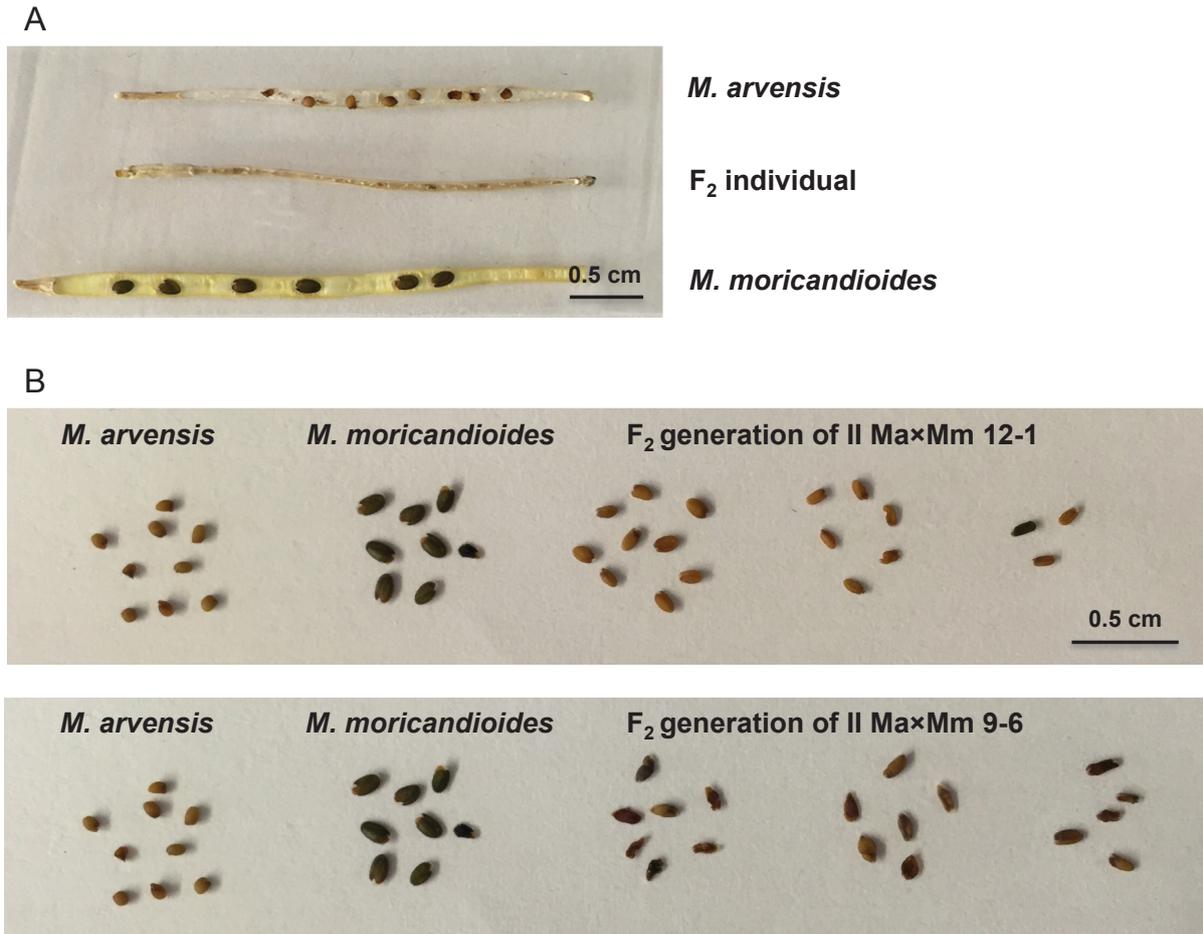


Figure 5. Comparison of parental species and F_2 line on (A) pods and (B) seeds.

Most F_2 pods were without viable seeds. Only few seeds from F_2 generations of II Ma x Mm 12-1 and II Ma x Mm 9-6 were obtained, which were more similar to seeds of *M. arvensis*. Additionally, some of F_2 seeds were shrink.

Discussion

Interspecific hybridizations between different photosynthetic types in *Moricandia*

Experimental hybrids within the *Moricandia* genus have been established through crossing of C₃-C₄ species (such as *M. suffruticosa* × *M. arvensis*) and of C₃ and C₃-C₄ species (*M. arvensis* × *M. moricandioides*) (Apel, 1984; Eduardo, 1997). The phenotypic information was very limited to the observation that the photosynthetic properties of interspecific hybrids of *M. arvensis* × *M. moricandioides* were likely to be influenced by both parents, such as an intermediate level of carbon compensation (Apel et al., 1984; Rawsthorne and Hylton, 1991). However, these historical crosses were no longer available, and we did not have the same lineages as previous studies. In order to further analyse *Moricandia* interspecific crosses with molecular tools, we reconstructed crosses using our collected species and retested the potential of interspecific hybridization of *Moricandia* species. In our study, crosses between C₃ species (*M. moricandioides* from Botanical Garden Osnabrück: 04-0393-10-00) and C₃-C₄ species (*M. arvensis* from IPK: MOR1 and *M. suffruticosa* from the seed bank at the Royal Kew Gardens: 0105433) proved again that species in *Moricandia* could be hybridized. In general, *Moricandia* interspecific F₁ hybrids showed intermediate morphological features, such as leaf shapes, between the two parental species (Supplemental Figure 1), indicating that hybrids inherited the phenotypic characters from parents. The most prolific combination was using *M. arvensis* as maternal genotype and *M. moricandioides* as paternal resource, whereas the other combinations produced few seeds. It has been shown that *M. moricandioides* was reproductively isolated from the other *Moricandia* C₃-C₄ species (Eduardo, 1997). The interspecific sterility might result from pre-fertilization and/or post-fertilization barriers, such as pollen-sigma incompatibility barriers, failed pollen germination, abnormal pollen tube development, and aborted pollen tube penetration to the embryo (Tuyl, 1997; Kathe et al., 2019). All in all, most Ma×Mm hybrids were pollen sterile and the self-pollination of F₁s produced barely F₂ seeds.

The issue of sterile hybrids might result from aborted pollens, causing either no pollen tube germination or failed endosperm development, thereby defective seeds. In the process of pollen development, a pollen mother cell divides to four pollen grains by meiosis. The division processes are based on precisely pairing of the chromosomes, therefore different structures of the chromosomes may lead to pairing failures especially when a hybrid is from two distant species (Harvey, 1988). Thus, the malformed pollen grains of Ma×Mm hybrids might indicate meiotic chromosome irregularities (Lin Manuscript II). It corresponded to genomic differences between *M. arvensis* and *M. moricandioides*: *M. arvensis* has a larger

genome and more duplicated sequences compared to *M. moricandioides*. Similar phenomena of reproductive disorder was reported very often in interspecific or intergeneric hybridization, which was caused by alternation in pollen fertility (Tuyl, 1997; Katche et al., 2019). In very few cases of obtaining F₂ hybrids in *Atriplex* and *Flaveria*, the subsequent construction of mapping populations has failed because of chromosome mispairing and abnormalities (Osmond et al., 1980; Covshoff et al., 2014). Taken together, *Moricandia* species could be crossed, but the crossability is influenced by cross direction. Many methods to overcome interspecific crossing barriers, such as *in vitro* fertilization, ovary culture, embryo rescue, and somatic fusion, have been utilized on hybridization between *Brassica* and *Moricandia* species and could be further applied on *Moricandia* interspecific hybridization.

Ma×Mm is the potential interspecific hybridization system for plant breeding and genetic researches

Interspecific or intergeneric hybridization between C₃ and C₃-C₄ species has been widely applied for addressing the inheritance of C₃-C₄ characteristics (Brown and Bouton, 1993; Kadereit et al., 2017). Also, C₃-C₄ characteristics with potentially beneficial aspects were introduced into crop plants through hybridization (Rawsthorne et al., 1998; Ueno et al., 2007). In this study, we showed that Ma×Mm was the most prolific hybridization, which was verified with the reliable molecular marker system (Table 2; Figure 4). The first interspecific hybridization of C₃ and C₃-C₄ species (Ma×Mm) in *Moricandia* showed that the CO₂ compensation points of Ma×Mm hybrids were between values of parental species (Apel, 1984). Discovery on more Ma×Mm hybridized events demonstrated that the hybrids' CO₂ compensation points as well as leaf anatomy patterns were generally between those of parents, but closer to C₃ species (Lin Manuscript II). Based on flow cytometry observation, the hybrids were diploid and had variation in genome sizes close to the parental species, resulting from chromosome pairing failures or the heterozygous state of parental species. Genomes (genes) of Ma×Mm hybrids comprised of one set of chromosomes from C₃ species and the other set from C₃-C₄ species, resulting in mixed physiological phenotypes (Lin Manuscript II). It confirmed that the C₃-C₄ characters of *M. arvensis*, including photosynthetic properties and leaf anatomical features, could be introduced into Ma×Mm hybrid plants. Therefore, Ma×Mm hybrid is the most potential system to study C₃-C₄ characters in *Brassicaceae*, which will also contribute to *Brassica* crop breeding program.

New aspects on genetic researches in *Moricandia*

Previous studies aimed at understanding the inheritance of C₃-C₄/C₄ characters and introducing them into C₃ crops through experimental interspecific or intergeneric hybridizations, which were somehow terminated because of the difficulties constructing further generations. In the high-throughput era of advanced sequencing technologies, such as whole genome sequencing and RNA-Seq through PacBio and Illumina platforms, genome and transcriptome are available for species of interest. Therefore, the preliminary genome and transcriptome of *M. arvensis* and *M. moricandioides* available in our lab, which could facilitate discovering the genetic control regarding C₃-C₄ characters, especially photorespiratory glycine shuttle and Kranz-like anatomy. Together with the Ma×Mm hybrid system, ASE analysis could be applied for understanding the transcriptional regulations between *Moricandia* C₃ and C₃-C₄ species. Moreover, the preliminary genome could contribute to investigate C₃-C₄ characters during leaf ontology through comparative transcriptomics.

Conclusions

In this study, we confirmed the genetic information of *M. arvensis* and *M. moricandioides*: both of them are diploid but differ in genome sizes. Four hybridized combinations between *Moricandia* C₃ and C₃-C₄ species (Ma×Mm, Mm×Ma, Ms×Mm, and Ms×Mm) were constructed and verified with a reliable molecular system. Ma×Mm was the most prolific hybridization and showed intermediate C₃-C₄ characters. Additionally, with the availability of genome and transcriptome, Ma×Mm served as the most potential system in *Brassicaceae* to study the genetic control of C₃-C₄ characters, which will also broaden genetic variability of *Brassica* crops breeding.

Author contributions

MY.L. designed and generated the interspecific hybridizations, performed all experiments, and wrote the manuscript.

N.K. assembled the genomes and discovered single sequence repeats in *M. moricandioides*.

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

References

- Alexander MP** (1969) Differential Staining of Aborted and Nonaborted Pollen. *Stain Technology* **44**: 117–122
- Apel P** (1984) Photosynthetic properties in the genus *Moricandia*. In *Kinetics of Photosynthetic Carbon Metabolism in C3- Plants*, ed J **1**: 141–45
- Apel P, Bauwe H, Bassüner B, Maass I** (1988) Photosynthetic Properties of *Flaveria cronquistii*, *F. palmeri*, and Hybrids between Them. *Biochemie und Physiologie der Pflanzen* **183**: 291–299
- Apel P, Bauwe H, Ohle H** (1984) Hybrids between *Brassica alboglabra* and *Moricandia arvensis* and their Photosynthetic Properties. *Biochemie und Physiologie der Pflanzen* **179**: 793–797
- Bansal V, Boucher C** (2019) Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going? *iScience* **18**: 37–41
- Bellasio C, Farquhar GD** (2019) A leaf-level biochemical model simulating the introduction of C2 and C4 photosynthesis in C3 rice: gains, losses and metabolite fluxes. *New Phytologist* **223**: 150–166
- Björkman O, Gauthier E, Nobs M** (1969) Comparative studies of *Atriplex* species with and without β -carboxylation photosynthesis. *Carnegie Institution of Washington Yearbook* **68**: 620–633
- Bräutigam A, Gowik U** (2016) Photorespiration connects C3 and C4 photosynthesis. *J Exp Bot* **67**: 2953–2962
- Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Maß J, Lercher MJ, et al** (2011) An mRNA Blueprint for C4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiology* **155**: 142–156
- Brown HR, Bouton JH** (1993) Physiology and Genetics of Interspecific Hybrids Between Photosynthetic Types. *Annual Review of Plant Physiology and Plant Molecular Biology* **44**: 435–456

- Brown RH, Bassett CL, Cameron RG, Evans PT, Bouton JH, Black CC, Sternberg LO, Deniro MJ** (1986) Photosynthesis of F1 Hybrids between C4 and C3-C4 Species of Flaveria 1. *Plant Physiol* **82**: 211–217
- von Caemmerer S, Quick WP, Furbank RT** (2012) The Development of C4 Rice: Current Progress and Future Challenges. *Science* **336**: 1671–1672
- Cameron RG, Bassett CL, Bouton JoeH, Brown RH** (1989) Transfer of C4 Photosynthetic Characters through Hybridization of Flaveria Species 1. *Plant Physiol* **90**: 1538–1545
- Covshoff S, Burgess SJ, Kneřová J, Kümpers BMC** (2014) Getting the most out of natural variation in C4 photosynthesis. *Photosynth Res* **119**: 157–167
- Eduardo S** (1997) Interfertility in the genus *Moricandia* DC. *Lagascalia* **19**: 839–844
- Gowik U, Bräutigam A, Weber KL, Weber APM, 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**: 2087–2105
- Hatch MD** (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **895**: 81–106
- Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ** (2013) Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588
- Hibberd JM, Sheehy JE, Langdale JA** (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Curr Opin Plant Biol* **11**: 228–231
- Jiménez JF, Sánchez-gómez P** (2012) Molecular taxonomy and genetic diversity of *Moricandia moricandioides* subsp. *pseudofortida* compared to wild relatives. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **146**: 99–105
- Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW** (2017) C3–C4 intermediates may be of hybrid origin – a reminder. *New Phytologist* **215**: 70–76

- Katche E, Quezada-Martinez D, Katche EI, Vasquez-Teuber P, Mason AS (2019)** Interspecific Hybridization for Brassica Crop Improvement. *Crop Breeding, Genetics and Genomics*. doi: 10.20900/cbagg20190007
- Monson RK, Rawsthorne S (2000)** CO₂ Assimilation in C₃-C₄ Intermediate Plants. *In* RC Leegood, TD Sharkey, S von Caemmerer, eds, *Photosynthesis: Physiology and Metabolism*. Springer Netherlands, Dordrecht, pp 533–550
- Osmond CB, Björkman O, Anderson DJ (1980)** *Physiological Processes in Plant Ecology: Toward a Synthesis with Atriplex*. Springer-Verlag, Berlin Heidelberg
- Rawsthorne S, Hylton CM (1991)** The relationship between the post-illumination CO₂ burst and glycine metabolism in leaves of C₃ and C₃-C₄ intermediate species of *Moricandia*. *Planta* **186**: 122–126
- Rawsthorne S, Morgan CL, O'Neill CM, Hylton CM, Jones DA, Freen ML (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*. *Theor Appl Genet* **96**: 922–927
- Sage RF, Sage TL, Kocacinar F (2012)** Photorespiration and the Evolution of C₄ Photosynthesis. *Annu Rev Plant Biol* **63**: 19–47
- Sarwat M (2012)** ISSR: A Reliable and Cost-Effective Technique for Detection of DNA Polymorphism. *In* NJ Sucher, JR Hennell, MC Carles, eds, *Plant DNA Fingerprinting and Barcoding: Methods and Protocols*. Humana Press, Totowa, NJ, pp 103–121
- Sultmanis SD (2018)** Developmental and Expression Evolution in C₃ and C₄ Atriplex and their Hybrids.
- Tuyl J (1997)** Interspecific hybridization of flower bulbs: A review. *Acta Horticulturae* **430**: 465-476. doi: 10.17660/ActaHortic.1997.430.76
- Ueno O, Bang SW, 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**: 1550–1559

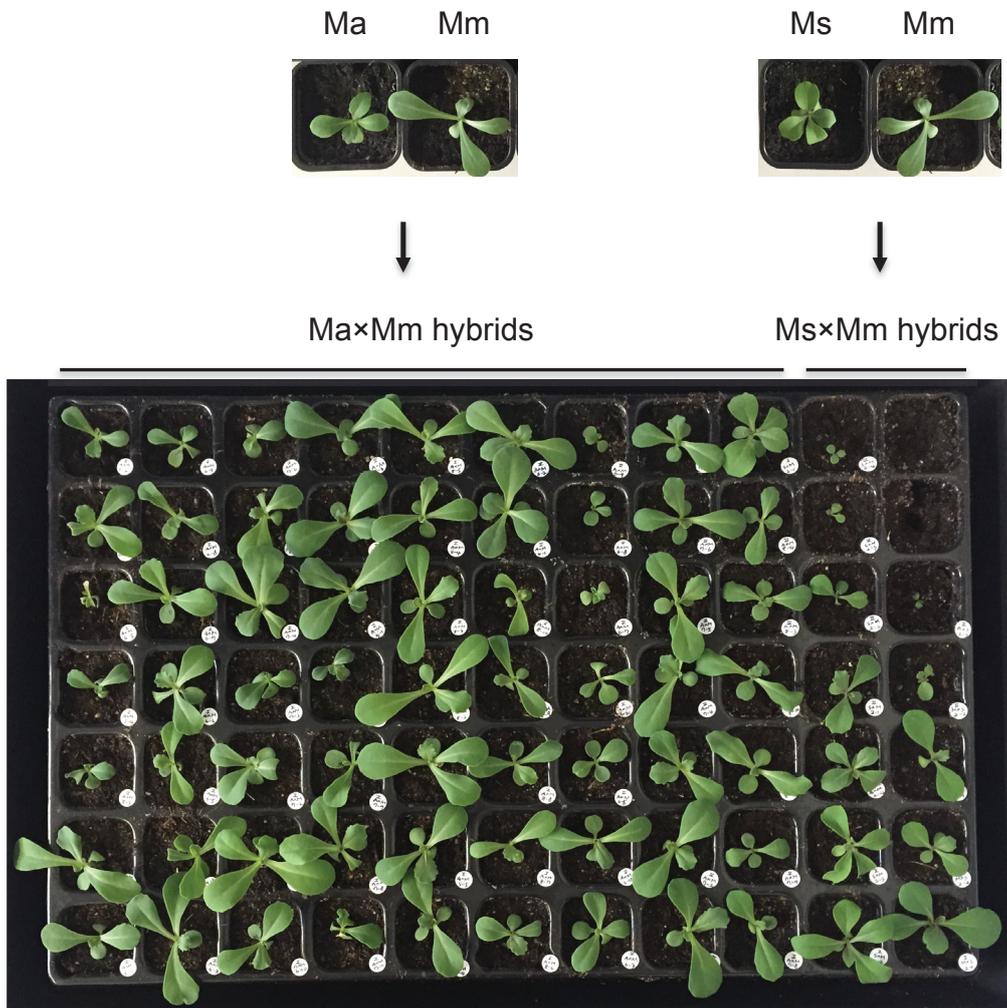
Ueno O, Woo Bang S, Wada Y, Kobayashi N, Kaneko R, Kaneko Y, Matsuzawa Y (2007)

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**: 68–79

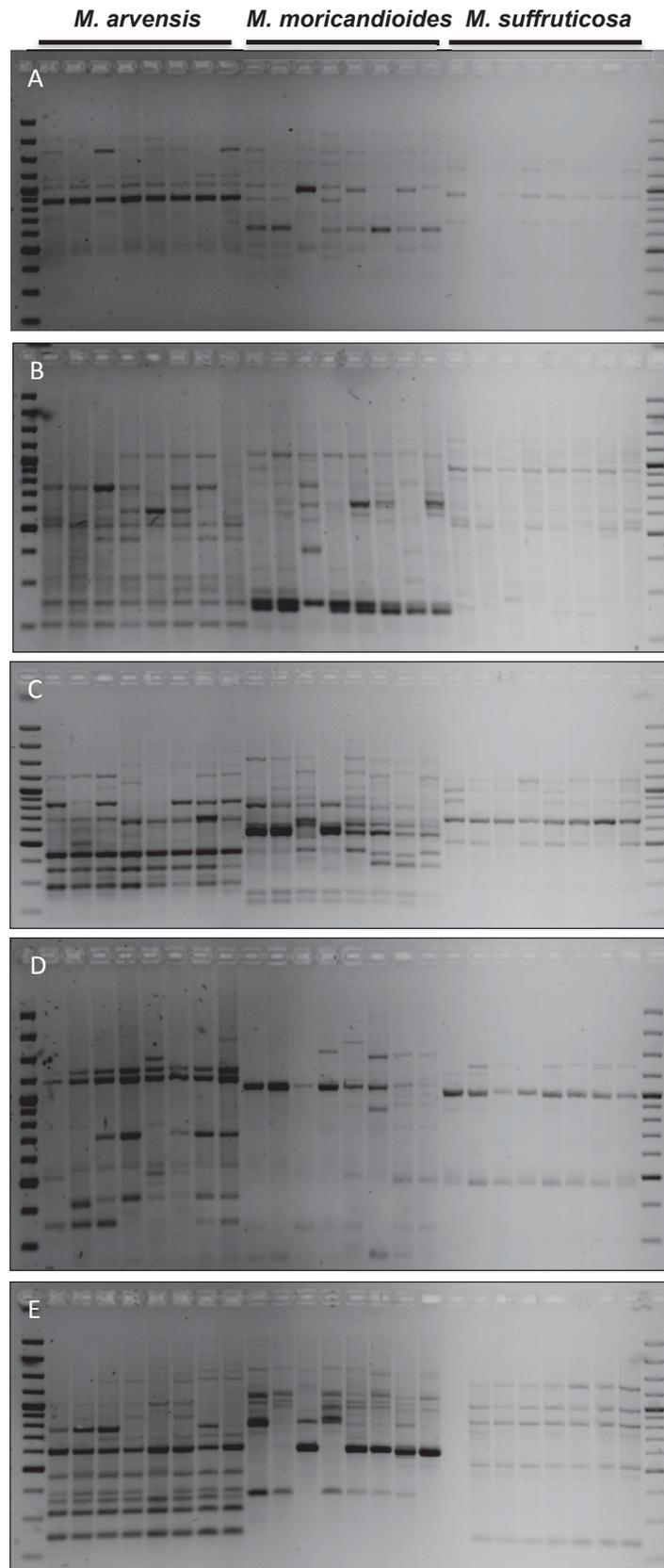
Williams BP, Johnston IG, Covshoff S, Hibberd JM (2013) Phenotypic landscape

inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife* **2**: e00961

Supplemental information



Supplemental Figure 1. Seedlings of Ma×Mm and Ms×Mm.



Supplemental Figure 2. Polymorphism of ISSR markers among *M. arvensis*, *M. moricandioides* and *M. suffruticosa*.

A-E, ISSR001, ISSR004, ISSR008, ISSR009, ISSR012, respectively.

V. Manuscript I



One day before
pollination, choose
flowers just flower



Carefully remove
anthers by a scissor



Bag the stigma one day
for being mature



The next morning, treat stigma
with pollens and cover by bag
to prevent contamination

Supplemental Figure 3. Procedure of hand hybridization.

Supplemental Table 1. A list of 15 selected ISSR markers.

Name	Nucleotide sequence(5'-3')	Resource
ISSR001	TCTCTCTCTCTCTCTCG	Jiménez and Sánchez-gómez, 2012
ISSR002	GTGTGTGTGTGTTG	Jiménez and Sánchez-gómez, 2012
ISSR003	CACACACACARY	Jiménez and Sánchez-gómez, 2012
ISSR004	GAGAGAGAGAGAGAGAYA	Jiménez and Sánchez-gómez, 2012
ISSR005	GAGAGAGAGAGAGAGAYC	Jiménez and Sánchez-gómez, 2012
ISSR006	AGAGAGAGAGAGAGAGT	UBC807
ISSR007	AGAGAGAGAGAGAGAGC	UBC808
ISSR008	GAGAGAGAGAGAGAGAC	UBC811
ISSR009	TCTCTCTCTCTCTCTCC	UBC823
ISSR010	AGAGAGAGAGAGAGAGYT	UBC834
ISSR011	AGAGAGAGAGAGAGAGYA	UBC836
ISSR012	GAGAGAGAGAGAGAGAYT	UBC840
ISSR013	GAGAGAGAGAGAGAGAYC	UBC841
ISSR014	CACACACACACACACART	UBC846
ISSR015	ACACACACACACACACYT	UBC855

Supplemental Table 2. A list of designed SSR markers.

Ma, *M. arvensis*; Mm, *M. moricandioides*.

Name	SSR sequence	Primer sequence (5'-3')	Product size (bp)	Result
SSR01	(TA)16	F AAGCGGTGTCTCACTCTCG R CGTACATGGGATGTGTCGCT	153	no product in Ma
SSR02	(TC)10	F CAAAACAACACTCTCCGGCG R ATGGAAGGGCAATTGAGGCA	143	polymorphism between Ma and Mm
SSR03	(AGG)5	F GTGCGTTTGAAGGAGGAGGA R CGAGGCGGGTTTAGATCCTC	100	primer dimer
SSR04	(ATA)5	F CGAGAATCTTGGCAATGGCG R TTCATAGTGAAGCAGCCGCA	101	polymorphism between Ma and Mm
SSR05	(AATC)5	F AAACCCACATCAACATCGT R TGACCTCGAAGCTCAAGACC	100	no product in Ma
SSR06	(TTCT)9	F ATGGCCGAAAAGGTGTTTGC R AATCAGTACACGGGCTTGGG	126	no product in Ma
SSR07	(GTTGA)6	F TAAGGGGGCGGTAAGATCGA R CTACAAAATTGCCGCTCGG	108	no product in Ma
SSR08	(TTGGT)9	F GCCAATCACCGTTCTCTGGA R CGTTTGAGACCTTGCAACCG	137	polymorphism between Ma and Mm
SSR09	(AATTTT)6	F ACGACTGTTTTATCTTTTTGCA R ACTCCAAAGTTTGAGTTGTGGAG	124	no product
SSR10	(TGAGTT)5	F TCCAAAGGGTGACAGAGGGA R CGCAGGTTGAACCTTTCGTA	105	polymorphism between Ma and Mm

2. Manuscript II

Transcriptional Regulation Associated with the CO₂ Concentrating Mechanism in *Moricandia*

Meng-Ying Lin¹, Urte Schlüter¹, Benjamin Stich², Andreas P.M. Weber¹

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

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

Abstract

Altered spatial gene expression patterns that are caused by regulatory divergences play an important role in the evolution of C₄ 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 regulatory divergences between C₃ and C₃-C₄ intermediates, using allele specific gene expression (ASE) analysis of *Moricandia arvensis* (C₃-C₄), *M. moricandioides* (C₃) and their interspecific hybrids. ASE analysis on SNP-level showed similar relative proportions of regulatory effects among hybrids: 30% and 7% of SNPs were controlled by *cis*-only and *trans*-only changes, respectively. GO terms involved in isopentenyl diphosphate biosynthesis, carbohydrate catabolic process, oxidoreduction coenzyme metabolic process, and chloroplast relocation were abundant in transcripts with common *cis*-SNPs. Transcripts with *cis*-specificity expressed biased toward C₃-C₄ intermediacy. Additionally, ASE evaluated on transcript-level indicated ~34% of transcripts exhibiting ASE in *Moricandia* hybrids. Promoter-GUS assays on selected genes confirmed altered spatial gene expression, which might result from regulatory divergences in promoter regions. With this approach, we assessed the ASE in *Moricandia* interspecific hybrids to gain more insight into the impact and importance of transcriptional regulations, participating in early evolutionary steps of C₄ photosynthesis.

Introduction

Ribulose-1,5-bisphosphate (RuBP) carboxylase-oxygenase (RuBisCO) is a bifunctional enzyme with affinity not only to CO₂ also to O₂. The oxidation of RuBP by RuBisCO generates a toxic intermediate, phosphoglycolate, and then proceeds to release CO₂ through the photorespiration pathway, which is energy-consuming and leads to release of carbon. C₄ plants overcome this problem with an efficient CO₂ pump, usually involving two cellular compartments, mesophyll (M) and bundle sheath (BS) cells, arranged as concentric layers of cells known as Kranz anatomy. The BS cells with abundant organelles located adjacent to vascular bundles are surrounded by M cells (Hatch, 1987). The CO₂ is fixed through PEPC in M cells and the generated C₄ acid is decarboxylated in BS cells, where the released CO₂ increases the CO₂:O₂ ratio around RuBisCO, resulting in high carboxylation rate (Bräutigam and Gowik, 2016; Hatch, 1987). The CO₂ compensation point was defined as the CO₂ concentration where photosynthetic CO₂ uptake equals respiratory CO₂ release. It has been shown that C₄ plants have much lower CO₂ compensation points relative to C₃ plants (Krenzer et al., 1975). The current model of C₄ evolution holds that C₄ plant species evolved from the ancestral C₃ state via C₃-C₄ intermediacy. C₃-C₄ intermediate species display Kranz-like leaf anatomy, including BS cells with centripetally localized mitochondrion and chloroplasts, and their CO₂ compensation points are between the values of C₃ and C₄ plants. The evolutionary trajectory from C₃ to C₄ photosynthesis includes a number of anatomical and biochemical adaptive steps, depicted in different models (Sage et al., 2012; Heckmann et al., 2013; Williams et al., 2013; Mallmann et al., 2014). Based on studies of various naturally occurring C₃-C₄ intermediates, a stepwise model was proposed: (1) the vein density increases; (2) the leaf proto-Kranz anatomy evolves; (3) a photorespiratory CO₂ pump built by the reduced M:BS ratio and the confinement of mitochondrial glycine decarboxylase (GDC) activity to BS cells; (4) enzymes of C₄ metabolic cycle are established with spatial or temporal expression adjustments of C₃ genes (Sage et al., 2012). The consensus trajectories of the statistical (Williams et al., 2013) and mechanistic (Heckmann et al., 2013; Mallmann et al., 2014) models confirmed these steps, but the order of steps was flexible and the path was smooth (Williams et al., 2013; Heckmann, 2016). All C₄ evolution models predicted that the photorespiratory CO₂ pump, resulting from the confinement of GDC activity to BS cells, is a crucial step.

The photorespiratory CO₂ pump in C₃-C₄ plants functions as an efficient CO₂ concentrating system for recycling of photorespiratory released CO₂, the so called glycine shuttle or C₂ photosynthesis (Sage et al., 2014; Schlüter and Weber, 2016; Kadereit et al., 2017). This

system evolved via confining the expression of the gene encoding the P-subunit of glycine decarboxylase (GLDP) to the BS cells. The GLDP is one of the four subunits of the GDC complex, and responsible for the decarboxylase activity (Bauwe, 2011). GLDP activity is absent from leaf M cells of C₃-C₄ plants (Monson and Rawsthorne, 2000), therefore, in order to complete the photorespiratory pathway, glycine must be shuttled to the BS cells, where CO₂ released from mitochondria can be efficiently recaptured by numerous, surrounding chloroplasts.

In the genera *Flaveria* (*Asteraceae*) and *Moricandia* (*Brassicaceae*), the molecular mechanisms by which *GLDP* expression becomes confined to BS cells during evolution of C₃-C₄ intermediacy have been resolved. The genomes of C₃ *Flaveria* species encode two isoforms of *GLDP*, one BS specific isoform (*GLDPA*) and the other ubiquitously expressed in all photosynthetic tissues (*GLDPB*). *GLDPB* becomes a pseudogene in C₃-C₄ intermediacy *Flaveria* and thereby GDC activity is lost from M cells during C₄ photosynthesis evolution (Schulze et al., 2013). A conceptually similar mechanism underpins the independent evolution of C₃-C₄ intermediacy in the *Brassicaceae*. In C₃ *Brassicaceae*, the promoter of the *GLDPI* gene carries two conserved *cis*-regulatory elements, one that drives expression the M cells (M-box), another that governs expression in the vasculature (V-box). The M-box is lost from the *GLDPI* promoter of C₃-C₄ *Brassicaceae*, which leads to the restriction of GDC activity to BS cells (Adwy et al., 2015). The M-box is conserved in the promoter region of *GLDPI* of C₃ *Moricandia* species, but lost in C₃-C₄ *Moricandia* species (Adwy, 2018). The establishment of the C₃-C₄ intermediate photorespiratory CO₂ pump very likely requires further metabolic adjustments and anatomical modifications, probably through altered transcriptional regulation. How these are achieved remains unclear.

The regulation of gene expression, governed by a complex network of *cis*-regulatory DNA sequences, *trans*-acting elements, epigenetic variations and post-transcriptional mechanisms, is responsible for phenotypic diversity within species or between closely related species. *cis*-acting regulations have been reported to play an important role in adaptive phenotypic evolution because they, compared to the nonsynonymous mutation in protein sequences, caused reduced deleterious pleiotropic effects (Wray, 2007; Stern and Orgogozo, 2008; Wittkopp and Kalay, 2012). For instance, *cis*-regulatory divergences dominated the positive selection and the adaptive improvement during maize domestication from teosinte, and genes with *cis*-regulatory effect demonstrated a directional bias toward maize (Lemmon et al., 2014). Additionally, *cis*-regulatory divergences have impacts on limiting gene expression to particular tissue or cellular compartments, to specific life stages or environments

(Prud'homme et al., 2007). However, studies on spatial-temporal gene expression pattern are infeasible by differential gene expression analysis and time-consuming through forward genetics. For example, the confinement of GDC to BS cells in leaves of C₃-C₄ intermediate and C₄ species was first discovered by immunogold labeling in *Moricandia*, *Panicum*, *Flaveria*, and *Mollugo* (Hylton et al., 1988). The genetic control of *GLDP* was assessed through the comparison of promoter sequences, further confirmed by GUS-promoter assays (Bauwe et al., 1995; Zhang et al., 2004; Adwy et al., 2015). Allele specific expression (ASE) analysis on heterozygote sites in diploid hybrids is considered as an effective method to identify *cis*-acting factors, as allelic expressions are under the same feedback control and sharing non-*cis*-elements. Comparing the allelic ratio between parental alleles and that in hybrids could distinguish the effect between *cis*- and *trans*-factors (Li et al., 2017). Progress in sequencing technologies, next-generation sequencing (NGS)-based approaches, such as RNA-Seq, enables analyzing ASE on a genome-scale. This strategy has been widely applied to yeast, fruit flies, and plants, including *Arabidopsis*, *Capsella*, *Atriplex*, maize, rice, millet (Tirosh et al., 2009; McManus et al., 2010; He et al., 2012; Lemmon et al., 2014; Steige et al., 2015; Rhoné et al., 2017; Sultmanis, 2018; Shao et al., 2019). To our knowledge, however, ASE has not yet been applied to interspecific hybrids of parents displaying different photosynthetic traits, such as C₃ and C₃-C₄ intermediate photosynthesis.

Anatomical, biochemical, physiological, and the phylogenetic evidence underpins the hypothesis that C₃-C₄ plants represent naturally occurring intermediates during C₄ evolution and thus ideal materials to unravel the early steps of C₄ evolution (Sage, 2004; Sage et al., 2012; Bräutigam and Gowik, 2016; Schlüter and Weber, 2016; Kadereit et al., 2017). In addition to experimental hybrids of C₃ and C₄, hybridizations between C₃-C₄ plants and C₃ or C₄ present promising systems to discover the genetic differences between photosynthetic types (Kadereit et al., 2017). C₃-C₄ plants have been reported in 21 plant lineages including dicot as well as monocot species, such as *Diploaxis*, *Flaveria*, *Moricandia*, *Neurachne*, and *Panicum* (Sage et al., 2011). Interspecific hybrids of C₃-C₄ and C₃ *Panicum* showed intermediate characteristics of CO₂ exchange and leaf anatomy between that of parental species (Brown et al., 1985). In intergeneric hybrids of *M. nitens* (C₃-C₄) × *Brassica napus* (C₃) (Rawsthorne et al., 1998), *D. tenuifolia* (C₃-C₄) × *Raphanus sativus* (C₃) (Ueno et al., 2003), and *M. arvensis* (C₃-C₄) × *B. oleracea* (C₃) and their reciprocal crosses (Ueno et al., 2007), GDC increased in BS cells, but was not completely confined to BS cells, and CO₂ compensation points were between the values observed in the parental lines. However, the leaf anatomy of *M. nitens* (C₃-C₄) × *B. napus* (C₃) resembled that of the C₃ parent

(Rawsthorne et al., 1998). Hybridizations of *Moricandia* with distant *Brassica* relatives have been reported in the literature, though embryo rescue, sexual crosses and somatic hybridizations (list in Warwick et al., 2009). Hybrids of *M. arvensis* × *M. moricandioides* (C₃) as well as *B. alboglabra* (C₃) × *M. arvensis* showed intermediate CO₂ compensation points between the values of parents (Apel et al., 1984; Brown and Bouton, 1993). However, analysis of the transcriptional regulation of interspecific hybrids in *Moricandia* has not been reported.

In this study, we assessed ASE on SNP- and transcript-level in *Moricandia* by means of RNA-Seq on *M. arvensis* (C₃-C₄), *M. moricandioides* (C₃), and six of their interspecific hybrids. Gene ontology assessments were introduced to discover genes participating in chloroplast relocation and demonstrating extreme allele imbalance. The spatial gene expression pattern of selected ASE genes was validated by promoter-GUS analysis in *A. thaliana*. Our results provide new insight into the impact of the *cis*-regulatory effect on early evolutionary steps of C₄ photosynthesis, especially the installation of the glycine shuttle.

Materials and Methods

Plant materials

Seeds from *Moricandia* were surface-sterilized using chloride gas and germinated on half MS medium for one week. Then, the seedlings were transferred individually to pots with soil and grown in the growth chamber under 12h/12h light/dark conditions with 23°C/20°C day/night. The anthers of *M. arvensis* (C₃-C₄ intermediates, as maternal plant) were removed and their stigma was bagged one day before the artificial cross-pollination. The pollens from *M. moricandioides* (C₃ species, as paternal plant) were collected and applied on the receptive stigma of *M. arvensis*. The reciprocal crosses were done in the same way. The two-week-old leaves were used for DNA extraction for genotyping and promoter region amplification. And the two youngest leaves from four-week-old plants were collected as materials for RNA-Seq. Moreover, the mature rosette leaves were taken for leaf anatomy and gas exchange analysis. Seeds from *Arabidopsis thaliana* wild-type plants ecotype Col-0 and the transgenic lines were surface sterilized by vapour-phase seed sterilization, further germinated on half MS medium with cold treatment for two days in the dark, and then transferred to the growth chamber under 10h/14h light/dark conditions with 22°C/20°C day/night for 10 days. The seedlings were later transferred individually to pots with soil and grown in the growth chamber. The two-week-old *Arabidopsis* T₁ plants and the wild-type plants were collected for further GUS staining analysis.

Leaf anatomy

The 2 mm² leaf sections were taken near the midrib of the top third of mature rosette leaves for the leaf ultrastructural analysis. The sections were fixed with fixation buffer (2% paraformaldehyde, 2% glutaraldehyde), dehydrated by an acetone series, and embedded with an araldite series. The sections were transferred to the mold filled with fresh araldite and polymerized at 65°C for two days. Semi-thin sections in 2.5 μm thickness obtained by cutting with a glass knife were mounted on slides, stained with 1% toluidine blue for 2 min and washed by distilled water. The leaf ultrastructure was examined under the light microscope, Zeiss Axiophot microscope (Carl Zeiss Microscopy GmbH, Göttingen, Germany).

Photosynthetic gas exchange

The mature rosette leaves were chosen to measure gas exchange characteristics using a LI-6400XT Portable Photosynthesis System (LI-COR Biosciences, Lincoln, USA) with the settings: the flow of 300 μmol s⁻¹, the light source of 1500 μmol m⁻² s⁻¹, the leaf

temperature of 25°C, and the vapor pressure deficit based on leaf temp less than 1.5 kPa. The CO₂ response curve, so called A-Ci curve, was captured by detecting net CO₂ assimilation rates under different intercellular CO₂ concentrations. A partial A-Ci curve obtained with measurements at 400, 100, 80, 65, 45, 25, 15, and 400 ppm CO₂ was used to calculate the CO₂ compensation points of parental species and hybrid lines.

Sample preparation and RNA sequencing

We selected 12 plants including three replicates of *M. arvensis*, three plants of *M. moricandioides* and six lines of F₁ interspecific hybrids (Ma×Mm) with divergent phenotype (table 1). Total RNA of parental species and interspecific F₁s was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Then, 17 µl total RNA (100 ng/µl) was added with 2 µl buffer and 0.5 µl RNase-free DNaseI enzyme (New England Biolabs GmbH, Frankfurt am Main, Germany) incubating on ice for 30 s. The treatment was stopped by adding 2 µl 50 mM EDTA and incubated at 65°C for 10 min. The quality of RNA and DNaseI treated RNA was assessed on a Bioanalyzer 2100 (Agilent, Santa Clara, USA) with an RNA Integrity Number (RIN) value greater than or equal to 8. Subsequently, cDNA libraries were prepared using 1 µg of total RNA with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, USA). The cDNA library was qualified on the Agilent Technologies 2100 Bioanalyzer to check the library quality and fragment size of the sample. RNA-Seq was performed on an Illumina HiSeq 3000 platform at the BMFZ (Biologisch-Medizinisches Forschungszentrum) of the Heinrich-Heine University (Düsseldorf, Germany) to gain 150 bp paired-end reads. In total, we obtained 54.97 Gb of RNA-Seq data, with an average of 4.58 Gb per sample. The sequencing quality was examined using FastQC v.0.11.5. Quality scores across all bases were generally good but showed lower quality at the end of reads observed in few samples.

Read mapping, and variant calling

The RNA-Seq reads were mapped on reference genome draft, *M. moricandioides* (unpublished data, assembled by Nils Koppers) using STAR v.2.5.2b (Dobin et al., 2013). 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* were from 61 to 81%. After duplication marking, base quality recalibration, we used a simulated set of SNPs as known variants for preparing analysis-ready RNA-Seq reads. The variant calling was conducted according to GATK best practices. Variant discovery was performed jointly the three *M. arvensis* replicates using the UnifiedGenotyper with *M.*

moriciandioides as the reference, which outputted a raw vcf file containing 1,748,436 SNP callings. The variant calling vcf file and aligned RNA-Seq reads were further inputted into ASEReadCounter from GATK to obtain read counts at each SNP site. Only SNP sites with more than 20 total read counts of parental species and less than 3 counts at the other parental species were processed for further allele specific expression (ASE) analysis. Transcriptome comparison between species was performed with the DESeq2 tool (Love et al., 2014) in R (www.R-project.org) using the Benjamini–Hochberg adjusted false discovery rate ≤ 0.01 as the cut-off for significant differential expression. The Chi-square test was applied to test the over-representation of upregulated transcripts in pathways, such as glycine shuttle, C_4 cycles, Calvin-Benson-cycle, and mitochondrial e^- transport.

Allele specific expression (ASE) analysis

In hybrid, parental alleles were expressed under the same genetic background which made it possible to distinguish between *cis*- or *trans*-regulatory effects by calculating and comparing the allele ratio of the parents (A: PA1/PA2) and that of F1 (B: F1A1/F1A2). The binomial test was applied to test if F1A1 is equal to F1A2 using adjust P-value by the Benjamini-Hochberg (BH) procedure; on the other hand, the Fischer exact test was used for the significant difference between the ratio of parental alleles (PA1/PA2) and the allele ratio of hybrids (F1A1/F1A2) with adjust P-value by the BH method. Four regulatory effects were defined according to the following conditions: *cis*- only, $B \neq 1$ and $A=B$; *trans*- only, $B=1$ and $A \neq B$; *cis*- plus *trans*-, $B \neq 1$ and $A \neq B$; no *cis*- no *trans*-, $A=B=1$. ASE analysis on SNP-level was conducted on six hybrids individually on a set of 120,200 SNPs, demonstrating polymorphisms on 14,004 transcripts.

In addition to evaluate the ASE on transcript-level, we implemented an R package, meta-analysis-based allele-specific expression detection (MBASED) to discover ASE on transcript-level (Mayba et al., 2014). MBASED applied the principles of meta-analysis on combining the information of every SNP site within a single transcript (a single unit of expression) in the absence of the prior information of phased data, the genetic information of hybrids. The ASE was evaluated based on the transcripts with at least one heterozygous locus. The pseudo-phasing based “major” haplotype of genes took the allele with higher counts as the major allele, resulting in the higher estimates of the major allele frequency (MAF, ranging from 0.5 to 1.0). To solve the problem, at least 10^6 simulations were carried out to make MAF estimates close to a null distribution.

ASE verification by qPCR

The DNaseI treated RNA was preceded to cDNA synthesis. First, a total of 1000 ng RNA was mixed with 1 μ l oligo-dT primer, 10 mM dNTP-Mix, 4 μ l 5X Firststrand-Buffer, and 2 μ l 0.1 M DTT and incubated at 42°C for 2 min. The mix was added with 1 μ l Invitrogen SuperScript™ II Reverse Transcriptase and incubated at 42°C for 50 min for cDNA synthesis. The heat inactivation of reverse transcripts was conducted with incubation for 15 min at 70°C. SNPs on *Moricandia* orthologs of GLDP1, ASP3, γ CA2, PPA2 were chosen for designing SNP-specific qPCR primers (Supplemental Table 12). The *Moricandia* ortholog (MSTRG.23175) of Arabidopsis housekeeping gene Helicase (AT1G58050) was tested and selected as reference house keeping gene. The qPCR amplification was carried out in a total reaction of 20 μ L containing 0.5 μ L forward primer (10 ng/ μ L), 0.5 μ L reverse primer (10 ng/ μ L), 5 μ L 5 ng/ μ L cDNA template, 4 μ L ddH₂O, and 10 μ L SYBR® Green qPCR SuperMix (Thermo Fisher Scientific, Schwerte, Germany). The qPCR reaction was conducted following the standard program. The delta CT value was calculated by normalized sample's CT value with that of the housekeeping gene.

Transcriptome annotation

After comparing the *M. moricandioides* predicted protein from TransDecoder (Haas et al., 2013) to UniProtKB (both Swiss-Prot and TrEMBL, ed on April 3, 2019) (Camacho et al., 2009) using BLASTP (UniProt, 2019) with e-value < 1e-5, we summarized the functional annotation in the form of "Human Readable Description" by the AHRD pipeline (<https://github.com/groupschoof/AHRD>). Then, to determine the phylogenetic relationships among *M. arvensis* and *M. moricandioides*, the predicted protein sequences of them together with *A. thaliana* were applied to OrthoFinder v.2.3.3 (Emms and Kelly, 2019).

Biased transcript with cis-specificity

The transcripts with *cis*-specificity (*cis*-SNPs or *cis plus trans*-SNPs) were classified to Ma-biased and Mm-biased transcript based on the gene expression direction. The biased transcripts were annotated with corresponding GO terms derived from *A. thaliana*. Afterward, the gene ontology comparison between Ma-biased and Mm-biased transcripts were conducted on WEGO 2.0 website (Ye J et al., 2018) and visualized using R.

Gene ontology term and pathway enrichment analysis

To recognize the function of corresponding transcript with common *cis*-SNPs and *trans*-SNPs, a custom mapping file was created, containing *Moricandia* transcripts name and the corresponding GO terms derived from *A. thaliana* genes. The 1,416 and 82 *Moricandia* transcripts with common *cis*-SNPs and common *trans*-SNPs, respectively, were processed with the custom mapping file by topGO R-package for gene set enrichment analysis of biological processes (Alexa et al., 2006).

Pathway enrichment analysis was conducted on the mRNA sequences of 1,416 and 82 *Moricandia* transcripts with common *cis*-SNPs and *trans*-SNPs, respectively, with the KEGG Orthology Based Annotation System (KOBAS) (Xie et al., 2011). A corrected P-value of 0.05 served as the threshold to define the significantly enriched pathways.

Promoter-GUS assay and plant transformation

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

The promoter-GUS constructs were transformed into *Agrobacterium tumefaciens* strain GV310::pMP90 (Koncz and Schell, 1986) by electroporation. All constructs were verified again by colony-PCR and DNA sequencing. The *Agrobacterium* introduced with the promoter-GUS constructs were transformed in 4 to 6-week old *A. thaliana* (col-0) by floral-dip method (Clough and Bent, 1998). The transformed T₁ seeds were collected in 4 to 6 weeks after transformation, and selected on Hygromycin B contained half MS plates for two weeks. The survival T₁ lines were further transferred to pots with soil and verified the insertion of T-DNA by PCR.

Primers used in promoter region amplification and colony-PCR were shown in Supplemental Table 14.

Measurement of GUS activity

Two to four week-old T₁ leaves were stained with GUS staining solution (100 mM Na₂HPO₄, 100 mM NaH₂PO₄, 1 mM Potassium-Ferricyanide K₄[Fe(N₆)], 1 mM Potassium-Ferrocyanide K₃[Fe(N₆)], 0.2% Triton X-100, 2mM X-Gluc) and incubated at 37°C in the dark for 2 to 72 hours. The GUS stained leaves were further fixed by the fixation solution (50% Ethanol, 5% Glacial acetic acid, 3.7% Formaldehyde) at 65°C for 10 min. Then, leaves were incubated in 80% Ethanol at room temperature in order to remove the chlorophyll.

Results

***Moricandia* interspecific hybrids demonstrated phenotypes between that of parents**

For the analysis presented here, interspecific hybridization in *Moricandia* was performed using *M. arvensis* (C₃-C₄) as maternal and *M. moricandioides* (C₃) as paternal species (Ma×Mm). The reciprocal cross, Mm×Ma, had produced five times less seeds than Ma×Mm hybridization. Additionally, the germination rate of seeds from Ma×Mm and Mm×Ma was 86% and 25%, respectively. In *M. arvensis* leaves, organelles are found in the BS and M cells along the inner tangential walls and are abundantly accumulated toward veins in the BS cells (Beebe and Evert, 1990; Schlüter et al., 2017). The same leaf anatomy was observed in this study: chloroplasts are not only arranged on the inner wall of M cells, but also abundantly accumulated toward veins in BS cells in *M. arvensis*; in contrast, few chloroplasts were found evenly distributed along the inner wall in BS cells, and some on the inner wall of M cells in *M. moricandioides* (Figure 1). The CO₂ compensation point of *M. arvensis* and *M. moricandioides* was measured at about 24 and 56 ppm, respectively, consistent with previous studies (Bauwe and Apel, 1979; Apel, 1980; Schlüter et al., 2017). The *Moricandia* interspecific hybrids displayed variation in their CO₂ compensation points, ranging from 39 to 55 ppm, generally between the parental lines, but closer to that of C₃ species (Figure 2). These hybrids further varied for the amount and arrangement of chloroplasts in BS cells (Table 1, Supplemental Figure 1). Increased vein density compared to C₃ species was found in *Heliotropium* and *Flaveria* C₃-C₄ species (Muhaidat et al., 2011; Sage et al., 2013), but not in *Moricandia* (Schlüter et al., 2017). The leaf venation was observed from the top view of cleared leaves under the light microscope and the vein density was calculated as the vein length per area. The vein density of *M. arvensis* was not significantly higher than that of *M. moricandioides* but broader veins were observed in *M. arvensis* because of more chloroplasts accumulating toward vascular bundles in BS cells (Supplemental Figure 1 and 2). The vein density of hybrids showed no differences compared to parental species and the leaf venation of hybrids was more similar to that of the C₃ parent with thinner veins corresponding to their leaf anatomy (Supplemental Figure 2). The *Moricandia* interspecific hybrids (F₁) were not uniform in CO₂ compensation points and organelle arrangements in BS cells, probably because of the heterozygosity of their parental species. They produced only very few F₂ seeds, likely because of abnormal pollen produced by the F₁s, resulting in sterility of hybrids (Supplemental Figure 3). Many interspecific or intergeneric hybrids were reported to be sterile as a result of abnormal chromosome pairing or irregular meiotic division of pollen mother cells (Apel et al., 1984; Brown and Bouton, 1993; Covshoff et al., 2014). Taken

together, the *Moricandia* interspecific hybrids demonstrated intermediate characteristics of CO₂ compensation points and leaf anatomy between that of C₃ and C₃-C₄ parents; however, they were more resembled to the C₃ species. The venation pattern of hybrids was similar to C₃ plants.

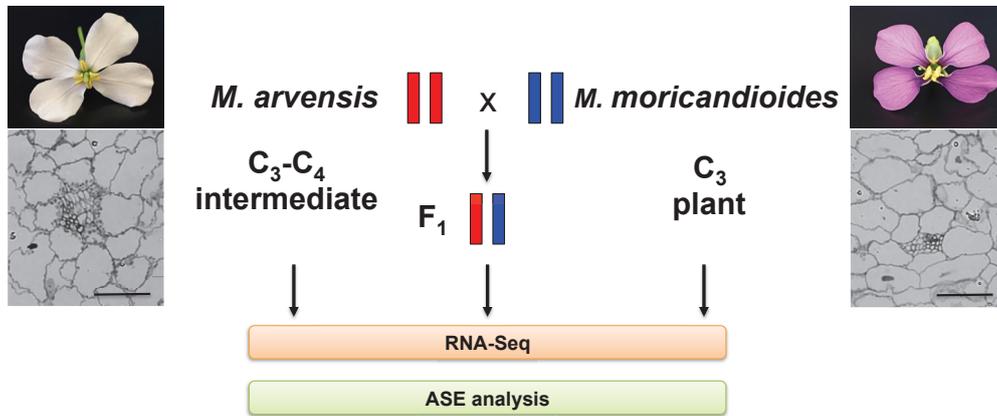


Figure 1. Experimental design.

The interspecific hybrids were obtained from the hybridization of *M. arvensis* as maternal and *M. moricandioides* as paternal species. RNA-Seq was processed on parents and selected hybrid lines, and further introduced to ASE analysis. Bar, 100 μm.

Table 1. Phenotypic characterization of *Moricandia* parental species and their interspecific hybrids applied to RNA-Seq analysis.

M. arvensis and *M. moricandioides* demonstrated typical C₃-C₄ and C₃ phenotypes, respectively. Interspecific hybrid lines indicated not uniform characteristics, generally intermediate between that of parents, but more resemble to C₃ parent.

Sample	Sample name	CO ₂ compensation points	Chloroplasts in BS cells
<i>M. arvensis</i>	Ma	24.99	Abundant, centripetally accumulated
<i>M. moricandioides</i>	Mm	54.91	Few, Centripetally and centrifugally accumulated
I Ma×Mm 9-1	Hybrid1	38.92	Centripetally and centrifugally accumulated
II Ma×Mm 12-1	Hybrid2	42.53	Centripetally and centrifugally accumulated
II Ma×Mm 9-1	Hybrid3	43.38	Nearly none
II Ma×Mm 12-3	Hybrid4	43.89	Nearly none
I Ma×Mm 9-6	Hybrid5	45.87	Centripetally and centrifugally accumulated
I Ma×Mm 11-8	Hybrid6	45.93	Centripetally and centrifugally accumulated

No strong transcriptional changes between *M. arvensis* and *M. moricandioides* on total leaf transcripts

The transcriptome of *M. arvensis* (C₃-C₄) and *M. moricandioides* (C₃) was assembled using STAR v.2.5.2b with the draft genome of *M. moricandioides* serving as the reference. Principle component analysis (PCA) showed that the first principle component (PC1) explained 72% of the variance and clearly separated samples by species. PC2 underlined the separation of three replicates of *M. moricandioides* (Supplemental Figure 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 Benjamini-Hochberg 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 C₃-C₄ 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 C₃-C₄ *Moricandia* (Supplemental table 1). The downregulated transcripts encompassed significantly overrepresented GO terms, such as telomere maintenance, meiotic/nuclear chromosome segregation, chromosome organization, and regulation of organelle organization (Supplemental table 2).

The metabolic difference between C₃ and C₃-C₄ plants is based on different intercellular arrangement of the photorespiratory process. Therefore, genes involved in pathways, such as glycine shuttle, C₄ cycle, Calvin-Benson cycle, and mitochondrial e⁻ transport, were screened for evidence of differential expression (Supplemental Figure 5). With the exception of transcripts encoding PLGG1, GLDT, GS2, alphaCA1, betaCA5, gammaCA2, PPK regulatory protein, AspAT3, AlaAT, rbcS1B, RPE, and UPC1, most of the selected transcripts did not show significant differences between C₃ and C₃-C₄ species. These results are consistent with earlier work showing that transcriptional changes did not indicate a strong preference to C₃-C₄ plants (Schlüter et al., 2017). However, upregulated transcripts were over-represented in glycine shuttle, C₄ cycles, and mitochondrial e⁻ transport relative to upregulated genes in all pathways in C₃-C₄ intermediate species, indicating a tendency of enhanced C₃-C₄ characteristic gene expression during early evolutionary steps of C₄ photosynthesis (Supplemental Figure 5).

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

In this study, F₁s were utilized for discovering regulatory divergence through ASE analysis. Gene expression at the transcriptional level is governed by the interaction of regulatory effects (*cis*- and *trans*-acting factors). In hybrids, the two alleles inherited from the parental lines are under the same cellular condition with a common set of *trans*-acting factors. Therefore, characteristic SNPs in transcripts could be used for ASE analysis and for comparing the allele ratio between hybrids and parents, which enables us to distinguish between *cis*- and *trans*-regulatory effects (Li et al., 2017). SNPs indicating *cis*-only regulatory effects (*cis*-SNP) are those that the two alleles are expressed unequally in hybrids and the allele ratio is the same between parents and hybrids; SNPs with *trans*-only regulatory effects (*trans*-SNP) are defined by equal allele expression in hybrid, but unequal in parents. To assess ASE between *M. arvensis* and *M. moricandioides*, three replicates of each parental species and six lines of their interspecific hybrids were selected for RNA-Seq (Figure 1). The six hybrids demonstrated relative lower CO₂ compensation points among hybrids and differing leaf anatomy patterns (Figure 2; Table 1). Compared to the parents, hybrids 1, 2, 5, 6 had fewer organelles in BS cells, whereas hybrid 3 and 4 showed very little organelle in BS cells (Supplemental Figure 1). The ASE analysis on six hybrids was conducted individually on a set of 120,200 SNPs, demonstrating polymorphisms on an average of 14,004 transcripts in *Moricandia* (Supplemental Table 3). A qPCR assay using SNP-specific primers for four selected genes

was used to validate the RNA-Seq and ASE results (Supplemental Table 4). The six hybrid lines displayed similar relative proportions of regulatory effects (Figure 3). On average, 30% of SNPs showed regulatory divergence by *cis*-only regulatory effect (*cis*-SNP), and 7% of SNPs indicated *trans*-only effects (*trans*-SNP). Furthermore, 30% of SNPs indicated mixed effects (*cis*- plus *trans*-SNP) and 33% of SNPs showed neither *cis*- nor *trans*-regulatory effects (no *cis*- no *trans*-SNP). *GLDPI* of *Moricandia*, known for BS specific expression regulated by the M-box in the promoter region of *M. arvensis*, was tagged by *cis*-SNPs in all hybrids. Although relative proportions of regulatory effects were similar among hybrids, most cases of ASE-SNPs were specific to hybrid lines. Around 8.7% of *cis*-SNPs were common among hybrids (3,142 common *cis*-SNPs), and only 1% of *trans*-SNPs (82 common *trans*-SNPs) were shared in all hybrids. Overall, *cis*-acting effects dominated the regulation of gene expression in *Moricandia* interspecific hybrids.

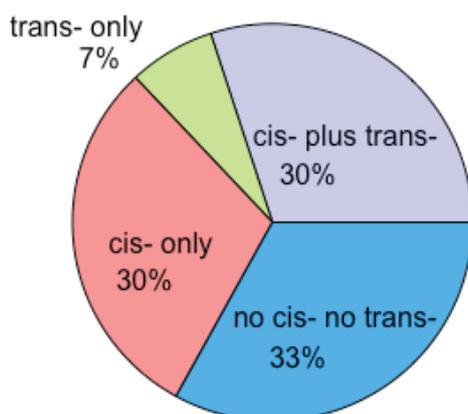


Figure 3 Average relative proportions of regulatory effects among six *Moricandia* interspecific hybrids.

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

GO and pathway enrichment analysis were used to annotate the functions of transcripts with common *cis*-SNPs and *trans*-SNPs. A custom-mapping file was created, containing *Moricandia* transcript names and the corresponding GO terms derived from *A. thaliana* genes. The 1,416 and 82 *Moricandia* transcripts with common *cis*-SNPs and common *trans*-SNPs, respectively, 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 (Supplemental Table 5). GO terms related to nucleosome assembly, RNA methylation, organophosphate biosynthetic process, and peptide metabolic process were abundant in transcripts with common *trans*-SNPs (Supplemental Table 6).

To further decipher biosynthetic pathways where transcripts with common *cis*-SNPs and common *trans*-SNPs participate, pathway enrichment analysis was conducted on the mRNA sequences of *Moricandia* transcripts with common *cis*-SNPs and common *trans*-SNPs, using the KEGG Orthology Based Annotation System (KOBAS) (Xie et al., 2011). Transcripts with common *cis*-SNPs were significantly enriched in 27 pathways, including carbon metabolism, protein processing in endoplasmic reticulum, carbon fixation in photosynthetic organisms, porphyrin and chlorophyll metabolism, glyoxylate and dicarboxylate metabolism, and nitrogen metabolism (Supplemental Table 7). In contrast, transcripts with common *trans*-SNPs were significantly enriched in 9 pathways, which were related to ribosome, carbon metabolism, biosynthesis of amino acid and secondary metabolites, and fatty acid metabolism/degradation/biosynthesis (Supplemental Table 8).

Therefore, our results suggested that *Moricandia* transcripts with *cis* mechanisms play a more prominent role in major photosynthetic pathways and chloroplast relocation, whereas transcripts with *trans* mechanisms are involved in more general biological pathways, such as ribosome, biosynthesis of amino acid and secondary metabolites, nucleosome assembly, and RNA methylation.

Transcripts with *cis*-specificity expressed biased toward C₃-C₄ species in *Moricandia*

cis-regulatory divergences dominate adaptive evolution because they tend to cause fewer deleterious pleiotropic effects than nonsynonymous mutations in protein-coding sequences (Wray, 2007; Stern and Orgogozo, 2008; Wittkopp and Kalay, 2012). Further, they frequently cause altered spatiotemporal gene expression patterns (Prud'homme et al., 2007). The compartmentation of CO₂ assimilatory enzymes between BS and M cells in C₄ plants results from modifications in regulatory sequences, accommodating the transcriptional changes in localization and abundance (Sheen, 1999; Gowik et al., 2004). Thus, genes with *cis*-specificity (*cis*-SNPs or *cis*- plus *trans*-SNPs) were candidates for selections of direct targets or promotions of spatial gene expression during C₄ evolutionary trajectories. In hybrid 1, there were 4,684 *cis*-specificity SNPs (1,105 transcripts) expressed toward *M. arvensis* (Ma-biased) and 3,871 SNPs (820 transcripts) expressed toward *M. moricandioides* (Mm-biased). Similar proportions were observed in the other hybrid lines (Supplemental Table 9). There were 513 Ma-biased and 326 Mm-biased transcripts with *cis*-specificity found in all hybrids.

To understand the gene function of biased transcripts with *cis*-specificity, the common Ma-biased and common Mm-biased transcripts were further investigated by GO term classification using Web Gene Ontology Annotation (WEGO) software (Ye J et al., 2018). These transcripts were classified into GO terms under cellular component, molecular function, and biological process categories and highlighted the differential abundance of categories between common Ma-biased and common Mm-biased transcripts (Figure 4). Common Ma-biased transcripts were detected with significantly more genes in GO terms, for instance, biosynthetic process, localization, anatomical structure morphogenesis, endomembrane system, catalytic activity, transmembrane transporter activity, and small molecular. In addition, under anatomical structure morphogenesis, more Ma-biased transcripts were found in the GO term leaf morphogenesis. On the contrary, common Mm-biased transcripts were detected more abundant in GO terms, such as photosynthetic membrane and positive regulation. Therefore, this directional bias indicated that *cis*-specificity caused upregulation of C₃-C₄ genes in categories such as anatomical structure morphogenesis, and transmembrane transporter activity that might be involved in early C₄ photosynthesis evolution.

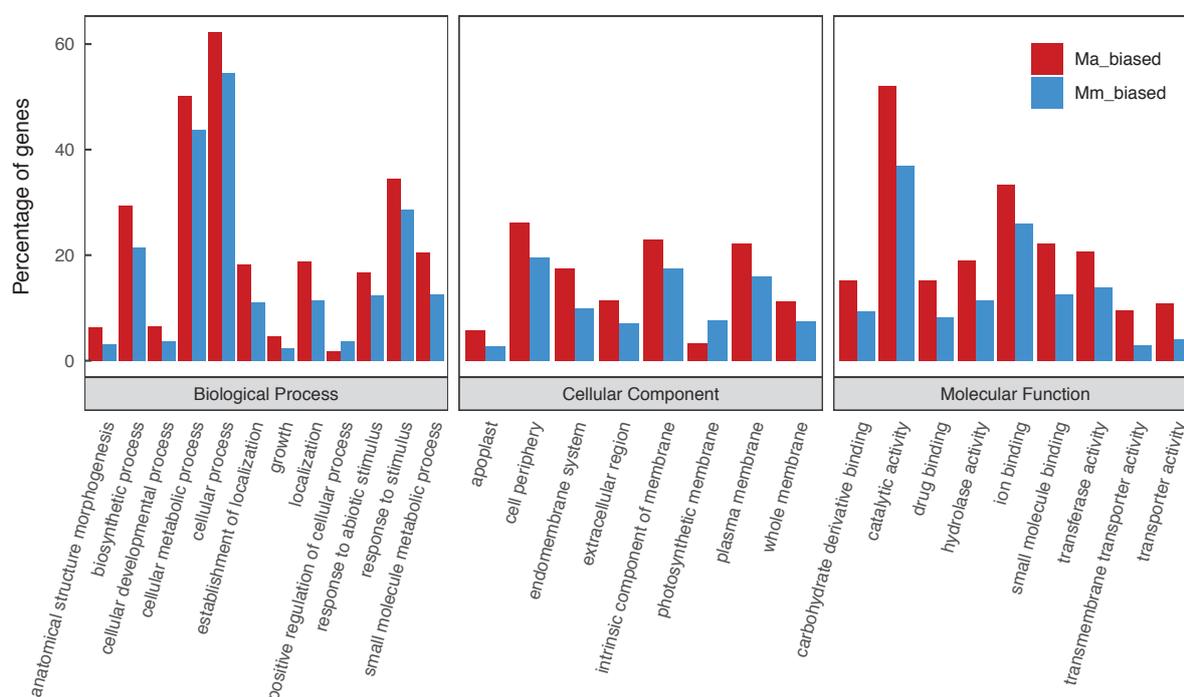


Figure 4. GO terms comparison between transcripts with *cis*-specificity biased toward *M. arvensis* (Ma-biased) and *M. moricandioides* (Mm-biased).

All GO terms shown here were significantly different in gene percentages between Ma-biased and Mm-biased transcripts.

Evaluating ASE on transcript-level in *Moricandia* hybrids

Assessing ASE on SNP-level could define SNPs into four categories of regulatory effects; however, SNPs within a transcript might indicate different regulatory effects. For instance, *GLDPI* in hybrid 1 revealed 5 *cis*-SNPs, 16 *cis*-plus *trans*-SNPs, and 2 *trans*-SNPs. Therefore, studies evaluating ASE on SNP-level require either an agreement across SNPs in the same transcript (Shao et al., 2019) or the phased information of read counts at SNPs from genomic data of hybrids (Skelly et al., 2011; He et al., 2012; Steige et al., 2015; Rhoné et al., 2017). However, in our study, the genomic information of hybrids is not available. In order to discover the intensity of allelic imbalance in hybrids, an ASE detection gene-level statistic method, meta-analysis based allele-specific expression detection (MBASED), was implemented to combine information on SNPs within a transcript in the absence of phasing data. It assigns the allele with more read counts at each SNP to the major allele as a pseudo-phasing approach under the assumption of the consistent direction of ASE within a transcript, and the 10^6 time simulation was employed to adjust the significance levels of ASE. The ASE level of transcripts was estimated with the major allele frequency (MAF), ranging from 0.5 to 1.0 (Mayba et al., 2014). Most of transcripts showed mild allelic imbalance with MAF 0.6. Transcripts with $MAF \geq 0.7$ and adjusted P-value ≤ 0.05 were defined as ASE-transcripts. On average in the six hybrid lines, out of 14,000 transcripts, around 34% of transcripts demonstrated ASE evidence; 7% of them had extreme allelic imbalance ($MAF \geq 0.9$ and adjusted P-value ≤ 0.05) (Figure 5). *GDLPI* in *Moricandia*, known for ASE regulation, showed an average MAF with 0.76. Interestingly, some transcripts showed extreme allele bias to one of the parental species in hybrids (Supplemental Table 10). For instance, *MSTRG.16015* encoding chloroplast RNA binding showed an average MAF of 0.93 toward *M. arvensis* allele, and *MSTRG.5109* encoding ATP synthase subunit d revealed an average MAF with 0.96 biased to *M. arvensis* allele. Overall, 34% of assayed transcripts were defined as ASE-transcript and a group of transcripts had strong allelic imbalance in hybrids.

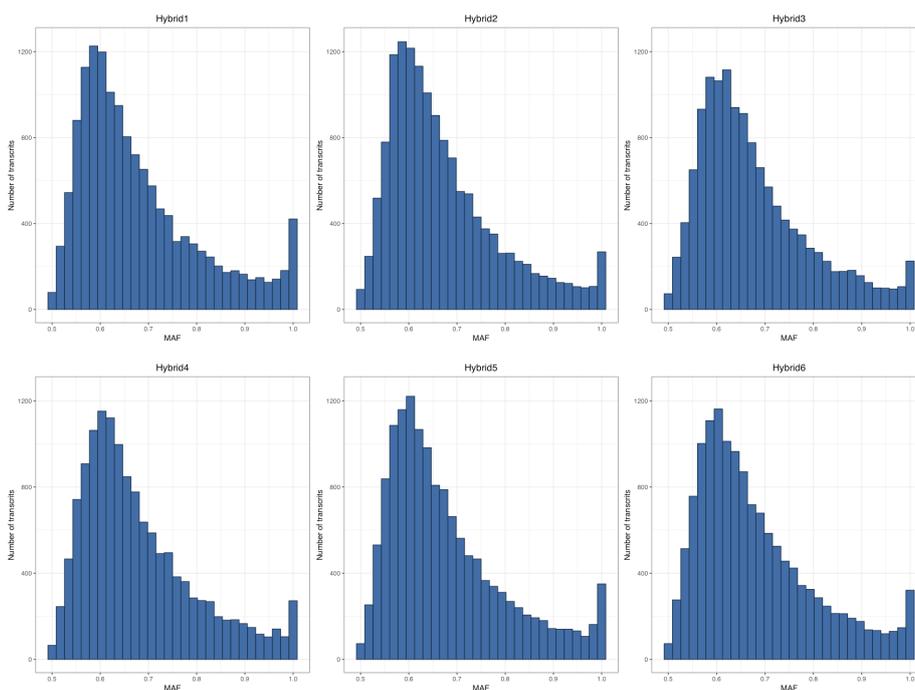


Figure 5. Distribution of allelic imbalance for all assayed transcripts among six *Moricandia* hybrids.

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. Transcripts with $MAF \geq 0.7$ and adjusted p-value ≤ 0.05 were defined as ASE-transcripts. Around 34% of transcripts were with $MAF \geq 0.7$ and 7% of them showed $MAF \geq 0.9$.

Enrichment of regulatory divergences in selected pathways

Most of *Moricandia* transcripts indicated no significant differential expression in comparative transcriptome studies using total leaf extracts. However, *cis*-acting factors, regulating spatial gene expressions, play a crucial role in adaptive phenotype evolution (Wray, 2007; Lemmon et al., 2014). Therefore, we examined the enrichment of regulatory effects on transcripts in selected pathways through detecting *cis*-SNPs and evaluating allelic imbalance in transcripts. The expression of many genes in glycine shuttle, C₄ cycle, and Calvin-Benson cycle were regulated by at least one *cis*-SNP or common *cis*-SNP among hybrids (Supplemental Table 11). In addition, the ASE evaluated on transcript-level demonstrated the intensity of allelic imbalance in hybrids. Transcripts with major alleles frequency ≥ 0.7 were considered to possess ASE. Genes involved in the C₃-C₄ glycine shuttle were regulated by allele specific expression, like *PLGG*, *GOX*, *GGAT*, *SHMT*, *GDC* complex, *DIT*, and *GS2* (Figure 6A).

Evaluating ASE of C_4 cycle genes revealed that gene expressions of *CA*, *PEPC*, *PPT*, *NADP-MDH*, *DIT*, *NADP-ME*, *BASS2*, *PPDK*, *PPT*, *AspAT*, and *PEPCK* were dominated by regulatory elements (Figure 6B). Calvin-Benson cycle genes were also found to be regulated by ASE, except *SBPase* (Figure 6C). Taken together, most of genes involving in selected pathways revealed no strong differential gene expression between C_3 and C_3 - C_4 species in *Moricandia*, but regulatory divergences play an important role in different photosynthetic types by spatial or temporal expression of critical genes involving in early C_4 photosynthesis evolution.

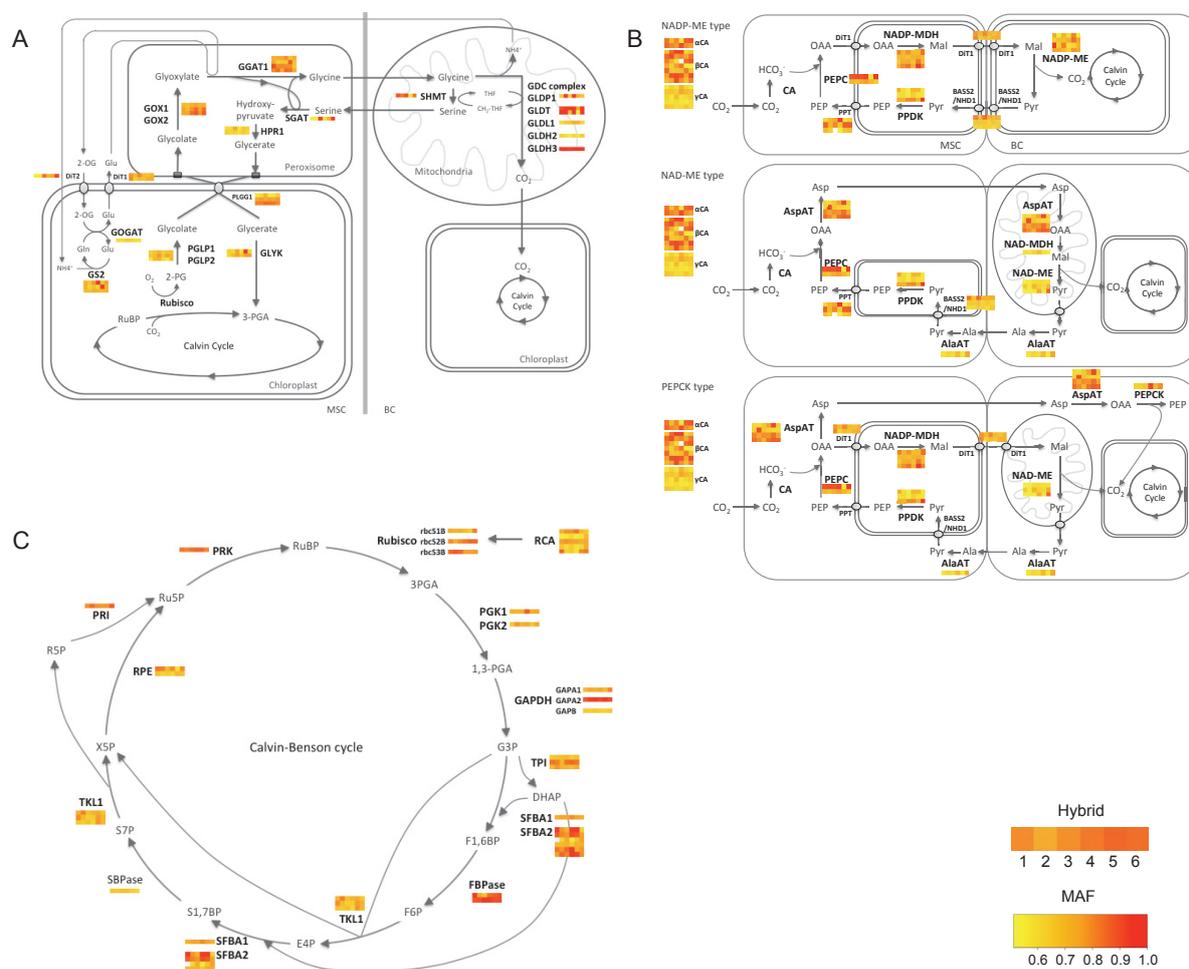


Figure 6. Overview on allelic imbalance of genes involving in selected pathways among six *Moricandia* hybrids.

Selected pathways were the C_3 - C_4 glycine shuttle (A), the C_4 photosynthesis (B) and the Calvin-Benson cycle (C). The six blocks in each gene bar presented the major allele frequency (MAF) of Hybrid1 to Hybrid6 from left to right.

Rubisco, ribulose 1,5-bisphosphate carboxylase/oxygenase; PGLP, 2-PG phosphatase; PLGG1, plastidial glycolate/glycerate transporter; GOX, glycolate oxidase; GGAT, glutamate:glyoxylate aminotransferase; SGAT, serine:glyoxylate aminotransferase; HPR1, peroxisomal hydroxypyruvate reductase; GLYK, glycerate 3-kinase; SHMT, serine hydroxymethyltransferase; GDC, glycine decarboxylase complex; GS, plastidial glutamine synthetase; GOGAT, ferredoxin-dependent glutamate synthase; CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; DiT1, dicarboxylate transporter 1 (oxaloacetate/malate transporter); PPT, phosphoenolpyruvate/phosphate translocator; MDH, Malate dehydrogenase; PPK, pyruvate, orthophosphate dikinase; BASS2/NHD1, plasma membrane pyruvate transport; NADP-ME, NADP-malic enzyme; NAD-ME, NAD-malic enzyme; AspAT, aspartate; PEPCK, PEP carboxykinase; RCA, Rubisco activase; PGK, Phosphoglycerate kinase; GAPDH, Glyceraldehyde 3-phosphate dehydrogenase; GAPA, Glyceraldehyde 3-phosphate dehydrogenase A subunit; GAPB, Glyceraldehyde 3-phosphate dehydrogenase B subunit; TPI, Triosephosphate isomerase; SFBA, Sedoheptulose/Fructose-biphosphate aldolase ; FBPase, Fructose biphosphatase; TKL, Transketolase; SBPase, Sedoheptulose-1,7-biphosphatase; RPE, Ribulose-5-phosphate 3-epimerase; PRI, Ribulose-5-phosphate isomerase; PRK, Phosphoribulokinase; RuBP, ribulose-1,5- bisphosphate; 3-PGA, 3-phosphoglycerate; 2-PG, 2-phosphoglycolate; Gln, glutamine; Glu, glutamate; 2-OG, 2-oxoglutarate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; Mal, Malate; Pyr, pyruvate; Asp, Aspartate; Ala, Alanine; 3PGA, 3-phosphoglycerate; 1,3-PGA, 1,3-bisphosphoglycerate; G3P, glyceraldehyde-3-phosphate; DHAP, dihydroxy-acetone-phosphate; F1,6BP, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; E4P, erythrose-4-phosphate; S1,7BP, sedoheptulose-1,7-bisphosphate; S7P, sedoheptulose-7-phosphate; X5P, xylulose-5-phosphate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate.

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 C₃ and C₃-C₄ *Moricandia* species. To confirm the ASE results, promoter-GUS assays were used. Genes were selected from the GO term chloroplast relocation (GO:0009902) obtained from (1) GO analysis on transcripts with common *cis*-SNPs, and (2) ASE-transcripts with extreme allelic imbalance toward *M. arvensis* (MAF \geq 0.9) among all six hybrids. *GLDPI* was characterized to be regulated by spatial expression between C₃ and C₃-C₄ *Moricandia* species, loss of M-box in promoter

region of *M. moricandioides* resulting in confinement of GLDP in BS cells (Adwy, 2018). *GLDP1* possessed *cis*-SNPs across hybrids and the average MAF of 0.76 (Supplemental Table 11; Table 2). *PHOT2* and *CHUP1* selected from the GO term chloroplast relocation (GO:0009902) had common *cis*-SNPs among six hybrids and an average MAF of 0.62 and 0.72, respectively. *CRB* and *ATPQ* were selected from transcripts with extreme allelic imbalance toward *M. arvensis*: *CRB* participating in chloroplast organization showed an average MAF of 0.93; *ATPQ* was annotated as ATP synthesis coupled proton transport and response to salt stress and with an average MAF of 0.96 (Supplemental Table 10; Table 2).

Approximately 2 kb upstream promoter region of genes from *M. arvensis* and *M. moricandioides* were amplified, fused to GUS, and the recombinant constructs were transformed into *A. thaliana* (Supplemental Table 13). GUS staining results demonstrated the spatial gene expression between *Moricandia* species (Figure 7). *GLDP1* was selected as control, which showed cell-specific regulation promoter of *M. arvensis GLDP1* (*MaGLDP1*) expressed in cells surrounding veins and GUS staining of that of *M. moricandioides* (*MmGLDP1*) was observed on the whole leaf. The *MaPHOT2* expression was observed in roots and slightly in shoots of two-week-old seedlings; however, GUS staining of *MmPHOT2* promoter region was found mainly surrounding the leaf mid rib, trichomes in leaves, and shoots. *MaCHUP1* was expressed in the whole leaf and slightly in shoots; *MmCHUP1* was more strongly expressed towards the veins and showed stronger signal in shoots compared to that of *MaCHUP1*. *MaCRB* expression was detected in parts of the leaf, shoots, and roots; in contrast, expression of *MmCRB* was only found in roots. *MaATPQ* displayed GUS staining in leaf with high specificity around veins, roots and shoots; *MmATPQ* showed similar pattern in leaves, but different expression pattern was detected between leaves. Thus, except *ATPQ*, promoter-GUS assay confirmed ASE results from two strategies, ASE analysis on SNP- and transcript-level, by spatial gene expression between species.

Table 2. List of selected genes in *Moricandia*.

Gene	Ortholog in <i>Moricandia</i>	Regulatory effect	Major allele frequency from Hybrid1 to Hybrid6	GO Biological Process
<i>GLDP1</i>	Mm <i>MSTRG.31033</i> (14 exons) Ma <i>MSTRG.3691</i> (14 exons)	common <i>cis</i> -SNP	0.64; 0.96; 0.62; 0.95; 0.72; 0.66	glycine catabolic process
<i>PHOT2</i>	Mm <i>MSTRG.685</i> (22 exons) Ma <i>MSTRG.42108</i> (20 exons)	common <i>cis</i> -SNP	0.63; 0.62; 0.63; 0.63; 0.60; 0.61	response to blue light, phototropism, chloroplast relocation
<i>CHUP1</i>	Mm <i>MSTRG.24525</i> (9 exons) Ma <i>MSTRG.31056</i> (8 exons)	common <i>cis</i> -SNP	0.78; 0.73; 0.68; 0.74; 0.68; 0.69	chloroplast relocation
<i>CRB</i>	Mm <i>MSTRG.16015</i> (10 exons) Ma <i>MSTRG.34712</i> (10 exons)	<i>cis</i> - plus <i>trans</i> -SNP	0.94; 0.92; 0.90; 0.91; 0.91; 0.97	chloroplast organization, circadian rhythm, regulation of gene expression
<i>ATPQ</i>	Mm <i>MSTRG.5109</i> (5 exons) Ma <i>MSTRG.42827</i> (5 exons)	<i>cis</i> - plus <i>trans</i> -SNP	0.95; 0.90; 1.00; 0.95; 0.97; 1.00	ATP synthesis coupled proton transport, response to salt stress

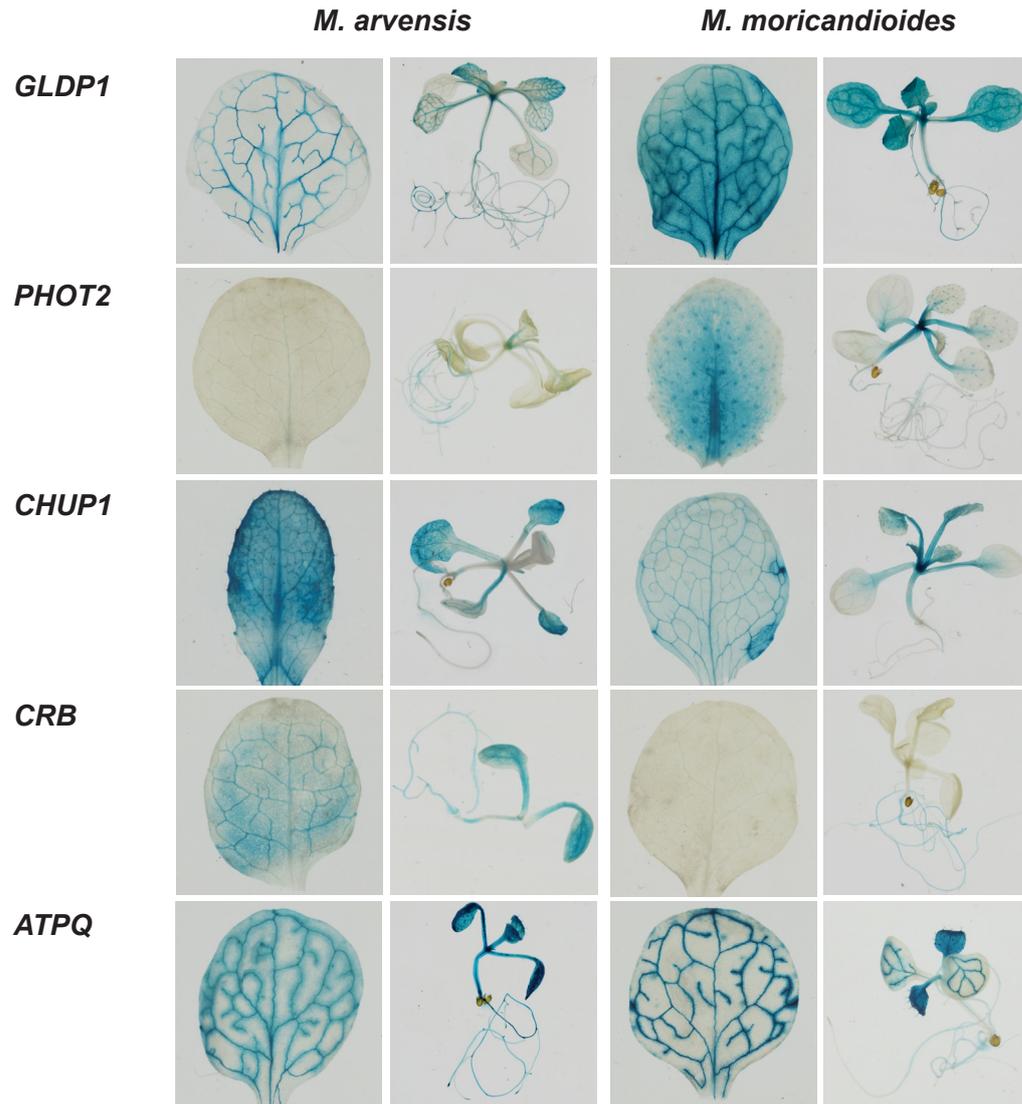


Figure 7. Promoter-GUS assay of *Moricandia* selected genes expressed in *A. thaliana*. *GLDP1* served as positive control, possessing common *cis*-SNP among hybrids and defined as ASE-transcript with major allele frequency (MAF) ≥ 0.7 ; *PHOT2* and *CHUP1* had common *cis*-SNP among hybrids; *CRB* and *ATPQ* demonstrated strong allelic imbalance with MAF ≥ 0.9 .

Discussion

C₃-C₄ × C₃ *Moricandia* hybrids showed intermediate phenotype between that of parental species, but more resemble to the C₃ parent

To dissect regulatory factors governing early evolutionary steps of C₄ photosynthesis, the ASE was assessed on SNP- and transcript-level utilizing RNA-Seq data of *M. arvensis* (C₃-C₄), *M. moricandioides* (C₃), and six of their interspecific hybrids. Ma×Mm hybrids were sterile because of irregular meiotic division of pollen mother cells. This sterile outcome was consistent with trials using species in *Panicum* and *Flaveria* (Bouton et al., 1986; Brown and Bouton, 1993). However, the interspecific hybrids of *Atriplex prostrata* (C₃) and *A. rosea* (C₄) were fertile (Sultmanis, 2018). The six Ma×Mm hybrids demonstrated intermediate characteristics between parents', but not completely uniform in the hybridization event. The CO₂ compensation points and leaf anatomy were thereby more similar to C₃ parent's (Figure2; Table 1; Supplemental Figure 1&2). Additionally, the PCA showed that the gene expressions of *Moricandia* interspecific hybrids were also closer to that of C₃ parents (Supplemental Figure 4B). Similar intermediate characteristics of CO₂ compensation points and organelle quantities in BS cells were observed in interspecific hybrids between C₃-C₄ and C₃ *Panicum* species (Brown et al., 1985). The leaf anatomy of hybrids from *M. nitens* (C₃-C₄) × *B. napus* (C₃) was resemble to the C₃ parent (Rawsthorne et al., 1988). The dominance of C₃-C₄ phenotypes (CO₂ exchange and confined GDC in BS cells) increased with the C₃-C₄ genome constitution in *D. tenuifolia* (C₃-C₄) × *R. sativus* (C₃) hybrids (Ueno et al., 2003), and the same phenomena was found in backcross of *B. alboglabra* (C₃) × *M. arvensis* (C₃-C₄) to the C₃-C₄ parent (Apel et al., 1984). These results, together with our observation, suggest that genes showed additive effects in C₃-C₄ × C₃ hybrids with no obvious maternal effect. Thereby, C₃-C₄ × C₃ hybrids possessed more C₃ genetic dosages because C₃-C₄ species had also the C₃ genetic background.

The impact of *cis* mechanisms on early evolutionary steps of C₄ photosynthesis

cis-regulatory effects presented larger impacts than *trans*-acting divergences on *Moricandia* interspecific hybrids: 30% and 7% of assayed SNPs were discovered as *cis*-SNP and *trans*-SNP, respectively. The abundance of *cis*-regulatory changes is often detected in interspecific hybrids (long evolutionary time-scales), whereas *trans*-regulatory divergence are more common in intraspecific hybrids (short evolutionary time-scales) (Stern and Orgogozo, 2008; McManus et al., 2010; Rhoné et al., 2017). The difference of interspecific expression between closely related *Drosophila* species was mainly caused by *cis*-acting changes (28 out of 29

differential expressed genes) (Wittkopp et al., 2004). The expression divergence between C_3 and C_4 *Atriplex* species was also dominated by *cis*-regulatory changes (Sultmanis, 2018). However, only 15% of transcribed genes in intraspecific selected wild pearl millet demonstrated *cis*-regulatory divergence, and no excess of *cis*-acting mutation was observed among adaptive traits associated genes (Rhoné et al., 2017). Relative proportions of regulatory effects were similar in all six Ma×Mm hybrids, but many of them were sample-specific. Only 8.7% of 36,000 *cis*-SNPs and 1% of 8,500 *trans*-SNPs were assessed consistent among hybrids. It was also found in the literature that ASE genes are often unique in different hybrid lines (Lemmon et al., 2014; Steige et al., 2015). Owing to the heterozygous level of parental species, hybrids didn't show uniform characteristics. The six selected hybrids had relative low CO₂ compensation points among all tested hybrids differed in their leaf anatomy. Transcripts with *cis*-SNPs were further filtered with phenotypic characteristics: 2671 common transcripts in hybrid 1, 2, 5, 6 were corresponding to the leaf anatomy category, and 2376 common transcripts in hybrid 1, 2, 3, 4 were corresponding to the CO₂ compensation point category. However, transcripts with *cis*-SNPs of these two characteristics showed no functional differences (Supplemental Figure 6). This might be because the value of CO₂ compensation points was in association with leaf anatomy in C_3 - C_4 plants, and selected hybrids in this study didn't possess very distinct phenotypes. The correlation of these two characteristics could be further confirmed with more hybrids showed distinct leaf anatomy.

The mechanism of *cis*- and *trans*-regulatory divergence has different impacts on the inheritance and evolution of gene expressions, and also different biological processes (McManus et al., 2010; Meiklejohn et al., 2014). This is related to the genetic nature of transcriptional regulations: *cis*-regulatory sequences, located in promoter regions, UTRs, and introns, modulate the binding of *trans*-acting factors to DNA, therefore affecting the transcription of nearby genes; while *trans*-element, like transcription factors and long noncoding RNA, are able to affect the expression of many genes (Wray, 2007). In our study, GO and pathway enrichment analysis on transcripts with common *cis*-SNPs and common *trans*-SNPs indicated that *cis*-regulatory changes dominated major photosynthetic pathways and chloroplast relocation; whereas, transcripts with *trans* mechanisms are involved in more general biological pathways. For instance, transcripts with common *cis*-SNP were overrepresented in isopentenyl diphosphate biosynthesis. In higher plants, the formation of isopentenyl diphosphate, the central intermediate of all isoprenoids, was compartmentalized in cytosol (sterols) or plastids (carotenoid, phytol) (Lange and Croteau, 1999), which seemed to be regulated with *cis*-acting divergences. The abundant and centripetal accumulation of

chloroplasts in BS cells was a prominent character for higher photosynthetic efficiency in C₃-C₄ than that in C₃ plants, and was reported as an early step in the C₄ photosynthesis evolution model (Sage et al., 2012). Genes for chloroplast movement in C₃ *Atriplex* species compared to C₄ type were upregulated and showed evidences of regulatory effects (Sultmanis, 2018). Regulatory divergences between C₃ and C₃-C₄ plants govern vary biological processes; however, the enrichment of photosynthesis and chloroplast related GO terms in *cis*-transcripts supports the involvement of *cis*-regulation in the installation of the glycine shuttle during early C₄ photosynthesis evolution.

cis-regulatory changes were considered to play a crucial role in evolution of adaptive traits, due to their less deleterious pleiotropic effects than nonsynonymous mutations in protein sequences (Wray, 2007; Stern and Orgogozo, 2008; Wittkopp and Kalay, 2012). Additionally, *cis*-regulatory divergences are additive in heterozygotes, which were visible and preferentially accumulated during evolution (Lemos et al., 2008; Wittkopp et al., 2008; Meiklejohn et al., 2014). Some regulatory divergences have been characterized to enhance expression of core C₄ genes in specific compartments (Hibberd and Covshoff, 2010; Reeves et al., 2017; Williams et al., 2012). Examples of such include: 41 bp *cis*-element in the *PEPC* promoter from *F. trinervia* for M specific expression (Gowik et al., 2004), the region -301 to -296 in *ZmPPDK* promoter crucial for M accumulation (Matsuoka and Numazawa, 1991) and the region including the 5'UTR to part of the second exon from *Zoysia japonica* *PEPCK* for BS accumulation (Nomura et al., 2005). In our study, transcripts with *cis*-specificity showed higher expression of *M. arvensis* alleles than *M. moricandioides* alleles in hybrids (Supplemental Table 9). Similar observations have been reported: during maize domestication, *cis*-evolution often favored alleles with enhanced expression in maize than in teosinte (Lemmon et al., 2014) and during selfing syndrome evolution in *Capsella*, alleles from selfing species *C. rubella* expressed at higher level than the outcrossing *C. rubella* alleles in their F₁s (Steige et al., 2015). Transcripts with *cis*-specificity showing a directional bias toward *M. arvensis* were more abundant in GO terms, involving in anatomical structure morphogenesis, transmembrane transporter activity and localization, compared to transcripts expressed biased to *M. moricandioides*. Therefore, the upregulation of specific genes through *cis*-regulation is important for establishing leaf anatomy and transporting substances (such as carbohydrates, amino acid, pyruvate) on evolutionary path from C₃ to C₃-C₄ plants.

All in all, *cis*-regulatory effects, favoured in adaptive phenotypic traits during evolution, were additive in C₃-C₄ × C₃ *Moricandia* hybrids. It is consistent with the predicted long and smooth path of C₄ photosynthesis (Williams et al., 2013). *cis* mechanisms had larger impact on

regulation of major photosynthetic pathways and chloroplast relocation. In addition, *cis*-specificity caused a directional bias toward C₃-C₄ species in *Moricandia* hybrids, upregulating genes participating in biological functions, such as anatomical structure morphogenesis, transmembrane transporter activity and localization. Comparative transcriptome analysis of *Flaveria* showed that C₄ cycle genes are upregulated in C₄ species, and photorespiration is downregulated in C₄ species, but upregulated in C₃-C₄ species (Gowik et al., 2011). We hypothesize that the *cis* mechanism upregulated phenomenon also happened on the evolutionary path from C₃-C₄ to C₄ plants; however, downregulation of photorespiratory genes might occur in later optimized process of C₄ evolution.

ASE analysis is a powerful strategy for understanding C₄ photosynthesis evolution

Two strategies used in this study, ASE analysis on SNP-level and on transcript-level, gave us accesses to understand regulatory divergences in different photosynthetic species in *Moricandia*. Genes selected for promoter-GUS assay were from: (1) the GO term, chloroplast relocation, discovered from transcripts with common *cis*-SNPs, and (2) transcripts with extreme allelic imbalance ($MAF \geq 0.9$), obtained from the ASE evaluation on the transcript scale. Except for *ATPQ*, the promoter-GUS assay with 2 kb upstream of *Moricandia* genes confirmed the ASE results, suggesting that there are regulatory divergences between species at least in the promoter region, causing spatial gene expression (Figure 7). However, in addition to promoter regions, elements in untranslated gene regions (5'UTR and 3'UTR) and introns have been reported to regulate genes at the transcriptional level (Barrett et al., 2012; Hernandez-Garcia and Finer, 2014). It was reported that a small part of C-terminus *PHOT2* was necessary for chloroplast avoidance response (Kong et al., 2013). Therefore, a more detailed research on regulatory elements governing C₃ and C₃-C₄ photosynthetic differences should be conducted.

We didn't observe large changes in the transcript abundance of mature leaves between C₃ and C₃-C₄ *Moricandia* (Supplemental Figure 5; Schlüter et al., 2017). Comparative transcriptome analysis between C₃ and C₃-C₄ plants using total leaf extracts couldn't detect either spatial gene expression or important transcriptional changes such as relocation of P-protein of GDC (Schlüter et al., 2017). However, based on our ASE results, in addition to 67% assayed SNPs possessing *cis*-SNPs, *trans*-SNPs, or *cis*- plus *trans*-SNPs, around 34% of assayed transcripts showed allelic imbalance, suggesting the critical role of ASE between different photosynthetic types. For instance, genes in the glycine shuttle (*GOX*, *GGAT*, *GDC* complex, *SHMT*, *GS2*, *DiT*, *AspAT*) contained at least one *cis*-SNP among hybrids and demonstrated

high ASE value (Supplemental Table 11; Fig. 6). *GDC* is exclusively expressed in BS cells because of the absence of M-box in upstream of *GLDPI*, resulting in the CO₂ pump in C₃-C₄ *Moricandia* species (Rawsthorne, 1992; Adwy, 2018). C₄ genes (*CA*, *PEPC*, *PPT*, *NADP-MDH*, *DIT*, *NADP-ME*, *BASS2*, *PPDK*, *PPT*, *AspAT*, and *PEPCK*) showed evidence of *cis*-SNPs and ASE regulation in *Moricandia* hybrids as well (Supplemental Table 11; Fig. 6). This pattern was also reported in the ASE analysis on interspecific C₃ × C₄ *Atriplex* hybrids: changes in *cis*-regulatory divergence accumulated in photorespiration-related pathways on the evolutionary path of C₄ photosynthesis (Sultmanis, 2018). These results are consistent with previous researches regarding regulatory effects of cell-specific gene expression, restricting gene expression to either BC or M cells, during C₄ photosynthesis evolution (Hibberd and Covshoff, 2010; Reeves et al., 2017). Therefore, ASE analysis on SNP- and transcript-level provides a powerful way to uncover cell-specific transcriptional regulations during C₄ photosynthesis evolution, which is missing in comparative transcriptome analysis. Studies on comparative transcriptome during leaf development of closely related C₃ and C₄ species have revealed the importance of genetic control in early leaf developmental stages on the evolution of C₄ photosynthesis (Wang et al., 2013; K ulahoglu et al., 2014). In the future work, ASE analysis could be applied on discovering regulatory changes during leaf development of C₃ and C₃-C₄ species in *Moricandia* to capture early key regulations of C₃-C₄ characteristics.

Conclusions

In our study, we showed that *cis*-regulatory divergences had a large impact on *Moricandia* interspecific hybrids and the corresponding transcripts enriched in major photosynthetic pathways and chloroplast relocation. We further observed that *cis*-specificity caused upregulation of C₃-C₄ genes in categories such as anatomical structure morphogenesis and transmembrane transporter activity, playing an important role in the early evolutionary steps of C₄ photosynthesis, especially for the installation of the glycine shuttle. With the genetic information of parental species, the RNA-Seq dataset and ASE approaches, we investigated *cis*- and *trans*-acting divergences on a genome scale, which gave us broad aspects on understanding the role of transcriptional regulations on shaping C₄ photosynthesis during evolution.

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.** supervised the experimental design and participated in drafting the manuscript.

References

- Adwy W** (2018) Identification and characterization of a frequent genetic alteration toward the evolution of C₂-photosynthesis in the genus *Moricandia*. doi: <http://dx.doi.org/10.15488/3095>
- Adwy W, Laxa M, Peterhansel C** (2015) A simple mechanism for the establishment of C₂-specific gene expression in Brassicaceae. *Plant J* **84**: 1231–1238
- Apel P** (1980) CO₂ compensation concentration and its O₂ dependence in *Moricandia spinosa* and *Moricandia moricandioides* (Cruciferae). *Biochemie und Physiologie der Pflanzen* **175**: 386–388
- Apel P, Bauwe H, Ohle H** (1984) Hybrids between *Brassica alboglabra* and *Moricandia arvensis* and their Photosynthetic Properties. *Biochemie und Physiologie der Pflanzen* **179**: 793–797
- Barrett LW, Fletcher S, Wilton SD** (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* **69**: 3613–3634
- Bauwe H** (2011) Chapter 6 Photorespiration: The Bridge to C₄ Photosynthesis. *In* AS Raghavendra, RF Sage, eds, *C₄ Photosynthesis and Related CO₂ Concentrating Mechanisms*. Springer Netherlands, Dordrecht, pp 81–108
- 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**: 251–254
- Beebe DU, Evert RF** (1990) The Morphology and Anatomy of the Leaf of *Moricandia arvensis* (L.) DC. (Brassicaceae). *Botanical Gazette* **151**: 184–203
- Bouton JH, Brown RH, Evans PT, Jernstedt JA** (1986) Photosynthesis, Leaf Anatomy, and Morphology of Progeny from Hybrids between C₃ and C₃/C₄ Panicum Species 1. *Plant Physiol* **80**: 487–492
- Bräutigam A, Gowik U** (2016) Photorespiration connects C₃ and C₄ photosynthesis. *J Exp Bot* **67**: 2953–2962

- Brown HR, Bouton JH** (1993) Physiology and Genetics of Interspecific Hybrids Between Photosynthetic Types. *Annual Review of Plant Physiology and Plant Molecular Biology* **44**: 435–456
- Brown RH, Bouton JH, Evans PT, Malter HE, Rigsby LL** (1985) Photosynthesis, Morphology, Leaf Anatomy, and Cytogenetics of Hybrids between C3 and C3/C4 Panicum Species. *Plant Physiol* **77**: 653–658
- Covshoff S, Burgess SJ, Kneřová J, Kümpers BMC** (2014) Getting the most out of natural variation in C4 photosynthesis. *Photosynth Res* **119**: 157–167
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21
- Emms DM, Kelly S** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *bioRxiv* 466201
- Gowik U, Bräutigam A, Weber KL, Weber APM, 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**: 2087–2105
- 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. *Plant Cell* **16**: 1077–1090
- Hatch MD** (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **895**: 81–106
- 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. *Mol Biol Evol* **29**: 3385–3395
- Heckmann D** (2016) C4 photosynthesis evolution: the conditional Mt. Fuji. *Current Opinion in Plant Biology* **31**: 149–154

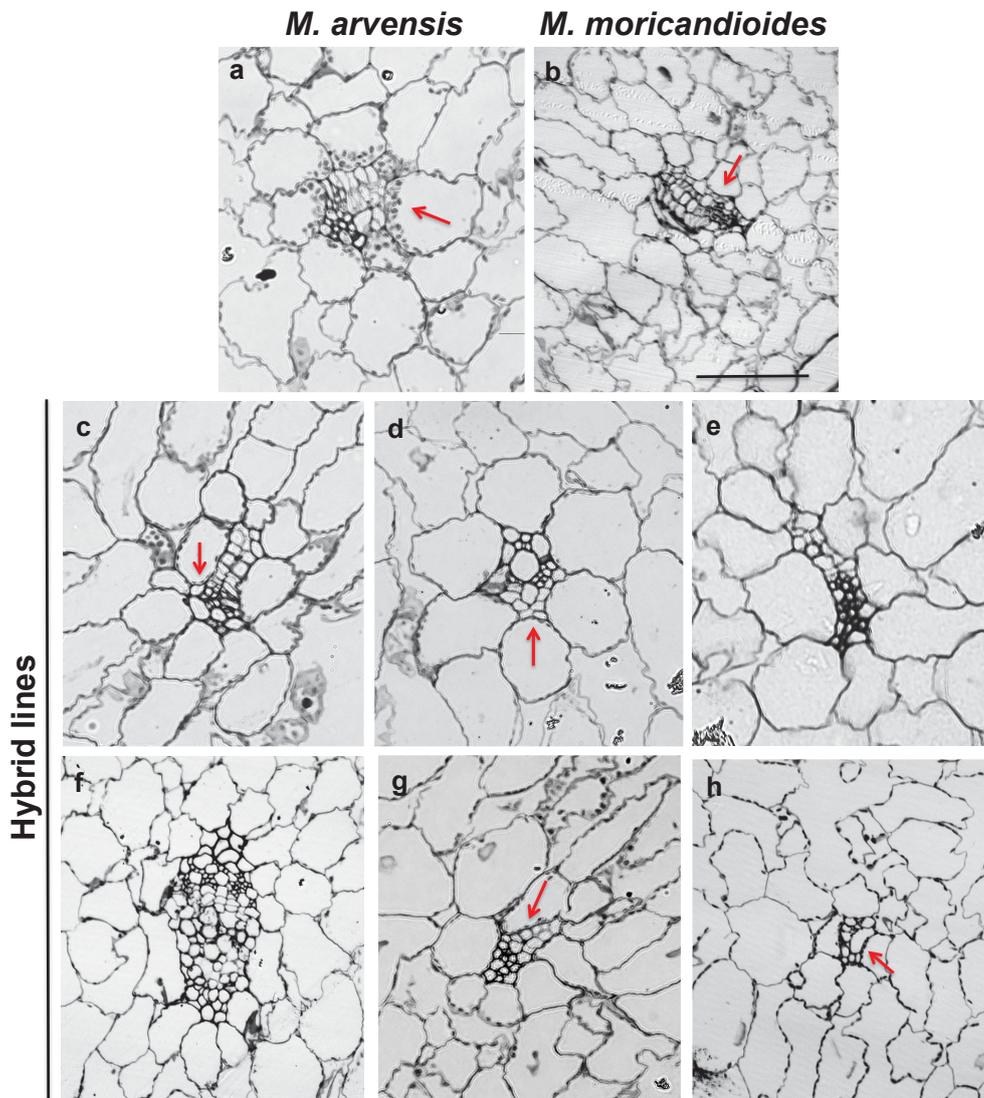
- Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ** (2013) Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588
- Hernandez-Garcia CM, Finer JJ** (2014) Identification and validation of promoters and cis-acting regulatory elements. *Plant Science* **217–218**: 109–119
- Hibberd JM, Covshoff S** (2010) The Regulation of Gene Expression Required for C4 Photosynthesis. *Annual Review of Plant Biology* **61**: 181–207
- Kadereit G, Bohley K, Lauterbach M, Tefarikis DT, Kadereit JW** (2017) C3–C4 intermediates may be of hybrid origin – a reminder. *New Phytologist* **215**: 70–76
- Kong S-G, Kagawa T, Wada M, Nagatani A** (2013) A C-terminal membrane association domain of phototropin 2 is necessary for chloroplast movement. *Plant Cell Physiol* **54**: 57–68
- Krenzer EG, Moss DN, Crookston RK** (1975) Carbon dioxide compensation points of flowering plants. *Plant Physiol* **56**: 194–206
- Lange BM, Croteau R** (1999) Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. *Proc Natl Acad Sci U S A* **96**: 13714–13719
- Lemmon ZH, Bukowski R, Sun Q, Doebley JF** (2014) The Role of cis Regulatory Evolution in Maize Domestication. *PLOS Genetics* **10**: e1004745
- Lemos B, Araripe LO, Fontanillas P, Hartl DL** (2008) Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci U S A* **105**: 14471–14476
- 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 HK** (1997) EVOLUTION OF CAROTENOID AND ISOPRENOID BIOSYNTHESIS IN PHOTOSYNTHETIC AND NON-PHOTOSYNTHETIC ORGANISMS. 14

- Love MI, Huber W, Anders S** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550
- Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U** (2014) The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *eLife* **3**: e02478
- Mayba O, Gilbert HN, Liu J, Haverty PM, Jhunjhunwala S, Jiang Z, Watanabe C, Zhang Z** (2014) MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome Biol*. doi: 10.1186/s13059-014-0405-3
- McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ** (2010) Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res* **20**: 816–825
- Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ** (2014) The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res* **24**: 84–95
- Monson RK, Rawsthorne S** (2000) CO₂ Assimilation in C₃-C₄ Intermediate Plants. *In* RC Leegood, TD Sharkey, S von Caemmerer, eds, *Photosynthesis: Physiology and Metabolism*. Springer Netherlands, Dordrecht, pp 533–550
- Prud'homme B, Gompel N, Carroll SB** (2007) Emerging principles of regulatory evolution. *PNAS* **104**: 8605–8612
- Rawsthorne S** (1992) C₃–C₄ intermediate photosynthesis: linking physiology to gene expression. *The Plant Journal* **2**: 267–274
- Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW** (1988) Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C₃-C₄ intermediate species *Moricandia arvensis* (L.) DC. *Planta* **176**: 527–532
- Reeves G, Grangé-Guermente MJ, Hibberd JM** (2017) Regulatory gateways for cell-specific gene expression in C₄ leaves with Kranz anatomy. *J Exp Bot* **68**: 107–116
- Rhoné B, Mariac C, Couderc M, Berthouly-Salazar C, Ousseini IS, Vigouroux Y** (2017) No Excess of Cis-Regulatory Variation Associated with Intraspecific Selection in Wild Pearl Millet (*Cenchrus americanus*). *Genome Biol Evol* **9**: 388–397

- Sage RF** (2004) The evolution of C4 photosynthesis. *New Phytologist* **161**: 341–370
- Sage RF, Khoshravesh R, Sage TL** (2014) From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *J Exp Bot* **65**: 3341–3356
- Sage RF, Sage TL, Kocacinar F** (2012) Photorespiration and the Evolution of C4 Photosynthesis. *Annu Rev Plant Biol* **63**: 19–47
- Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin P-A, Kurz S, Mettler-Altmann T, Weber AP** (2017) Photosynthesis in C3–C4 intermediate *Moricandia* species. *J Exp Bot* **68**: 191–206
- Schlüter U, Weber APM** (2016) The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol* **57**: 881–889
- Shao L, Xing F, Xu C, Zhang Q, Che J, Wang X, Song J, Li X, Xiao J, Chen L-L, et al** (2019) Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *PNAS* **116**: 5653–5658
- Sheen J** (1999) C4 Gene Expression. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**: 187–217
- Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM** (2011) A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res* **21**: 1728–1737
- Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T** (2015) Cis-Regulatory Changes Associated with a Recent Mating System Shift and Floral Adaptation in *Capsella*. *Mol Biol Evol* **32**: 2501–2514
- Stern DL, Orgogozo V** (2008) The loci of evolution: how predictable is genetic evolution? *Evolution* **62**: 2155–2177
- Sultmanis SD** (2018) Developmental and Expression Evolution in C3 and C4 Atriplex and their Hybrids.
- Tirosh I, Reikhav S, Levy AA, Barkai N** (2009) A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* **324**: 659–662

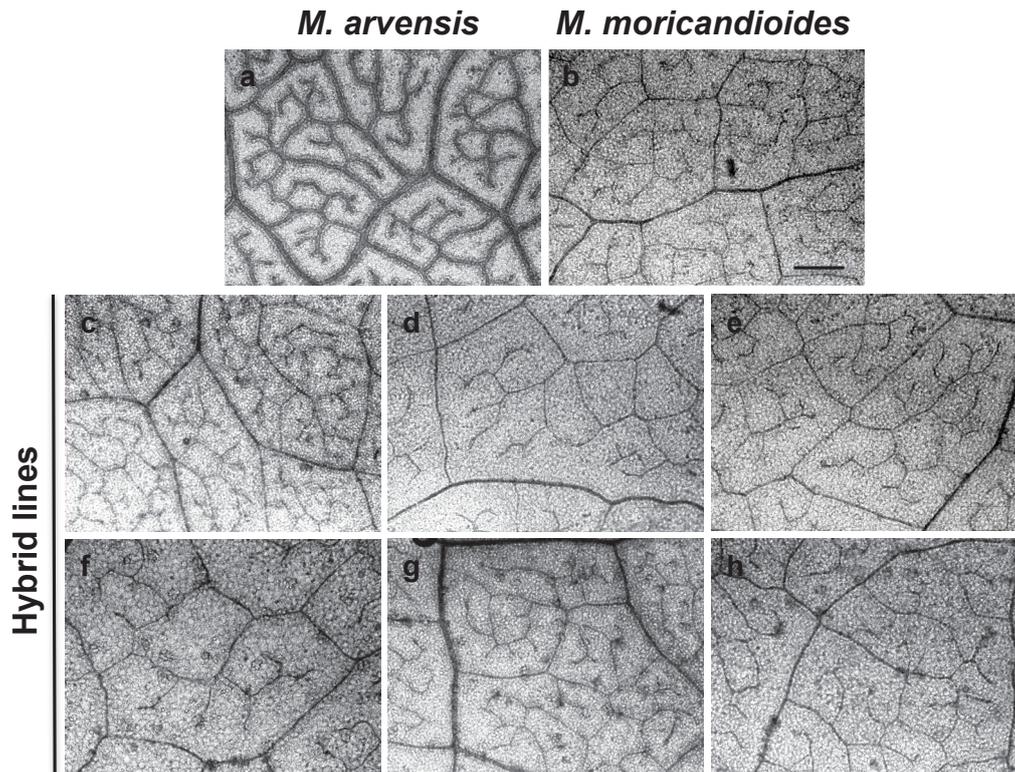
- Ueno O, Woo Bang S, Wada Y, Kobayashi N, Kaneko R, Kaneko Y, Matsuzawa Y** (2007) Inheritance of C3-C4 Intermediate Photosynthesis in Reciprocal Hybrids Between *Moricandia arvensis* (C3-C4) and *Brassica oleracea* (C3) That Differ in Their Genome Constitution. *Plant Production Science* **10**: 68–79
- Williams BP, Johnston IG, Covshoff S, Hibberd JM** (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife* **2**: e00961
- Wittkopp PJ, Haerum BK, Clark AG** (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**: 85–88
- Wittkopp PJ, Haerum BK, Clark AG** (2008) Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genetics* **40**: 346–350
- Wittkopp PJ, Kalay G** (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**: 59–69
- Wray GA** (2007) The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216
- 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 Res* **39**: W316–W322

Supplemental information



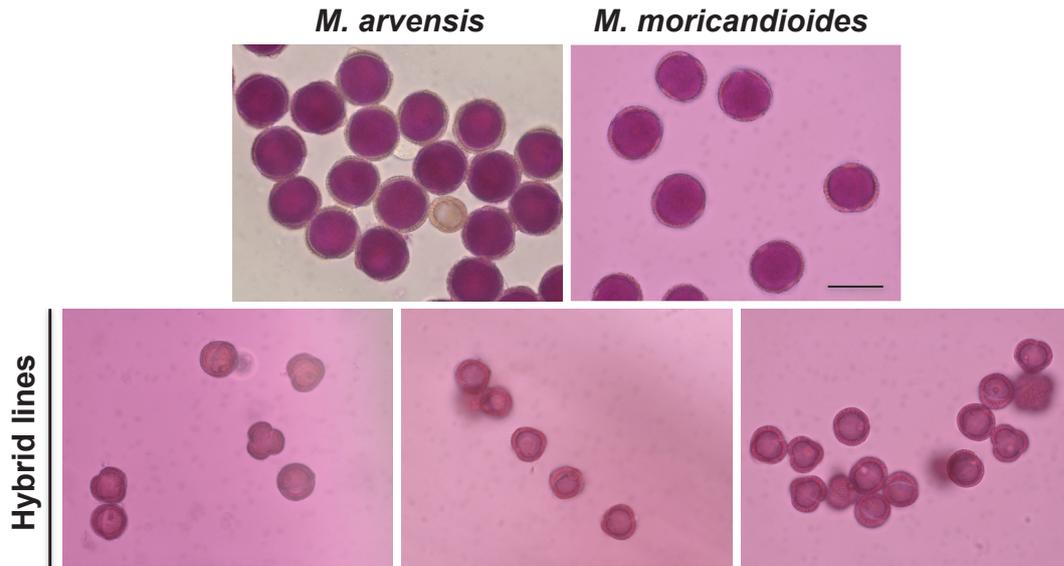
Supplemental Figure 1. Leaf micrographs of transverse sections of *M. arvensis* and *M. moricandioides* and their hybrid lines.

a, *M. arvensis*; b, *M. moricandioides*; c-h, Hybrid1-6. Arrow, chloroplasts. Bar, 100 μ m.



Supplemental Figure 2. Leaf venations of *M. arvensis* and *M. moricandioides* and their hybrid lines.

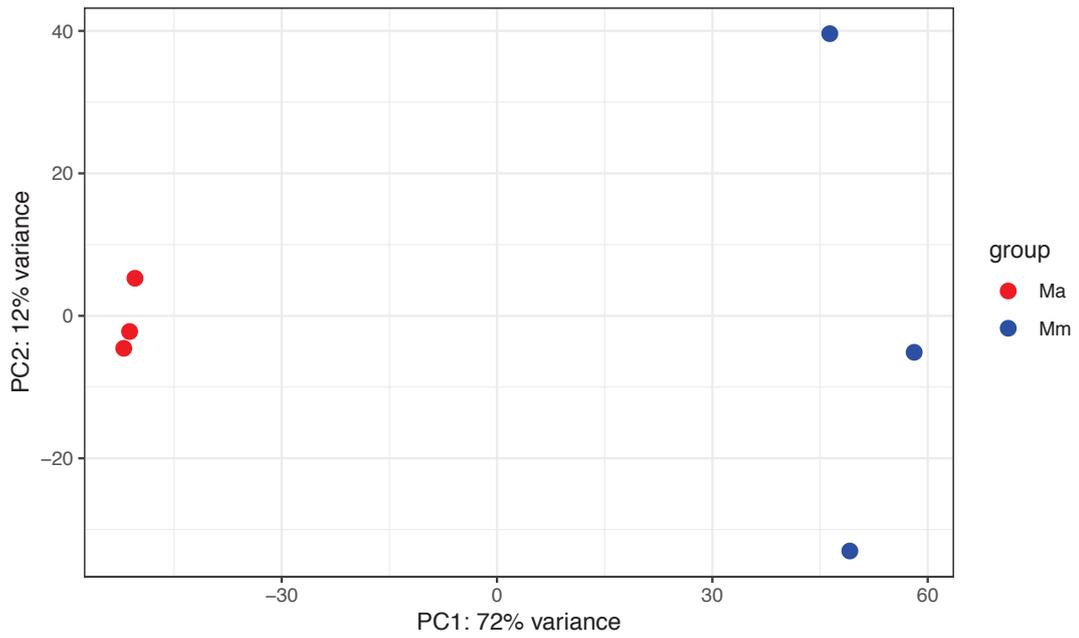
a, *M. arvensis*; b, *M. moricandioides*; c-h, Hybrid1-6. Bar, 500 um.



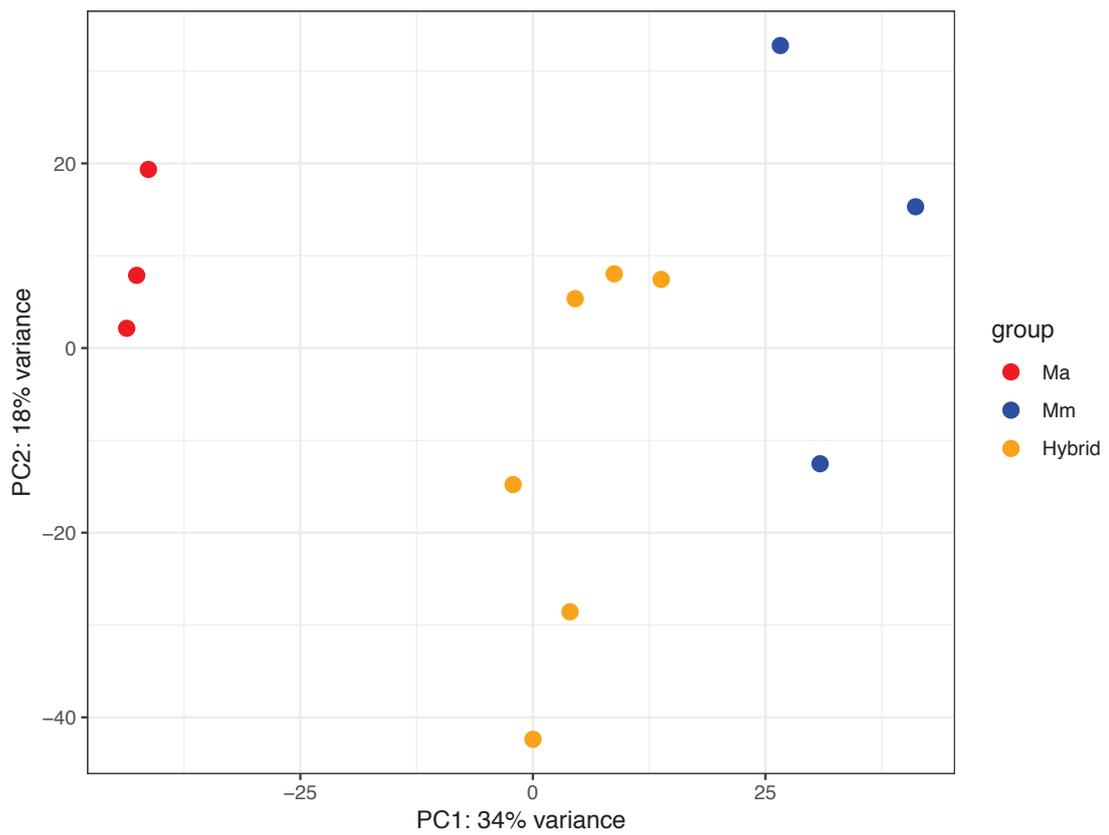
Supplemental Figure 3. The pollen activity test of *M. arvensis* and *M. moricandioides* and their hybrid lines dyed by Alexander staining method.

Pollens from hybrid lines were stained red, but demonstrated abnormal shapes compared to parents, which were round. Aborted pollen grains are stained blue-green, and non-aborted pollen grains are stained magenta-red. Bar, 20 μm .

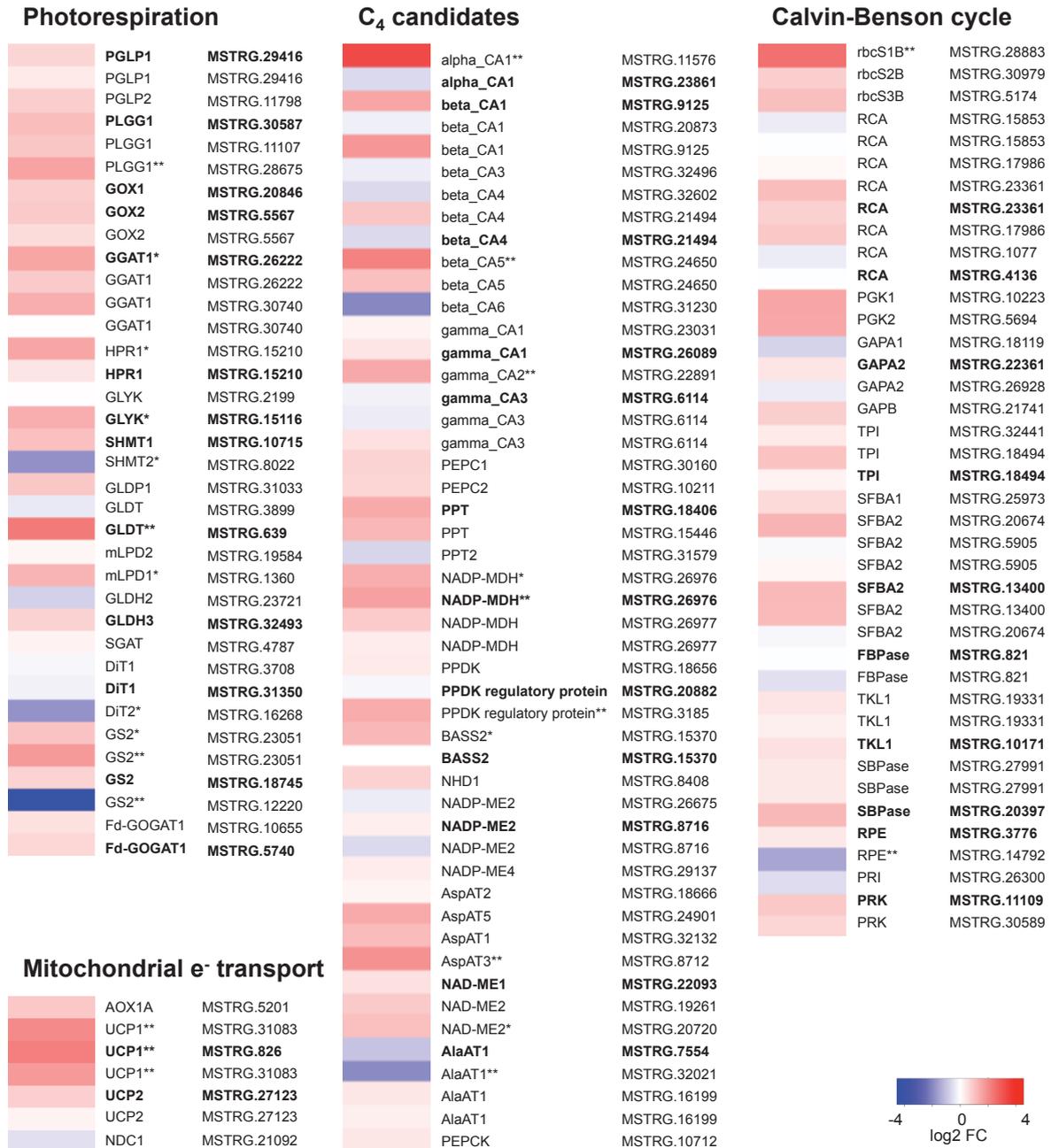
A



B



Supplemental Figure 4. Principle component analysis on (A) parental species, *M. arvensis* (Ma) and *M. moricandioides* (Mm) and (B) parental species and their interspecific hybrids.



Supplemental Figure 5. Transcriptional changes in selected pathways.

The heatmap indicated the log₂-fold changes in transcript level of C₃–C₄ species *M. arvensis* compared to the C₃ species *M. moricandioides*. Blue and red indicates reduced and enhanced transcript abundance in C₃–C₄, respectively. *, adjusted P-value < 0.05; **, adjusted P-value < 0.01. Bold, transcripts with the highest expression among isoforms.

Supplemental Table 1. GO analysis on Ma-upregulated transcripts using topGO.

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	q_value
1	GO:0044281	small molecule metabolic process	6113	192	116.4	1.3e-13	1.6e-11
2	GO:0044710	single-organism metabolic process	9956	272	189.57	1.4e-12	6.4e-11
3	GO:0044723	single-organism carbohydrate metabolic process	2775	107	52.84	1.6e-12	6.4e-11
4	GO:0006082	organic acid metabolic process	4432	148	84.39	2.8e-12	7.4e-11
5	GO:0005975	carbohydrate metabolic process	3074	114	58.53	3.1e-12	7.4e-11
6	GO:0019752	carboxylic acid metabolic process	3784	129	72.05	2.8e-11	5.3e-10
7	GO:0043436	oxoacid metabolic process	3994	134	76.05	3.1e-11	5.3e-10
8	GO:0046686	response to cadmium ion	1364	63	25.97	1.1e-10	1.7e-09
9	GO:0044724	single-organism carbohydrate catabolic process	927	48	17.65	5.3e-10	6.6e-09
10	GO:0010038	response to metal ion	1595	68	30.37	5.5e-10	6.6e-09
11	GO:0016052	carbohydrate catabolic process	1035	51	19.71	8.2e-10	8.9e-09
12	GO:0006833	water transport	412	29	7.84	2.1e-09	1.9e-08
13	GO:0042044	fluid transport	412	29	7.84	2.1e-09	1.9e-08
14	GO:0007030	Golgi organization	555	34	10.57	3.1e-09	2.7e-08
15	GO:0010256	endomembrane system organization	610	35	11.62	9.8e-09	7.8e-08
16	GO:0044283	small molecule biosynthetic process	2871	98	54.67	1.1e-08	8.2e-08
17	GO:0044712	single-organism catabolic process	2558	90	48.71	1.2e-08	8.5e-08
18	GO:0044763	single-organism cellular process	14828	351	282.34	1.6e-08	1.1e-07
19	GO:0032787	monocarboxylic acid metabolic process	3088	102	58.8	2.7e-08	1.7e-07
20	GO:0044711	single-organism biosynthetic process	6322	175	120.38	5.8e-08	3.5e-07
21	GO:0006970	response to osmotic stress	2193	78	41.76	8.0e-08	4.6e-07
22	GO:0016053	organic acid biosynthetic process	2159	77	41.11	8.8e-08	4.8e-07
23	GO:0006732	coenzyme metabolic process	1496	59	28.49	1.2e-07	5.6e-07
24	GO:1901575	organic substance catabolic process	3606	112	68.66	1.3e-07	5.6e-07
25	GO:0009144	purine nucleoside triphosphate metabolic process	709	36	13.5	1.3e-07	5.6e-07
26	GO:0009199	ribonucleoside triphosphate metabolic process	709	36	13.5	1.3e-07	5.6e-07
27	GO:0009205	purine ribonucleoside triphosphate metabolic process	709	36	13.5	1.3e-07	5.6e-07
28	GO:0046034	ATP metabolic process	709	36	13.5	1.3e-07	5.6e-07
29	GO:0009123	nucleoside monophosphate metabolic process	742	37	14.13	1.4e-07	5.6e-07
30	GO:0009161	ribonucleoside monophosphate metabolic process	742	37	14.13	1.4e-07	5.6e-07

Supplemental Table 2. GO analysis on Ma-downregulated transcripts using topGO.

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	q_value
1	GO:0043247	telomere maintenance in response to DNA damage	184	68	11.22	< 1e-30	NA
2	GO:0032204	regulation of telomere maintenance	187	68	11.4	< 1e-30	NA
3	GO:0032844	regulation of homeostatic process	191	68	11.65	< 1e-30	NA
4	GO:0000723	telomere maintenance	206	70	12.56	< 1e-30	NA
5	GO:0032200	telomere organization	206	70	12.56	< 1e-30	NA
6	GO:0060249	anatomical structure homeostasis	209	70	12.74	< 1e-30	NA
7	GO:0010212	response to ionizing radiation	261	77	15.92	< 1e-30	NA
8	GO:0010332	response to gamma radiation	229	72	13.96	< 1e-30	NA
9	GO:0007129	synapsis	202	66	12.32	1.1e-30	3.6e-28
10	GO:0045143	homologous chromosome segregation	202	66	12.32	1.1e-30	3.6e-28
11	GO:0000724	double-strand break repair via homologous recombination	215	68	13.11	1.3e-30	3.6e-28
12	GO:0000725	recombinational repair	215	68	13.11	1.3e-30	3.6e-28
13	GO:0070192	chromosome organization involved in meiotic cell cycle	210	66	12.81	1.5e-29	3.4e-27
14	GO:0043687	post-translational protein modification	301	78	18.36	2.2e-28	4.1e-26
15	GO:0006302	double-strand break repair	277	73	16.89	3.8e-27	6.1e-25
16	GO:0042138	meiotic DNA double-strand break formation	232	65	14.15	6.0e-26	8.4e-24
17	GO:0033044	regulation of chromosome organization	342	80	20.86	7.1e-26	8.4e-24
18	GO:0007062	sister chromatid cohesion	369	80	22.5	1.4e-23	1.6e-21
19	GO:0000819	sister chromatid segregation	378	81	23.05	1.6e-23	1.6e-21
20	GO:0009887	animal organ morphogenesis	354	78	21.59	1.7e-23	1.6e-21
21	GO:0045132	meiotic chromosome segregation	328	74	20	5.1e-23	4.4e-21
22	GO:0048513	animal organ development	477	91	29.09	1.6e-22	1.3e-20
23	GO:0090305	nucleic acid phosphodiester bond hydrolysis	333	72	20.31	2.7e-21	2.0e-19
24	GO:0098813	nuclear chromosome segregation	430	83	26.22	5.6e-21	3.9e-19
25	GO:0007131	reciprocal meiotic recombination	327	70	19.94	1.7e-20	1.1e-18
26	GO:0035825	reciprocal DNA recombination	327	70	19.94	1.7e-20	1.1e-18
27	GO:0007127	meiosis I	330	70	20.12	3.0e-20	1.8e-18
28	GO:0007059	chromosome segregation	444	83	27.08	4.6e-20	2.6e-18
29	GO:0007126	meiotic nuclear division	464	84	28.3	2.2e-19	1.2e-17
30	GO:0010638	positive regulation of organelle organization	492	87	30	2.4e-19	1.2e-17

Supplemental Table 3. Regulatory effects among six *Moricandia* interspecific hybrids.

Sample name	Total SNP site (transcript)	only cis (transcript)	only trans (transcript)	cis plus trans (transcript)	no cis no trans (transcript)
Hybrid1	123,684 (14,312)	36,936 (8,708)	8,680 (3,565)	37,078 (8,558)	40,990 (8,597)
Hybrid2	123,797 (14,315)	38,372 (9,089)	8,595 (3,606)	35,623 (8,452)	41,207 (8,798)
Hybrid3	112,328 (13,254)	34,589 (8,521)	8,158 (3,493)	33,460 (8,189)	36,121 (8,212)
Hybrid4	117,222 (13,781)	34,973 (8,639)	8,413 (3,506)	35,731 (8,428)	38,105 (8,499)
Hybrid5	122,693 (14,256)	37,511 (9,072)	8,444 (3,460)	36,276 (8,499)	40,462 (8,699)
Hybrid6	121,478 (14,108)	35,152 (8,657)	8,837 (3,658)	38,878 (8,841)	38,611 (8,430)

Supplemental Table 4. Confirmation of RNA-Seq data by allele-specific RT-PCR of *M. arvensis* × *M. moricandioides* hybrid.

<i>Moricandia</i> gene	Hybrid	Allele-specific transcript level and allelic ratio in RNA-Seq			Allelic ratio in qPCR	
		Ma allele in F1	Mm allele in F1	Allelic ratio Ma:Mm in F1	Allelic ratio Ma:Mm in F1	
GLK2 (MSTRG.5234)	1	77	145	0.53	0.18	
	2	52	151	0.34	0.37	
	3	38	128	0.30	0.21	
	4	67	122	0.55	0.12	
	5	61	148	0.41	0.18	
	6	87	137	0.64	0.20	
ASP3 (MSTRG.33199)	1	283	74	3.82	35.39	
	2	404	100	4.04	58.21	
	3	180	65	2.77	31.86	
	4	466	88	5.30	1.30	
	5	295	51	5.78	41.20	
	6	395	70	5.64	55.84	
gamma_CA2 (MSTRG.26888)	1	180	85	2.12	3.10	
	2	173	89	1.94	5.20	
	3	121	47	2.57	3.01	
	4	151	46	3.28	3.82	
	5	159	66	2.41	3.74	
	6	113	61	1.85	2.37	
PPA2 (MSTRG.30130)	1	299	121	2.47	2.57	
	2	331	170	1.95	5.72	
	3	171	85	2.01	2.83	
	4	713	359	1.99	2.41	
	5	380	173	2.20	2.24	
	6	346	158	2.19	2.58	

Supplemental Table 5. GO analysis on common *cis*-SNPs using topGO.

1, 2, 3 and 4 showed related GO terms of isopentenyl diphosphate biosynthesis, carbohydrate catabolic process, oxidoreduction coenzyme metabolic process and chloroplast relocation, respectively.

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	q_value
1	GO:0051186	³ cofactor metabolic process	2019	183	76.7	1.7e-28	3.8e-26
2	GO:0006090	¹ pyruvate metabolic process	1188	123	45.13	4.8e-24	4.3e-22
3	GO:0006081	¹ cellular aldehyde metabolic process	1067	115	40.54	5.8e-24	4.3e-22
4	GO:0019682	¹ glyceraldehyde-3-phosphate metabolic process	840	97	31.91	1.8e-22	1.0e-20
5	GO:0044723	single-organism carbohydrate metabolic process	2775	208	105.42	7.7e-22	3.5e-20
6	GO:0009056	² catabolic process	4056	271	154.09	2.3e-21	8.6e-20
7	GO:1901575	² organic substance catabolic process	3606	247	136.99	9.3e-21	3.0e-19
8	GO:0022607	cellular component assembly	2041	164	77.54	4.4e-20	1.2e-18
9	GO:0005975	² carbohydrate metabolic process	3074	216	116.78	2.5e-19	6.2e-18
10	GO:0044281	¹ small molecule metabolic process	6113	359	232.24	3.5e-19	7.8e-18
11	GO:0006732	³ coenzyme metabolic process	1496	131	56.83	3.8e-19	7.8e-18
12	GO:0044710	single-organism metabolic process	9956	521	378.23	1.6e-18	3.0e-17
13	GO:0009250	glucan biosynthetic process	671	78	25.49	1.9e-18	3.2e-17
14	GO:0019288	¹ isopentenyl diphosphate biosynthetic process, MEP pathway	572	71	21.73	2.0e-18	3.2e-17
15	GO:0006091	generation of precursor metabolites and energy	1410	124	53.57	2.6e-18	3.9e-17
16	GO:0009240	¹ isopentenyl diphosphate biosynthetic process	581	71	22.07	4.6e-18	6.0e-17
17	GO:0046490	¹ isopentenyl diphosphate metabolic process	581	71	22.07	4.6e-18	6.0e-17
18	GO:0006733	³ oxidoreduction coenzyme metabolic process	1199	110	45.55	1.2e-17	1.5e-16
19	GO:1901135	¹ carbohydrate derivative metabolic process	2850	199	108.27	1.7e-17	1.9e-16
20	GO:0008299	¹ isoprenoid biosynthetic process	849	88	32.25	1.8e-17	1.9e-16
21	GO:0043623	cellular protein complex assembly	1515	128	57.56	1.8e-17	1.9e-16
22	GO:0070271	protein complex biogenesis	1609	133	61.13	2.4e-17	2.5e-16
23	GO:0006720	¹ isoprenoid metabolic process	871	89	33.09	2.8e-17	2.7e-16
24	GO:0006461	protein complex assembly	1600	132	60.78	3.7e-17	3.5e-16
25	GO:0009902	⁴ chloroplast relocation	234	42	8.89	4.6e-17	4.0e-16
26	GO:0051667	⁴ establishment of plastid localization	234	42	8.89	4.6e-17	4.0e-16
27	GO:0009658	⁴ chloroplast organization	554	67	21.05	7.0e-17	5.7e-16
28	GO:0019750	⁴ chloroplast localization	237	42	9	7.4e-17	5.7e-16
29	GO:0051644	⁴ plastid localization	237	42	9	7.4e-17	5.7e-16
30	GO:0016052	² carbohydrate catabolic process	1035	98	39.32	1.1e-16	8.2e-16

Supplemental Table 6. GO analysis on common *trans*-SNPs using topGO.

1, 2, 3, 4 and 5 showed related GO terms of nucleosome assembly, RNA methylation, organophosphate biosynthetic process, peptide metabolic process and ribonucleotide metabolic process, respectively.

	GO.ID	Term	Annotated	Significant	Expected	classicFisher	q_value
1	GO:0019637	³ organophosphate metabolic process	2720	17	4.75	2.3e-06	2.8e-04
2	GO:0006333	¹ chromatin assembly or disassembly	149	5	0.26	6.4e-06	3.1e-04
3	GO:1901564	² organonitrogen compound metabolic process	5173	23	9.03	7.8e-06	3.1e-04
4	GO:0009451	² RNA modification	860	9	1.5	1.7e-05	3.6e-04
5	GO:0009117	⁵ nucleotide metabolic process	1889	13	3.3	1.7e-05	3.6e-04
6	GO:0006753	⁵ nucleoside phosphate metabolic process	1898	13	3.31	1.8e-05	3.6e-04
7	GO:0055086	⁵ nucleobase-containing small molecule metabolic process	1967	13	3.43	2.7e-05	4.5e-04
8	GO:1901566	organonitrogen compound biosynthetic process	3705	18	6.46	3.6e-05	4.5e-04
9	GO:0044281	small molecule metabolic process	6113	24	10.67	3.8e-05	4.5e-04
10	GO:0006334	¹ nucleosome assembly	38	3	0.07	4.1e-05	4.5e-04
11	GO:0034728	¹ nucleosome organization	38	3	0.07	4.1e-05	4.5e-04
12	GO:0019693	⁵ ribose phosphate metabolic process	1505	11	2.63	5.0e-05	5.0e-04
13	GO:1901135	⁵ carbohydrate derivative metabolic process	2850	15	4.97	8.0e-05	6.6e-04
14	GO:0006081	cellular aldehyde metabolic process	1067	9	1.86	9.0e-05	6.6e-04
15	GO:0001510	² RNA methylation	618	7	1.08	9.8e-05	6.6e-04
16	GO:0065004	¹ protein-DNA complex assembly	51	3	0.09	9.9e-05	6.6e-04
17	GO:0071824	¹ protein-DNA complex subunit organization	51	3	0.09	9.9e-05	6.6e-04
18	GO:0006796	⁵ phosphate-containing compound metabolic process	4392	19	7.66	1.0e-04	6.6e-04
19	GO:0006793	⁵ phosphorus metabolic process	4423	19	7.72	1.1e-04	7.0e-04
20	GO:0090407	³ organophosphate biosynthetic process	1720	11	3	1.6e-04	9.6e-04
21	GO:0031497	¹ chromatin assembly	62	3	0.11	1.8e-04	1.0e-03
22	GO:0006323	¹ DNA packaging	67	3	0.12	2.2e-04	1.2e-03
23	GO:0071704	organic substance metabolic process	17880	44	31.2	3.7e-04	2.0e-03
24	GO:0006518	⁴ peptide metabolic process	1656	10	2.89	5.3e-04	2.5e-03
25	GO:0009987	cellular process	21493	49	37.5	5.3e-04	2.5e-03
26	GO:0008152	metabolic process	18806	45	32.82	5.8e-04	2.6e-03
27	GO:0043603	² cellular amide metabolic process	1676	10	2.92	5.8e-04	2.6e-03
28	GO:0019682	⁵ glyceraldehyde-3-phosphate metabolic process	840	7	1.47	6.3e-04	2.7e-03
29	GO:0009259	⁵ ribonucleotide metabolic process	1114	8	1.94	6.7e-04	2.8e-03
30	GO:0034641	² cellular nitrogen compound metabolic process	9857	29	17.2	8.5e-04	3.4e-03

Supplemental Table 7. Significantly enriched pathways identified in transcripts with common *cis*-SNPs using KOBAS database.

Pathway	Pathway ID	Input number	Background number	P-Value	Corrected P-Value
Carbon metabolism	ath01200	49	262	4.6e-13	2.4e-11
Metabolic pathways	ath01100	172	1910	5.9e-12	2.7e-10
Oxidative phosphorylation	ath00190	35	162	3.3e-11	1.4e-09
Proteasome	ath03050	21	58	1.3e-10	5.0e-09
Protein processing in endoplasmic reticulum	ath04141	33	212	1.4e-07	2.7e-06
Phagosome	ath04145	20	86	1.9e-07	3.5e-06
Biosynthesis of amino acids	ath01230	33	255	5.5e-06	7.0e-05
Biosynthesis of secondary metabolites	ath01110	91	1076	8.1e-06	9.8e-05
Carbon fixation in photosynthetic organisms	ath00710	15	69	1.3e-05	1.5e-04
Citrate cycle (TCA cycle)	ath00020	14	63	2.0e-05	2.2e-04
Porphyrin and chlorophyll metabolism	ath00860	12	48	2.8e-05	2.9e-04
Pyruvate metabolism	ath00620	16	85	3.1e-05	3.2e-04
Glycolysis / Gluconeogenesis	ath00010	18	117	1.1e-04	9.6e-04
Glyoxylate and dicarboxylate metabolism	ath00630	13	74	3.1e-04	2.4e-03
Spliceosome	ath03040	22	192	8.6e-04	6.0e-02
2-Oxocarboxylic acid metabolism	ath01210	12	74	9.6e-04	6.6e-02
Arginine biosynthesis	ath00220	8	35	1.0e-03	6.8e-02
Pentose phosphate pathway	ath00030	10	58	1.7e-03	1.0e-02
Ubiquitin mediated proteolysis	ath04120	17	142	2.1e-03	1.3e-02
mRNA surveillance pathway	ath03015	14	114	4.0e-03	2.1e-02
Endocytosis	ath04144	16	142	4.8e-03	2.4e-02
RNA degradation	ath03018	13	112	8.3e-03	3.7e-02
Cysteine and methionine metabolism	ath00270	13	112	8.3e-03	3.7e-02
Fatty acid biosynthesis	ath00061	7	41	8.4e-03	3.7e-02
Peroxisome	ath04146	11	87	8.4e-03	3.7e-02
Nitrogen metabolism	ath00910	7	42	9.4e-03	4.0e-02
Fatty acid metabolism	ath01212	9	67	1.2e-02	4.9e-02

Supplemental Table 8. Significantly enriched pathways identified in transcripts with common *trans*-SNPs using KOBAS database.

Pathway	Pathway ID	Input number	Background number	P-Value	Corrected P-Value
Ribosome	ath03010	7	363	2.2e-05	8.7e-04
Carbon metabolism	ath01200	5	262	3.7e-04	5.7e-03
Fatty acid metabolism	ath01212	3	67	5.6e-04	8.2e-03
Metabolic pathways	ath01100	12	1910	1.2e-03	1.6e-02
Propanoate metabolism	ath00640	2	33	2.9e-03	2.3e-02
Biosynthesis of amino acids	ath01230	4	255	3.0e-03	2.3e-02
Biosynthesis of secondary metabolites	ath01110	8	1076	3.2e-03	2.3e-02
Fatty acid degradation	ath00071	2	41	4.4e-03	3.5e-02
Fatty acid biosynthesis	ath00061	2	41	4.4e-03	3.5e-02

Supplemental Table 9. Number of biased transcript with *cis*-specificity among six hybrids.

Sample name	Transcripts with <i>cis</i> -specificity	
	Ma_biased	Mm_biased
Hybrid1	1,105	820
Hybrid2	1,128	822
Hybrid3	1,050	758
Hybrid4	1,086	793
Hybrid5	1,124	847
Hybrid6	1,057	825

Supplemental Table 10. Transcripts showing extreme allelic imbalance with major allele frequency ≥ 0.9 in all hybrids.

Transcript	Gene model	Bias	Description
MSTRG.1939	AT1G52300	Ma	Zinc-binding ribosomal protein family protein
MSTRG.30345	AT2G22170	Ma	Lipase/lipoxygenase, PLAT/LH2 family protein
MSTRG.19197	AT2G31670	Ma	Stress responsive alpha-beta barrel domain protein
MSTRG.25443	AT1G31330	Ma	Encodes subunit F of photosystem I
MSTRG.25261	AT4G04020	Ma	Fibrillin precursor protein
MSTRG.25262	AT4G04040	Ma	Phosphofructokinase family protein
MSTRG.31918	AT5G09810	Ma	Member of Actin gene family
MSTRG.21568	AT1G32400	Ma	TOBAMOVIRUS MULTIPLICATION 2A, TOM2A
MSTRG.22263	AT1G79040	Ma	Encodes for the 10 kDa PsbR subunit of photosystem II (PSII)
MSTRG.30384	AT1G75560	Ma	zinc knuckle (CCHC-type) family protein
MSTRG.18535	AT4G26080	Ma	Involved in abscisic acid (ABA) signal transduction
MSTRG.16015	AT1G09340	Ma	Encodes CHLOROPLAST RNA BINDING (CRB), a putative RNA-binding protein
MSTRG.27651	AT5G59310	Ma	Encodes a member of the lipid transfer protein family
MSTRG.5109	AT3G52300	Ma	ATP synthase subunit d, mitochondrial
MSTRG.26182	AT1G29930	Ma	Subunit of light-harvesting complex II (LHCII)
MSTRG.30822	AT5G50340	Mm	DNA repair protein RadA-like protein
MSTRG.12252	AT3G27100	Mm	ENY2 is a component of the deubiquitination module of the SAGA complex
MSTRG.27711	AT3G63410	Mm	Encodes a MPBQ/MSBQ methyltransferase located in the chloroplast inner envelope membrane
MSTRG.20464	AT2G27030	Mm	Encodes a calmodulin that has higher affinity to kinesin-like calmodulin binding motor protein than CAM4 or CAM6
MSTRG.1986	AT2G44060	Mm	Late embryogenesis abundant protein
MSTRG.8677	AT1G16880	Mm	Encodes a ACT domain-containing protein
MSTRG.30005	AT1G60950	Mm	Encodes a major leaf ferredoxin
MSTRG.8677	AT1G16880	Mm	Encodes a ACT domain-containing protein
MSTRG.26616	AT5G10140	Mm	MADS-box protein encoded by FLOWERING LOCUS C
MSTRG.12357	AT4G05320	Mm	One of five polyubiquitin genes in A. thaliana
MSTRG.12357	AT4G05320	Mm	One of five polyubiquitin genes in A. thaliana
MSTRG.26378	AT1G30380	Mm	Encodes subunit K of photosystem I reaction center
MSTRG.10014	AT4G01310	Mm	Ribosomal L5P family protein
MSTRG.7243	AT3G27830	Mm	50S ribosomal protein L12-A The mRNA is cell-to-cell mobile
MSTRG.7244	AT3G27850	Mm	50S ribosomal protein L12-C The mRNA is cell-to-cell mobile
MSTRG.19445	AT2G47320	Mm	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein
MSTRG.7438	AT1G31330	Mm	Encodes subunit F of photosystem I

Supplemental Table 11. Enrichment of regulatory effects in selected pathways.

0, no *cis*-SNP; 1, at least one *cis*-SNP found in hybrid line; 2, common *cis*-SNP among hybrid.

Pathway	Gene		<i>A. thaliana</i>	<i>M. moricandioides</i>	Info	
Glycine shuttle	2-phosphoglycolate phosphatase	PGLP1	AT5G36700	MSTRG.29416	1	
		PGLP2	AT5G36700	MSTRG.29416	1	
	Plastidial glycolate/glycerate transporter 1	PGLP2	AT5G47760	MSTRG.11798	1	
		PLGG1	AT1G32080	MSTRG.30587	1	
		PLGG1	AT1G32080	MSTRG.11107	1	
	Glycolate oxidase	PLGG1	AT1G32080	MSTRG.28675	2	
		GOX1	AT3G14420	MSTRG.20846	1	
		GOX2	AT3G14415	MSTRG.5567	1	
		GOX2	AT3G14415	MSTRG.5567	1	
	Glutamate:glyoxylate aminotransferase	GGAT1	AT1G23310	MSTRG.26222	1	
		GGAT1	AT1G23310	MSTRG.26222	1	
		GGAT1	AT1G23310	MSTRG.30740	1	
	Hydroxypyruvate reductase	GGAT1	AT1G23310	MSTRG.30740	1	
		HPR1	AT1G68010	MSTRG.15210	0	
		HPR1	AT1G68010	MSTRG.15210	1	
	Glycerate kinase	GLYK	AT1G80380	MSTRG.2199	1	
		GLYK	AT1G80380	MSTRG.15116	1	
	Serine hydroxymethyltransferase	SHMT1	AT4G37930	MSTRG.10715	1	
		SHMT2	AT5G26780	MSTRG.8022	0	
	GDC complex	GLDP1	GLDP1	AT4G33010	MSTRG.31033	1
			GLDT	AT1G11860	MSTRG.3899	1
		GLDT	GLDT	AT1G11860	MSTRG.639	0
			mLPD2	AT3G17240	MSTRG.19584	1
		mLPD1	AT1G48030	MSTRG.1360	0	
		GLDH2	GLDH2	AT2G35120	MSTRG.23721	1
			GLDH3	AT1G32470	MSTRG.32493	1
			SGAT	AT2G13360	MSTRG.4787	1
		Serine:glyoxylate aminotransferase	DiT1	AT5G12860	MSTRG.3708	2
			DiT1	AT5G12860	MSTRG.31350	1
	Dicarboxylate transporter	DiT2	AT5G64280	MSTRG.16268	1	
		GS2	AT5G35630	MSTRG.23051	1	
		GS2	AT5G35630	MSTRG.23051	0	
	Plastidial glutamine synthetase	GS2	AT5G35630	MSTRG.18745	2	
		GS2	AT5G35630	MSTRG.12220	0	
		Ferredoxin-dependent glutamate synthase	Fd-GOGAT1	AT5G04140	MSTRG.10655	2
			Fd-GOGAT1	AT5G04140	MSTRG.5740	2
	C4 candidates	Carbonic anhydrase	alpha_CA1	AT3G52720	MSTRG.11576	1
			alpha_CA1	AT3G52720	MSTRG.23861	1
			beta_CA1	AT3G01500	MSTRG.9125	1
			beta_CA1	AT3G01500	MSTRG.20873	0
			beta_CA1	AT3G01500	MSTRG.9125	1
			beta_CA3	AT1G23730	MSTRG.32496	0
			beta_CA4	AT1G70410	MSTRG.32602	1
			beta_CA4	AT1G70410	MSTRG.21494	1
			beta_CA4	AT1G70410	MSTRG.21494	1
			beta_CA5	AT4G33580	MSTRG.24650	2
			beta_CA5	AT4G33580	MSTRG.24650	0
beta_CA6			AT1G58180	MSTRG.31230	0	
gamma_CA1			AT1G19580	MSTRG.23031	0	
gamma_CA1			AT1G19580	MSTRG.26089	1	
gamma_CA2			AT1G47260	MSTRG.22891	2	
gamma_CA3			AT5G66510	MSTRG.6114	1	
gamma_CA3			AT5G66510	MSTRG.6114	1	
gamma_CA3		AT5G66510	MSTRG.6114	2		
Phosphoenolpyruvate carboxylase		PEPC1	AT1G53310	MSTRG.30160	1	
		PEPC2	AT2G42600	MSTRG.10211	1	
Phosphoenolpyruvate/phosphate translocator		PPT	AT5G33320	MSTRG.18406	1	
		PPT	AT5G33320	MSTRG.15446	1	
		PPT2	AT3G01550	MSTRG.31579	1	
NADP-Malate dehydrogenase		NADP-MDH	AT5G58330	MSTRG.26976	1	
		NADP-MDH	AT5G58330	MSTRG.26976	1	
		NADP-MDH	AT5G58330	MSTRG.26977	1	
		NADP-MDH	AT5G58330	MSTRG.26977	1	
Pyruvate, orthophosphate dikinase		PPDK	AT4G15530	MSTRG.18656	1	
		PPDK regulatory protein	AT3G01200	MSTRG.20882	0	
		PPDK regulatory protein	AT4G21210	MSTRG.3185	0	
Plasma membrane pyruvate transport		BASS2	AT2G26900	MSTRG.15370	1	
		BASS2	AT2G26900	MSTRG.15370	2	
		NHD1	AT3G19490	MSTRG.8408	2	
NADP-malic enzyme		NADP-ME2	AT5G11670	MSTRG.26675	1	
		NADP-ME2	AT5G11670	MSTRG.8716	1	
		NADP-ME2	AT5G11670	MSTRG.8716	1	
		NADP-ME4	AT1G79750	MSTRG.29137	2	
Aspartate aminotransferase		AspAT2	AT5G19550	MSTRG.18666	2	
		AspAT5	AT4G31990	MSTRG.24901	1	
		AspAT1	AT2G30970	MSTRG.32132	1	
		AspAT3	AT5G11520	MSTRG.8712	2	
NAD-malic enzyme		plastidic_NAD-MDH	AT3G47520	MSTRG.14459	1	
		NAD-ME1	AT2G13560	MSTRG.22093	1	
		NAD-ME2	AT4G00570	MSTRG.19261	1	
		NAD-ME2	AT4G00570	MSTRG.20720	1	
Alanine aminotransferase		AlaAT1	AT1G17290	MSTRG.7554	1	
		AlaAT1	AT1G17290	MSTRG.32021	0	
	AlaAT1	AT1G17290	MSTRG.16199	0		
	AlaAT1	AT1G17290	MSTRG.16199	0		
Phosphoenolpyruvate carboxykinase	PEPCK	AT4G37870	MSTRG.10712	1		

Supplemental Table 11. Enrichment of regulatory effects in selected pathways. Continued.

0, no *cis*-SNP; 1, at least one *cis*-SNP found in hybrid line; 2, common *cis*-SNP among hybrid.

Pathway	Gene		<i>A. thaliana</i>	<i>M. moricandioides</i>	Info
Calvin Benson cycle	Rubisco	rbcS1A	AT1G67090	MSTRG.9519	0
		rbcS1B	AT5G38430	MSTRG.28883	1
		rbcS2B	AT5G38420	MSTRG.30979	1
		rbcS3B	AT5G38410	MSTRG.5174	2
	Rubisco activase	RCA	AT2G39730	MSTRG.15853	0
		RCA	AT2G39730	MSTRG.15853	0
		RCA	AT2G39730	MSTRG.17986	1
		RCA	AT2G39730	MSTRG.23361	0
		RCA	AT2G39730	MSTRG.23361	0
		RCA	AT2G39730	MSTRG.17986	1
		RCA	AT1G73110	MSTRG.1077	1
		RCA	AT1G73110	MSTRG.4136	1
	Phosphoglycerate kinase	PGK1	AT3G12780	MSTRG.10223	1
		PGK2	AT1G56190	MSTRG.5694	1
	Glyceraldehyde 3-phosphate dehydrogenase A subunit	GAPA1	AT3G26650	MSTRG.18119	1
		GAPA2	AT1G12900	MSTRG.22361	0
		GAPA2	AT1G12900	MSTRG.26928	0
	Glyceraldehyde 3-phosphate dehydrogenase B subunit	GAPB	AT1G42970	MSTRG.21741	2
	Triosephosphate isomerase	TPI	AT2G21170	MSTRG.32441	1
		TPI	AT2G21170	MSTRG.18494	1
	Sedoheptulose/Fructose-biphosphate aldolase	SFBA1	AT2G21330	MSTRG.25973	1
		SFBA2	AT4G38970.1	MSTRG.20674	1
		SFBA2	AT4G38970.1	MSTRG.5905	1
		SFBA2	AT4G38970.1	MSTRG.5905	1
		SFBA2	AT4G38970.1	MSTRG.13400	1
		SFBA2	AT4G38970.1	MSTRG.13400	1
		SFBA2	AT4G38970.2	MSTRG.20674	1
	Fructose biphosphatase	FBPase	AT3G54050	MSTRG.821	1
		FBPase	AT3G54050	MSTRG.821	1
	Transketolase	TKL1	AT3G60750	MSTRG.19331	1
		TKL1	AT3G60750	MSTRG.19331	1
		TKL1	AT3G60750	MSTRG.10171	2
Sedoheptulose-1,7-biphosphatase	SBPase	AT3G55800	MSTRG.27991	0	
	SBPase	AT3G55800	MSTRG.27991	0	
	SBPase	AT3G55800	MSTRG.20397	0	
Ribulose-5-phosphate 3-epimerase	RPE	AT5G61410	MSTRG.3776	1	
	RPE	AT5G61410	MSTRG.14792	1	
Ribulose-5-phosphate isomerase	PRI	AT3G04790	MSTRG.26300	1	
Phosphoribulokinase	PRK	AT1G32060	MSTRG.11109	1	
	PRK	AT1G32060	MSTRG.30589	1	
Mitochondrial e- transport	Alternative oxidase	AOX1A	AT3G22370	MSTRG.5201	1
		AOX1D	AT1G32350	MSTRG.17569	0
	Uncoupling protein	UCP1	AT3G54110	MSTRG.31083	0
		UCP1	AT3G54110	MSTRG.826	1
		UCP1	AT3G54110	MSTRG.31083	0
		UCP2	AT5G58970	MSTRG.27123	1
	Dicarboxylate carriers	UCP2	AT5G58970	MSTRG.27123	0
		DIC1	AT2G22500	MSTRG.20154	0
		DIC1	AT2G22500	MSTRG.20154	0
		DIC1	AT2G22500	MSTRG.32374	0
		DIC2	AT4G24570	MSTRG.32295	0
	NADH dehydrogenase C1	NDC1	AT5G08740	MSTRG.21092	1

Supplemental Table 12. qPCR primer list for ASE verification.

Gene	Ortholog		Primer name	Sequence (5'-3')	Product size (bp)
	<i>A. thaliana</i>	<i>M. moricandioides</i>			
Helicase	AT1G58050	MSTRG.4472	mori_Helicase_10F mori_Helicase_10R	CGGATGCCATTGGTAGAACT CTTCACTCGGAGGTCCAAA	97
GLDP1	AT4G33010	MSTRG.37856	GLDP1_SNP5_a_5R GLDP1_SNP5_m_5R GLDP1_SNP5_g_5F	TGCTTGACGAAGGGACTGAA TGCTTGACGAAGGGACTGAG AGAACCACACACATCACCA	208
ASP3	AT5G11520	MSTRG.33199	ASP3_SNP2_a_1F ASP3_SNP2_m_1F ASP3_SNP2_g_1R	AATGTA CTCAAATCCTCCGAGC AATGTA CTCAAATCCTCCGAGT GAGCACGTAATGCCTCGAA	152
γCA2	AT1G47260	MSTRG.26888	gamma_CA2_SNP2_a_1F gamma_CA2_SNP2_m_1F gamma_CA2_SNP2_g_1R gamma_CA2_SNP2_a_2F gamma_CA2_SNP2_m_2F gamma_CA2_SNP2_g_2,3R gamma_CA2_SNP2_a_3F gamma_CA2_SNP2_m_3F	CTTCTCAACCACCACACCATAA CTTCTCAACCACCACACCATAG GATGAGGCATTTGTTGGCAT TTCTCAACCACCACACCATAA TTCTCAACCACCACACCATAG TGTTGAGGATGAGGCATTTG TCTCAACCACCACACCATAA TCTCAACCACCACACCATAG	57 63
PPA2	AT2G18230	MSTRG.30130	PPA2_SNP2_a_1F PPA2_SNP2_m_1F PPA2_SNP2_g_1R	CATTCTCGAACCATTGTGTGTA CATTCTCGAACCATTGTGTG GAAGGAATGACCCAGTCAGC	194

Supplemental Table 13. Selected gene list for promoter-GUS assay.

Primer index was listed in supplemental table 14.

Gene	Promoter region	(+1) position	Primers for promoter region amplification		size (bp)	Final vector	Construct resistance	Agrobacteria strain	Agrobacteria resistance	Plant resistance	Primers for colony PCR	
			F	R							F	R
MmGLDP1	(-)2011..(+26)	ATG	1	3	2037	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	22	23
MaGLDP1	(-)2550..(+26)	ATG	2	3	2576	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	22	23
MmPHOT2	(-)2193..(+24)	Transcription start site	4	5	2223	pGWB3	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg, Kan	18	19
MaPHOT2	(-)2230..(+3)	Transcription start site	6	7	2231	pGWB3	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg, Kan	20	21
MmCHUP1	(-)2000..(-)1	Putative ATG	8	9	2000	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	22	23
MaCHUP1	(-)2000..(-)1	Putative ATG	10	11	2000	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	22	23
MmCRB	(-)2491..(+23)	Putative ATG	12	14	2514	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	24	25
MaCRB	(-)2231..(+23)	Putative ATG	13	14	2254	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	24	25
MmATPQ	(-)2301..(+28)	Putative ATG	15	17	2329	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	24	25
MaATPQ	(-)2293..(+28)	Putative ATG	16	17	2321	pC1381	Kan	GV3101(pMP90RK)	Rif, Gen	Hyg	24	25

Supplemental Table 14. Primer list for promoter-GUS assay.

Primer name	Sequence (5'-3')
1. MmGLDP1_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCCCGGTAACCTTTAAATTGCTTG
2. MaGLDP1_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCGGAGCGGAACCTTTACGAG
3. GLDP1_univ_Gibson_R	CGTAAACTAGTCAGATCTACCATGGTAAGCAAGCCTACGTGCG
4. MmPHOT2_Gateway_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCACTATCATTCCCTCACCAT
5. MmPHOT2_Gateway_R	GGGGACCACCTTTGTACAAGAAAGCTGGGTCTGAAGGACCACACACTCTGTT
6. MaPHOT2_Gateway_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGACAAAGGCAGAAGACTGAC
7. MaPHOT2_Gateway_R	GGGGACCACCTTTGTACAAGAAAGCTGGGTCTCCTTTTCTCTTTTACTCC
8. MmCHUP1_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCCTAGAATCTCTGCTTTGATAAAAAG
9. MmCHUP1_Gibson_R	CGTAAACTAGTCAGATCTACCATGGATATTAACACCTTGAATTGTGAATAAC
10. MaCHUP1_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCAGAGGCTAACAACGGATAAATC
11. MaCHUP1_Gibson_R	CGTAAACTAGTCAGATCTACCATGGATATTAACACCTTGAATTATGAATAAC
12. MmCRB_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCGCGATATTGGGCTTTTGTG
13. MaCRB_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCCATGGCGTTGCTTATGGG
14. CRB_univ_Gibson_R	CGTAAACTAGTCAGATCTACCATGGTGCATCATCATCTTCCG
15. MmATPQ_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCGTTCCCTGTTTCAGCTTCTTGAG
16. MaATPQ_Gibson_F	CGGCGCGCCGAATCCCAGGGATCCGTTTCATGTTTGTGGATCACAAAG
17. ATPQ_univ_Gibson_R	CGTAAACTAGTCAGATCTACCATGGCCGCCACTTTCTTACCCG
18. p207_F	TGCGGTTAACGCTAGCATGGATCTC
19. p207_R	GTAACATCAGAGATTTTGAGACAC
20. pGWB3_F	GCCTGCAGGTCGACTCTAAT
21. pGWB3_R	GGTTGGGTTTCTACAGGAC
22. pC1381_F	CGTGCTCCACCATGTTGG
23. pC1381_R	CTGCATCGGCGAACTGATC
24. pC1381_new_F	CCACCATGTTATCACATCAA
25. pC1381_new_R	CCCGCATAATTACGAATATC
26. MmGLDP1_F	CATTTTCGTCCACCAAATCC
27. MaGLDP1_F	CCCAGCTCGCTTCTCAAGTA
28. MmPHOT2_R	CTCCTCAGGAAGCTCATGCT
29. MaPHOT2_R	TCTTGGAGTTGGGACTTCGT
30. MmCHUP1_F	ATTCACGAACTGGGTTTGC
31. MaCHUP1_F	CCACTTCCTCCTCCTCTCT
32. MmCRB_F	CCACTAGGGTCATGTTTATT
33. MaCRB_F	CAAGGACTGATGCATACAAA
34. ATPQ_F	CTTGCTTCTCCTCTTCTCT

Supplemental method

Pollen activity assay

The pollen viability of *M. arvensis*, *M. moricandioides* and their hybrids was observed followed modified Alexander's staining method (Alexander, 1969). The primary inflorescences with mature pollens were collected one day after flowering and then incubated in 1:50 staining solution for 5 min. The stock solution was comprised of 10 ml 96% ethanol, 1 ml 1% malachite green (w/v, in 96% ethanol), 25 ml glycerol, 5 ml 1% acid fuchsin (w/v, in dH₂O), 4 ml glacial acetic acid, and 100 ml dH₂O. The phase contrast images of dyed pollens were obtained under inverted microscopy (Eclipse Ti, Nikon).

3. Manuscript III

Comparative Transcriptome Analysis during Leaf Development on Closely Related C₃ and C₃-C₄ *Moricandia* Species

Meng-Ying Lin¹, Alisandra K Denton¹, Urte Schlüter¹, Andreas P.M. Weber¹

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

Abstract

C₃-C₄ species are considered to be naturally occurring intermediates on the evolutionary path from C₃ to C₄ photosynthesis. The photorespiratory glycine shuttle is a CO₂ concentrating mechanism that functions in a C₃-C₄ Kranz-like leaf anatomy, where organelles accumulate centripetally around the vein within bundle sheath cells. This system evolved through confinement of P-protein of glycine decarboxylase activity to the bundle sheath cells, which is predicted to be both an early evolutionary step on the path to C₄ and beneficial for engineering high-yielding crops. However, the genetic architecture of C₃-C₄ characteristics remains largely elusive. To identify genetic mechanisms governing divergences between C₃ and C₃-C₄ species, we performed a transcriptomics comparison of leaf developmental trajectories in C₃ *M. moricandioides* and C₃-C₄ *M. arvensis*. The general gene expression profiles were similar between C₃ and C₃-C₄ species. Across the leaf developmental gradient the expression of genes from the categories of photosynthesis and protein increased and decreased, respectively. Through cluster analysis, we identified that transcription factors involved in vein initiation (*PLT3*), auxin response (*ARF6*), growth regulating (*GRFs*), and leaf differentiation (*TCP4*), and chloroplast biosynthesis (*SIG6*) were more abundant in C₃-C₄ species in early stages and showed a delayed decrease in expression relative to C₃ species. Transcripts of core genes involved in photorespiration and C₄ cycle increased in C₃-C₄ species, associated with enhanced transcript level of a group of transporters. The vasculature pattern and plastid accumulation in C₃-C₄ species were observed earlier than in C₃ species, and were associated with genes functioning in vein initiation, procambium formation, xylem formation, *SCR/SHR* pathway, and vein patterning. Additionally, early organelle accumulation in C₃-C₄ BS cells was associated with increased transcripts of plastid fission genes in *Moricandia*. With this approach, we gained a deeper understanding of genetic control of early leaf development and organelle accumulation in *Moricandia* C₃-C₄ BS cells and discovered transcription factors showing potential to mediate C₃-C₄ characteristics, and we thereby contributed to understanding the development of Kranz-like anatomy.

Introduction

Photosynthetic productivity is the key factor of global challenges, such as food security and bioenergy production (Lobell and Gourdjji, 2012; Popp et al., 2014). Introducing C₄ characteristics into C₃ crops to improve the yield is one of the solutions to the yield loss caused by climate change. C₄ photosynthesis is more efficient than C₃ type under high temperature and drought because of its CO₂ concentrating mechanism, which (with a few exceptions) pumps carbon between two cell types, bundle sheath (BS) and mesophyll (M) cells. Leaves of C₄ species exhibit high vein density and a specific leaf pattern, the so called Kranz anatomy with one layer of BS containing abundant organelles and two layers of M between veins (V-BS-M-M-BS-V) (Haberlandt, 1904). The CO₂ is first fixed by phosphoenolpyruvate carboxylase (PEPC) in M cells. The produced four-carbon acid is converted to malate and/or aspartate, which then diffused to BS cells, where the released CO₂ is efficiently metabolized by ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO). C₃-C₄ species are considered as evolutionary intermediates on the trajectory from C₃ to C₄ photosynthesis based on anatomical, physiological, biochemical, and modeling evidence. These intermediate species play an important role in understanding the evolution of C₄ photosynthesis (Sage et al., 2014; Bräutigam and Gowik, 2016; Schlüter and Weber, 2016). C₃-C₄ plants have been reported in 21 plant lineages including eudicot as well as monocot species, such as *Diploaxis*, *Cleome*, *Flaveria*, *Salsoleae*, *Moricandia*, *Neurachne*, and *Panicum* (Sage et al., 2011). They possess a Kranz-like anatomy, BS cells with numerous organelles arranged centripetally around the veins. Additionally, C₃-C₄ plants have a CO₂ concentrating mechanism, the so called photorespiratory glycine shuttle, which is caused by confining the P-protein of glycine decarboxylase complex (GLDP) to the BS cells. In order to complete the photorespiratory pathway, glycine is then shuttled to BS cells, where the released CO₂ can be efficiently recaptured by numerous, adjacent chloroplasts.

In spite of the complexity of leaf anatomy and biochemistry, C₄ photosynthesis has independently evolved more than 66 times in different families from C₃ ancestors (Sage et al., 2012; Sage, 2016). Hundreds of genes have been reported to be differentially expressed in mature leaves between closely related C₃ and C₄ species in *Flaveria* and *Cleome* (Bräutigam et al., 2011; Gowik et al., 2011). Many studies targeting genetic factors of the Kranz anatomy used the C₄ monocot maize. For instance, comparing the transcriptional dynamics of ontogeny (from leaf primordia to mature leaves) of Maize foliar (Kranz) and husk (non-Kranz) leaves uncovered early leaf primordia genes associated with vein patterning (Wang et al., 2013b). Comparison of two independent C₄ lineages, *Gynandropsis gynandra* (formerly known as

Cleome gynandra) and *Zea mays*, identified a set of 18 homologous transcription factors showing the same behaviour as C₄-related genes, which could be key targets for C₄ engineering (Aubry et al., 2014). Quantifying transcripts in a set of leaf sections from base to tip of maize showed that genes for primary cell wall and basic cellular metabolism were abundant at the base of the leaf, while genes for secondary cell wall biosynthesis and C₄ photosynthesis development were abundant in the tip of the leaf (Li et al., 2010). Several studies leveraged comparative transcriptomics to investigate genetic mechanisms governing C₃ and C₄ leaf development. These studies used different strategies, such as series of developmental leaves of *G. gynandra* (C₄) and *Tarenaya hassleriana* (C₃) (Kulahoglu et al., 2014) and of *Flaveria bidentis* (C₄) and *F. robusta* (C₃) (Billakurthi et al., 2018), or a series of leaf sections covering different developmental stages within leaves of C₃ and C₄ *Flaveria* (Kümpers et al., 2017). However, the genetic mechanisms regulating C₄ Kranz anatomy remain largely unknown. Whether C₃-C₄ intermediates share similar genetic mechanisms for mediating Kranz-like anatomy with C₄ species, or implement anatomical changes in a C₃-C₄ specific manner is of special interest.

Current C₄ evolution models predicted that the glycine pump, resulting from the confinement of GLDP in BS cells, is a crucial early step during the evolution of C₄ photosynthesis (Heckmann et al., 2013; Mallmann et al., 2014; Sage et al., 2012; Williams et al., 2013; Bräutigam and Gowik, 2016). C₃-C₄ intermediates perform the glycine shuttle with a requisite of Kranz-like anatomy. One potential way to dissect the genetic control of C₃-C₄/C₄ photosynthesis features is to include C₃-C₄ intermediate species into the comparative transcriptomics. In *Flaveria* and Salsoleae, the differential gene expression analysis on mature leaves has been conducted among C₃, C₃-C₄, and C₄ species (Gowik et al., 2011; Mallmann et al., 2014; Lauterbach et al., 2017). No C₄ species was found in Brassicaceae, but several C₃-C₄ intermediate species exist mainly in *Moricandia*, which makes *Moricandia* a promising system for investigating the transition from C₃ to C₃-C₄ photosynthesis. The comparative transcriptome analysis on mature leaves of *M. arvensis* (C₃-C₄), *M. suffruticosa* (C₃-C₄) and *M. moricandioides* (C₃) revealed that transcriptional patterns showed no strong C₃-C₄ specific signature (Schlüter et al., 2017). In addition, comparison of transcriptional activity on mature leaves between C₃ and C₃-C₄ species provides limited information for understanding the establishment of C₃-C₄ leaf anatomy. Therefore, capturing the genetic dynamic during leaf development could provide us new insights into genetic control of C₃-C₄ characteristics.

In this study, we investigated the genetic mechanisms underlying C₃-C₄ characteristics in *Moricandia*. To do this we compared the transcriptomes of closely related *M. moricandioides*

V. Manuscript III

(C₃) and *M. arvensis* (C₃-C₄) during leaf development. Cluster analysis was applied to discover potential transcriptional regulators as candidate targets for C₃-C₄ engineering. Furthermore, the transcriptional comparisons elucidated differences in leaf structural development as well as chloroplast biosynthesis, division and movement between C₃ and C₃-C₄ *Moricandia* species.

Materials and Methods

Plant materials

Seeds of *Moricandia arvensis* (IPK: MOR1) and *M. moricandioides* (Botanical Garden Osnabrück: 04-0393-10-00) were surface-sterilized using chloride gas and germinated on half MS medium for one week. Then, the seedlings were transferred individually to pots containing a mixture of soil and sand at a ratio of 2:1 and grown in a climate chamber under 12 h light/12 h dark conditions with 23 °C day/20 °C night temperatures. For leaf ultrastructural analysis and for comparative transcriptomics, series of developmental leaves from leaf emergence to maturation were collected at the 6th or 7th leaves in the rosette, defined from stage 0 to 5. Leaf materials from stage 0 to 5 were the emerging leaf, the first visible leaf with leaf length of 4 mm, the leaf collected when next leaf emerges, the leaf with leaf length of 16 mm, the leaf with half size of the mature leaf, and the mature leaf, respectively.

Leaf anatomical studies

For ultrastructural analysis, sections (2 × 2 mm) from series of developmental leaves were proceeded with fixation, dehydration, and embedding. The sections were fixed with fixation buffer (2% paraformaldehyde, 2% glutaraldehyde), dehydrated by an acetone series, and embedded with an araldite series. The sections were transferred to the mold filled with fresh araldite and polymerized at 65°C for two days. Semi-thin sections in 2.5 µm thickness obtained by cutting with a glass knife were mounted on slides, stained with 1% toluidine blue at 60°C for 2 min and washed by distilled water. The leaf ultrastructure was examined under the light microscope, Zeiss Axiophot microscope (Carl Zeiss Microscopy GmbH, Göttingen, Germany).

RNA extraction, cDNA library construction, and sequencing

Total RNA of four biological replicates of leaves from each developing stage was extracted using the GeneMatrix Universal RNA purification kit (Roboklon GmbH, Berlin, Germany). Then, the RNA was treated with RNase-free DNaseI enzyme (New England Biolabs GmbH, Frankfurt am Main, Germany) for few seconds only. The quality of RNA and DNaseI treated RNA was assessed on a Bioanalyzer 2100 (Agilent, Santa Clara, USA) with an RNA Integrity Number (RIN) value greater than or equal to 8. Subsequently, cDNA libraries were prepared using 1 µg of total RNA with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, USA). The cDNA library was qualified on the Agilent Technologies 2100 Bioanalyzer to check the library quality and fragment size of the sample. RNA-Seq was performed on an

Illumina HiSeq 3000 platform at the BMFZ (Biologisch-Medizinisches Forschungszentrum) of the Heinrich-Heine University (Düsseldorf, Germany) to gain 150 bp paired-end reads. In total, we obtained RNA-Seq data with an average of 3.53 Gb per sample. The sequencing quality was examined using FastQC v.0.11.5. Quality scores across all bases were generally good but showed lower quality at the end of reads observed in few samples.

Transcriptome annotation and gene expression profiling

For transcript annotation of *M. arvensis* and *M. moricandioides*, we performed a pipeline on all samples separately and further merged them by species, following mapping with HISAT2, transcriptome assembly with StringTie. The produced transcript general feature format (gff) file and the draft genome sequence were used to generate the transcriptome sequence file (fasta). Orthologs between *M. arvensis* and *M. moricandioides* were discovered by grouping with *Moricandia* species and *Arabidopsis thaliana* using OrthoFinder v.2.3.3 (Emms and Kelly, 2019). To have an overview of gene distribution during leaf development in *M. arvensis* and *M. moricandioides*, genes were categorized on customized MapMan terms (<http://www.mapman.gabipd.org/>) (Supplemental Table 1).

The RNA-Seq reads were mapped on reference genome draft, *M. moricandioides* (unpublished data, assembled by Nils Koppers) using STAR v.2.5.2b (Dobin et al., 2013). The average mapping rate of *M. arvensis* and *M. moricandioides* were 78% and 91%, respectively. Differential expression testing between *M. arvensis* and *M. moricandioides* for each leaf developmental stages was performed with the DESeq2 tool (Love et al., 2014) in R (www.R-project.org). A Benjamini–Hochberg adjusted false discovery rate of ≤ 0.01 was used as the cut-off for significant differential expression.

Cluster analysis of transcript abundances

Clustering transcript abundances of two species during leaf development reveals gene expression patterns between species. We normalized expression values by variance stabilizing transformation. Clustering using K-means clustering algorithm attempts to classify all genes to a limited set of clusters, which raises the possibility to contain not only co-expressed, but also non co-expressed genes in the cluster. Therefore, to assign genes to clusters with expected biological properties, we applied *clust* on the dynamic gene expression changes during leaf development between *M. arvensis* and *M. moricandioides* (Abu-Jamous and Kelly, 2018).

Functional enrichment assessment and transcriptional factor discovery

The genes in 19 clusters generated from *clust* were further processed to functional category assessment using Araport11 MapMan functional classification system (from <http://mapman.gabipd.org>). The genes in each cluster were then classified into the MapMan functional categories, and the Fisher's exact test using the Benjamini–Hochberg adjusted false discovery rate ≤ 0.01 as the cut-off was applied to identify the significantly over-represented categories for each cluster. Heatmaps of annotated clusters were generated by hierarchical clustering. Transcription factors including in annotated clusters were defined based on two transcription factor databases, PlnTFDB (<http://plntfdb.bio.uni-potsdam.de>) and PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>).

Results

Comparison of anatomical features during leaf development between C₃ and C₃-C₄ *Moricandia* species

In mature leaves, organelles were found in both BS and M cells along the inner tangential walls in C₃ and C₃-C₄ *Moricandia* species. In C₃-C₄ *Moricandia* species, numerous organelles in the BS cells were additionally accumulated centripetally toward veins, the so called Kranz-like anatomy (Beebe and Evert, 1990; Schlüter et al., 2017). It has been shown that the leaf anatomy of C₃-C₄ *M. arvensis* was initially C₃-like, no difference in plastid number and area between BS and M cells, whereas 4-fold increase of plastid number was observed in BS cells at the leaf length between 6 and 12 mm in C₃-C₄ species (Rylott et al., 1998). However, our knowledge on the genetic mechanism behind the leaf anatomical development of C₃-C₄ Kranz-like anatomy is still limited. In order to unravel the photosynthetic development between C₃ and C₃-C₄ *Moricandia* species, a series of leaf development stages was defined and collected from leaf emergence to maturation: stage 0, the emerging leaf; stage 1, the first visible leaf with leaf length of 4 mm; stage 2, the leaf collected when next leaf emerges; stage 3, the leaf with leaf length of 16 mm; stage 4, the leaf with half size of the mature leaf; stage 5, the mature leaf. Leaf materials were investigated through cross sections (Figure 1). In both species, M cells of stage 0 leaves were undifferentiated, and the palisade parenchyma layers were visible in stage 1. In addition, the vasculature pattern in C₃-C₄ leaves was more visible since stage 0 than in C₃ leaves. The organelle accumulation towards the vein could be observed in C₃-C₄ BS cells in stage 1, and the number of organelles kept increasing till the mature stage; on the contrary, in stage 2 of C₃ leaves, plastids distributed along the BS cell wall, and no strong accumulation was observed. At the mature stage, C₃-C₄ species had two layers of palisade parenchyma, whereas C₃ species had three palisade parenchyma layers. The mature leaves of C₃-C₄ species were thinner than those of C₃ species.

Taken together, the vasculature patterns and plastid accumulation were observed earlier in C₃-C₄ species than in C₃ species, and the centripetal accumulation of organelles in BS cells towards the vein was only observed in C₃-C₄ species, starting from stage 1. Mature leaves of C₃-C₄ *M. arvensis* were thinner and possessed one layer fewer of palisade parenchyma than those of C₃ *M. moricandioides*.

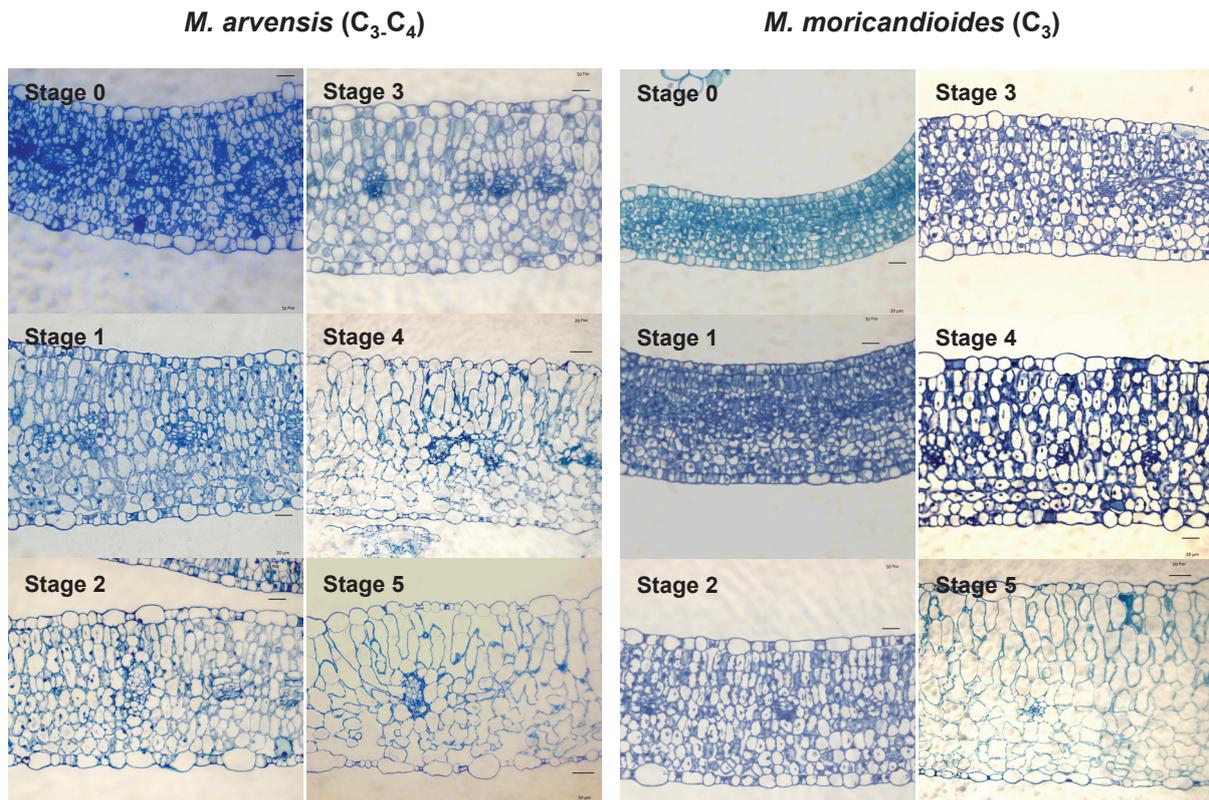


Figure 1. Overview of cross sections during leaf development in *M. moricandioides* and *M. arvensis*.

Transcriptomes of C_3 and C_3 - C_4 *Moricandia* species were comparable and showed similar developmental patterns

Mapping RNA-Seq reads against closely related species in the same genus provides a more informative data set than using species from different genera. The C_3 *M. moricandioides* genome assembly possesses a higher quality than the genome of C_3 - C_4 *M. arvensis* in that it has a more reliable amount of repetitive elements. Therefore, all the RNA-Seq reads in this study were mapped against the preliminary assembly of C_3 *M. moricandioides* using STAR v.2.5.2b. The average mapping rates of *M. arvensis* and *M. moricandioides* reads on *M. moricandioides* preliminary genome were 78% and 91%, respectively, which are higher than in previous work, mapping *Moricandia* RNA-Seq reads against the transcriptome of *Arabidopsis thaliana* (Schlüter et al., 2017). The principle component analysis (PCA) showed that the first PC (PC1) explained 62% of the total variance and separated the two species, C_3 - C_4 and C_3 *Moricandia*. PC2 explained 21% of the total variance and separated leaf developmental stages in sorted order from immature to mature leaves (Figure 2). The functional categories with total gene expression revealed the conservation between

transcriptomes of C₃-C₄ and C₃ *Moricandia* (Figure 3). During leaf ontogeny of both species, gene expression in the photosynthesis category increased; on the other hand, gene expression in the protein category decreased. Hierarchical clustering of C₃-C₄ and C₃ *Moricandia* revealed that not only cohorts of genes showed the same expression patterns along the leaf gradient with or without differences in expression level, but also modules with gene upregulation in C₃-C₄ and C₃ *Moricandia*, respectively, were observed (Figure 4). In summary, transcriptomes during leaf ontology of C₃-C₄ and C₃ *Moricandia* species were comparable and well paired, which also demonstrated groups of genes differing in expression pattern and level between these two species. These transcriptomes were then further introduced to clustering and differential gene expression analysis.

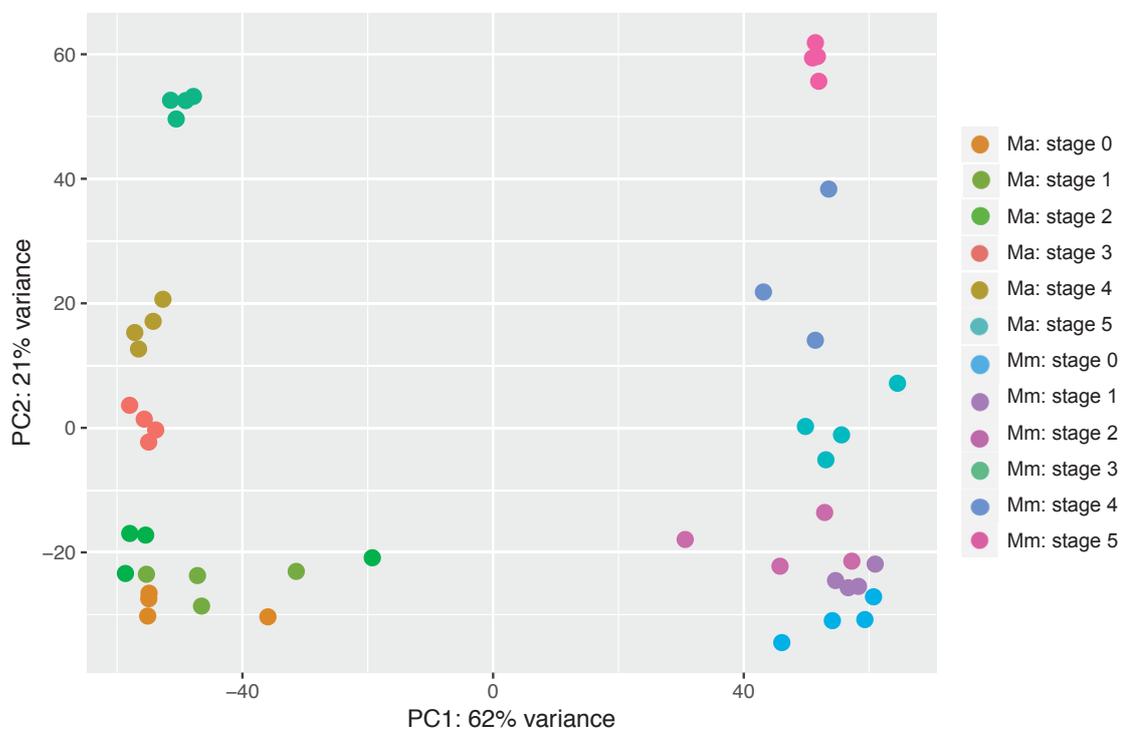


Figure 2. Principle component analysis (PCA) on leaf development series of *M. moricandioides* (Mm) and *M. arvensis* (Ma).

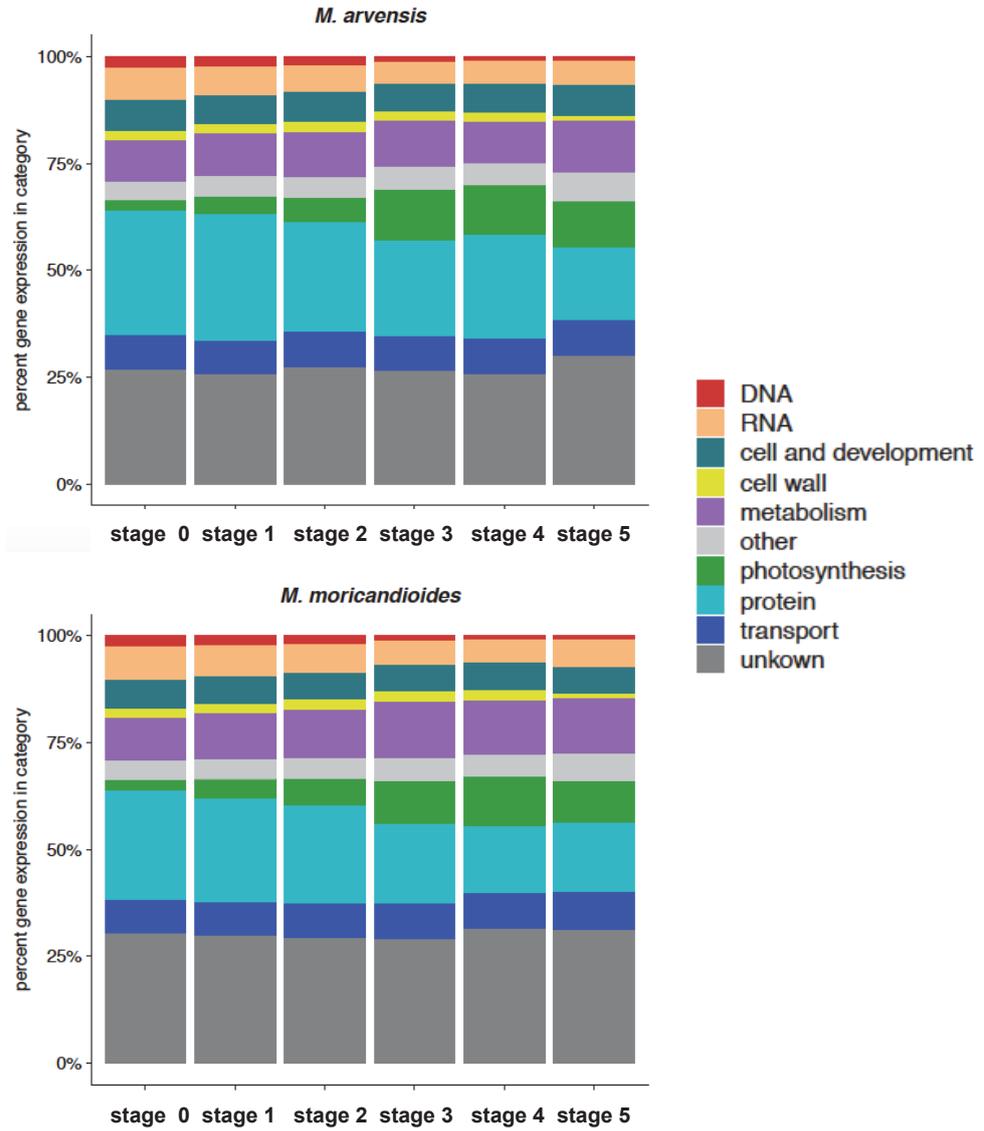


Figure 3. Gene distribution during leaf development in *M. arvensis* and *M. moricandioides*.

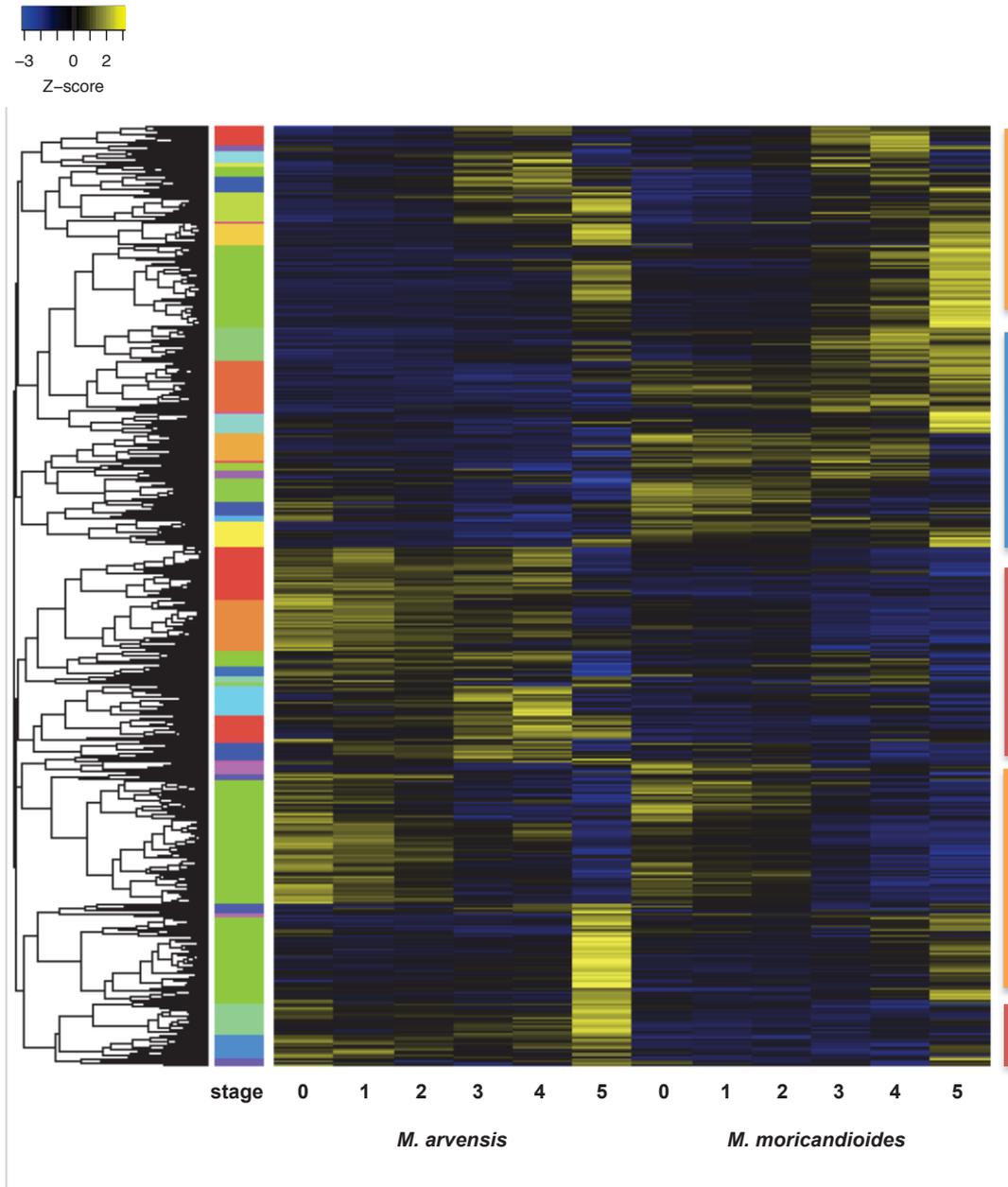


Figure 4. Pearson's correlation hierarchical cluster of Z-scores of all genes during leaf development.

Red bar, upregulation in *M. arvensis*; blue bar, upregulation in *M. moricandioides*; orange bar, genes showing the same expression pattern during leaf ontology with or without differences in expression level between *M. arvensis* and *M. moricandioides*.

Clustering of transcriptional changes during leaf development

C_3 - C_4 *M. arvensis* demonstrated earlier development of vasculature pattern and organelle accumulation relative to C_3 *M. moricandioides* (Figure 1). Together with the observation in hierarchical clustering of all expressed genes, groups of transcripts showed differences in expression pattern and level between these two species (Figure 4), we hypothesized an association between cohorts of transcripts and phenotypic differences between C_3 - C_4 and C_3 *Moricandia* species. To capture the dynamic gene expression changes regarding these phenotypic differences between C_3 - C_4 and C_3 *Moricandia* species, genes were clustered with *clust* (Abu-Jamous and Kelly, 2018). Of 19 clusters, five were divergent in expression pattern between the two species (cluster 7, 10, 11, 18, and 19), whereas patterns of remaining clusters were similar between C_3 - C_4 and C_3 *Moricandia* species (Supplemental Figure 1). Six clusters showed descending pattern during leaf development (cluster 1, 2, 3, 4, 5, and 16), while five clusters were ascending clusters (cluster 8, 9, 10, 14, and 15). With MapMan category enrichment analysis, ten clusters revealed specific functional patterns (Figure 5). Generally, genes involved in protein biosynthesis and translocation were found in descending clusters (cluster 1, 2, 12, and 17), and photosynthesis related genes were found in ascending clusters (cluster 8 and 14).

Transcripts of genes in descending cluster 1 were more abundant in C_3 - C_4 relative to C_3 *Moricandia* species, and showed a delayed decrease during leaf development in C_3 - C_4 species (Figure 5). Genes in cluster 1 were enriched in functional categories, such as protein biosynthesis, cytoskeleton, nucleotide metabolism, amino acid metabolism, cell cycle, chromatin organization, RNA processing, and protein translocation. On the other hand, transcripts of genes in descending cluster 4 were less abundant in C_3 - C_4 relative to C_3 *Moricandia* species, and showed earlier downregulation during leaf development in C_3 - C_4 species (Figure 5). Genes in cluster 4 were enriched in the functions of cell wall and cytoskeleton.

The expression of genes in ascending cluster 8 was lower and demonstrated a delayed increase in C_3 - C_4 species. Genes in cluster 8 were enriched in solute transport, protein degradation, vesicle trafficking, phytochromes, photosynthesis, and coenzyme metabolism. In contrast, transcripts of genes in ascending cluster 14, which were more abundant and showed a delayed increase in C_3 - C_4 species, were enriched in solute transport, nutrient uptake, and photosynthesis functions.

Genes in cluster 15 showed higher transcript abundance in C_3 - C_4 species with a peak between stage 3 and stage 4, and were enriched in the functions of cell wall, carbohydrate metabolism,

solute transport, and cellular respiration. In cluster 5, genes showed lower expression in C₃-C₄ species with a descending expression pattern, however gene expression in mature leaves was higher than in stage 4. Genes in cluster 5 enriched in functions of RNA biosynthesis, RNA processing, and chromatin organization. Genes in slightly descending cluster 12 showed enhanced transcript abundance in all leaf developmental stages in C₃-C₄ species. MapMan categories, such as protein biosynthesis, RNA processing, vesicle trafficking, and cellular respiration, were enriched in cluster 12 genes.

The transcriptional patterns of genes in descending cluster 2 were similar in C₃-C₄ and C₃ species; these genes were enriched in functions such as cell cycle, RNA processing, protein translocation. In cluster 16, genes showed a transcriptional activity drop in stage 4 in both species, and were enriched in the function of cell wall. Genes in cluster 17, with a transcriptional peak between stage 3 and stage 4, were enriched in the functions of protein biosynthesis and coenzyme metabolism.

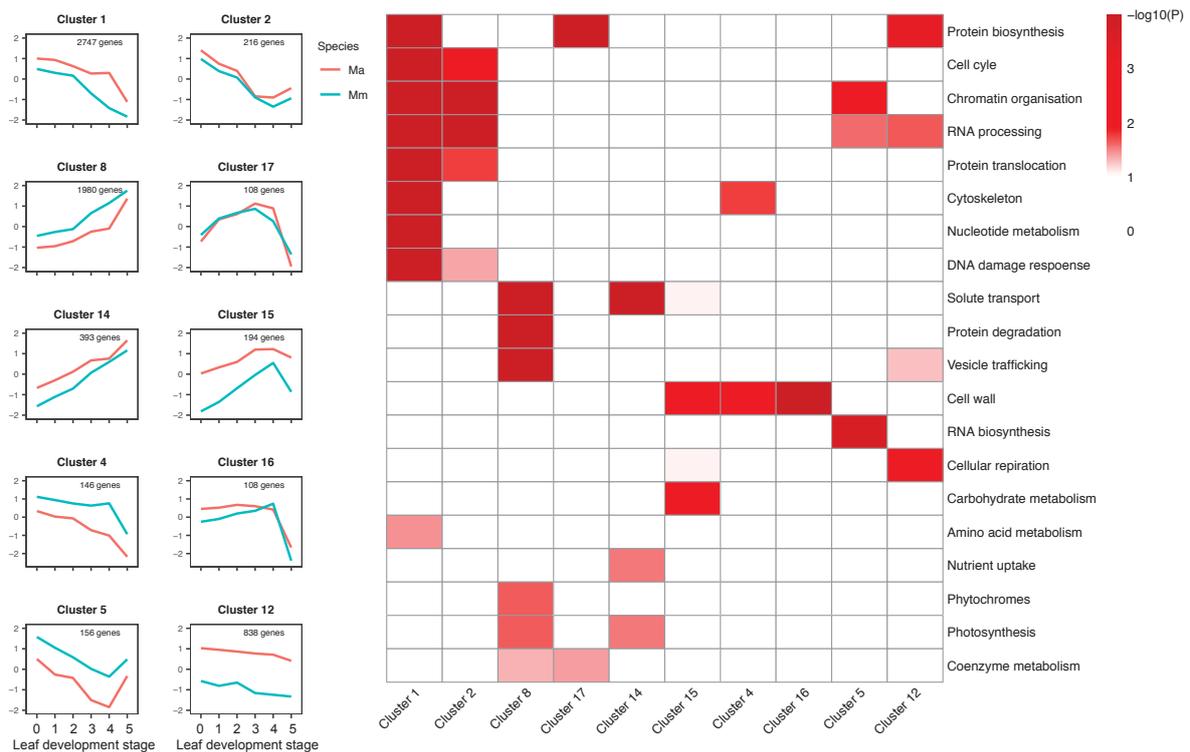


Figure 5. Clustering pattern and Mapman category enrichment analysis of selected clusters.

Ma, *M. arvensis*; Mm, *M. moricandioides*.

Discovery of transcription factors governing C₃-C₄ characteristics

A few transcription factors are known to regulate genes involved in either photosynthesis or leaf development in maize, rice, sorghum, *Flaveria*, and *Cleome*, such as *GOLDEN2-LIKE* (*GLK*) genes and *auxin response factor5* (*ARF5*) (Scarpella et al., 2006; Wang et al., 2017; Waters et al., 2009; Wenzel et al., 2007). To unravel potential molecular mechanisms governing C₃-C₄ characteristics in *Moricandia*, we investigated the transcript activities of known transcription factors of *Arabidopsis* participating in leaf development and organelle formation in *Moricandia*, as well as novel transcription factors derived from cluster analysis. Transcription factors governing leaf development and regulation of cell-cycle leaf development, such as gene-regulating factor (*GRF1*, *GRF2*, *GRF4*, *GRF5*, and *GRF7*), *AP2-EREBP* transcription factors (*PLT3* and *ANT*), and homeobox-leucine zipper family protein (adaxial/abaxial pattern specification, *ATHB9* and *ATHB14*), were found in descending clusters, either cluster 1 or 2 (Supplemental Table 2 and 3). Genes from cluster 1 showed an increased abundance and a delayed decrease in C₃-C₄ relative to C₃ species; however there was no difference in gene expression between two species in cluster 2 genes. *PLETHORA 3* (*PLT3*) belonged to cluster 1, and participated in auxin mediated signaling pathway through maintaining high expression level of an auxin transporter *PINI*, which functions in vein initiation (Prasad et al., 2011). The *TEOSINTE BRANCHED 1*, *CYCLOIDEA*, *PROLIFERATING CELL FACTOR 4* (*TCP4*) belonged to cluster 1, and promoted the onset of leaf differentiation (Sarvepalli and Nath, 2011). Auxin response factors (ARFs) are transcription factors known to mediate distinct growth and developmental processes, such as early embryogenesis, root formation and hypophysis specification (Rademacher et al., 2012). *ARF5/MP* (*MSTRG.32093*) transcription factor with functions in leaf initiation through the upregulation of *PINI* was abundant in C₃-C₄ species in stage 0. The role of *ARF6* in flower development has been described in *Arabidopsis* and tomato, however *ARF6* might also regulate vegetative development (Liu et al., 2014). *ARF6* (*MSTRG.13343*) was assigned to cluster 1, and was more abundant in C₃-C₄ than C₃ species in the leaf gradient, suggesting the role in auxin mediated vein development in early leaf developing stages.

Of 162 transcription factors in ascending cluster 8, where transcript abundance was increased in C₃ species, we found ARFs (*ARF2*, *ARF7*, and *ARF16*), G2-like transcription factors (*GLK1* and *GLK2*), which have a great impact on photosynthesis genes (Chen et al., 2016), and four sigma70-like transcription factors (Supplemental Table 4). It has been shown that *ARF2* regulated cell division and cell expansion, resulting in enlarged leaf size in *Arabidopsis* (Gonzalez et al., 2010; Okushima et al., 2005). The enhanced abundance of *ARF2*,

corresponded to the larger leaf size of C₃ relative to C₃-C₄ *Moricandia* species (Figure 1; Supplemental Table 4). The other ARF genes (*ARF7* and *ARF16*) found in ascending cluster 8 were also assumed to play an important role in regulating different signalling pathways involved diverse developmental processes. *ZmGlk1* and *ZmG2* regulate chloroplast organization in M and BS cells, respectively, in maize; however the *GLK* orthologs in C₃ species, rice (*OsGLK1* and *OsGLK2*) and Arabidopsis (*AtGLK1* and *AtGLK2*) are functionally redundant (Fitter et al., 2002; Rossini et al., 2001). In *Moricandia*, *GLK1* (*MSTRG.21425* and *MSTRG.10931*) and *GLK2* (*MSTRG.25869*) showed enhanced transcript abundance in late developmental stages of C₃ *Moricandia* species. Sigma-like factors are proteins in association with sigma factor activity belonging to the superfamily of sigmaA or sigma70, which regulate plastid gene expression (Chi et al., 2015). *SIG3*, *SIG5*, and *SIGA* were more abundant in C₃ *Moricandia* species relative to C₃-C₄ species, and belonged to the ascending cluster 8, where genes enriched in the function of photosynthesis. Sigma70-like factor, *SIG1*, belonging to cluster 17, showed no transcriptional difference between C₃-C₄ and C₃ *Moricandia* species. Another known sigma70-like factor, *SIG6*, mediates the early chloroplast biosynthesis (Ishizaki et al., 2005), which was more abundant in C₃-C₄ *Moricandia* species from stage 0 to stage 4.

Taken together, earlier leaf development in C₃-C₄ *Moricandia* species, including leaf differentiation and organelle biosynthesis, was associated with increased and long-lasting abundance of transcription factors in early to mid development stages, such as *GRFs*, *PLT3*, *TCP4*, *ARF6*, and *SIG6*.

Enhanced transcriptional activities in photorespiration and C₄ cycle candidates in *Moricandia* C₃-C₄ species

Genes in ascending clusters, cluster 8 and cluster 14, were enriched in the MapMan category, photosynthesis, including subcategories, such as light reaction, Calvin-Benson cycle, and photorespiration (Figure 5). *Moricandia* C₃-C₄ species possess a carbon-concentrating mechanism, the photorespiratory glycine shuttle, which results in efficient recapture of photorespiratory released CO₂ in BS cells. Additionally, non-photosynthetic isoforms of C₄ cycle genes are expressed not only in non-photosynthetic tissues of C₄ species, but also in C₃ species, which are considered as the starting point for the C₄ gene compartmentalization (Monson, 1999). However, there is no full C₄ species in the genus *Moricandia*. Thus, we were interested in which genetic mechanisms underlying C₃-C₄ evolution were common in independent C₃ lineages and which were specific to *Moricandia*. Previous work found no

strong C₃-C₄ signature in the transcriptional changes of photorespiratory and C₄ cycle genes in mature leaves between C₃ *M. moricandioides* and C₃-C₄ *Moricandia* intermediates (*M. arvensis* and *M. arvensis* and *M. suffruticosa*) (Schlüter et al., 2017). Only *AspAT*, *PEPC2*, *PPDK*, and *PEPCK* showed enhanced abundance in both C₃-C₄ *Moricandia* intermediates (Schlüter et al., 2017). In order to understand the genetic dynamic of photorespiratory genes between *Moricandia* C₃ and C₃-C₄ species as well as the role of C₄ genes in *Moricandia* C₃-C₄ species, the transcriptional dynamic of genes involving in photorespiration and C₄ cycle were compared during leaf ontology (Supplemental Table 5). In general, transcripts of photorespiratory genes were more abundant in C₃-C₄ *M. arvensis* during middle to mature stage, such as genes in glutamine synthetase-glutamate synthase pathway (*GS2* and *Fd-GOGAT1*), *GDC/SHMT* system (*mLPD1*, *GLDT*, *GLDH3*, and *SHMT1*), hydroxypyruvate reductase (*HPR1* and *HPR2*), and plastidic glycolate glycerate transporter (*PLGG1*) (Figure 6A). *SHMT2*, involved in glycine biosynthesis process, showed higher transcriptional abundance in C₃-C₄ species in the early leaf developmental stages. Interestingly, glycerate kinase (*GLYK*), which catalyzes the regeneration of 3-phosphoglycerate, showed lower expression during late stages in C₃-C₄ *M. arvensis* relative to C₃ *M. moricandioides*. Transcripts of key C₄ cycle genes were more abundant in C₃-C₄ *M. arvensis* either in mid-late or late stages, except *alpha-CA1s* (*MSTRG.4736* and *MSTRG.23861*), *DiT2s* (*MSTRG.32249* and *MSTRG.16268*), *AlaAT1s* (*MSTRG.32021* and *MSTRG.7554*), *beta-CA5* (*MSTRG.24856*) (Figure 6B). Increased transcript abundance of *PEPC1*, *AspAT*, decarboxylation genes (*NAD-ME*, *NADP-ME*, and *PEPCK*), pyruvate transporters (*BASS2/NHD1*), and *PPDK* was observed in mature leaves in C₃-C₄ species. *PEPC2*, *AlaAT*, *BASS4*, triose phosphate translocator *TPT*, and plastidic oxaloacetate/malate transporter *DiT1* revealed enhanced abundance in stage 3 and stage 4 in C₃-C₄ species.

All in all, the transcript abundance of key genes in photorespiration and C₄ cycles was enhanced in C₃-C₄ *Moricandia* species during leaf development compared to the C₃ species.

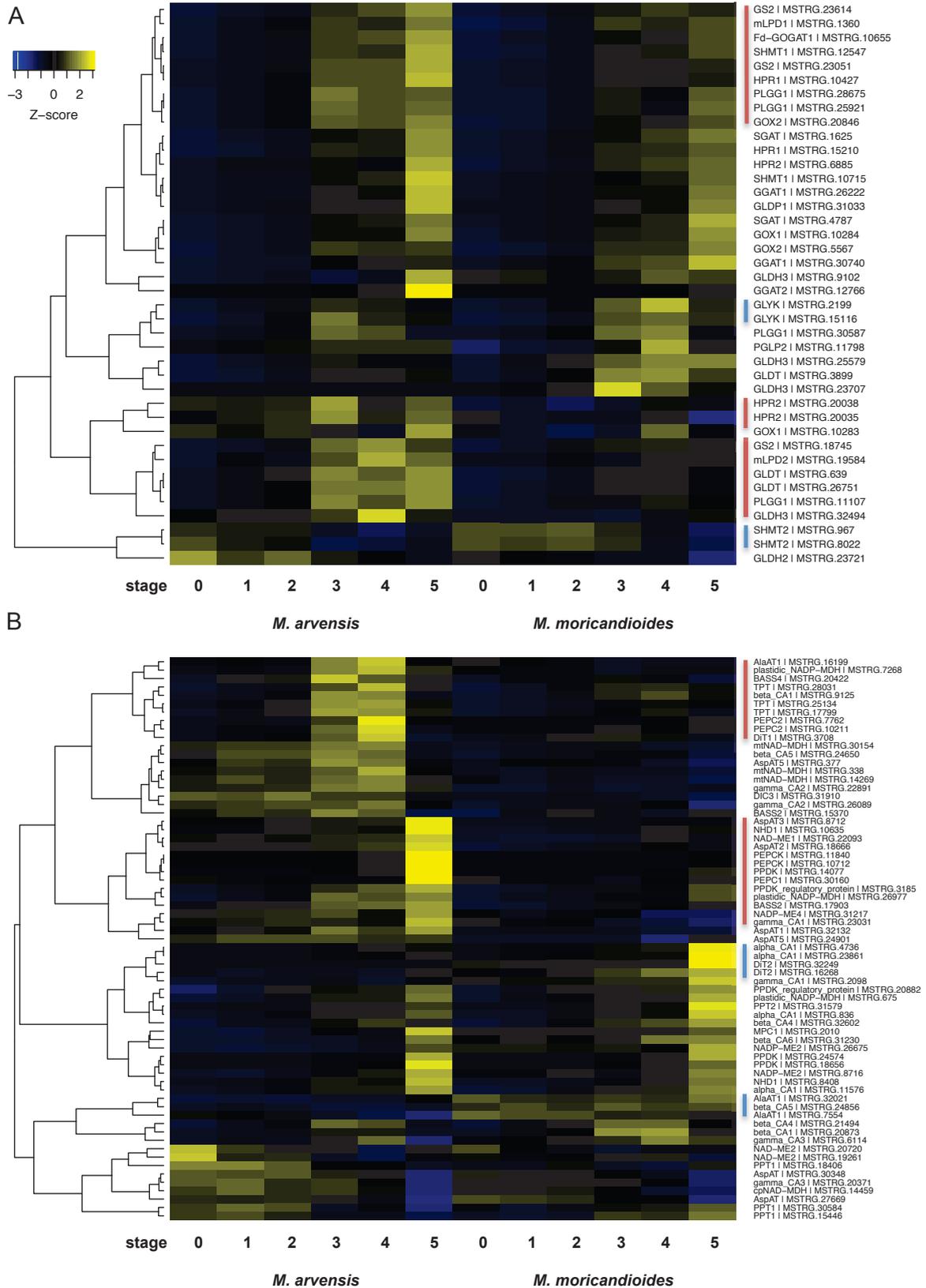


Figure 6. Pearson's correlation hierarchical cluster of Z-scores of genes involved in photorespiration (A) and C₄ cycle (B).

Red bar, upregulation in *M. arvensis*; blue bar, upregulation in *M. moricandioides*.

Numerous chloroplasts accumulated in C₃-C₄ BS cells was associated with chloroplast fission in early stages of leaf development

In descending cluster 1, transcript abundance of *M. arvensis* genes was higher and demonstrated a delayed decrease relative to that of *M. moricandioides*, where we found genes enriched in functional categories, cell cycle and cytoskeleton, including subcategories, such as organelle machinery, organelle fission, and plastid division (Figure 5). *FtsZ1* and *FtsZ2*, which regulate the formation of the inner FtsZ division ring in organelle division, were found in cluster 1. In addition, centripetally accumulated chloroplasts were observed in C₃-C₄ BS cells from stage 2 and kept accumulating till mature stage (Figure 1). The abundance and centripetal position of organelles in C₃-C₄ BS cells are especially important C₃-C₄ characteristic for efficient recapture of CO₂ released during photorespiration. Therefore, we examined the dynamic gene expression of *Moricandia* orthologs of known genes in Arabidopsis involving in chloroplast development, division, and movement during leaf development (Supplemental Table 6). *FtsZ1* and *FtsZ2* showed increased transcript abundance starting from stage 0 and lasted till mature stage in C₃-C₄ *Moricandia* species (Figure 7A). The transcript abundance was enhanced in stage 3 and stage 4 in C₃-C₄ *Moricandia* species: those involving in the placement of the FtsZ ring (*MinD1* and *ARC3*), upstream regulator of outer ring positioning (*ARC6*), chloroplast development (*GNC* and *CGAI*), and chlorophyll biosynthesis (*PORB* and *HEMA1*). Chloroplast organization genes, *GLK1* (*MSTRG.21425* and *MSTRG.10931*) and *GLK2* (*MSTRG.25869*), showed enhanced transcript abundance in late development stages of C₃ *Moricandia* species. Most genes involved in chloroplast movement showed no significant difference between C₃ and C₃-C₄ *Moricandia* species. Only, *PMIR1* (involved in chloroplast relocation) and *WEB1* genes (regulating the velocity of chloroplast photorelocation movement) showed increased transcript abundance in mature leaves in C₃-C₄ *Moricandia* species. The phototropism gene, *PHOT1*, showed increased transcript abundance in mid to late stages in C₃-C₄ *Moricandia* species.

Taken together, transcriptional comparison of chloroplast development, division, and movement during leaf ontology of C₃ and C₃-C₄ *Moricandia* species revealed a connection of chloroplast division genes, *FtsZ1* and *FtsZ2*, and early chloroplast proliferation in C₃-C₄ BS cells.

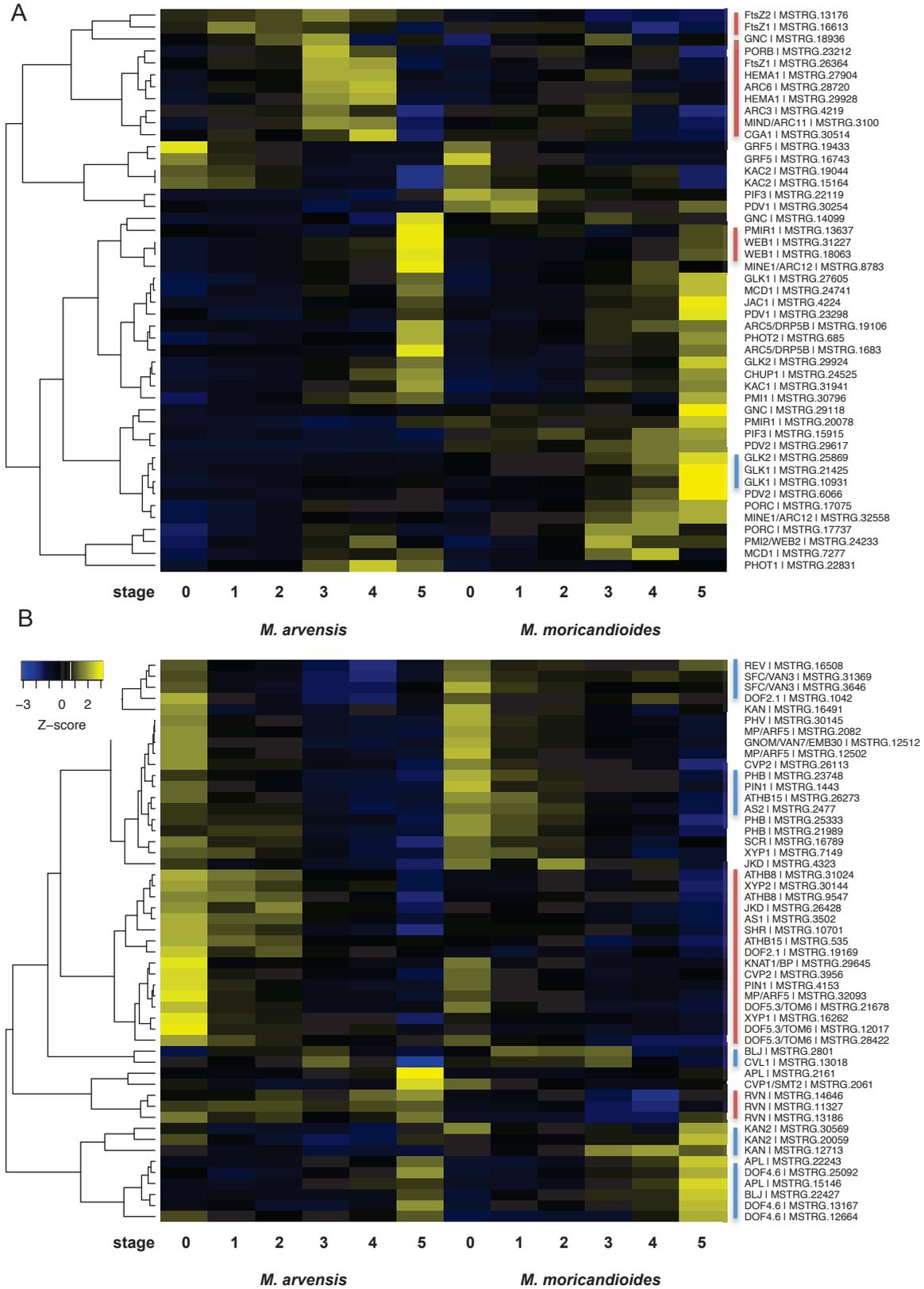


Figure 7. Pearson's correlation hierarchical cluster of Z-scores of genes involved in chloroplast development/division/movement (A) and leaf development (B).

Red bar, upregulation in *M. arvensis*; blue bar, upregulation in *M. moricandioides*.

Transcriptional changes in leaf development genes were associated with early vein development in the C₃-C₄ *M. arvensis*

Of 213 transcription factors in descending cluster 1 and cluster 2, many are known to be involved in auxin mediated vein development, leaf development and regulation of cell-cycle leaf development (*GRFs*, *ARFs*, *AP2-EREBPs*, and *HD-ZIPs*) (Supplemental Table 2 and 3). *TCP4* involved in heterochronic regulation of leaf differentiation was found in cluster 1. The leaf vascular development became visible and clear earlier in *M. arvensis* (C₃-C₄) than in *M. moricandioides* (C₃) (Figure 1). The vasculature pattern could be identified in the smallest investigated stage of C₃-C₄ *M. arvensis* but not C₃ *M. moricandioides*. We hypothesized this early leaf development in *M. arvensis* is associated with altered gene expression of leaf and vasculature developmental genes during leaf ontology. Therefore, we focused on the dynamic transcript abundance of auxin mediated vein initiation, vein patterning, and *SCARECROW/SHORT ROOT* (*SCR/SHR*) pathway between *M. arvensis* and *M. moricandioides* during leaf ontology (Supplemental Table 7). A number of genes revealed enhanced transcript abundance in early stages in C₃-C₄ *Moricandia* species: vein initiation and procambium formation (*MP/ARF5*, *CVP2*, *ASI*), xylem formation (*XYP1*, *XYP2*, *ATHB8*, and *ATHB15*), *SCR/SHR* pathway (*SHR*, *JKD*), and vein patterning (*DOF2.1* and *DOF5.3*) (Figure 7B). Phloem differentiation gene (*APL*, *MSTRG.2161*) transcription was more abundant in mature C₃-C₄ *Moricandia* leaves. *RVN*s (*MSTRG.14646* and *MSTRG.11327*) involved in *SCR/SHR* pathway (BS cell development) were abundant during all leaf development stages in C₃-C₄ *Moricandia*. Transcript abundances of *KAN*s, suppressing xylem differentiation, showed reduced abundance in middle to mature stages in C₃-C₄ *Moricandia* species. In C₃ *M. moricandioides*, a delay of leaf vasculature pattern formation was observed (Figure 1), which was correlated with later transcript abundance relative to C₃-C₄ species, such as *REV* (xylem maintaining), *SFC* (vein pattern formation), *APL* (phloem differentiation) and *DOF4.6* (vein patterning) (Figure 7A). Additionally, *BLJ1* and *CVL1*, which are involved in *SHR* pathway and establishment of foliar vein patterns, respectively, were upregulated from stage 1 to stage 3 in C₃ *M. moricandioides*. To summarize, the early leaf vascular development in C₃-C₄ *M. arvensis* relative to that in C₃ *M. moricandioides* corresponded with early transcript abundance of a cohort of genes known to participate in vein initiation, formation, *SCR/SHR* pathway, and vein patterning. It also implies that the development of BS cells controlled through *SCR/SHR* pathway was also upregulated in the early stage of C₃-C₄ *M. arvensis* leaf development, in which especially the transcriptional abundance of *RVN* showed enhanced abundance through all stages of leaf development in C₃-C₄ *M. arvensis*.

Discussion

Consideration of comparative transcriptome during leaf development is necessary for capturing dynamic genetic control between C₃ and C₃-C₄ *Moricandia* species

In this study, we investigated the genetic control of C₃-C₄ characteristics through comparative transcriptomics during leaf development of *M. arvensis* (C₃-C₄) and *M. moricandioides* (C₃). We utilized *M. moricandioides* preliminary assembly as reference genome for RNA-Seq read mapping and verified the comparability of the two *Moricandia* transcriptomes through gene distribution during leaf ontology (Figure 3). It showed that the transcript activities of photosynthesis and protein related genes increased and decreased from immature to mature leaves, respectively. This result was consistent with the comparative transcriptome between C₃ and C₄ species in *Cleome* and *Flaveria* (Külahoglu et al., 2014; Kümpers et al., 2017).

It has been shown that the abundances of 874 transcripts were commonly reduced and 797 transcripts commonly enhanced in C₃-C₄ *Moricandia* species (*M. arvensis* and *M. suffruticosa*) relative to C₃ *M. moricandioides*, based on comparative transcriptome analysis on mature leaves between *Moricandia* C₃ and C₃-C₄ species (Schlüter et al., 2017). In our study, the same pattern—where more transcripts were transcriptionally reduced than increased in C₃-C₄ species relative to C₃ species in *Moricandia*—was observed not only in mature leaves, but also in all leaf developing stages (Supplemental Table 8). Interestingly, in stage 4, the number of significantly differentially expressed genes was the greatest during leaf development, implying the importance of considering comparative transcriptome during leaf development. That might be a reason that no strong C₃-C₄ signature was discovered in photorespiration pathway and Calvin-Benson cycle in previous study (Schlüter et al., 2017). The transcripts of C₄ cycle and photorespiration genes were increased during leaf development in C₃-C₄ *Moricandia* species compared to C₃ species. Similar results have been reported in comparative transcriptome analysis on mature leaves between C₃ and C₃-C₄ species in *Flaveria* and Salsoleae (Gowik et al., 2011; Lauterbach et al., 2017).

Implementation of C₄-like features in *Moricandia* C₃-C₄ species

There is no C₄ species in the genus *Moricandia*, however most of key genes in C₄ cycle were upregulated during mid-late leaf developing stages in C₃-C₄ species compared with C₃ species. Not only NADP-ME type C₄ genes (*plastidic-NADP-MDH* and *NADP-ME4*), but also NAD-ME type (*mtNAD-MDH* and *NAD-ME1*) and PEPCK type (*PEPCK*) increased in expression in C₃-C₄ species, indicating the intermediate state of *M. arvensis* as well as a certain extent that C₄ cycle works in C₃-C₄ *M. arvensis*. In the case of *F. ramosissima* (C₃-C₄),

only NADP-ME type and typical NAD-ME type C_4 genes were enhanced compared with the C_3 plant *F. robusta* (Gowik et al., 2011). In *Salsola dicaricata* (C_3 - C_4), transcripts of *NADP-ME*, *NAD-MDH*, and *PEPCK* significantly increased relative to *S. webbia* (C_3) (Lauterbach et al., 2017). Additionally, C_4 species require high metabolite transport rates across the chloroplast membrane. In our study, transcripts of C_4 -related transporters, including pyruvate transporters (*BASS2*, *NHD1*), dicarboxylate transporter (*DiT1*), and another gene belonging to the family of bile acid sodium symporter (*BASS4*), increased in C_3 - C_4 *M. arvensis*. Except transcript of *DiT2* decreased in C_3 - C_4 *M. arvensis*, the enhanced transcript level of C_4 -related transporters was consistent with previous study in *Flaveria* (Gowik et al., 2011). The enhanced transcript level of *BASS4* and *NHD2* was also found in C_3 - C_4 *Salsola* species (Lauterbach et al., 2017).

Enhanced transcript level of solute transporters was associated with increased transcript of photorespiratory and C_4 cycle genes in *M. arvensis*

C_3 - C_4 intermediates are characterized by much more labor division between M and BS, requiring more metabolite transport within the leaf. For instance, NH_3 and serine are released from GDC/SHMT coupled system in the mitochondria of BS cells, and then need to be shuttled to the M cells and imported into the chloroplasts and peroxisomes, respectively. We hypothesized a connection of higher demand and higher expression of transporters in C_3 - C_4 *M. arvensis*. On the regulatory level the stronger differentiation of both cell types will also demand differences in the regulation of these processes. Therefore, we looked for differential expression in transporters between C_3 and C_3 - C_4 *Moricandia* species.

In cluster 14 and cluster 15, we found a group of solute transporters, showing higher transcript level in C_3 - C_4 *M. arvensis* (Figure 5; Supplemental table 2). Transcripts of genes in ascending cluster 14 were upregulated in stages 0 to 3 in C_3 - C_4 *M. arvensis*. In addition to two PLGG1 glycerate:glycolate transporters and one NHD proton:sodium cation transporter, there were amino acid transporters, sulfate transporters, an auxin transporter, a tonoplast dicarboxylate transporter TDT, NRT1/PTR anion transporters, and monosaccharide transporters found in cluster 14 (Supplemental Table 9). It has been shown that chloroplast-located sulfate transporters facilitate the integration of photorespiration and sulfate metabolism (Eisenhut et al., 2015). In addition, C_3 - C_4 glycine shuttle results in serine production in mitochondria in BS cells, which needs to be shuttled back to the peroxisome in M cells. Therefore, the enhanced transcript level of sulfate transporters was consistent with increased transcripts of photorespiratory genes in C_3 - C_4 *Moricandia* as well as suggested candidate transporters for

sulfate. We also found a *PIN*-like auxin efflux carrier family protein (MSTRG.9764), which might be associated with the vein development in C_3 - C_4 *M. arvensis*. Enhanced transcript level of *TDT* in C_3 - C_4 *Moricandia* suggested that they produced more malate, consistent with increased transcripts of C_4 cycle genes, such as *NADP-MDH* and *NAD-MDH*, and parts of malate were subsequently stored in the vacuole, which was considered essential for crassulacean acid metabolism and stomatal opening (Emmerlich et al., 2015). In addition, enhanced transcript level of monosaccharide transporters in C_3 - C_4 *M. arvensis* might be associated with the higher rate of recapture of photorespiration released CO_2 relative to C_3 plants. Genes in ascending cluster 15 showed increased transcripts with a transcriptional activity peak between in C_3 - C_4 *M. arvensis*. Of 16 transporters in cluster 15, BASS2, MC-type solute transporters, NRT1/PTR anion transporters, metabolite transporter belonging to MATE family, and subunits of vascular-type H^+ (V-type) ATPase complex were discovered (Supplemental Table 9). NRT1/PTR anion transporters are nitrate transporters known to regulate plant growth and development, such as lateral root development and leaf growth (Fan et al., 2017). The possible substrates for NRT1/PTR family transporters are not only nitrate, but also phytohormones, amino acid and peptide (Corratgé-Faillie and Lacombe, 2017). Therefore, enhanced transcript level of nitrate transporters in C_3 - C_4 *M. arvensis* indicated the possible increase of shuttling of these substrates, regulating plant development. Increased transcripts of V-type ATPases of C_3 - C_4 *M. arvensis* relative to C_3 *M. moricandioides* suggested that C_3 - C_4 *M. arvensis* is more active in acidification of intracellular compartments to transport metabolites (Padmanaban et al. 2004).

Taken together, the enrichment of solute transports in cluster 14 and cluster 15 implied that C_3 - C_4 *M. arvensis* operates a great transporting system to fulfill the metabolites shuttling among three different organelles (chloroplast, peroxisome, and mitochondrion) and two cell types (M and BS cells), consistent with transcriptional upregulation of genes involving in photorespiratory glycine shuttle and C_4 cycle.

What regulates chloroplast development in *Moricandia* C_3 - C_4 BS cells?

One important feature of C_3 - C_4 species is centripetally accumulated chloroplasts in BS cells, which facilitate the efficient recapture of photorespiratory released CO_2 from mitochondria. Organelles in *Moricandia* C_3 - C_4 BS cells developed earlier and accumulated centripetally through all leaf developing stages, whereas the formation of chloroplasts in C_3 BS cells was discovered later than that of C_3 - C_4 species and no centripetal accumulation was observed. We discovered that the early abundance of chloroplasts in C_3 - C_4 BS cells was more relevant to

early upregulation of chloroplast fission genes (*FtsZ1* and *FtsZ2*) than chloroplast development gene. In C₃-C₄ *M. arvensis*, only *GRF5* (*MSTRG.19433*) expression was increased since stage 0, and then the chloroplast development was controlled through increased transcripts of chloroplast development (*GNC* and *CGA1*), biosynthesis (*PORB* and *HEMA1*), and division (*MinD1*, *ARC6*, and *ARC3*) genes in mid-late stages. In the late stage, transcripts of chloroplast movement genes (*PMIR1* and *WEB1*) increased in C₃-C₄ *M. arvensis*. *GLK1* and *GLK2* in maize and *Sorghum bicolor* are known to regulate the transition from proplastids to chloroplasts in BS and M cells, respectively (Rossini et al., 2001; Wang et al., 2013; Waters et al., 2009). However, transcripts of both *GLK* genes increased in C₃ *M. moricandioides*, which might function redundantly in *Moricandia* as in rice and *Arabidopsis* (Fitter et al., 2002; Rossini et al., 2001). *PHOT1* and *PHOT2* are blue-light receptors, which regulate the blue-light-induced chloroplast photorelocation movement. *PHOT2* expressed in mature leaves in C₃-C₄ *M. arvensis* and C₃ *M. moricandioides*, and demonstrated no differential expression between the two species. However, it has been shown that *PHOT2* was regulated by allele specific expression, resulting in different spatial expression patterns between C₃-C₄ *M. arvensis* and C₃ *M. moricandioides* (Lin Manuscript II). On the other hand, transcript of *PHOT1*, which is involved in the accumulation response of chloroplast movement, increased in mid and late stages of C₃-C₄ *M. arvensis*. We assumed that *PHOT1* plays an important role in chloroplast positioning of mid and late stages in C₃-C₄ *M. arvensis*, keeping BS chloroplasts accumulated centripetally.

All in all, in the early leaf development stage of C₃-C₄ *M. arvensis*, chloroplast fission genes governed the organelle abundance, followed by a cohort of genes involved in chloroplast development, biosynthesis, and division genes in the mid-late stages. As to organelle positioning in C₃-C₄ *M. arvensis*, *PHOT1*, *WEB1*, *PMI1*, and *PMIR1* have an effect on chloroplasts positioning in the mid-late stages. However, what regulated the BS chloroplast positioning in C₃-C₄ *M. arvensis* from early to mid stages remains elusive. In cluster 11 (337 genes), a group of genes increased in expression in all leaf developmental stages in C₃-C₄ *M. arvensis*, which contains potential candidate for genetic control of chloroplast positioning. Thus, further gene ontology and functional analysis are recommended.

Early enhanced transcript level of leaf development genes was associated with early establishment of leaf anatomy of C₃-C₄ *M. arvensis*

C₄ carbon concentrating mechanism is fulfilled through the specific leaf structure, Kranz anatomy. Likewise, C₃-C₄ *M. arvensis* require a Kranz-like anatomy to operate the efficient

photorespiratory glycine shuttle. Transcripts of genes involved in vein initiation and procambium formation (*MP/ARF5*, *CVP2*, *ASI*), xylem formation (*XYPI*, *XYP2*, *ATHB8*, and *ATHB15*), *SCR/SHR* pathway (*SHR*, *JKD*), and vein patterning (*DOF2.1* and *DOF5.3*) significantly increased in early leaf development stages in C₃-C₄ *M. arvensis*. Additionally, *RVNs* involved in *SCR/SHR* pathway, which regulates BS cell development, increased in expression through all leaf development stages in C₃-C₄ *M. arvensis*. In C₃ *M. moricandioides*, transcripts of another group of genes increased either in early stages (such as *REV*, *SFC/VAN3*, and *MP/ARF5*) or in mature leaves (such as *KAN*, *KAN2*, and *DOF4.6*). The late upregulation of *DOF4.6* in C₃ *M. moricandioides* implied the late vein patterning. Taken together, it indicated that the implementation of Kranz-like anatomy of C₃-C₄ *M. arvensis* started in early leaf development stages, which were genetically regulated by different groups of genes compared with C₃ *M. moricandioides*. For instance, different *DOF* genes governed the vein patterning between C₃-C₄ *M. arvensis* and C₃ *M. moricandioides*. In *Moricandia*, transcripts of *DOF2.1* and *DOF5.3* increased in early leaf developmental stages in C₃-C₄ *M. arvensis*, whereas transcripts of *DOF4.6* increased in late leaf development stages in C₃ *M. moricandioides*.

The adaxial layer of the leaf comprises of upper epidermis and tightly packed palisade mesophyll, which contains more chloroplasts and shows higher photosynthetic activity compared with loosely packed spongy mesophyll, forming air spaces allowing uptake of CO₂ and release of O₂. In mature leaves, C₃ *M. moricandioides* had one more layer of palisade relative to C₃-C₄ *M. arvensis*, which might be associated with enhanced transcript level in early stages in C₃ *M. moricandioides*, such as *AS2*, involved in adaxial fate specification, and adxial determinat (*HD-ZIPIII* genes family, *REV*, *PHB* and *ATHB15*). Distinct palisade pattern was observed in C₃, C₃-C₄, C₄-like, and C₄ *Flaveria* species: C₃ and C₃-C₄ species possessed distinct palisade layer; C₄-like species (*F. trinervia*) had a palisade layer; C₄ species (*F. brownii*) showed almost no palisade (Brown and Hattersley, 1989). With the evidence of less photosynthetic palisade layer in C₃-C₄ *M. arvensis*, it indicated that photosynthetic pathway was gradually implemented in BS cells on the evolutionary path from C₃ to C₃-C₄ photosynthesis.

C₃-C₄ *M. arvensis* demonstrated an efficient glycine shuttle in mid-late stages with transcriptomic evidences of enhanced transcript level of glutamine synthetase-glutamate synthase pathway (*GS2* and *Fd-GOGATI*), *GDC/SHMT* system (*mLPD1*, *mLPD2*, *GLDT*, *GLDH3*, and *SHMT1*), hydroxypyruvate reductase (*HPR1* and *HPR2*), glutamate: glyoxylate aminotransferase (*GGAT2*), and plastidic glycolate glycerate transporter (*PLGG1*). It

corresponded to the early formation of Kranz-like anatomy in C₃-C₄ *M. arvensis*, such as centripetally accumulated chloroplasts in BS cells and vasculature pattern formation. We hypothesized that C₃-C₄ *M. arvensis* evolved to develop the leaf vascular system as well as organelle abundance and accumulation earlier in the leaf development relative to C₃ *M. moricandioides*, in order to proceed with the photorespiratory glycine shuttle in mid-late leaf development stages.

Possible scenario for development of C₃-C₄ Kranz-like anatomy

Combing transcriptomics and anatomical observation from leaf primordia to maturation between C₃-C₄ *M. arvensis* and C₃ *M. moricandioides*, we observe that C₃-C₄ *M. arvensis* developed vein and BS cells earlier than C₃ *M. moricandioides*. The establishment of C₃-C₄ vasculature pattern and plastid accumulation starts from very early leaf development stage with the evidence of early and long-lastingly enhanced transcript level in *M. arvensis* of (1) vein initiation transcription factor *PLT3* as well as a cohort of genes functioning in vein initiation, procambium formation, (2) transcription factors, such as leaf development (*GRFs*), leaf cell proliferation (*TCP4*), and vegetative development (*ARF6*), (3) *SCR/SHR* pathway (*RVN*), involved in BS cell development. In addition to organelle development, increased transcript abundance of chloroplast division genes (*FtsZ1* and *FtsZ2*) from early stages could be associated with earlier formation and higher abundance of C₃-C₄ BS chloroplasts. It is particularly important for C₃-C₄ *M. arvensis* from leaf development stage 0 to stage 4, that we observed a long-lasting transcript increase of cell cycle related genes. Therefore, early BS development serves as a prerequisite for organelle accumulation in C₃-C₄ *M. arvensis*, which forms Kranz-like anatomy and thereby supports efficient operation of the photorespiratory glycine shuttle.

Conclusions

In this study, we compared the transcriptomics along the leaf development between closely related C₃ *M. moricandioides* and C₃-C₄ *M. arvensis*. The comparability was shown through the gene expression profiles, where the expression of genes from the categories of photosynthesis and protein increased and decreased, respectively. Through transcription factor discovery in annotated clusters, *GRFs*, *PLT3*, *TCP4*, *ARF6*, *SIG6*, and a G2-like transcription factor (*MSTRG.15393*) were potential candidates to mediate early development of C₃-C₄ characteristics. Key photorespiratory and C₄ cycle genes showed enhanced transcript level in C₃-C₄ *M. arvensis*, associated with the increased transcripts of solute transporter genes. We assumed that plastids fission plays an important role in early organelle accumulation in C₃-C₄ BS cells. Transcripts of genes, which function in vein initiation, procambium formation, xylem formation, *SCR/SHR* pathway, and vein patterning, increase in C₃-C₄ *M. arvensis* in early leaf development stages, which correspond to the early C₃-C₄ leaf development. With this approach, we gained insights into genetic mechanisms of early development of C₃-C₄ Kranz-like anatomy in *Moricandia*, such as early formation of vasculature pattern and organelle accumulation in C₃-C₄ BS cells, which could be further benefit to breeding new *Brassica* varieties with high yield.

Author contributions

MY.L. performed all data analysis and wrote the manuscript.

U.S. designed and performed the leaf gradient experiments, generated the leaf anatomical figures, and participated in drafting the manuscript.

A.K.D. helped with data analysis and participated in drafting the manuscript.

A.P.M.W. participated in drafting the manuscript.

References

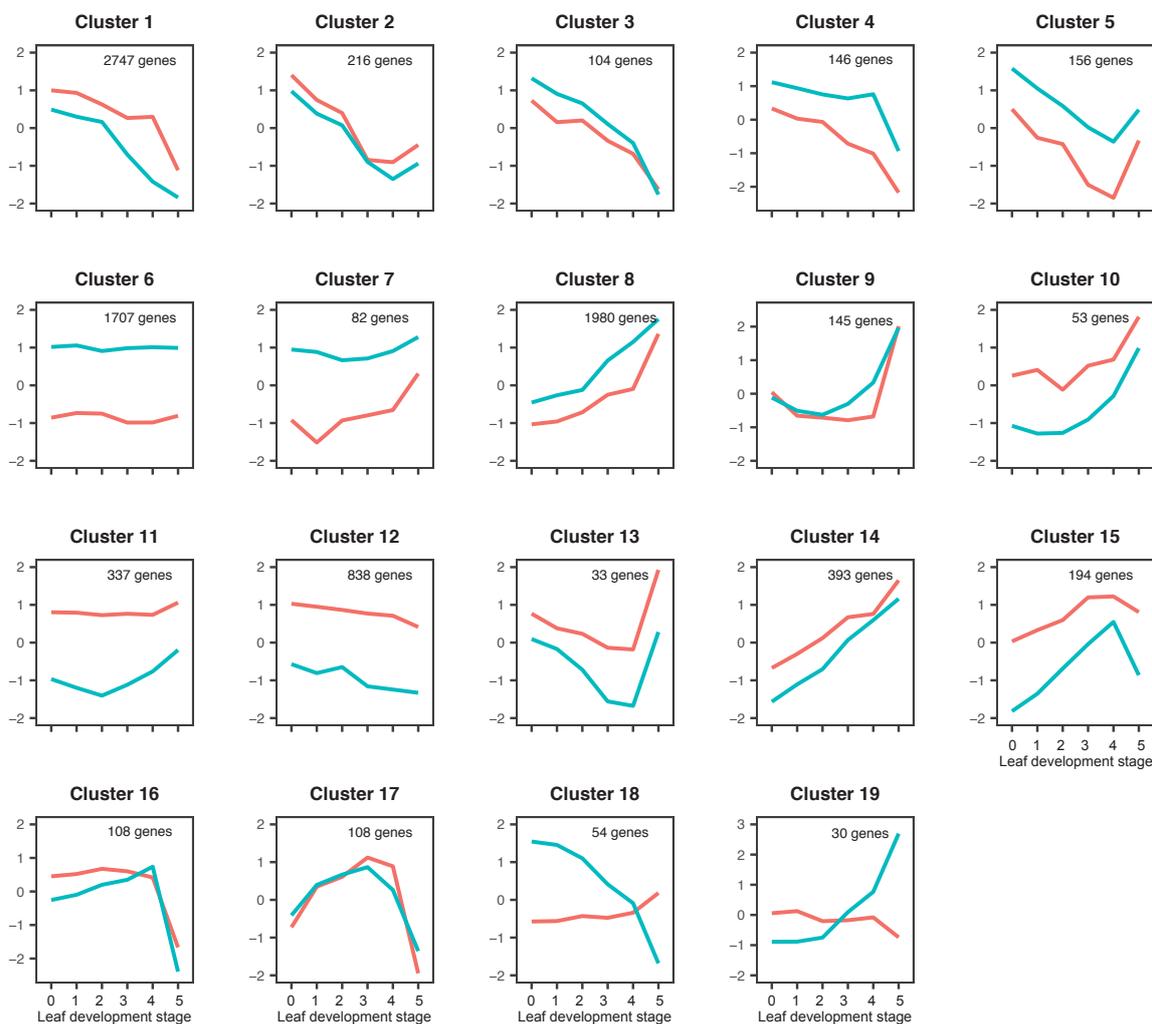
- Abu-Jamous B, Kelly S** (2018) Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biology* **19**: 172
- Aubry S, Kelly S, Kümpers BMC, Smith-Unna RD, Hibberd JM** (2014) Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C4 Photosynthesis. *PLOS Genetics* **10**: e1004365
- Beebe DU, Evert RF** (1990) The Morphology and Anatomy of the Leaf of *Moricandia arvensis* (L.) DC. (Brassicaceae). *Botanical Gazette* **151**: 184–203
- Billakurthi K, Wrobel TJ, Bräutigam A, Weber APM, Westhoff P, Gowik U** (2018) Transcriptome dynamics in developing leaves from C3 and C4 Flaveria species reveal determinants of Kranz anatomy. *bioRxiv* 473181
- Bräutigam A, Gowik U** (2016) Photorespiration connects C3 and C4 photosynthesis. *J Exp Bot* **67**: 2953–2962
- Chi W, He B, Mao J, Jiang J, Zhang L** (2015) Plastid sigma factors: Their individual functions and regulation in transcription. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1847**: 770–778
- Corratgé-Faillie C, Lacombe B** (2017) Substrate (un)specificity of Arabidopsis NRT1/PTR FAMILY (NPF) proteins. *J Exp Bot* **68**: 3107–3113
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21
- Emms DM, Kelly S** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *bioRxiv* 466201
- Fan X, Naz M, Fan X, Xuan W, Miller AJ, Xu G** (2017) Plant nitrate transporters: from gene function to application. *J Exp Bot* **68**: 2463–2475
- Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA** (2002) GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J* **31**: 713–727

- Gonzalez N, De Bodt S, Sulpice R, Jikumaru Y, Chae E, Dhondt S, Van Daele T, De Milde L, Weigel D, Kamiya Y, et al** (2010) Increased leaf size: different means to an end. *Plant Physiol* **153**: 1261–1279
- Gowik U, Bräutigam A, Weber KL, Weber APM, 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**: 2087–2105
- Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ** (2013) Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. *Cell* **153**: 1579–1588
- Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, Shinmyo A, Kobori M, Takeba G, Nakahira Y, Shiina T** (2005) A nuclear-encoded sigma factor, *Arabidopsis* SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *The Plant Journal* **42**: 133–144
- Külahoglu C, Denton AK, Sommer M, Maß J, Schliesky S, Wrobel TJ, Berckmans B, Gongora-Castillo E, Buell CR, Simon R, et al** (2014) Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two *Cleomaceae* C3 and C4 Plant Species. *The Plant Cell* **26**: 3243–3260
- Kümpers BMC, Burgess SJ, Reyna-Llorens I, Smith-Unna R, Bournnell C, Hibberd JM** (2017) Shared characteristics underpinning C4 leaf maturation derived from analysis of multiple C3 and C4 species of *Flaveria*. *J Exp Bot* **68**: 177–189
- Lauterbach M, Schmidt H, Billakurthi K, Hankeln T, Westhoff P, Gowik U, Kadereit G** (2017) De novo Transcriptome Assembly and Comparison of C3, C3-C4, and C4 Species of Tribe *Salsoleae* (*Chenopodiaceae*). *Front Plant Sci*. doi: 10.3389/fpls.2017.01939
- Liu N, Wu S, Van Houten J, Wang Y, Ding B, Fei Z, Clarke TH, Reed JW, van der Knaap E** (2014) Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. *J Exp Bot* **65**: 2507–2520
- Lobell DB, Gourджи SM** (2012) The Influence of Climate Change on Global Crop Productivity. *Plant Physiology* **160**: 1686–1697

- Love MI, Huber W, Anders S** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550
- Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U** (2014) The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *eLife* **3**: e02478
- Monson RK** (1999) The origins of C4 genes and evolutionary pattern in the C4 metabolic phenotype. In RF Sage, RK Monson, eds, *C4 Plant Biology*. Academic Press, San Diego, pp 377–410
- Okushima Y, Mitina I, Quach HL, Theologis A** (2005) AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. *The Plant Journal* **43**: 29–46
- Popp J, Lakner Z, Harangi-Rákos M, Fári M** (2014) The effect of bioenergy expansion: Food, energy, and environment. *Renewable and Sustainable Energy Reviews* **32**: 559–578
- Prasad K, Grigg SP, Barkoulas M, Yadav RK, Sanchez-Perez GF, Pinon V, Blilou I, Hofhuis H, Dhonukshe P, Galinha C, et al** (2011) Arabidopsis PLETHORA transcription factors control phyllotaxis. *Curr Biol* **21**: 1123–1128
- Rossini L, Cribb L, Martin DJ, Langdale JA** (2001) The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. *The Plant Cell* **13**: 1231–1244
- Rylott EL, Metzloff K, Rawsthorne S** (1998) Developmental and Environmental Effects on the Expression of the C3-C4 Intermediate Phenotype in *Moricandia arvensis*. *Plant Physiol* **118**: 1277–1284
- Sage RF, Khoshravesh R, Sage TL** (2014) From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *J Exp Bot* **65**: 3341–3356
- Sage RF, Sage TL, Kocacinar F** (2012) Photorespiration and the Evolution of C4 Photosynthesis. *Annu Rev Plant Biol* **63**: 19–47
- Sarvepalli K, Nath U** (2011) Hyper-activation of the TCP4 transcription factor in *Arabidopsis thaliana* accelerates multiple aspects of plant maturation. *The Plant Journal* **67**: 595–607

- Scarpella E, Marcos D, Friml J, Berleth T** (2006) Control of leaf vascular patterning by polar auxin transport. *Genes Dev* **20**: 1015–1027
- Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin P-A, Kurz S, Mettler-Altmann T, Weber AP** (2017) Photosynthesis in C3–C4 intermediate Moringa species. *J Exp Bot* **68**: 191–206
- Schlüter U, Weber APM** (2016) The Road to C4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant Cell Physiol* **57**: 881–889
- Wang P, Kelly S, Fouracre JP, Langdale JA** (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *The Plant Journal* **75**: 656–670
- Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA** (2017) Re-creation of a Key Step in the Evolutionary Switch from C3 to C4 Leaf Anatomy. *Curr Biol* **27**: 3278-3287.e6
- Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA** (2009) GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. *The Plant Cell* **21**: 1109–1128
- Wenzel CL, Schuetz M, Yu Q, Mattsson J** (2007) Dynamics of MONOPTEROS and PINFORMED1 expression during leaf vein pattern formation in Arabidopsis thaliana. *The Plant Journal* **49**: 387–398
- Williams BP, Johnston IG, Covshoff S, Hibberd JM** (2013) Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife* **2**: e00961

Supplemental information



Supplemental Figure 1. Clustering of leaf development gene expression data through *clust* algorithm. 19 clusters were generated.

Supplemental Table 1. Customized MapMan terms for gene distribution analysis.

Mapman category	Name	Assignment
1	photosynthesis	photosynthesis
2	cellular respiration	cell and development
3	carbohydrate metabolism	metabolism
4	amino acid metabolism	metabolism
5	lipid metabolism	metabolism
6	nucleotide metabolism	metabolism
7	coenzyme metabolism	metabolism
8	polyamine metabolism	metabolism
9	secondary metabolism	metabolism
10	redox homeostasis	other
11	phytohormones	other
12	chromation organisation	DNA
13	cell cyle	cell and development
14	DNA damage response	DNA
15	RNA biosynthesis	RNA
16	RNA processing	RNA
17	protein biosynthesis	protein
18	protein modification	protein
19	protein degradation	protein
20	cytoskeleton	cell and development
21	cell wall	cell wall
22	vesicle trafficking	transport
23	protein translocation	transport
24	solute transport	transport
25	nutrient uptake	cell and development
26	external stimuli response	other
27	multi-process regulation	other
35	not assigned	unkown
50	enzyme classification	protein

Supplemental Table 2. Transcription factor list of descending cluster 1.database 1: PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>);database 2: PlnTFDB (<http://plntfdb.bio.uni-potsdam.de>).

Gene name	Arabidopsis		TF family		Name
	orthologue	database 1	database 2		
MSTRG.10062	AT2G46040		ARID		ARID1
MSTRG.1018	AT4G24540	MIKC_MADS	MADS		AGL24
MSTRG.10232	AT3G12980		TAZ		HAC5
MSTRG.1044	AT4G12750	HB-other			
MSTRG.10479	AT1G14685	BBR-BPC	BBR/BPC		BPC2
MSTRG.10485	AT1G14600	G2-like	G2-like		
MSTRG.10636	AT3G19360	C3H	C3H		
MSTRG.10679	AT5G03740	C2H2	C2H2		HD2C
MSTRG.10702	AT4G37670		GNAT		
MSTRG.10705	AT4G37740	GRF	GRF		GRF2
MSTRG.1087	AT1G72740	MYB_related	MYB-related		
MSTRG.10999	AT1G49560	G2-like	G2-like		
MSTRG.11088	AT4G00260	B3	ABI3VP1		
MSTRG.11460	AT4G35590	Nin-like	RWP-RK		
MSTRG.11586	AT3G52910	GRF	GRF		GRF4
MSTRG.11705	AT3G53340	NF-YB	CCAAT		
MSTRG.11998	AT3G20670		CCAAT		
MSTRG.12667	AT4G24150	GRF	GRF		GRF8
MSTRG.13179	AT2G36340	GeBP	GeBP		
MSTRG.13343	AT1G30330	ARF	ARF		ARF6
MSTRG.13361	AT1G31040		PLATZ		ORE15
MSTRG.13642	AT1G14510		Alfin-like		AL7
MSTRG.13695	AT3G62100		AUX/IAA		IAA30
MSTRG.14240	AT1G10470		Orphans		ARR4
MSTRG.14566	AT4G34290		SWI/SNF-BAF60b		
MSTRG.14741	AT1G49560	G2-like	G2-like		
MSTRG.14748	AT1G72050	C2H2	C2H2		
MSTRG.15261	AT2G28340	GATA	C2C2-GATA		
MSTRG.15470	AT3G49940	LBD	LOB		
MSTRG.1557	AT3G62100		AUX/IAA		IAA30
MSTRG.15585	AT1G54690		CCAAT		
MSTRG.15724	AT1G19050		Orphans		ARR7
MSTRG.15744	AT3G06010		SNF2		
MSTRG.15932	AT1G09770	MYB	MYB		
MSTRG.16068	AT2G44020		mTERF		
MSTRG.16313	AT2G45190	YABBY	C2C2-YABBY		
MSTRG.16364	AT3G20670		CCAAT		
MSTRG.16421	AT1G74250	C2H2			
MSTRG.16566	AT4G36020		CSD		
MSTRG.16594	AT1G66350	GRAS	GRAS		RGL1
MSTRG.16782	AT3G54320	ERF	AP2-EREBP		WRI1
MSTRG.16911	AT2G33610		MYB-related		SWI3B
MSTRG.16930	AT2G33290		SET		
MSTRG.17095	AT2G28450	C3H	C3H		
MSTRG.17237	AT3G12280		RB		
MSTRG.1762	AT4G30860		SET		
MSTRG.17838	AT2G01830		Orphans		
MSTRG.1811	AT2G45190	YABBY	C2C2-YABBY		
MSTRG.18126	AT3G26744	bHLH	bHLH		
MSTRG.18512	AT3G60670		PLATZ		
MSTRG.18551	AT1G08010	GATA	C2C2-GATA		
MSTRG.19026	AT5G63420	Trihelix	Trihelix		
MSTRG.19142	AT4G00260	B3	ABI3VP1		
MSTRG.19156	AT5G46690	bHLH	bHLH		
MSTRG.19433	AT3G13960	GRF	GRF		GRF5
MSTRG.19541	AT3G48160	E2F/DP	E2F-DP		
MSTRG.19656	AT2G36010	E2F/DP	E2F-DP		
MSTRG.19689	AT2G18850		SET		

Supplemental Table 2. Transcription factor list of descending cluster 1. Continued.

Gene name	Arabidopsis		TF family		Name
	orthologue	database 1	database 2		
MSTRG.19999	AT5G17240			SET	
MSTRG.20144	AT5G18620			SNF2	
MSTRG.20272	AT5G53210	bHLH		bHLH	
MSTRG.20384	AT5G53660	GRF		GRF	GRF7
MSTRG.20610	AT2G36010	E2F/DP		E2F-DP	
MSTRG.20725	AT4G00480	bHLH		bHLH	
MSTRG.20742	AT4G00180	YABBY		C2C2-YABBY	
MSTRG.20775	AT1G02065	SBP		SBP	
MSTRG.2087	AT1G19790	SRS		SRS	
MSTRG.20975	AT5G11340			GNAT	
MSTRG.21578	AT1G32730	SRS			
MSTRG.21678	AT5G60200	Dof		C2C2-Dof	
MSTRG.21723	AT1G49720	bZIP		bZIP	
MSTRG.21781	AT3G22760	CPP		CPP	
MSTRG.21918	AT5G41020	MYB_related		MYB-related	
MSTRG.21989	AT2G34710	HD-ZIP		HB	ATHB14
MSTRG.22234	AT3G19184	B3		ABI3VP1	
MSTRG.22604	AT3G60390	HD-ZIP		HB	HAT3
MSTRG.22730	AT1G62360	TALE		HB	
MSTRG.22869	AT5G43250	NF-YC		CCAAT	
MSTRG.2323	AT5G53660	GRF		GRF	
MSTRG.23296	AT5G53210	bHLH		bHLH	
MSTRG.2333	AT5G52600	MYB		MYB	
MSTRG.23485	AT2G16390			SNF2	
MSTRG.23500	AT3G52910	GRF		GRF	
MSTRG.23677	AT2G36050			OFF	
MSTRG.23679	AT2G36010	E2F/DP		E2F-DP	
MSTRG.23858	AT1G14440	ZF-HD		zf-HD	
MSTRG.24132	AT5G25190	ERF		AP2-EREBP	
MSTRG.24355	AT2G45190	YABBY		C2C2-YABBY	
MSTRG.24366	AT2G44910	HD-ZIP		HB	ATHB4
MSTRG.2442	AT3G20670			CCAAT	
MSTRG.24585	AT2G17870			CSD	
MSTRG.25172	AT4G37540	LBD		LOB	
MSTRG.25433	AT5G53660	GRF		GRF	
MSTRG.25612	AT1G62120			mTERF	
MSTRG.25768	AT3G20670			CCAAT	
MSTRG.25847	AT5G43250	NF-YC		CCAAT	
MSTRG.25849	AT3G49940	LBD		LOB	
MSTRG.25901	AT1G62120			mTERF	
MSTRG.26164	AT1G07360			C3H	
MSTRG.26194	AT1G32730	SRS			
MSTRG.26260	AT1G26780	MYB		MYB	
MSTRG.26384	AT4G31270	Trihelix			
MSTRG.26428	AT5G03150	C2H2		C2H2	
MSTRG.26521	AT4G37740	GRF		GRF	
MSTRG.26539	AT5G07400			FHA	
MSTRG.26574	AT5G08430			SWI/SNF-BAF60b	
MSTRG.26629	AT5G10510	AP2		AP2-EREBP	
MSTRG.26718	AT5G13060			TRAF	
MSTRG.26817	AT5G13960			SET	
MSTRG.2746	AT4G28190			ULT	
MSTRG.27780	AT1G22490	bHLH		bHLH	
MSTRG.27962	AT3G15030	TCP		TCP	TCP4
MSTRG.27968	AT3G42790			Alfin-like	
MSTRG.2804	AT1G14510			Alfin-like	
MSTRG.2805	AT1G14410	Whirly		PBF-2-like	
MSTRG.2839	AT3G11260	WOX		HB	

Supplemental Table 2. Transcription factor list of descending cluster 1. Continued.

Gene name	Arabidopsis		TF family		Name
	orthologue	database 1	database 2		
MSTRG.28894	AT1G51060			CCAAT	
MSTRG.29061	AT3G51120	C3H		C3H	
MSTRG.29415	AT5G36740			PHD	
MSTRG.29693	AT5G51980	C3H		C3H	
MSTRG.29828	AT1G34355			FHA	
MSTRG.30063	AT1G62830			SWI/SNF-SWI3	
MSTRG.30077	AT1G63100	GRAS		GRAS	
MSTRG.30257	AT5G53210	bHLH		bHLH	
MSTRG.3042	AT3G62100			AUX/IAA	
MSTRG.30635	AT3G50890	ZF-HD		zf-HD	
MSTRG.30702	AT4G14770	CPP		CPP	
MSTRG.3076	AT5G24330			PHD	
MSTRG.30884	AT1G31760			SWI/SNF-BAF60b	
MSTRG.31582	AT3G01600	NAC		NAC	
MSTRG.31669	AT3G44750	C2H2			
MSTRG.3172	AT4G21430			Orphans	
MSTRG.31746	AT2G47210	MYB_related		MYB-related	
MSTRG.31823	AT5G56780	HRT-like		HRT	
MSTRG.31853	AT5G07810			SNF2	
MSTRG.31899	AT5G08630			DDT	
MSTRG.31944	AT5G10510	AP2		AP2-EREBP	PLT3
MSTRG.32531	AT1G68810	bHLH		bHLH	
MSTRG.3350	AT1G25580	NAC		NAC	
MSTRG.3502	AT2G37630	MYB		MYB	
MSTRG.3585	AT1G31760			SWI/SNF-BAF60b	
MSTRG.3720	AT5G13060			TRAF	
MSTRG.4078	AT1G10240	FAR1		FAR1	
MSTRG.4349	AT5G03680	Trihelix		Trihelix	
MSTRG.4512	AT5G14000	NAC		NAC	
MSTRG.4568	AT5G15150	HD-ZIP		HB	ATHB3
MSTRG.4683	AT5G66750			SNF2	
MSTRG.5041	AT3G15030	TCP		TCP	TCP4
MSTRG.535	AT1G52150	HD-ZIP		HB	
MSTRG.5891	AT5G06110	MYB		MYB	
MSTRG.591	AT2G21060			CSD	
MSTRG.6082	AT5G65640	bHLH		bHLH	
MSTRG.6381	AT5G63950			SNF2	
MSTRG.6820	AT2G24645	B3			
MSTRG.7034	AT1G51060			CCAAT	
MSTRG.7427	AT1G68550	ERF		AP2-EREBP	CRF10
MSTRG.7689	AT2G41070	bZIP		bZIP	
MSTRG.7703	AT2G41450			GNAT	
MSTRG.7856	AT1G05120			SNF2	
MSTRG.8135	AT3G12280			RB	
MSTRG.8188	AT2G17410			ARID	
MSTRG.8285	AT5G24330			PHD	
MSTRG.8389	AT3G19210			SNF2	
MSTRG.8443	AT3G20010			SNF2	
MSTRG.8475	AT3G18100	MYB		MYB	
MSTRG.880	AT5G46880	HD-ZIP		HB	HB7
MSTRG.8855	AT5G46880	HD-ZIP		HB	HB7
MSTRG.9346	AT3G52910	GRF		GRF	GRF4
MSTRG.9387	AT1G68120	BBR-BPC		BBR/BPC	
MSTRG.9440	AT2G36010	E2F/DP		E2F-DP	
MSTRG.9523	AT2G22840	GRF		GRF	GRF1
MSTRG.9627	AT4G23800			HMG	
MSTRG.963	AT1G71260	Whirly		PBF-2-like	
MSTRG.9740	AT1G16070			TUB	

Supplemental Table 3. Transcription factor list of descending cluster 2.database 1: PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>);database 2: PlnTFDB (<http://plntfdb.bio.uni-potsdam.de/>).

Gene name	<i>A. thaliana</i> orthologue	TF family		Name
		database 1	database 2	
MSTRG.10697	AT3G22780	CPP	CPP	
MSTRG.10706	AT4G37750	AP2	AP2-EREBP	ANT
MSTRG.11458	AT2G17560		HMG	
MSTRG.12093	AT4G34430		MYB-related	
MSTRG.13070	AT4G27910		PHD	
MSTRG.13196	AT4G15180		SET	
MSTRG.13225	AT4G18890	BES1	BES1	
MSTRG.14559	AT4G34430		MYB-related	
MSTRG.15743	AT3G06010		SNF2	
MSTRG.1622	AT4G29940	HB-PHD	HB	
MSTRG.16704	AT3G05380		MYB-related	
MSTRG.18801	AT2G25170		PHD	
MSTRG.1913	AT2G23740		C2H2	
MSTRG.20327	AT5G65410	ZF-HD	zf-HD	
MSTRG.20835	AT4G16310		SWI/SNF-SWI3	
MSTRG.20875	AT3G01460		PHD	
MSTRG.21297	AT5G05130		SNF2	
MSTRG.22249	AT1G79350		PHD	
MSTRG.23486	AT2G16390		SNF2	
MSTRG.25160	AT4G37740	GRF	GRF	GRF2
MSTRG.25673	AT5G57390	AP2	AP2-EREBP	
MSTRG.26071	AT1G30330	ARF	ARF	ARF6
MSTRG.29074	AT4G29000	CPP	CPP	
MSTRG.29334	AT1G10170	NF-X1		
MSTRG.29348	AT1G09770	MYB	MYB	
MSTRG.30030	AT1G62085		mTERF	
MSTRG.30145	AT1G30490	HD-ZIP	HB	ATHB9
MSTRG.31269	AT1G20910		ARID	
MSTRG.31380	AT5G44800		PHD	
MSTRG.31603	AT2G30470	B3	ABI3VP1	
MSTRG.31668	AT5G35210		DDT	
MSTRG.32716	AT1G47870	E2F/DP	E2F-DP	
MSTRG.4520	AT5G14170		SWI/SNF-BAF60b	
MSTRG.6987	AT4G32730	MYB	MYB	
MSTRG.71	AT1G59890		Orphans	
MSTRG.7203	AT4G20400		Jumonji	
MSTRG.8052	AT4G32551		LUG	
MSTRG.8191	AT1G59890		Orphans	
MSTRG.9576	AT2G02470		Alfin-like	

Supplemental Table 4. Transcription factor list of ascending cluster 8.database 1: PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>);database 2: PlnTFDB (<http://plntfdb.bio.uni-potsdam.de/>).

Gene name	<i>A. thaliana</i> orthologue	TF family		Name
		database 1	database 2	
MSTRG.10222	AT2G42280	bHLH	bHLH	
MSTRG.10931	AT2G20570	G2-like	G2-like	GLK1
MSTRG.11029	AT1G64280		TRAF	
MSTRG.11279	AT2G46790		Pseudo	
MSTRG.11517	AT5G14370		Orphans	
MSTRG.11716	AT5G24930	CO-like	C2C2-CO-like	
MSTRG.11783	AT5G59780	MYB_related	MYB-related	
MSTRG.11917	AT2G17150	Nin-like	RWP-RK	
MSTRG.12380	AT1G77920	bZIP	bZIP	
MSTRG.12854	AT1G06040	DBB	Orphans	
MSTRG.12884	AT3G61890	HD-ZIP	HB	
MSTRG.13041	AT4G27310		Orphans	
MSTRG.13044	AT4G27410	NAC	NAC	
MSTRG.13102	AT4G28640		AUX/IAA	
MSTRG.13157	AT4G30080	ARF	ARF	ARF16
MSTRG.13401	AT4G38960	DBB	Orphans	
MSTRG.13506	AT1G74930	ERF	AP2-EREBP	
MSTRG.13518	AT1G74650	MYB	MYB	
MSTRG.13581	AT5G48150	GRAS	GRAS	
MSTRG.13850	AT3G49800		BSD	
MSTRG.14078	AT1G75410	TALE	HB	
MSTRG.14170	AT5G22220	E2F/DP	E2F-DP	
MSTRG.14456	AT3G47500	Dof	C2C2-Dof	
MSTRG.14525	AT4G16750	ERF	AP2-EREBP	
MSTRG.1474	AT3G06590	bHLH		
MSTRG.14791	AT5G61420	MYB_related	MYB	
MSTRG.14980	AT2G46680	HD-ZIP	HB	
MSTRG.15209	AT1G67970	HSF	HSF	
MSTRG.15416	AT2G42040		Orphans	
MSTRG.15637	AT5G18550	C3H	C3H	
MSTRG.15648	AT4G35270	Nin-like	RWP-RK	
MSTRG.15806	AT3G57040		Orphans	
MSTRG.15813	AT2G39000		GNAT	
MSTRG.15892	AT2G38090	MYB	MYB	
MSTRG.15928	AT1G09710	MYB_related	MYB-related	
MSTRG.1634	AT4G20970	bHLH	bHLH	
MSTRG.17069	AT1G03350		BSD	
MSTRG.17246	AT3G11280	MYB	MYB	
MSTRG.17570	AT1G32360	C3H	C3H	
MSTRG.17602	AT3G53920		Sigma70-like	SIG3
MSTRG.17712	AT5G47390	MYB_related	MYB-related	
MSTRG.18224	AT3G28917	ZF-HD	zf-HD	
MSTRG.18365	AT3G59060	bHLH	bHLH	
MSTRG.18935	AT5G56840	MYB_related	MYB-related	
MSTRG.19279	AT4G00050	bHLH	bHLH	
MSTRG.19620	AT4G19990	FAR1	FAR1	
MSTRG.1963	AT2G44730	Trihelix	Trihelix	
MSTRG.19657	AT2G35940	TALE	HB	
MSTRG.19780	AT4G30080	ARF	ARF	ARF16
MSTRG.20111	AT2G16720	MYB	MYB	
MSTRG.2034	AT3G23030		AUX/IAA	
MSTRG.20605	AT2G35940	TALE	HB	
MSTRG.20874	AT3G01470	HD-ZIP	HB	
MSTRG.21040	AT5G10140	MIKC_MADS	MADS	
MSTRG.21064	AT5G09740		C2H2	
MSTRG.21263	AT5G05790	MYB	MYB	
MSTRG.21580	AT1G32700		PLATZ	
MSTRG.21584	AT4G31550	WRKY	WRKY	

Supplemental Table 4. Transcription factor list of ascending cluster 8. Continued.

Gene name	<i>A. thaliana</i> orthologue	TF family		Name
		database 1	database 2	
MSTRG.21738	AT1G42990	bZIP	bZIP	
MSTRG.220	AT5G56270	WRKY	WRKY	
MSTRG.2235	AT1G25440	CO-like	C2C2-CO-like	
MSTRG.22385	AT1G13450	Trihelix	Trihelix	
MSTRG.2250	AT3G08505	C3H	C3H	
MSTRG.22852	AT3G11100	Trihelix	Trihelix	
MSTRG.23012	AT1G19000	MYB_related	MYB-related	
MSTRG.23201	AT4G27310		Orphans	
MSTRG.23259	AT3G53920		Sigma70-like	SIG3
MSTRG.23525	AT3G23030		AUX/IAA	
MSTRG.23682	AT2G35940	TALE	HB	
MSTRG.2369	AT1G50600	GRAS	GRAS	SCL5
MSTRG.23747	AT2G34720	NF-YA	CCAAT	
MSTRG.23950	AT5G49450	bZIP	bZIP	
MSTRG.24264	AT4G20380	LSD		
MSTRG.24334	AT2G45660	MIKC_MADS	MADS	
MSTRG.24676	AT1G34190	NAC	NAC	
MSTRG.24851	AT3G57230	MIKC_MADS	MADS	
MSTRG.25107	AT1G34190	NAC	NAC	
MSTRG.25157	AT4G37790	HD-ZIP	HB	
MSTRG.26122	AT1G05690		TRAF	
MSTRG.26174	AT1G07640	Dof	C2C2-Dof	
MSTRG.26259	AT1G26790	Dof	C2C2-Dof	
MSTRG.26583	AT5G08790	NAC	NAC	
MSTRG.26613	AT5G10030	bZIP	bZIP	
MSTRG.26655	AT5G11060	TALE	HB	
MSTRG.26723	AT5G13180	NAC	NAC	
MSTRG.26797	AT1G10720		BSD	
MSTRG.27029	AT3G02150	TCP	TCP	
MSTRG.27054	AT1G76890	Trihelix	Trihelix	
MSTRG.27085	AT1G77920	bZIP	bZIP	
MSTRG.27388	AT1G61660	bHLH	bHLH	
MSTRG.27415	AT1G78080	ERF	AP2-EREBP	
MSTRG.27436	AT2G17820		Orphans	
MSTRG.27500	AT1G24625	C2H2	C2H2	
MSTRG.27582	AT4G18880	HSF	HSF	
MSTRG.27605*	AT2G20570	G2-like	G2-like	GLK1
MSTRG.27817	AT1G51700	Dof	C2C2-Dof	
MSTRG.28162	AT3G62260		DBP	
MSTRG.28185	AT3G61890	HD-ZIP	HB	
MSTRG.28189	AT1G76890	Trihelix	Trihelix	
MSTRG.28221	AT1G76350	Nin-like	RWP-RK	
MSTRG.28241	AT1G76110		ARID	
MSTRG.28371	AT2G28550	AP2	AP2-EREBP	
MSTRG.28575	AT1G49130	CO-like	C2C2-CO-like	
MSTRG.29087	AT3G24050	GATA	C2C2-GATA	
MSTRG.29090	AT1G32700		PLATZ	
MSTRG.29203	AT4G02640	bZIP	bZIP	
MSTRG.29216	AT4G03250	HB-other	HB	
MSTRG.29239	AT1G30500	NF-YA	CCAAT	
MSTRG.2931	AT4G22950	MIKC_MADS	MADS	
MSTRG.29381	AT2G06025		GNAT	
MSTRG.29635	AT2G43500	Nin-like	RWP-RK	
MSTRG.29924	AT5G44190	G2-like	G2-like	GLK2
MSTRG.3001*	AT3G62420	bZIP	bZIP	
MSTRG.30020	AT1G61660	bHLH	bHLH	
MSTRG.30247	AT5G53420		Orphans	
MSTRG.30273	AT5G52660	MYB_related	MYB-related	

Supplemental Table 4. Transcription factor list of ascending cluster 8. Continued.

Gene name	<i>A. thaliana</i> orthologue	TF family		Name
		database 1	database 2	
MSTRG.30350	AT2G22300	CAMTA	CAMTA	
MSTRG.30376	AT1G75410	TALE	HB	
MSTRG.30555	AT1G32700		PLATZ	
MSTRG.3092	AT5G24120		Sigma70-like	SIG5
MSTRG.31123	AT1G43700	bZIP	bZIP	
MSTRG.31164	AT3G01470	HD-ZIP	HB	
MSTRG.31289	AT1G21450	GRAS	GRAS	
MSTRG.31464	AT3G05690	NF-YA	CCAAT	
MSTRG.3162	AT5G60850	Dof	C2C2-Dof	
MSTRG.31684	AT1G64860		Sigma70-like	SIGA
MSTRG.31771	AT2G46830	MYB_related	MYB-related	
MSTRG.31775	AT2G46790		Pseudo	
MSTRG.31808	AT2G46400	WRKY	WRKY	
MSTRG.31903	AT5G08790	NAC	NAC	
MSTRG.32665	AT1G72450		Tify	
MSTRG.3381	AT3G21270	Dof	C2C2-Dof	
MSTRG.3440	AT5G20730	ARF	ARF	ARF7
MSTRG.3468	AT1G35460	bHLH	bHLH	
MSTRG.3485	AT4G34680	GATA	C2C2-GATA	
MSTRG.373	AT4G32040	TALE	HB	
MSTRG.4204	AT1G74650	MYB	MYB	
MSTRG.4244	AT5G59780	MYB_related	MYB-related	
MSTRG.4302	AT5G59780	MYB_related	MYB-related	
MSTRG.4313	AT5G02840	MYB_related	MYB-related	
MSTRG.4355	AT1G22070	bZIP		
MSTRG.4673	AT5G67030		FHA	
MSTRG.5049	AT1G21000		PLATZ	
MSTRG.5484	AT5G25220	TALE	HB	
MSTRG.5664	AT2G29065	GRAS		
MSTRG.583	AT1G27320		Orphans	
MSTRG.6	AT3G19860	bHLH	bHLH	
MSTRG.6142	AT5G67030		FHA	
MSTRG.6524	AT5G26749	C3H		
MSTRG.6772	AT4G31550	WRKY	WRKY	
MSTRG.7022	AT5G35750		Orphans	
MSTRG.7286	AT1G20640	Nin-like	RWP-RK	
MSTRG.7347	AT5G15850	CO-like	C2C2-CO-like	
MSTRG.7383	AT1G69580	G2-like	G2-like	
MSTRG.7839	AT1G03970	bZIP	bZIP	
MSTRG.8167	AT5G45710	HSF	HSF	
MSTRG.8494	AT3G17860		Tify	
MSTRG.8550	AT5G14370		Orphans	
MSTRG.9167	AT5G62000	ARF	ARF	ARF2
MSTRG.9481	AT4G19660		TRAF	
MSTRG.9543	AT3G58710	WRKY	WRKY	
MSTRG.9902	AT1G01060	MYB_related	MYB-related	

Supplemental Table 5. Gene list of photorespiration and C₄ cycle.

Gene name	<i>A. thaliana</i>	Gene
Photorespiration		
PGLP2	AT5G47760	2-PHOSPHOGLYCOLATE PHOSPHATASE 2
PLGG1	AT1G32080	PLASTIDIAL GLYCOLATE/GLYCERATE TRANSPORTER 1
GOX1	AT3G14420	GLYCOLATE OXIDASE 1
GOX2	AT3G14415	GLYCOLATE OXIDASE 2
GGAT1	AT1G23310	GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE 1
GGAT2	AT1G70580	GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE 2
HPR1	AT1G68010	HYDROXYPYRUVATE REDUCTASE 1
HPR2	AT1G79870	HYDROXYPYRUVATE REDUCTASE 2
GLYK	AT1G80380	GLYCERATE KINASE
SHMT1	AT4G37930	SERINE HYDROXYMETHYLTRANSFERASE 1
SHMT2	AT5G26780	SERINE HYDROXYMETHYLTRANSFERASE 2
GLDP1	AT4G33010	GLYCINE DECARBOXYLASE P-PROTEIN
GLDT	AT1G11860	GLYCINE CLEAVAGE T-PROTEIN FAMILY
mLPD2	AT3G17240	LIPOAMIDE DEHYDROGENASE 2
mLPD1	AT1G48030	LIPOAMIDE DEHYDROGENASE 1
GLDH2	AT2G35120	GLYCINE DECARBOXYLASE H-PROTEIN 2
GLDH3	AT1G32470	GLYCINE DECARBOXYLASE H-PROTEIN 3
SGAT	AT2G13360	SERINE:GLYOXYLATE AMINOTRANSFERASE
GS2	AT5G35630	PLASTIDIAL GLUTAMINE SYNTHETASE
Fd-GOGAT1	AT5G04140	FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE
C₄ cycle		
alpha_CA1	AT3G52720	alpha-CARBONIC ANHYDRASE 1
beta_CA1	AT3G01500	beta-CARBONIC ANHYDRASE 1
beta_CA4	AT1G70410	beta-CARBONIC ANHYDRASE 4
beta_CA5	AT4G33580	beta-CARBONIC ANHYDRASE 5
beta_CA6	AT1G58180	beta-CARBONIC ANHYDRASE 6
gamma_CA1	AT1G19580	gamma-CARBONIC ANHYDRASE 1
gamma_CA2	AT1G47260	gamma-CARBONIC ANHYDRASE 2
gamma_CA3	AT5G66510	gamma-CARBONIC ANHYDRASE 3
PEPC1	AT1G53310	PHOSPHOENOLPYRUVATE CARBOXYLASE 1
PEPC2	AT2G42600	PHOSPHOENOLPYRUVATE CARBOXYLASE 2
PPT1	AT5G33320	PHOSPHOENOLPYRUVATE/PHOSPHATE TRANSLOCATOR 1
PPT2	AT3G01550	PHOSPHOENOLPYRUVATE/PHOSPHATE TRANSLOCATOR 2
plastidic_NADP-MDH	AT5G58330	NADP-MALATE DEHYDROGENASE
PPDK	AT4G15530	PYRUVATE, ORTHOPHOSPHATE DIKINASE
PPDK_regulatory_protein1	AT3G01200	PPDK REGULATORY PROTEIN 2
PPDK_regulatory_protein2	AT4G21210	PPDK REGULATORY PROTEIN 1
BASS2	AT2G26900	PLASMA MEMBRANE PYRUVATE TRANSPORT 2
BASS4	AT3G56160	PLASMA MEMBRANE PYRUVATE TRANSPORT 4
NHD1	AT3G19490	SODIUM:HYDROGEN ANTIporter 1
DiT1	AT5G12860	DICARBOXYLATE TRANSPORTER 1
DiT2	AT5G64280	DICARBOXYLATE TRANSPORTER 2
AspAT2	AT5G19550	ASPARTATE AMINOTRANSFERASE 2
AspAT5	AT4G31990	ASPARTATE AMINOTRANSFERASE 5
AspAT1	AT2G30970	ASPARTATE AMINOTRANSFERASE 1
AspAT3	AT5G11520	ASPARTATE AMINOTRANSFERASE 3
AspAT	AT2G22250	ASPARTATE AMINOTRANSFERASE
cpNAD-MDH	AT3G47520	PLASTIDIC NAD-DEPENDENT MALATE DEHYDROGENASE
mtNAD-MDH	AT1G53240	MITOCHONDRIAL NAD-DEPENDENT MALATE DEHYDROGENASE
NAD-ME1	AT2G13560	NAD-MALIC ENZYME 1
NAD-ME2	AT4G00570	NAD-MALIC ENZYME 2
AlaAT1	AT1G17290	ALANINE AMINOTRANSFERASE
PEPCK	AT4G37870	PHOSPHOENOLPYRUVATE CARBOXYKINASE
TPT	AT5G46110	TRIOSE-PHOSPHATE /PHOSPHATE TRANSLOCATOR
DIC3	AT5G09470	DICARBOXYLATE CARRIER
MPC1	AT5G20090	MITOCHONDRIAL PYRUVATE CARRIER 1,

Supplemental Table 6. Gene list of chloroplast development.

name	<i>A. thaliana</i>	Gene
Chloroplast movement		
PHOT1	AT3G45780	PHOTOTROPIN 1
PHOT2	AT5G58140	PHOTOTROPIN 2
CHUP1	AT3G25690	CHLOROPLAST UNUSUAL POSITIONING 1
KAC1	AT5G10470	KINESIN LIKE PROTEIN FOR ACTIN BASED CHLOROPLAST MOVEMENT 1
KAC2	AT5G65460	KINESIN LIKE PROTEIN FOR ACTIN BASED CHLOROPLAST MOVEMENT 2
JAC1	AT1G75100	J-DOMAIN PROTEIN REQUIRED FOR CHLOROPLAST ACCUMULATION RESPONSE 1
WEB1	AT2G26570	WEAK CHLOROPLAST MOVEMENT UNDER BLUE LIGHT 1
PMI1	AT1G42550	PLASTID MOVEMENT IMPAIRED1
PMI2/WEB2	AT1G66840	PLASTID MOVEMENT IMPAIRED2
PMIR1	AT5G20610	PLASTID MOVEMENT IMPAIRED1-RELATED1
Chloroplast division		
ARC5/DRP5B	AT3G19720	ACCUMULATION AND REPLICATION OF CHLOROPLAST 5/DYNAMIN RELATED PROTEIN 5B
FtsZ1	AT5G55280	HOMOLOG OF BACTERIAL CYTOKINESIS Z-RING PROTEIN
FtsZ2	AT2G36250	HOMOLOG OF BACTERIAL CYTOKINESIS Z-RING PROTEIN
MCD1	AT1G20830	MULTIPLE CHLOROPLAST DIVISION SITE 1
MIND/ARC11	AT5G24020	ACCUMULATION AND REPLICATION OF CHLOROPLASTS 11
MINE1/ARC12	AT1G69390	HOMOLOGUE OF BACTERIAL MINE 1
PDV1	AT5G53280	PLASTID DIVISION1
PDV2	AT2G16070	PLASTID DIVISION2
ARC6	AT5G42480	ACCUMULATION AND REPLICATION OF CHLOROPLAST 6
ARC3	AT1G75010	ACCUMULATION AND REPLICATION OF CHLOROPLASTS 3
Chloroplast development		
GLK1	AT2G20570	GOLDEN2-LIKE 1
GLK2	AT5G44190	GOLDEN2-LIKE 2
GNC	AT5G56860	GATA NITRATE-INDUCIBLE CARBON-METABOLISM-INVOLVED
CGA1	AT4G26150	CYTOKININ-RESPONSIVE GATA1
GRF5	AT3G13960	GROWTH REGULATING FACTOR5
PIF3	AT1G09530	PHYTOCHROME INTERACTING FACTOR 3
PORB	AT4G27440	PROTOCHLOROPHYLLIDE OXIDOREDUCTASE B
PORC	AT1G03630	PROTOCHLOROPHYLLIDE OXIDOREDUCTASE C
HEMA1	AT1G58290	HEMA1

Supplemental Table 7. Gene list of vein development.

name	<i>A. thaliana</i>	Gene
Vein patterning		
DOF2.1	AT2G28510	DOF-TYPE ZINC FINGER DNA-BINDING FAMILY PROTEIN 2.1
DOF4.6	AT4G24060	DOF-TYPE ZINC FINGER DNA-BINDING FAMILY PROTEIN 4.6
DOF5.3/TOM6	AT5G60200	DOF-TYPE ZINC FINGER DNA-BINDING FAMILY PROTEIN 5.3
Polar auxin transport, auxin response and vascular development		
MP/ARF5	AT1G19850	MONOPTEROS/AUXIN RESPONSE FACTOR 5
PIN1	AT1G73590	PIN-FORMED1
GNOM/VAN7/EMB30	AT1G13980	GUANINE-NUCLEOTIDE-EXCHANGE FACTOR FOR ADP-RIBOSYLATION FACTOR G PROTEIN
SFC/VAN3	AT5G13300	SCARFACE/VASCULAR NETWORK 3
CVP1/SMT2	AT1G20330	COTYLEDON VASCULAR PATTERN 1
CVP2	AT1G05470	COTYLEDON VASCULAR PATTERN 2
CVL1	AT2G32010	CVP2 LIKE 1
AS1	AT2G37630	ASYMMETRIC LEAVES 1
AS2	AT1G65620	ASYMMETRIC LEAVES 2
KNAT1/BP	AT4G08150	KNOTTED-LIKE FROM ARABIDOPSIS THALIANA/BREVIPEDICELLUS
XYP1	AT5G64080	XYLOGEN PROTEIN 1
XYP2	AT2G13820	XYLOGEN PROTEIN 2
REV	AT5G60690	REVOLUTA
PHB	AT2G34710	PHABULOSA
PHV	AT1G30490	PHAVOLUTA
ATHB8	AT4G32880	HOMEODOMAIN GENE 8
ATHB15	AT1G52150	HOMEODOMAIN GENE 15
KAN	AT5G16560	KANADI
KAN2	AT1G32240	KANADI 2
APL	AT1G79430	ALTERED PHLOEM DEVELOPMENT (APL)
SCR/SHR pathway		
SCR	AT3G54220	SCARECROW
SHR	AT4G37650	SHORT ROOT
RVN	AT2G02070	RAVEN
JKD	AT5G03150	JACKDAW
BLJ	AT1G14580	BLUEJAY

Supplemental Table 8. Number of significantly differentially expressed genes during leaf development.

Ma, *M. arvensis*; Mm, *M. moricandioides*.

	Leaf development series					
	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Transcripts decreased in Ma relative to Mm	5.075	3.489	4.634	6.587	7.004	6.493
Transcripts increased in Ma relative to Mm	3.233	2.213	2.302	4.496	5.366	5.086

Supplemental Table 9. Solute transporter list of cluster 14 and cluster 15.

Gene name	Arabidopsis orthologue	Name	MapMan bin	MapMan function
Cluster 14				
MSTRG.1060	AT2G40420		24.2.3.4.8	Solute transport.carrier-mediated transport.APC superfamily.AAAP family.AAAP-type transporter
MSTRG.2476	AT5G49630		24.2.3.4.1	Solute transport.carrier-mediated transport.APC superfamily.AAAP family.amino acid transporter (AAP-type)
MSTRG.6374	AT5G63850		24.2.3.4.1	Solute transport.carrier-mediated transport.APC superfamily.AAAP family.amino acid transporter (AAP-type)
MSTRG.10264	AT3G11900		24.2.3.4.4	Solute transport.carrier-mediated transport.APC superfamily.AAAP family.amino acid transporter (ANT-type)
MSTRG.10924	AT3G13620		24.2.3.5.1	Solute transport.carrier-mediated transport.APC superfamily.APC family.amino acid transporter (LAT-type)
MSTRG.19600	AT4G19960		24.2.3.11	Solute transport.carrier-mediated transport.APC superfamily.HAK/KUP/KT potassium cation transporter
MSTRG.22742	AT4G19960		24.2.3.11	Solute transport.carrier-mediated transport.APC superfamily.HAK/KUP/KT potassium cation transporter
MSTRG.16225	AT1G23090		24.2.3.1.1	Solute transport.carrier-mediated transport.APC superfamily.SulP family.sulfate transporter (SULTR-type)
MSTRG.18967	AT5G10180		24.2.3.1.1	Solute transport.carrier-mediated transport.APC superfamily.SulP family.sulfate transporter (SULTR-type)
MSTRG.26620	AT5G10180		24.2.3.1.1	Solute transport.carrier-mediated transport.APC superfamily.SulP family.sulfate transporter (SULTR-type)
MSTRG.27741	AT1G23090		24.2.3.1.1	Solute transport.carrier-mediated transport.APC superfamily.SulP family.sulfate transporter (SULTR-type)
MSTRG.9764	AT2G17500		24.2.5.2.2	Solute transport.carrier-mediated transport.BART superfamily.AEC family.auxin transporter (PILS-type)
MSTRG.15396	AT2G38170		24.2.9.1.1	Solute transport.carrier-mediated transport.CDF superfamily.CaCA family.cation antiporter (CAX-type)
MSTRG.22237	AT1G79520		24.2.9.2.2	Solute transport.carrier-mediated transport.CDF superfamily.CDF family.manganese cation transporter (Mn-CDF-type)
MSTRG.9753	AT1G16310		24.2.9.2.2	Solute transport.carrier-mediated transport.CDF superfamily.CDF family.manganese cation transporter (Mn-CDF-type)
MSTRG.32593	AT1G70260		24.2.1.5	Solute transport.carrier-mediated transport.DMT superfamily.UmamiT-type solute transporter
MSTRG.7684	AT2G40900		24.2.1.5	Solute transport.carrier-mediated transport.DMT superfamily.UmamiT-type solute transporter
MSTRG.19360	AT5G47560	TDT	24.2.7.1.1	Solute transport.carrier-mediated transport.IT superfamily.DASS family.di-/tricarboxylate transporter (TDT-type)
MSTRG.10635	AT3G19490	NHD	24.2.7.3	Solute transport.carrier-mediated transport.IT superfamily.proton:sodium cation antiporter (NHD-type)
MSTRG.31814	AT2G46320		24.2.13	Solute transport.carrier-mediated transport.MC-type solute transporter
MSTRG.14005	AT2G32040		24.2.2.5	Solute transport.carrier-mediated transport.MFS superfamily.BT1 small solute transporter
MSTRG.13189	AT2G02040		24.2.2.9	Solute transport.carrier-mediated transport.MFS superfamily.NRT1/PTR anion transporter
MSTRG.17041	AT3G53960		24.2.2.9	Solute transport.carrier-mediated transport.MFS superfamily.NRT1/PTR anion transporter
MSTRG.21563	AT1G32450	NRT1.5	24.2.2.9	Solute transport.carrier-mediated transport.MFS superfamily.NRT1/PTR anion transporter
MSTRG.530	AT1G52190	NRT1.11	24.2.2.9	Solute transport.carrier-mediated transport.MFS superfamily.NRT1/PTR anion transporter
MSTRG.15581	AT1G08930	EDR6	24.2.2.1.5	Solute transport.carrier-mediated transport.MFS superfamily.SP family.monosaccharide transporter (ERD6-type)
MSTRG.2607	AT1G19450		24.2.2.1.5	Solute transport.carrier-mediated transport.MFS superfamily.SP family.monosaccharide transporter (ERD6-type)
MSTRG.4048	AT3G05400		24.2.2.1.5	Solute transport.carrier-mediated transport.MFS superfamily.SP family.monosaccharide transporter (ERD6-type)
MSTRG.9945	AT5G27350		24.2.2.1.5	Solute transport.carrier-mediated transport.MFS superfamily.SP family.monosaccharide transporter (ERD6-type)
MSTRG.16456	AT4G36670	PTL6	24.2.2.1.7	Solute transport.carrier-mediated transport.MFS superfamily.SP family.polyol/monosaccharide transporter (PLT-type)
MSTRG.5533	AT4G26590		24.2.10.2	Solute transport.carrier-mediated transport.OPT family.oligopeptide transporter (OPT-type)
MSTRG.25921	AT1G32080	PLGG1	24.2.19	Solute transport.carrier-mediated transport.PLGG1 glycerate:glycolate transporter
MSTRG.28675	AT1G32080	PLGG1	24.2.19	Solute transport.carrier-mediated transport.PLGG1 glycerate:glycolate transporter
MSTRG.8418	AT3G19640		24.3.3.1	Solute transport.channels.CorA family.MRS/MGT metal cation transporter
MSTRG.3625	AT2G45960		24.3.1.2	Solute transport.channels.MIP family.plasma membrane intrinsic protein (PIP-type)
MSTRG.13344	AT1G30360		24.3.7	Solute transport.channels.OSCA calcium-permeable channel
MSTRG.25365	AT4G04340		24.3.7	Solute transport.channels.OSCA calcium-permeable channel
MSTRG.17486	AT1G12480		24.3.10	Solute transport.channels.SLAC anion channel
MSTRG.17462	AT5G54250		24.3.2.4	Solute transport.channels.VIC superfamily.cyclic nucleotide-gated cation channel (CNGC-type)
MSTRG.232	AT2G28260		24.3.2.4	Solute transport.channels.VIC superfamily.cyclic nucleotide-gated cation channel (CNGC-type)
MSTRG.27242	AT4G03560		24.3.2.1	Solute transport.channels.VIC superfamily.voltage-gated calcium cation channel (TPC-type)
MSTRG.20677	AT4G38920		24.1.1.1.2	Solute transport.primary active transport.V-type ATPase complex.membrane V0 subcomplex.subunit c
MSTRG.10877	AT4G11150		24.1.1.2.5	Solute transport.primary active transport.V-type ATPase complex.peripheral V1 subcomplex.subunit E
Cluster 15				
MSTRG.17903	AT2G26900	BASS2	24.2.5.1.1	Solute transport.carrier-mediated transport.BART superfamily.BASS family.BAT 2-keto acid transporter
MSTRG.18120	AT3G26670		24.2.1.4	Solute transport.carrier-mediated transport.DMT superfamily.NIPA-type solute transporter
MSTRG.3286	AT4G35335		24.2.1.1.7	Solute transport.carrier-mediated transport.DMT superfamily.NST-TPT group.CSTLP nucleotide sugar transporter
MSTRG.10909	AT5G46800		24.2.13	Solute transport.carrier-mediated transport.MC-type solute transporter
MSTRG.16793	AT5G48970		24.2.13	Solute transport.carrier-mediated transport.MC-type solute transporter
MSTRG.1328	AT5G62730		24.2.2.9	Solute transport.carrier-mediated transport.MFS superfamily.NRT1/PTR anion transporter
MSTRG.16457	AT4G36670	PTL6	24.2.2.1.7	Solute transport.carrier-mediated transport.MFS superfamily.SP family.polyol/monosaccharide transporter (PLT-type)
MSTRG.26625	AT5G44050		24.2.4.1.1	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX-type)
MSTRG.2953	AT1G73700		24.2.4.1.1	Solute transport.carrier-mediated transport.MOP superfamily.MATE family.metabolite transporter (DTX-type)
MSTRG.12499	AT1G19910		24.1.1.1.2	Solute transport.primary active transport.V-type ATPase complex.membrane V0 subcomplex.subunit c
MSTRG.22331	AT4G38920		24.1.1.1.2	Solute transport.primary active transport.V-type ATPase complex.membrane V0 subcomplex.subunit c
MSTRG.8050	AT2G25610		24.1.1.1.2	Solute transport.primary active transport.V-type ATPase complex.membrane V0 subcomplex.subunit c
MSTRG.1650	AT1G12840		24.1.1.2.3	Solute transport.primary active transport.V-type ATPase complex.peripheral V1 subcomplex.subunit C
MSTRG.26494	AT3G58730		24.1.1.2.4	Solute transport.primary active transport.V-type ATPase complex.peripheral V1 subcomplex.subunit D
MSTRG.14323	AT4G02620		24.1.1.2.6	Solute transport.primary active transport.V-type ATPase complex.peripheral V1 subcomplex.subunit F
MSTRG.27351	AT4G23710		24.1.1.2.7	Solute transport.primary active transport.V-type ATPase complex.peripheral V1 subcomplex.subunit G

VI. Acknowledgements

Studying PhD is a process of not only acquiring knowledge, but also self-development and improvement. I'd like to express my sincere thanks to those who have helped, supported, and believed in me through this journey.

I am thankful to my supervisor Prof. Dr. Andreas P.M. Weber for giving me the opportunity to pursue my PhD in his lab and for believing in my capabilities.

I appreciate the scientific guidance from my post-doctoral advisor Dr. Urte Schlüter and her time investigated in my annual reports, posters, and thesis.

I'd like to thank my committee members: Prof. Dr. Miltos Tsiantis for his advice and guidance through committee meetings; Prof. Dr. Benjamin Stich for the perspective he brought to my research and for reviewing my thesis.

Further, I appreciate the community support from the International Max Planck Research Schools and the iGRADplant. It was a great benefit to me to attend the well-organized workshops, annual retreats, and seminars with IMPRS and iGRADplant peers. I'd like to thank iGRADplant coordinator Dr. Petra Fackendahl for her support during my studies.

I'm especially thankful to Dr. Otho Mantegazza for his patient and kind guidance that I learned a lot about bioinformatics.

I'm really grateful to Dr. Alisandra Denton for her kind help with my bioinformatics questions and for reviewing the manuscript for leaf development.

For nice lunches in Mensa, sharing thoughts, and social activities discovering nature and food, I'd like to thank Dr. Florian Hahn and Dr. Dagbegnon Angelo Romeo Agossou Yao as well as Dr. Shirin Zamani Nour, Dr. Tatjana Goss, Dr. Dagmar Lyska, and Raja Stracke.

I also appreciate the support from Weber lab: Frau Anja Nöcker helped me with all the administrative work and always cared about my health; Samantha Flachbart introduced me details of lab rules.

For interesting quiz time during lunch and chats, I appreciate the cool team of Marc-Sven Röhl, Nils Koppers, Vanessa Reichel-Deland, Lennart Charton, Vanessa Reichel-Deland, Thomas Wrobel, and Anastasija Plett. Special thank to Eva Maleckova and Franziska Kuhnert for being so great officemates and for sharing thoughts and sweets.

VI. Acknowledgements

I'd like to sincerely thank my great and thoughtful friends: Dr. Kumari Billakurthi and Daniela Lichtblau for the wonderful time we shared and cares about me. For nice chats and warm cares, I'm thankful to Dr. Anastasiia Bovdilova.

For always encouraging, supporting and believing in me, I'm especially grateful to Ying-Shuan Chen, Tzu-Yi Yang, Su-Ting Cheng, Po-Ya Wu, Hsin-Ya Huang, Hsin-Chieh Lu, Jou-Yi Li, and Ying-Ru Li. I'm thankful to Shu-Min Kao for sharing his expertise in bioinformatics and for going through the difficult phases of PhD studies together. I'd like to thank Joshua for the endless support and encouragement during the hard time of my studies.

Lastly, I could not have fulfilled my dream and made it so far without my mom, Tsui-Hua Hsiao. I deeply appreciate her unconditional love and support. I'm thankful to my brothers, Ting-Chun Lin and Cheng-Wei Lin, and all the other friends who I don't have space to name.