

University of Sheffield

Drivers of C₄ evolution in the grasses

by

Ahmed Salem F Alenazi

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Sciences

School of Biosciences

June 2024

Table of contents

Acknowledgements	vi
Summary	vii
1. General Introduction	2
1.1. The evolution of complex traits.....	2
1.2. How genetic variation can contribute to complex trait evolution.....	3
1.3. Photosynthesis.....	5
1.4. Photorespiration.....	7
1.5. C ₄ photosynthesis.....	10
1.6. C ₃ -C ₄ intermediates bridge the transition from C ₃ to C ₄	12
1.7. Identifying the genetic basis of C ₄ photosynthesis.....	13
1.7.1. Gene duplication and C ₄ photosynthesis.....	14
1.8. Using quantitative genetics to identify novel C ₄ genes.....	15
1.9. Using grasses as a study system for C ₄ evolution.....	16
1.10. <i>Alloteropsis semialata</i> as a study system.....	17
1.11. Thesis aims.....	19
1.12. References.....	23
Chapter 2. Leaf anatomy explains the strength of C₄ activity within the grass species <i>Alloteropsis semialata</i>	34
2.1. Abstract.....	35
2.2. Introduction.....	36
2.3. Materials and Methods.....	39
2.3.1. Plant sampling and carbon isotopes.....	39
2.3.2. Genome scan and phylogenetic analyses.....	41
2.3.3. Microscopy and leaf anatomical measurements.....	42
2.3.4. Statistical analyses.....	44
2.4. Results.....	45
2.4.1. Plant sampling captures a range of carbon isotope ratios.....	45
2.4.2. Ecology affects the strength of the C ₄ cycle.....	46
2.4.3. Phylogenetic lineages of non-C ₄ <i>A. semialata</i> occupy different regions across the Central Zambebian miombo woodlands.....	48
2.4.4. Variation in leaf anatomy explains carbon isotope ratios.....	50
2.5. Discussion.....	52
2.5.1. Anatomical variation is related to the strength of the C ₄ cycle in non-C ₄ plants.....	52
2.5.2. Transition to stronger C ₄ is associated with high temperatures.....	56
2.6. Conclusion.....	58

2.7. References.....	60
Chapter 3. Identifying genomic regions associated with C₄ photosynthetic activity and leaf anatomy in <i>Alloteropsis semialata</i>.....	73
3.1. Abstract.....	75
3.2. Introduction.....	76
3.3 Materials and Methods.....	79
3.3.1 Genome data, δ ¹³ C values, and population genetic analyses.....	79
3.3.2 Leaf anatomical traits of C ₃ +C ₄ <i>A. semialata</i>	81
3.3.3 Estimating trait heritability.....	82
3.3.4 Genome-wide association study of photosynthetic traits.....	83
3.3.5 Linkage disequilibrium.....	83
3.3.6 Identification of candidate genes.....	84
3.4 Results.....	85
3.4.1 Population structure.....	85
3.4.2 Identifying regions of the genome correlated with the strength of the C ₄ cycle.....	87
3.4.3 Identifying regions of the genome associated with C ₄ leaf anatomy in the C ₃ +C ₄ intermediates.....	92
i. Bundle sheath distance.....	92
ii. Inner bundle sheath fraction (IBSF).....	93
3.5. Discussion.....	95
3.5.1. Genetic basis of the carbon isotope ratio (δ ¹³ C) in <i>Alloteropsis semialata</i>	96
3.5.2. The genetic basis of C ₄ leaf anatomy.....	99
3.5.3. Plasmodesmata and reduced distance between bundle sheaths.....	100
3.5.4 The genetic basis of the inner bundle sheath fraction in <i>Alloteropsis semialata</i>	101
3.6. Conclusion.....	102
3.7. References.....	104
Chapter 4. Gene duplication and the repeated origin of C₄ photosynthesis in PACMAD grasses.....	120
4.1. Abstract.....	121
4.2. Introduction.....	122
4.3. Materials and methods.....	127
4.3.1. Tissue collection and DNA extractions.....	127
4.3.2. Organelle assembly.....	127
4.3.3. Nuclear assembly.....	128
4.3.4. Genome annotation.....	129
4.3.5. Identifying orthologous genes and identifying duplication events.....	130
4.3.6. Gene ontology.....	130
4.4. Results.....	131
4.4.1. Assembly of two early diverging PACMAD grasses.....	131
4.4.2. Duplication events at the base of PACMAD clade.....	133

4.5. Discussion.....	135
4.5.1. Duplication of a photorespiratory enzyme and a C ₂ pump.....	136
4.5.2. Cell wall modifications may limit CO ₂ diffusion and water use efficiency.....	136
4.6. Conclusion.....	139
4.7. References.....	140
5. General discussion.....	149
5.1. Understanding the evolutionary steps to C ₄ carbon fixation.....	150
5.2. Will these results be useful for engineering C ₄ photosynthesis in C ₃ species?.....	153
5.3. Anatomical measurements of C ₃ and C ₄ would enrich the analysis.....	154
5.4. What is the role of phenotypic plasticity in C ₄ leaf anatomy?.....	155
5.5. Would increasing the density of genome wide markers identify new C ₄ loci?.....	156
5.6. Could structural variation play a role in C ₄ leaf anatomy evolution?.....	157
5.7. Verifying the function of our candidate genes in the evolution of C ₄ leaf anatomy.....	157
5.8. Conclusions.....	158
5.9. References.....	159

Acknowledgements

First, and foremost, I would like to express my gratitude to my primary supervisor Dr. Luke Dunning. I am deeply thankful for his generous support over the years. His constant enthusiasm and passion for science have been truly inspiring. I greatly appreciate his dedication and unwavering readiness to assist, which has significantly contributed to the progress of this work. I would like to extend my gratitude to my co-supervisor, Professor Colin Osborne, for his valuable advice in every aspect. His generosity and willingness to assist have made a significant impact, and I truly appreciate his valuable discussions and insightful comments that have contributed to my research.

I sincerely appreciate Dr. Lara Pereira for dedicating her valuable time that she spent guiding and supporting me and for generously providing data and ideas, which significantly enhanced this work. Special thanks to Dr. Pascal-Antoine Christin, my primary supervisor. He provided me with all the support and advice I needed when I first arrived and set me on the right path to continue in graduate research. He did not hesitate to offer any kind of assistance he could. I am deeply grateful to Dr. Matheus Bianconi for his assistance during my first year. He kindly trained me in histology lab work.

I owe a profound debt of gratitude to my beloved family for their unwavering support and immense sacrifices throughout my studies, particularly my wife and children, who continually encouraged me to complete this project. I would like to express my deep gratitude to the Ministry of Education in Saudi Arabia and Northern Border University in Saudi Arabia for their generous financial support for my PhD studies. Additionally, I extend heartfelt thanks to the University of Sheffield for their hospitality throughout my research, and to the staff in the School of Biosciences for their kind assistance.

Summary

Biologists have long been fascinated by the biological complexity that exists today, with organisms displaying a remarkable diversity of adaptations that have enabled them to thrive in almost every environment. Many of the most impressive adaptations would be classified as complex traits, with structural and metabolic elements working in synergy. To evolve, multiple elements need to be rewired, which can be achieved by incremental modifications over successive generations.

Understanding evolutionary steps leading to the emergence of a complex trait can help solve the evolutionary complexity puzzle. The C_4 photosynthetic pathway is a complex adaptation involving multiple biochemical and anatomical changes that have allowed certain plant lineages to succeed in specific environmental conditions, including warm and arid climates. Here, we aimed to investigate the evolutionary modifications required to construct the C_4 cycle, a complex trait. I initially focused on examining the remarkable photosynthetic variation in the grass *Alloteropsis semialata*, a species unique in having both C_4 and non- C_4 photosynthetic genotypes. In my research, I conducted a comparative analysis to identify leaf traits associated with the proportion of carbon fixed using the C_4 cycle. The findings revealed that plants with higher C_4 activity generally have a greater ratio of photosynthetically active bundle sheath tissue. Subsequently, I employed a genome-wide association study (GWAS) to identify several candidate genes associated with the strength of the C_4 cycle and enhancing the proportion of bundle sheath tissue. Finally, in the final chapter of my research, I adopted a more comprehensive perspective to investigate the genetic precursors that may have enabled the repeated evolution of C_4 photosynthesis in the PACMAD grasses. I found that genes associated with cell wall modifications and stomatal aperture were duplicated at the base of the PACMAD clade, potentially facilitating the repeated convergent evolution of Kranz anatomy. Collectively, all three data chapters highlight the importance of anatomical modifications in the emergence of the C_4 cycle. Overall my thesis identifies several key changes required for C_4

evolution, and the results may be relevant to engineering C_4 in C_3 species, such as rice.

Chapter 1. General Introduction

1. General Introduction

1.1. The evolution of complex traits

Biologists generally try to explain the evolution of complex traits via a series of gradual changes that happen slowly, with natural selection favouring each consecutive intermediate stage (Darwin 1859). In this manner organisms have evolved countless complex traits to adapt to their surroundings and increase their fitness (Suzuki, 2017), with textbook examples including wings for flight and the camera eye to see. The development of complex traits typically relies on a suite of individual features that must all be present for the trait to function as a whole, such as biochemical and morphological adaptations that are co-opted to function as a unit (Geary et al., 2010; Stayton, 2015; Lau and Oakley, 2020).

Complex traits do not appear instantaneously, and the transition from the ancestral state to a complex one requires passing through intermediate stages where only some of the constituent elements may be present (Shapiro, 2023). Theoretically, each intermediate stage should bridge the gap between the ancestral and complex trait by offering an incremental advantage, or at least not reducing fitness, to be retained by natural selection (Darwin, 1859). The development of complex traits is therefore more analogous to the sequential evolution of a series of simple traits, with random mutations that modify a phenotype that is selected for, and over time these incremental mutations accumulate (Barton and Turelli, 1989; Flatt, 2020). Whilst it is known theoretically how complex traits develop, there are relatively few instances where the evolutionary path that a specific complex trait took has been resolved (Danowitz et al., 2015). Phenotypic analysis can help our understanding of how intermediate stages are connected, particularly when looking at whole organism traits that are documented in the fossil record, e.g. the intermediate forms of aquatic mammals (Jablonski and Shubin, 2015). However, the genetic basis of these changes, or how less

visible phenotypes occur (e.g. the evolution of complex metabolic pathways) is still poorly understood.

1.2. How genetic variation can contribute to complex trait evolution

All traits are predominately established by natural selection selecting for genetic variation that can increase fitness. The simplest form of DNA variation among individuals are point mutations, where the substitution of one single nucleotide for another occurs, often referred to as a single nucleotide polymorphism' (SNP). Genome wide, a majority of these SNPs occur in non-coding regions and are either selectively neutral or deleterious (Halushka et al., 1999). However, occasionally nonsynonymous SNPs that are a type of genetic variation where a single nucleotide change in the DNA sequence leads to an alteration in the amino acid sequence of a protein that can alter functions can be positively selected for, such as those that resulted in adaptive amino acid substitutions in vertebrate opsin genes that are associated with improved colour vision (Hagen et al., 2022). Whilst we know that point mutations can alter phenotypes, the rate of mutation means that developing a new phenotype purely through adaptive amino acid substitution seems impossibly slow, especially in large populations. Other sources of genetic variation that can indirectly accelerate the evolution of complex traits include changes to gene expression levels and gene duplication.

Gene expression variation can modify phenotypes and is often associated with rapid adaptive divergence requiring relatively few genetic changes (King and Wilson, 1975; Wray, 2003; Duarte et al., 2006; Ganko et al., 2007; Fay and Wittkopp, 2008; Babbitt, Haygood, Nielsen, and Wray, 2017). In short, gene expression can be a bridge between genotype (genetic makeup) and phenotype (observable traits), mediating the complex interactions with environmental factors that together culminate in an organism's trait expression (Hamann et al., 2012). A high gene expression can

facilitate adaptive transitions and typically arises from duplication events and gene dosage effects (Ohno, 1970; True and Carroll 2002; Zhang, 2003; Kliebenstein, 2008; Kaessmann et al., 2009; Bianconi et al., 2020).

Duplication events occur through a variety of processes, and they can affect individual loci (e.g. unequal crossing over) or even whole genomes (polyploidisation). Once duplicated, these genes evolve along separate evolutionary paths (Lynch and Conery, 2003). The duplicated gene can be lost from the genome, it can work in concert with the original copy to increase expression (dosage effect), split the function of the ancestral gene (subfunctionalization) or even diverge to generate a novel function (neofunctionalization). Gene duplication is known to be particularly important for complex trait evolution, e.g. the visual diversification of visual pigments in fish (Hofmann and Carleton, 2009), and Homeobox (HOX) gene duplication and the diversification of animal body plans (Akam, 1995).

For my thesis I focus on the evolution of C_4 photosynthesis, a superb model of convergent evolution of a complex trait that has originated independently over 60 times in plants (Figure 1.1 adapted from Sage et al., 2011). C_4 photosynthesis is believed to have evolved via the gradual accumulation of multiple biochemical and anatomical components, increasing the efficiency of photosynthetic CO_2 fixation in high temperature environments (Hatch and Slack, 1966; Hatch 1978). However, our understanding of this process is hindered by the fact that photosynthetic types are typically segregated among different species that diverged a long time ago. This makes it difficult to distinguish changes that occurred during the origin of C_4 from those which accumulated later (Heyduk et al., 2019).

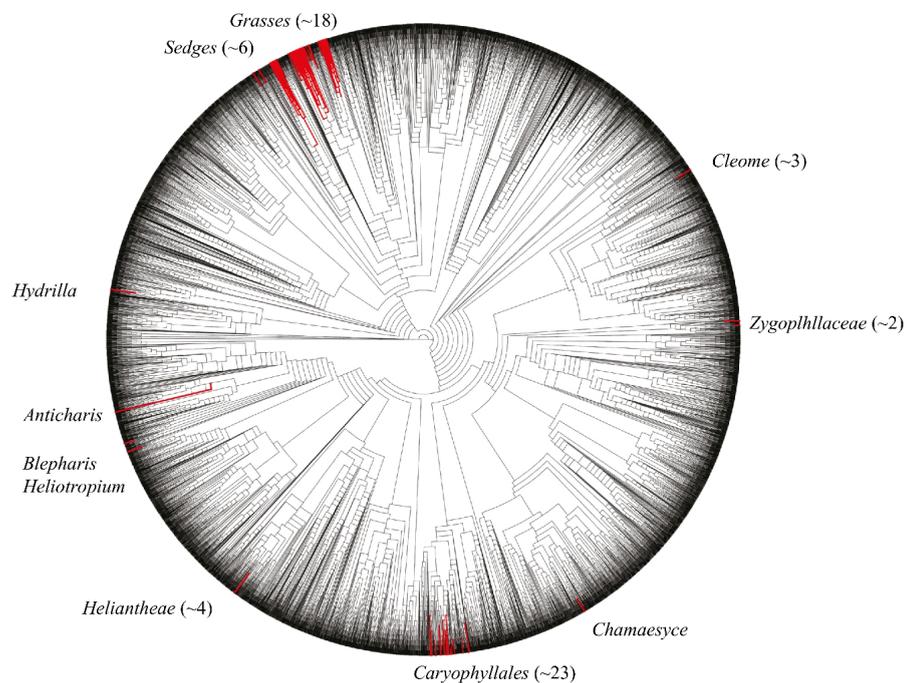


Figure 1.1

The convergent evolution of C₄ photosynthesis. The phylogenetic distribution of C₄ photosynthetic lineages in angiosperms is shown in red. The number of independent origins is indicated between parentheses within each taxonomic group. The figure was adapted from Sage et al., (2011).

1.3. Photosynthesis

Solar energy is the ultimate origin of almost all energy on earth, which serves as a fundamental cornerstone for sustaining life on the planet, driving the processes of primary production. The primary mechanism by which autotrophic organisms convert solar energy into chemical energy in the form of organic compounds is called photosynthesis. Autotrophic organisms such as plants, algae and cyanobacteria carry out photosynthesis to convert light energy into chemical energy using water (H₂O), and carbon dioxide (CO₂) as inputs to produce simple sugars and oxygen. In plants, photosynthesis primarily occurs in the mesophyll cells of leaves, where CO₂ is fixed directly in the chloroplasts, a photosynthetic organelle. In C₃ photosynthesis, which is the most common pathway

among plants, CO₂ is initially fixed into a three-carbon compound, 3-phosphoglycerate (3PG), hence the name C₃. All parts of C₃ photosynthesis occur within the leaf mesophyll cells. The reactions can be separated into two major steps, the light and dark reactions (Calvin–Benson–Bassham). C₃ photosynthesis is characterised by its reliance on the Calvin cycle that occurs in mesophyll, where the enzyme Rubisco plays a critical role in fixing CO₂. This pathway is best suited to environments where temperature and light intensity are moderate, particularly in temperate regions.

Light-dependent reactions take place in the thylakoid membranes of the chloroplasts and use light energy to generate ATP from ADP and inorganic phosphate, as well as NADPH from NADP⁺ (Romanowska and Wasilewska-Dębowska, 2022). The dark reaction takes place in the chloroplast stroma and depend on the enzyme ribulose -1,5 bisphosphate carboxylase/oxygenase (Rubisco), which is the primary photosynthetic carboxylase, enzyme that is crucial for photosynthesis, and is the most abundant enzyme on earth (Ellis, 1979). The first resulting molecule after CO₂ is fixed by RuBisco is an unstable 6-carbon compound that immediately is converted to two 3-phosphoglycerate (3PG) molecules (Merlo et al., 1993). Subsequently, multiple steps requiring NADPH and ATP produced by light reactions lead to the storage of solar energy as glyceraldehyde-3-phosphate (G3P), the precursor for sugar molecule synthesis (Miziorko and Loimer, 1983). Anatomically, C₃ plants typically have a leaf structure with abundant mesophyll tissue that is the primary site of photosynthesis, with chloroplasts evenly distributed. There is no distinction between cells involved in the initial CO₂ fixation and those where the Calvin cycle operates (Figure 1.2; Sage, 2004).

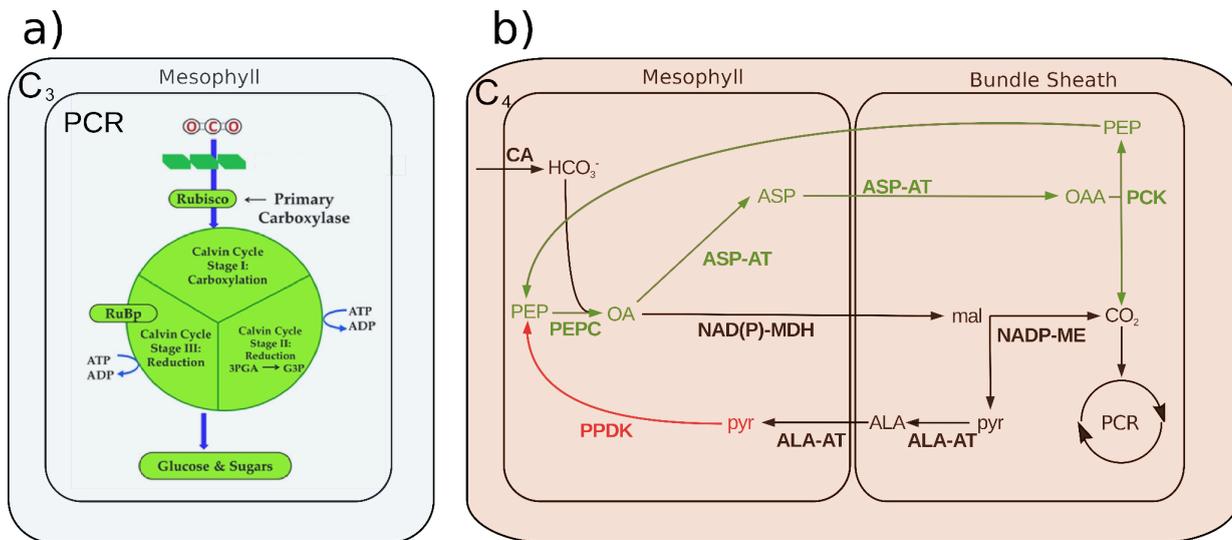


Figure 1.2

(a) Calvin cycle in C₃ plants. (b) pathway of metabolism during C₄ in *Alloteropsis semialata*. C₃ diagram is adapted from Ghazzawy et al., 2024 and C₄ diagram is adapted from Dunning et al., 2019.

Under certain environmental conditions, including high light intensity, drought and high temperature, the concentration of O₂ relative to CO₂ concentration inside the leaf increases due to stomatal closure to save water. This favours the oxygenase activity of RuBisCO, which can also catalyse O₂ fixation, as well CO₂ assimilation. This results in the wasteful photorespiratory pathway that may reduce photosynthetic efficiency (Bowes et al., 1971; Jordan and Ogren, 1984). The first result of the photorespiratory cycle, through a series of reactions that consume ATP, is glycine that is converted to serine, NH₃ and CO₂ (Bowes et al., 1971; Lorimer and Andrews, 1973; Lorimer and Andrews, 1973; Chollet and Ogren 1975 Ogren, 1984). Due to photorespiration, C₃ photosynthesis is inefficient in hot and dry environments due to the lost energy (Sage and Monson, 1999).

1.4. Photorespiration

Photorespiration is initiated by the enzyme RuBisCO, which typically facilitates the fixation of CO₂ in the Calvin–Benson–Bassham cycle but can also act as an oxygenase under conditions of high O₂

concentrations. In such cases, RuBisCO catalyses the reaction with oxygen, leading to the formation of 3-phosphoglycerate (3-PGA) and 2-phosphoglycolate (2-PG), the latter being a harmful byproduct that cannot proceed through the Calvin–Benson–Bassham cycle. To mitigate this, 2-phosphoglycolate is dephosphorylated in the chloroplast by the enzyme phosphoglycolate phosphatase, producing glycolate. This glycolate is then transported to the peroxisome, where it is oxidised by glycolate oxidase to generate glyoxylate and hydrogen peroxide (H_2O_2). Glyoxylate is subsequently converted into glycine through the action of glutamate aminotransferase. The glycine is then shuttled to the mitochondria, where two glycine molecules are converted into one molecule of serine, along with the release of CO_2 and ammonia (NH_3), through the activities of the glycine decarboxylase complex (GDC) and serine hydroxymethyltransferase (SHMT). Although the CO_2 released can be recaptured by RuBisCO in the chloroplasts, this process requires significant energy, resulting in a net carbon loss and the consumption of ATP and reducing power. Subsequently, serine is transported back to the peroxisome, where it is converted into hydroxypyruvate by serine aminotransferase and then reduced to glycerate by hydroxypyruvate reductase. Glycerate is then returned to the chloroplast, where it is phosphorylated by glycerate kinase to reform 3-phosphoglycerate (3-PGA), allowing it to re-enter the Calvin–Benson–Bassham cycle. The energy demands of photorespiration are substantial, potentially using up to 50% of the energy obtained from photosynthesis, thus rendering it an inefficient process under certain environmental conditions (Bauwe et al., 2010; Sage and Sage, 2009). However, recent studies have redefined this perspective by demonstrating that photorespiratory intermediates, such as glycine, play a vital role in supporting optimal leaf function. During light transitions, glycine accumulation enhances net carbon assimilation, thereby contributing to the efficiency and adaptability of photosynthesis under fluctuating environmental conditions. This emerging evidence underscores the functional

importance of photorespiration in maintaining the stability and performance of the photosynthetic apparatus (Figure 1.3; Busch et al., 2013; Busch et al., 2018; Fu and Walker, 2024).

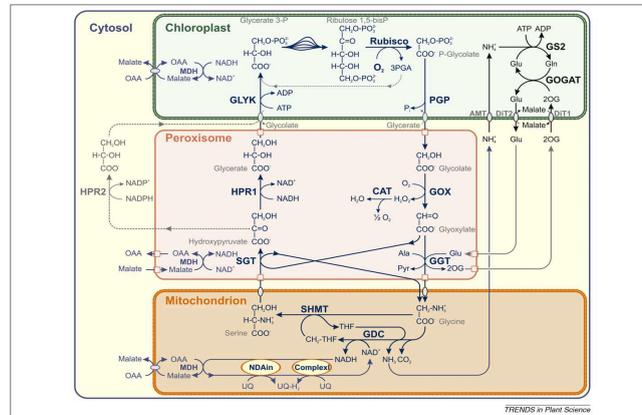


Figure 1.3

The metabolic steps across multiple sub-cellular compartments of photorespiration. The figure was adapted from Bauwe et al., 2010.

To potentially reduce energy waste associated with photorespiration, some plants have developed CO₂-concentrating mechanisms (CCMs), which may enhance photosynthetic efficiency under certain conditions (Hatch, 1971; Edwards and Ku, 1987; Leegood et al. 1995), which improve the efficiency of photosynthesis during conditions that favour photorespiration (Figure 1.3, adapted from Yamori et al., 2013). The main role of CCMs is to transport and release CO₂ around Rubisco either through temporal separation of primary -CO₂ fixation and release, such as crassulacean acid metabolism (CAM), or spatial separation of carbon-fixation and release, such as C₄ photosynthesis, thereby reducing the energy cost of photorespiration (Figure 1.4). In contrast to C₃ photosynthesis, these mechanisms allow certain plants to thrive in environments where C₃ plants would suffer significant reductions in photosynthetic efficiency due to high rates of photorespiration.

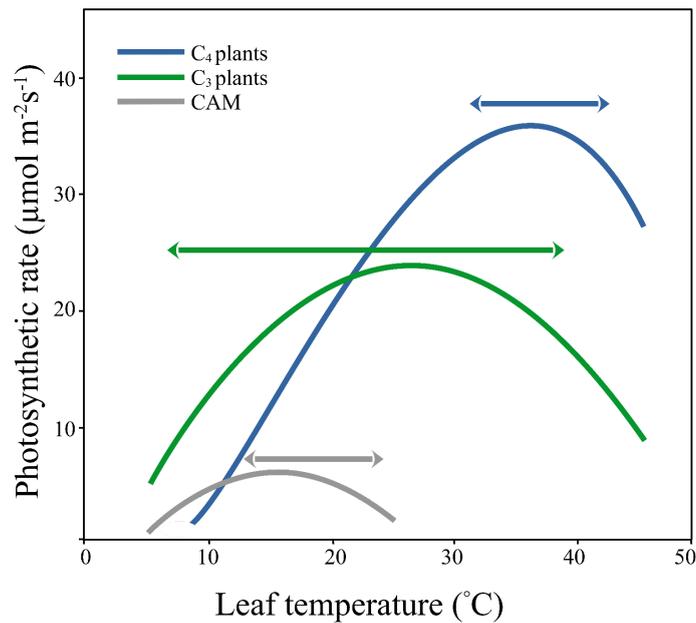


Figure 1.4

C_3 is less efficient at high temperatures. Standard temperature reactions of photosynthesis in C_3 , C_4 , and CAM plants, arrows represent the potential range of T_{opt} . This figure is adapted from Yamori et al., (2013).

1.5. C_4 photosynthesis

The C_4 photosynthetic pathway takes place in two neighbouring cell types, therefore a modified leaf anatomy is needed to support the spatial segregation of the C_4 cycle and enable rapid metabolite transfer (Dengler et al., 1994; Hattersley, 1984; Lundgren et al., 2014). According to Hattersley and Watson (1975), the mesophyll and bundle sheath must be close to each other, ideally with no more than a single cell between them (Figure 1.5). Moreover, the large fraction of the leaf compartment (bundle sheath) that is housing the Calvin–Benson–Bassham cycle must be expanded to fix CO_2 that is released from decarboxylation (Brown and Hattersley, 1989; Dengler and Nelson, 1999). Finally, there are many enzymes that need to be co-opted to transport C_4 cycle metabolites to the bundle sheath cells, release CO_2 and regenerate the intermediate compounds, all requiring heightened activity in specific leaf cells (Hatch, 1971; Hatch and Osmond, 1976).

In C₄ species, atmospheric CO₂ in the mesophyll is first converted into HCO₃⁻ by carbonic anhydrase, which catalyses the rapid interconversion of CO₂ and water to bicarbonate (HCO₃⁻) and protons, before being fixed by phosphoenolpyruvate carboxylase (PEPC) (Figure 1.2; Hatch and Slack, 1966; Burnell and Hatch, 1988; Ermakova et al., 2020). PEPC, unlike Rubisco, does not have an affinity for O₂ and it is also a superior carboxylase relative to Rubisco (Hatch and Slack, 1966; Webster et al., 1975). The first result of the CO₂ fixation by PEPC is a compound with four carbon atoms (hence its name, C₄ photosynthesis), which is oxaloacetate or malate, that then is transported into the bundle sheath cell via plasmodesmata. In the bundle sheath, CO₂ is released via one of decarboxylation enzymes within this compartment isolated from the atmosphere, and the intermediate compounds of the decarboxylation reactions diffuse back to the mesophyll and are regenerated (Figure 1.2; Hatch, 1987). Effective C₄ plants need to possess a large proportion of bundle sheath that contains chloroplasts with Rubisco (Hattersley and Watson 1975; Lundgren et al. 2014, Alenazi et al., 2023), and a short distance between consecutive bundle sheath cells which might have resulted from reducing the number of mesophyll cells (Christin et al., 2012; Alenazi et al., 2023).

There are three known decarboxylating enzymes in the bundle sheath that can catalyse the release of CO₂ from C₄ acids, and the different subtypes of the C₄ pathway are named based on the enzyme used (Hattersley, 1982; 1992; Hatch, 1987). Most C₄ species use NADP-malic enzyme (NADP-ME) as their primary decarboxylating enzyme (Sage, 1999). The other enzymes are mitochondrial NAD-malic enzyme (NAD-ME) and PEP carboxykinase (PCK) (Hatch et al., 1975). Even within these C₄ subtypes, there are some anatomical differences (Dengler and Nelson, 1999; Edwards and Voznesenskaya, 2011), but it is now acknowledged that all C₄ plants use the three enzymes in different proportions (Furbank, 2011). *Alloteropsis semialata*, for example, predominantly uses the

NADP-ME pathway for its C_4 photosynthetic metabolism. However, this pathway is complemented by the incorporation of a PCK shuttle, which can constitute the principal route for carbon flux within the C_4 cycle of this species (Ueno and Sentoku, 2006). Consequently, while both NADP-ME and PCK pathways are operational, the PCK shuttle is notably significant in mediating the carbon flux (Figure 1.2b; Christin et al., 2013).

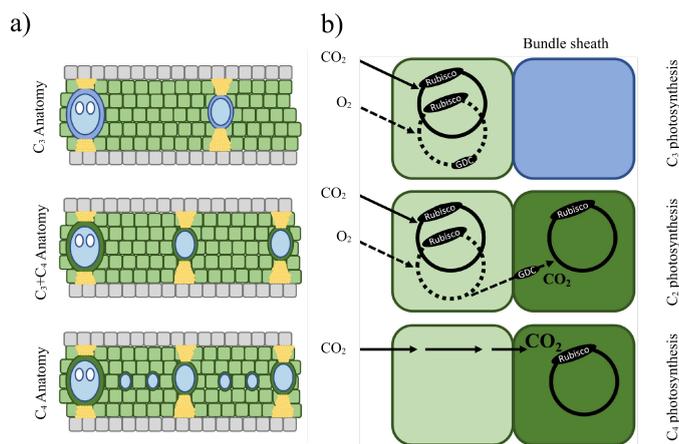


Figure 1.5

Leaf structure and photosynthetic pathways in C_3 , C_3 - C_4 , and C_4 grasses. In C_3 plants, both CO_2 assimilation and release occur in mesophyll cells, resulting in larger mesophyll areas compared to bundle sheath tissue. C_3 - C_4 plants utilise an intermediate physiology known as C_2 photosynthesis, with the Calvin-Benson cycle in mesophyll cells and glycine decarboxylase localised in bundle sheath cells, making a weak CO_2 -concentrating mechanism. C_4 plants employ a robust CO_2 -concentrating mechanism, transferring CO_2 from mesophyll to bundle sheath cells, minimising oxygenation and photorespiration. This process requires significant bundle sheath tissue but less mesophyll, achieved through minor vein insertion. This figure is adapted from Lundgren et al., (2019).

1.6. C_3 - C_4 intermediates bridge the transition from C_3 to C_4

C_3 - C_4 intermediates are taxonomically widely distributed, being found in several plant orders including Asterales, Capparales, Caryophyllales, Poales and *Flaveria* genus (Asterales). At least some of the C_3 - C_4 species are considered to be examples of an active evolutionary transition from C_3 to C_4 (Powell in 1978 ; Edwards and Ku, 1987). The existence of intermediary species exhibiting traits of both C_3 and C_4 photosynthesis allows this transition to be studied (Monson and Moore 1989; Sage, 2004; Sage and Sage et al., 2012). In general, C_3 - C_4 intermediate plants can be

classified into two distinct groups depending on whether the metabolism they exhibit is closer to C_3 or C_4 species. In ‘type I C_3 - C_4 ’ intermediates (Edwards and Ku, 1987), which is also called the C_2 Cycle, the Calvin-Benson cycle takes place in the mesophyll and the expression of a multi-protein complex GDC is enhanced. GDC converts glycine, which comes from the photorespiration cycle, to start CO_2 liberation in the bundle sheath cells. Losing GDC expression from the mesophyll cells separates photorespiration between two compartments and thus increases CO_2 in the bundle sheath cells (Figure 1.5; Rawsthorne, 1992; Sage et al., 2012; Schulze et al., 2016). Hence, this localization drives the photorespiratory cycle to be split between mesophyll and bundle sheath cells, resulting in a modest CO_2 -concentrating mechanism

The ‘type II C_3 + C_4 intermediates’ fix part of their carbon using the C_4 cycle through having increased PEPC, pyruvate phosphate dikinase (PPDK) and NADP-ME activities, but these enzymatic activities remain notably lower than those observed in integrated full C_4 plant (Edwards and Ku, 1987). Thus, the transition from type II to fully C_4 largely happens via the modification to the expression levels of key enzymes of C_4 cycle and the compartmentalisation of the different steps between mesophyll and bundle sheath cells (Engelmann et al., 2003).

1.7. Identifying the genetic basis of C_4 photosynthesis

Many researchers are engaged in studying the genetic basis of C_4 photosynthesis, and the core metabolic enzymes are well characterised (Broglie et al., 1984; Glackin and Grula, 1990; Burnell and Hatch, 1988; Hermans and Westhoff 1990; Lepiniec et al., 1993). It has been found that the underlying C_4 genes exhibit cell-type specific expression and over time, selection favours the gradual increase in their expression (Heckmann et al., 2013).

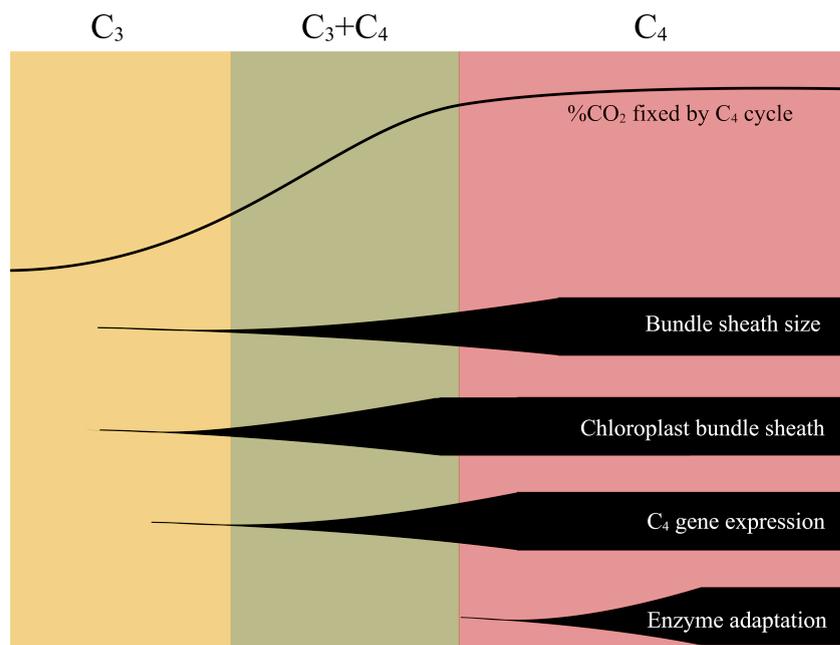


Figure 1.6

The order of the theoretical modifications during the evolution shift from C_3 to C_4 (Adapted from Dunning et al., 2017).

Once these enzymes have been co-opted for a function in C_4 photosynthesis, several have been found to have undergone positive selection for improved function (Kapralov et al., 2012; Rosnow et al., 2014); for instance, Rubisco, PEPC and NADP-ME associated with C_4 have all been shown to be evolving under positive selection (Wang et al., 2009; Rosnow et al., 2014, Orr et al., 2016). Positive selection is believed to have led to convergent amino acid substitutions in independent origins of C_4 (Christin et al., 2007; Watson-Lazowski et al., 2018) and also been used to infer when a full C_4 physiologically evolved (Dunning et al., 2019).

1.7.1. Gene duplication and C_4 photosynthesis

Gene duplication events have been observed in all angiosperm lineages, and they are a prominent feature of plant evolution (Adams and Wendel 2005; Soltis et al. 2009). However, gene duplication

occurs randomly, and the retention of duplicate genes is influenced by selection (Lynch and Conery 2000). If the duplicate genes are retained, they might keep the original function (potential dosage effect), part of it (subfunctionalisation) or evolve a new function (neofunctionalisation). Hence, it is essential to research the contribution of gene duplication events to the emergence of complex convergent phenotypes such as C_4 photosynthesis by including many independent origins.

The role of gene duplication in C_4 evolution is presently unclear. A few gene families have been duplicated and preserved in independent C_4 origins. However, the subsequent adoption of specific genes from these duplicated pairs for C_4 function was not always the same, and the adoption of a gene for C_4 function was more commonly associated with its ancestral expression levels (Emms et al., 2016; Moreno-Villena et al., 2018).

Gene duplication also has the potential to have a transient impact on short term evolutionary scales. For example, an initial rise in gene copies likely provided a quick way to boost expression during the early stages of C_4 evolution and physiological innovation, and subsequently duplicates may have been lost after the plant had acquired the genes encoding more suitable isoforms (Bianconi et al., 2018). Thus, the dosage effect of gene duplication might serve as a temporary process during the evolution of C_4 . Eventually the duplicated genes became unnecessary due to regulatory mutations that enhanced expression levels (Bianconi et al., 2018).

1.8. Using quantitative genetics to identify novel C_4 genes

Quantitative genetics analysis methods like genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping can shed light on the loci associated with a complex trait such as C_4 . Genome-wide association studies (GWAS) are a highly useful technique for identifying

genetic variations that are correlated with trait variation. GWAS uses SNPs to capture genetic variation across the genome (Wang et al., 2005; Uffelmann et al., 2021). Statistical evaluation of hundreds of thousands of SNPs is conducted to identify relationships with a distinct phenotype. It is crucial to consider an excessive rate of false positives when analysing GWAS data due to numerous statistical tests performed (Bureau et al., 2005).

The segregation of different photosynthetic types among different species can make it difficult to use methods to identify changes that led to the emergence of C_4 . However, quantitative genetics methods like GWAS and QTL mapping have been used to investigate the intraspecific variation of C_4 associated traits. For instance, in maize, Strigens et al. (2013) have studied photosynthetic performance during chilling. Also, Ferguson et al. (2021) have investigated genes associated with stomatal conductance and water use efficiency in sorghum. However, no specific region related to differences in C_4 carbon fixation or Kranz anatomy has been identified due to the lack of segregating populations (Simpson et al., 2021).

1.9. Using grasses as a study system for C_4 evolution

While the grass family (Poaceae) is generally well-defined, some relationships within it remain uncertain. Phylogenetic analyses have contributed to the classification of grasses into 12 subfamilies, three of which are small subfamilies with a few species (Figure 1.7; Clark and Judziewicz, 1996; Clark et al., 2000; Kellogg, 2015; Soreng et al., 2017). The remaining nine subfamilies form two large clades, PACMAD and BOP (Soreng et al., 2017). The PACMAD clade, which includes six subfamilies (**P**anicoideae, **A**ristidoideae, **C**hloridoideae, **M**icrairoideae, **A**rundinoideae, and **D**anthonioideae), is closely related to the BOP clade, comprising three subfamilies (**B**ambusoideae, **O**ryzoideae, and **P**ooideae), where C_4 species are notably absent

(Grass Phylogeny Working Group II, 2012). The PACMAD clade contains all C₄ grasses (Sage, 2004) including multiple C₄ origins (Christin et al., 2012; Edwards and Still, 2008); however, the relationships among these origins are not very clear and they vary depending on the data (nuclear or chloroplastic gene sequences) used to infer phylogenetic relationships (Figure 1.7; Grass Phylogeny Working Group II, 2012; Huang et al., 2022).

1.10. *Alloteropsis semialata* as a study system

Alloteropsis semialata, according to Ellis (1974), is the only known species with both C₄ and non-C₄ genotypes, so it has long been used as a model to study the evolutionary path of C₄ photosynthesis (reviewed by Pereira et al., 2023). A number of populations of *A. semialata* in the ground layer of the Central Zambebian miombo forests perform a weak C₄ cycle that has been classified as "C₃+C₄" (Lundgren et al., 2016; Dunning et al. 2017). Previous comparative studies claimed that the transition to full C₄ photosynthesis in *A. semialata* results from the overexpression of a few core C₄ enzymes (Figure 1.6; Dunning et al. 2019), and the acquisition of C₄-like anatomical traits such as minor veins (Figure 1.6; Lundgren et al. 2019). The strengthening of the C₄ cycle in the C₃+C₄ intermediates is associated with alterations in a number of leaf anatomical traits related to the preponderance of inner bundle sheath tissue fraction, the cellular location of the C₄ cycle, including the distance between consecutive bundle sheaths, the width of inner bundle sheath cells and the proportion of bundle sheath tissue in the leaf (Alenazi et al., 2023).

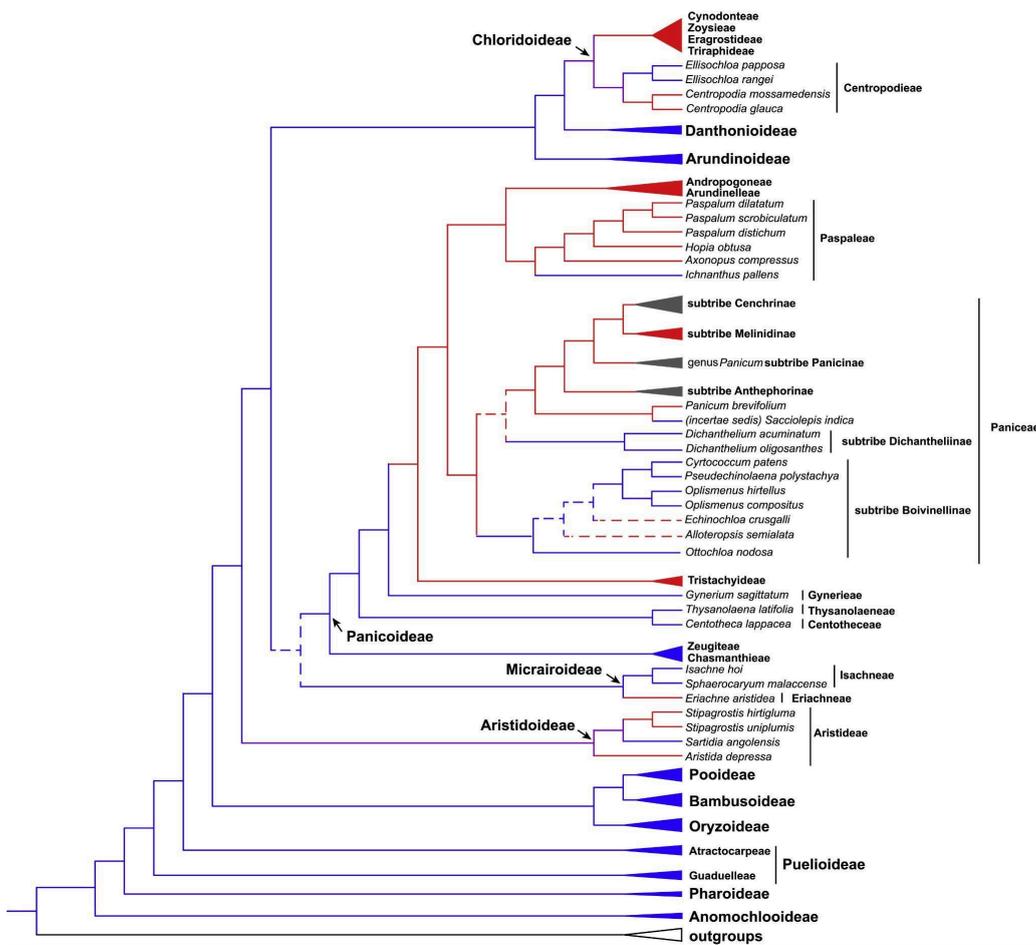


Figure 1.7

Summary of the most recent nuclear phylogeny of PACMAD grass subfamilies inferred from genome wide nuclear data, C_3 = blue, C_4 = red, and mixed = grey. This figure is taken from Huang et al., (2022).

Previous molecular phylogenetic work using nuclear-encoded gene sequences that *Alloteropsis semialata* could be demarcated into four distinct clades. The connections between the four clades differed according to the methods used, as the $C_3 + C_4$ clade II was either positioned as a sibling to clade III + IV, which are C_4 clades, or clade I that corresponds to C_3 accessions (Figure 1.8). In both scenarios, there was a comparable amount of evidence supporting each possible phylogenetic arrangement (Bianconi et al., 2020).

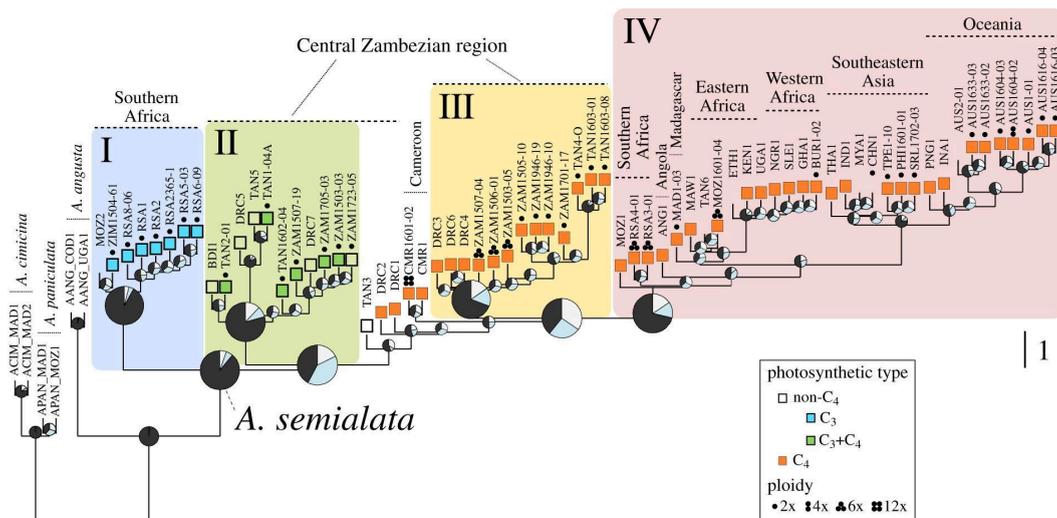


Figure 1.8

A coalescent species tree depicting the nuclear history of *Alloteropsis*. The four main nuclear clades of *A. semialata* are labelled with Roman numerals I-IV and highlighted in different colours. Key geographical regions are also indicated. I clade = C₃, II = C₃+C₄ intermediates and both III and IV are C₄. (The figure is taken from Bianconi et al., 2020)

1.11. Thesis aims

The main aim of the research described in this thesis was to determine how the C₄ cycle originated during its evolution using C₃+C₄ individuals of *A. semialata* as a study system and using independent C₄ origins in Poaceae to elucidate the evolutionary modifications while transitioning from non-C₄ to full C₄ photosynthesis. This overarching aim was investigated using three complementary studies that applied different approaches to overlapping sets of samples.

In the first chapter, the leaf anatomy of wild populations of C₃+C₄ *A. semialata* representing a range of carbon isotope values that showing the proportion of the heavier carbon isotope, ¹³C, that is discriminated by Rubisco enzyme not the PEPC in the leaf was studied to allow the quantification of the effect of leaf anatomical changes on the transition from weak to strong C₄ cycles. This aim was achieved using genomic data that helped to establish the demographic history of multiple C₃+C₄ populations spread across the Central Zambebian miombo woodlands. The leaf stable carbon isotope ratio ($\delta^{13}\text{C}$) was used to estimate the strength of the C₄ cycle in each individual (Farquhar et

al., 1989; O'Leary, 1981), and the corresponding leaf anatomy of each accession was quantified. These data were leveraged to assess the variation amongst these C₃+C₄ accessions for important C₄ leaf characters, and to test for the effects of anatomical variation on the strength of the C₄ cycle, while controlling for the species' evolutionary history.

The second chapter sought to elucidate the genomic architecture underlying variation in C₄-related leaf anatomical traits in *A. semialata*. Statistical methods were used to connect specific variation regions of the genome with leaf anatomical phenotypes and $\delta^{13}\text{C}$ by applying a genome-wide association study (GWAS) approach to *A. semialata*. Initially, a global analysis was performed to identify candidate genes associated with the strength of the C₄ cycle (as indicated by $\delta^{13}\text{C}$) using genomic data from 420 individuals representing C₃, C₃+C₄ and C₄ phenotypes. The C₃+C₄ intermediates to identify candidate genes associated with the relative expansion of bundle sheath tissue during the transition from a weak to a strong C₄ cycle. The high level of interspecific variation in *A. semialata* permits a fine-scale understanding of the genetic basis of C₄ evolution, including the intermediate steps involved in assembling this complex trait. It is crucially important were then chosen as a focus in order to identify the initial changes required for the emergence of this trait, something that may ultimately have applications in the engineering of C₄ photosynthesis in C₃ crops such as rice.

The third chapter is aimed at understanding why C₄ origins in the grasses are restricted to the PACMAD clade by identifying gene duplication events that occurred in the common ancestor of all of the independent C₄ origins. We do this by generating two new genome assemblies for Aristidoideae species *Aristida adscensionis* and *Stipagrostis hirtigluma*, and using comparative genomics to identify gene duplication events associated with the BOP and PACMAD split.

1.12. References

- Adams, K. L., and Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 2, 135-141.
- Alenazi, A. S., Bianconi, M. E., Middlemiss, E., Milenkovic, V., Curran, E. V., Sotelo, G., Lundgren, M. R., Nyirenda, F., Pereira, L., Christin, P., Dunning, L. T., and Osborne, C. P. (2023). Leaf anatomy explains the strength of C₄ activity within the grass species *Alloteropsis semialata*. *Plant, Cell and Environment*, 8, 2310–2322.
- Akam, M. (1995). Hox genes and the evolution of diverse body plans. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 349, 313-319.
- Barton, N. H., and Turelli, M. (1989). Evolutionary quantitative genetics: how little do we know?. *Annual Review of Genetics*, 23, 337-370.
- Bauwe, H., Hagemann, M., and Fernie, A. R. (2010). Photorespiration: players, partners and origin. *Trends in Plant Science*, 15, 330–336.
- Bender, M. M. (1968). Mass spectrometric studies of carbon 13 variations in corn and other grasses. *Radiocarbon*, 10, 468-472.
- Bianconi, M. E., Dunning, L. T., Curran, E. V., Hidalgo, O., Powell, R. F., Mian, S., Leitch, I. J., Lundgren, M. R., Manzi, S., Vorontsova, M. S., Besnard, G., Osborne, C. P., Olofsson, J. K., and Christin, P. A. (2020). Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B, Biological Sciences*, 287, 20201960.
- Bianconi, M. E., Dunning, L. T., Moreno-Villena, J. J., Osborne, C. P., and Christin, P. A. (2018). Gene duplication and dosage effects during the early emergence of C₄ photosynthesis in the grass genus *Alloteropsis*. *Journal of Experimental Botany*, 69, 1967-1980.
- Bowes, G., Ogren, W.L. and Hageman, R.H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications*, 45, 716–722.
- Brogliè, R., Coruzzi, G., Keith, B., and Chua, N. H. (1984). Molecular biology of C₄ photosynthesis in *Zea mays*: differential localization of proteins and mRNAs in the two leaf cell types. *Plant Molecular Biology*, 3, 431-444.
- Brown, R.H. and Hattersley, P.W. (1989). Leaf anatomy of C₃-C₄ species as related to evolution of C₄ photosynthesis. *Plant Physiology*, 91, 1543–1550.

- Buerkle, A. C., and Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, 22, 3028-3035.
- Bureau, A., Dupuis, J., Falls, K., Lunetta, K. L., Hayward, B., Keith, T. P., and Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 28, 171-182.
- Burnell, J. N., and Hatch, M. D. (1988). Low bundle sheath carbonic anhydrase is apparently essential for effective C₄ pathway operation. *Plant Physiology*, 86, 1252-1256.
- Burnell, J. N., and Ludwig, M. (1997). Characterisation of two cDNAs encoding carbonic anhydrase in maize leaves. *Functional Plant Biology*, 24, 451-458.
- Busch, F. A., Sage, T. L., Cousins, A. B., and Sage, R. F. (2013). C₃ plants enhance rates of photosynthesis by reassimilating photorespired and respired CO₂. *Plant, Cell & Environment*, 36, 200-212.
- Busch, F. A., Sage, R. F., and Farquhar, G. D. (2018). Plants increase CO₂ uptake by assimilating nitrogen via the photorespiratory pathway. *Nature Plants*, 4, 46-54.
- Cheung, V.G., Jen, K.Y., Weber, T., Morley, M., Devlin, J.L., Ewens, K.G., and Spielman, R.S. (2003). Genetics of quantitative variation in human gene expression. *Cold Spring Harb. Symp. Quant. Biol.* 68, 403-407.
- Christenhusz, M. J., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, 261, 201-217.
- Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R., and Besnard, G. (2007). C₄ photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology*, 17, 1241-1247.
- Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., Hartwell, J and Osborne, C. P. (2012). Adaptive evolution of C₄ photosynthesis through recurrent lateral gene transfer. *Current Biology*, 22, 445-449.
- Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013). Parallel recruitment of multiple genes into C₄ photosynthesis. *Genome Biology and Evolution*, 5, 2174-2187.
- Christin, P. A., and Osborne, C. P. (2014). The evolutionary ecology of C₄ plants. *New Phytologist*, 204, 765-781.
- Chollet, R., and Ogren, W. L. (1975). Regulation of photorespiration in C₃ and C₄ species. *The Botanical Review*, 41, 137-179.
- Clark, L. G., and Judziewicz, E. J. (1996). The grass subfamilies Anomochlooideae and Pharoideae (Poaceae). *Taxon*, 641-645.

- Clark, L. G., Kobayashi, M., Mathews, S., Spangler, R. E., and Kellogg, E. A. (2000). The Puelioideae, a new subfamily of Poaceae. *Systematic Botany*, 25, 181-187.
- Cui, H., Kong, D., Liu, X., and 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.
- Danowitz, M., Vasilyev, A., Kortlandt, V., and Solounias, N. (2015). Fossil evidence and stages of elongation of the Giraffa camelopardalis neck. *Royal Society Open Science*, 2, 150393.
- Darwin C. (1859). *On the origins of species by means of natural selection*. London: Murray.
- Dengler, N. (1994). Quantitative leaf anatomy of C₃ and C₄ grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Annals of Botany*, 73, 241–255.
- Dengler, N.G. and Nelson, T. (1999). Leaf structure and development in C₄ plants. In: R.F. Sage and R.K. Monson (Eds.) *C₄ plant biology*. San Diego: Academic Press, pp. 133–172.
- Duarte, J. M., Cui, L., Wall, P. K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., Altman, N., and dePamphilis, C. W. (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution*, 23, 469–478.
- Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P., and Christin, P. A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C₄ photosynthesis among close relatives. *Evolution*, 71, 1541–1555.
- Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J. K., Mapaura, A., Grundy, I. M., Kayombo, C. J., Dunning, L. A., Kentatchime, F., Ariyaratne, M., Yakandawala, D., Besnard, G., Quick, W. P., Bräutigam, A., Osborne, C. P., and Christin, P. A. (2019). Key changes in gene expression identified for different stages of C₄ evolution in *Alloteropsis semialata*. *Journal of Experimental Botany*, 70, 3255–3268.
- Edwards, G. E., Franceschi, V. R., and Voznesenskaya, E. V. (2004). Single-cell C₄ photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.*, 55, 173-196.
- Edwards, G.E. and Ku, M.S. (1987). Biochemistry of C₃–C₄ intermediates. In: M.D. Hatch and Boardman (Eds.) *The biochemistry of plants: a comprehensive treatise*, vol. 10. New York: Academic Press, pp. 275–325.
- Edwards, G. E., and Voznesenskaya, E. V. (2011). C₄ photosynthesis: Kranz forms and single-cell C₄ in terrestrial plants. In Raghavendra, A. S and Sage R. F., *C₄ photosynthesis and related CO₂ concentrating mechanisms*, 29-61.

- Ellis, R.P. (1974) Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades. *Bothalia*, 11, 273–275.
- Ellis, R. J. (1979). The most abundant protein in the world. *Trends in biochemical sciences*, 4, 241-244.
- Engelmann, S., Bläsing, O.E., Gowik, U., Svensson, P., and Westhoff, P. (2003). Molecular evolution of C₄ phosphoenolpyruvate carboxylase in the genus *Flaveria*—A gradual increase from C₃ to C₄ characteristics. *Planta* 217, 717–725.
- Ermakova, M., Danila, F. R., Furbank, R. T., and von Caemmerer, S. (2020). On the road to C₄ rice: advances and perspectives. *The Plant Journal*, 101 940-950.
- Ferguson, J. N., Fernandes, S. B., Monier, B., Miller, N. D., Allen, D., Dmitrieva, A., Schmuker, P., Lozano, R., Valluru, R., Buckler, E. S., Gore, M. A., Brown, P. J., Spalding, E. P., and Leakey, A. D. B. (2021). Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiology (Bethesda)*, 187, 1481–1500.
- Flatt, T. (2020). Life-history evolution and the genetics of fitness components in *Drosophila melanogaster*. *Genetics*, 214, 3-48.
- Furbank, R. T. (2011). Evolution of the C₄ photosynthetic mechanism: are there really three C₄ acid decarboxylation types? *Journal of Experimental Botany*, 62, 3103-3108.
- Ganko, E. W., Meyers, B. C., and Vision, T. J. (2007). Divergence in expression between duplicated genes in *Arabidopsis*. *Molecular Biology and Evolution*, 24, 2298-2309.
- Ghazzawy, H. S., Bakr, A., Mansour, A. T., and Ashour, M. (2024). Paulownia trees as a sustainable solution for CO₂ mitigation: assessing progress toward 2050 climate goals. *Frontiers in Environmental Science*, 12, 1307840.
- Geary, D. H., Hunt, G., Magyar, I., and Schreiber, H. (2010). The paradox of gradualism: phyletic evolution in two lineages of lymnocardiid bivalves (Lake Pannon, central Europe). *Paleobiology*, 36, 592-614.
- Glackin, C. A., and Grula, J. W. (1990). Organ-specific transcripts of different size and abundance derive from the same pyruvate, orthophosphate dikinase gene in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 3004-3008.
- Grass Phylogeny Working Group II. (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C₄ origins. *New Phytologist*, 193, 304-312.
- Griffiths, H., Weller, G., Toy, L. F., and Dennis, R. J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C₃ and C₄ plants. *Plant, Cell and Environment*, 36, 249-261.

- Kellogg, E. A. (2015). Flowering plants. Monocots: Poaceae (Vol. 13). In K. Kubitzki (Ed.), *The families and genera of vascular plants* (Vol. 13). Springer.
- King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees: Their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science*, 188, 107-116.
- Hagen, J. F., Roberts, N. S., and Johnston Jr, R. J. (2023). The evolutionary history and spectral tuning of vertebrate visual opsins. *Developmental Biology*, 493, 40-66.
- Hamann, E., Pauli, C. S., Joly-Lopez, Z., Groen, S. C., Rest, J. S., Kane, N. C., Purugganan, M. D., and Franks, S. J. (2021). Rapid evolutionary changes in gene expression in response to climate fluctuations. *Molecular ecology*, 3, 193–206.
- Hatch, M.D. (1971). The C₄ pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C₄ dicarboxylic acid. *Biochemical Journal*, 125, 425–432.
- Hatch, M. D., and Osmond, C. B. (1976). Compartmentation and transport in C₄ photosynthesis. In Lüttge, U. and Pitman, M. G. (Eds.), *Transport in plants III: intracellular interactions and transport processes* (pp. 144-184). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hatch, M. and Slack, C. (1966). Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochemical Journal*, 101, 103–111.
- Hattersley, P.W. (1984). Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Annals of Botany*, 53, 163–180.
- Hattersley, P.W. and Watson, L. (1975) Anatomical parameters for predicting photosynthetic pathways of grass leaves: the “maximum lateral cell count” and the “maximum cells distant count”. *Phytomorphology*, 25, 325–333.
- Hattersley, P. W., Watson, L., and Johnston, C. R. (1982). Remarkable leaf anatomical variations in *Neurachne* and its allies (Poaceae) in relation to C₃ and C₄ photosynthesis. *Botanical Journal of the Linnean Society*, 84, 265-272.
- Huang, P., Studer, A. J., Schnable, J. C., Kellogg, E. A., and Brutnell, T. P. (2017). Cross species selection scans identify components of C₄ photosynthesis in the grasses. *Journal of experimental botany*, 68, 127-135.
- Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P. A., and Edwards, E. J. (2019). The genetics of convergent evolution: insights from plant photosynthesis. *Nature reviews. Genetics*, 20, 485–493.
- Hermans, J., and Westhoff, P. (1990). Analysis of expression and evolutionary relationships of phosphoenol-pyruvate carboxylase genes in *Flaveria trinervia* (C₄) and *F. pringlei* (C₃). *Molecular and General Genetics MGG*, 224, 459-468.

- Hofmann, C. M., and Carleton, K. L. (2009). Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *Integrative and comparative biology*, 49, 630-643.
- Jablonski, D., and Shubin, N. H. (2015). The future of the fossil record: Paleontology in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 4852–4858.
- Jordan, D. B., and Ogren, W. L. (1984). The CO₂/O₂ specificity of ribulose 1, 5-bisphosphate carboxylase/oxygenase: dependence on ribulosebisphosphate concentration, pH and temperature. *Planta*, 161, 308-313.
- Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics*, 10, 19-31.
- Kapralov, M. V., Smith, J. A. C., and Filatov, D. A. (2012). Rubisco evolution in C₄ eudicots: an analysis of *Amaranthaceae sensu lato*. *PLoS one*, 7, e52974.
- Kortschak, H. P., Hartt, C. E., and Burr, G. O. (1965). Carbon dioxide fixation in sugarcane leaves. *Plant physiology*, 40, 209.
- Ku, M.S.B., Monson, R.K., Littlejohn Jr., R.O., Nakamoto, H., Fisher, D.B. and Edwards, G.E. (1983). Photosynthetic characteristics of C₃–C₄ intermediate *Flaveria* species: I. Leaf anatomy, photosynthetic responses to O₂ and CO₂, and activities of key enzymes in the C₃ and C₄ pathways. *Plant Physiology*, 71, 944–948.
- Langdale, J. A., Metzler, M. C., and Nelson, T. (1987). The argentia mutation delays normal development of photosynthetic cell-types in *Zea mays*. *Developmental Biology*, 122, 243–255.
- Langdale, J. A., Rothermel, B. A., and Nelson, T. (1988). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Genes and Development*, 1, 106–115.
- Lau, E. S., and Oakley, T. H. (2021). Multi-level convergence of complex traits and the evolution of bioluminescence. *Biological Reviews*, 96, 673-691.
- Leegood, R. C. (2008). Roles of the bundle sheath cells in leaves of C₃ plants. *Journal of Experimental Botany*, 59, 1663-1673.
- Lepiniec, L., Keryer, E., Philippe, H., Gadal, P., and Crépin, C. (1993). Sorghum phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution. *Plant Molecular Biology*, 21, 487-502.
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, 12, e1005767.
- Lorimer, G.H. and Andrews, T.J. (1973). Plant photorespiration—an inevitable consequence of the existence of atmospheric oxygen. *Nature*, 243, 359–360.

- Lundgren, M. R., Osborne, C. P., and Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C₄ evolution. *Journal of Experimental Botany*, *65*, 3357–3369.
- Lundgren, M. R., Besnard, G., Ripley, B. S., Lehmann, C. E., Chatelet, D. S., Kynast, R. G., Namaganda, M., Vorontsova, M. S., Hall, R. C., Elia, J., Osborne, C. P., and Christin, P. A. (2015). Photosynthetic innovation broadens the niche within a single species. *Ecology Letters*, *18*, 1021–1029.
- Lundgren, M. R., Christin, P. A., Escobar, E. G., Ripley, B. S., Besnard, G., Long, C. M., Hattersley, P. W., Ellis, R. P., Leegood, R. C., and Osborne, C. P. (2016). Evolutionary implications of C₃-C₄ intermediates in the grass *Alloteropsis semialata*. *Plant, Cell and Environment*, *39*, 1874–1885.
- Lundgren, M. R., Dunning, L. T., Olofsson, J. K., Moreno-Villena, J. J., Bouvier, J. W., Sage, T. L., Khoshravesh, R., Sultmanis, S., Stata, M., Ripley, B. S., Vorontsova, M. S., Besnard, G., Adams, C., Cuff, N., Mapaura, A., Bianconi, M. E., Long, C. M., Christin, P. A., and Osborne, C. P. (2019). C₄ anatomy can evolve via a single developmental change. *Ecology Letters*, *22*, 302–312.
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *5494*, 1151-1155.
- Merlo, L., Geigenberger, P., Hajirezaei, M., & Stitt, M. (1993). Changes of carbohydrates, metabolites and enzyme activities in potato tubers during development, and within a single tuber along astolon-apexgradient. *Journal of Plant Physiology*, *142*, 392-402.
- Miziorko, H. M., and Lorimer, G. H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annual review of biochemistry*, *52*, 507–535.
- Monson, R.K., Edwards, G.E., Ku, M.S.B. (1984) C₃-C₄ intermediate photosynthesis in plants. *Bioscience* *34*, 563-574.
- Monson, R.K. (1989) The relative contributions of reduced photorespiration, and improved water-and nitrogen-use efficiencies, to the advantages of C₃-C₄ intermediate photosynthesis in *Flaveria*. *Oecologia* *80*:215–221.
- Monson, R. K., and Moore, B. D. (1989). On the significance of C₃-C₄ intermediate photosynthesis to the evolution of C₄ photosynthesis. *Plant, Cell and Environment*, *12*, 689-699.
- Moore, J. H., and Williams, S. M. (2002). New strategies for identifying gene-gene interactions in hypertension. *Annals of Medicine*, *34*, 88-95.

- Ogren, W. L. (1984). Photorespiration: pathways, regulation, and modification. *Annual Review of Plant Physiology*, 35, 415-442.
- Olofsson, J. K., Curran, E. V., Nyirenda, F., Bianconi, M. E., Dunning, L. T., Milenkovic, V., Sotelo, G., Hidalgo, O., Powell, R. F., Lundgren, M. R., Leitch, I. J., Nosil, P., Osborne, C. P., and Christin, P. (2021) Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology*, 9, 2116–2130.
- Orr, D. J., Alcântara, A., Kapralov, M. V., Andralojc, P. J., Carmo-Silva, E., and Parry, M. A. (2016). Surveying Rubisco diversity and temperature response to improve crop photosynthetic efficiency. *Plant Physiology*, 172, 707-717.
- Pereira, L., Bianconi, M. E., Osborne, C. P., Christin, P.-A., and Dunning, L. T. (2023). *Alloteropsis semialata* as a study system for C₄ evolution in grasses. *Annals of Botany*, 131, 167-179
- Rawsthorne, S. (1992). C₃-C₄ intermediate photosynthesis: linking physiology to gene expression. *The Plant Journal*, 2, 267-274.
- Romanowska, E., and Wasilewska-Dębowska, W. (2022). Light-Dependent Reactions of Photosynthesis in Mesophyll and Bundle Sheath Chloroplasts of C₄ Plant Maize. How Our Views Have Changed in Recent Years. *Acta Societatis Botanicorum Poloniae*, 91.
- Rosnow, J. J., Edwards, G. E., and Roalson, E. H. (2014). Positive selection of Kranz and non-Kranz C₄ phosphoenolpyruvate carboxylase amino acids in Suaedoideae (Chenopodiaceae). *Journal of Experimental Botany*, 65, 3595-3607.
- Sage, R. F. (1999). Why C₄ photosynthesis. In Monson, R. and Sage, R. (Eds.), *C₄ plant biology*. Elsevier Science 3.
- Sage, R.F. (2004). The evolution of C₄ photosynthesis. *New Phytologist*, 161, 341–370.
- Sage, T. L., and Sage, R. F. (2009). The functional anatomy of rice leaves: implications for refixation of photorespiratory CO₂ and efforts to engineer C₄ photosynthesis into rice. *Plant and Cell Physiology*, 50, 756-772.
- Sage, R. F., Christin, P. A., and Edwards, E. J. (2011). The C₄ plant lineages of planet Earth. *Journal of Experimental Botany*, 62, 3155-3169.
- Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C₄ photosynthesis. *Annual Review of Plant Biology*, 63, 19-47.

- Sage, R. F., Khoshravesht, R., and Sage, T. L. (2014). From proto-Kranz to C₄ Kranz: building the bridge to C₄ photosynthesis. *Journal of Experimental Botany*, 65, 3341-3356.
- Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics*, 6, 552.
- Schadt, E.E., Monks, S.A., Drake, T.A., Lusk, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–30
- Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15, 749-763.
- Schulze, S., Westhoff, P., and Gowik, U. (2016). Glycine decarboxylase in C₃, C₄ and C₃-C₄ intermediate species. *Current Opinion in Plant Biology*, 31, 29-35.
- Shapiro, J. A., Corning, P. A., Kauffman, S. A., Noble, D., Shapiro, J. A., Vane-Wright, R. I., and Pross, A. (2023). Evolutionary change is naturally biological and purposeful. In Corning, P. A. and Shapiro, J. A. *Evolution 'on purpose'*, 275-298. Springer.
- Sheen, J. Y., and Bogorad, L. (1986). Differential expression of six light-harvesting chlorophyll a/b binding protein genes in maize leaf cell types. *Proceedings of the National Academy of Sciences*, 83, 7811-7815.
- Sheen, J. (1999). C₄ gene expression. *Annual Review of Plant Biology*, 50, 187-217.
- Simpson, C.J.C., Reeves, G., Tripathi, A., Singh, P. and Hibberd, J.M. (2021) Using breeding and quantitative genetics to understand the C₄ pathway. *Journal of Experimental Botany*, 73, 3072– 3084.
- Slewinski, T. L., Anderson, A. A., Zhang, C., and Turgeon, R. (2012). Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant and Cell Physiology*, 53(12), 2030-2037.
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., David Sankoff, Claude W. de Pamphilis, P. Kerr Wall, Pamela S. Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, 1, 336-348.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., Barberá, P., Gillespie, L., and Zuloaga, F. O. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and evolution*, 55, 259-290.
- Shastri, B. S. (2009). SNPs: impact on gene function and phenotype. *Single nucleotide polymorphisms: methods and protocols*, 3-22.

- Stayton, C. T. (2015). The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution*, 69, 2140-2153.
- Stata, M., Sage, T.L. and Sage, R.F. (2019). Mind the gap: the evolutionary engagement of the C₄ metabolic cycle in support of net carbon assimilation. *Current Opinion in Plant Biology*, 49, 27–34.
- Strigens, A., Schipprack, W., Reif, J. C., and Melchinger, A. E. (2013). Unlocking the Genetic Diversity of Maize Landraces with Doubled Haploids Opens New Avenues for Breeding. *PLoS One*, 2, e57234–e57234.
- Suzuki, T. K. (2017). On the origin of complex adaptive traits: progress since the Darwin versus Mivart debate. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 328, 304-320.
- Thoday, J. M. (1961). Location of polygenes.
- Troughton, J. H. (1979). $\delta^{13}\text{C}$ as an indicator of carboxylation reactions. In *Photosynthesis II: Photosynthetic Carbon Metabolism and Related Processes*, (pp. 140-149). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Von Caemmerer, S. (1992). Carbon isotope discrimination in C₃–C₄ intermediates. *Plant, Cell and Environment*, 15, 1063– 1072.
- Von Caemmerer, S. (2000). *Biochemical models of leaf photosynthesis*. Collingwood: CSIRO.
- Voznesenskaya EV, Koteyeva NK, Chuong SDX, Ivanova AN, Barroca J, Craven LA, Edwards GE. (2007). Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). *Functional Plant Biology*, 34, 247–267.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, C., Lappalainen, L. and Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1, 59.
- Wang, W. Y., Barratt, B. J., Clayton, D. G., and Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*, 6, 109-118.
- Wang, X., Gowik, U., Tang, H., Bowers, J. E., Westhoff, P., and Paterson, A. H. (2009). Comparative genomic analysis of C₄ photosynthetic pathway evolution in grasses. *Genome Biology*, 10, 1-18.
- Wang, L., Peterson, R. B., and Brutnell, T. P. (2011). Regulatory mechanisms underlying C₄ photosynthesis. *New Phytologist*, 190, 9-20.
- Ward, B. P., Brown-Guedira, G., Kolb, F. L., Van Sanford, D. A., Tyagi, P., Sneller, C. H., and Griffey, C. A. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. *PLoS One*, 14, e0208217.

- Watson-Lazowski, A., Papanicolaou, A., Sharwood, R., and Ghannoum, O. (2018). Investigating the NAD-ME biochemical pathway within C₄ grasses using transcript and amino acid variation in C₄ photosynthetic genes. *Photosynthesis Research*, 138, 233-248.
- Webster, G. L., Brown, W. V., and Smith, B. N. (1975). Systematics of photosynthetic carbon fixation pathways in Euphorbia. *Taxon*, 24, 27-33.
- Wingler, A., Lea, P. J., Quick, W. P., and Leegood, R. C. (2000). Photorespiration: metabolic pathways and their role in stress protection. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355, 1517-1529.
- Fu, X., and Walker, B. J. (2024). Photorespiratory glycine contributes to photosynthetic induction during low to high light transition. *Scientific Reports*, 14, 19365.
- Young, N. D. (1996). QTL mapping and quantitative disease resistance in plants. *Annual review of phytopathology*, 34, 479-501.
- Zhang H (2003). Evolution by gene duplication: an update. *Trends Ecology and Evolution* 18, 292.

**Chapter 2. Leaf anatomy explains the strength
of C₄ activity within the grass species
*Alloteropsis semialata***

Chapter 2. Leaf anatomy explains the strength of C₄ activity within the grass species *Alloteropsis semialata*

Ahmed S. Alenazi^{1,2}, Matheus E. Bianconi¹, Ella Middlemiss¹, Vanja Milenkovic¹, Emma V. Curran¹, Graciela Sotelo¹, Marjorie R. Lundgren¹, Florence Nyirenda³, Lara Pereira¹, Pascal-Antoine Christin¹, Luke T. Dunning¹, Colin P. Osborne⁴

Affiliations:

¹ Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

² Department of Biological Sciences, Northern Border University, Saudi Arabia

³ Department of Biological Sciences, University of Zambia, Lusaka, Zambia

⁴ Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

* Author for correspondence: c.p.osborne@sheffield.ac.uk, telephone: +44-1142220146

Author contributions

ASA, PAC, LTD, and CPO designed the study. EVC, FN, and PAC performed field work. ASA, MEB, EM, VM, EVC, GS and MRL generated data. ASA, EM, MEB, VM, EVC, LP, and GS analysed data. ASA, EM, PAC, LTD, and CPO wrote the paper, with the help of all authors.

This work was published in **Plant, Cell and Environment**, Volume46, Issue8, August 2023,

Pages 2310-2322.

Personal contribution: I generated cross section data, measured the anatomical traits and performed statistical analysis. I also performed phylogenetic analyses and then wrote the manuscript with help from co-authors.

2.1. Abstract

C₄ photosynthesis results from anatomical and biochemical characteristics that together concentrate CO₂ around ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), increasing productivity in warm conditions. This complex trait evolved through the gradual accumulation of components, and particular species possess only some of these, resulting in weak C₄ activity. The consequences of adding C₄ components have been modelled and investigated through comparative approaches, but the intraspecific dynamics responsible for strengthening the C₄ pathway remain largely unexplored. Here, we evaluate the link between anatomical variation and C₄ activity, focusing on populations of the photosynthetically diverse grass *Alloteropsis semialata* that fix various proportions of carbon via the C₄ cycle. The carbon isotope ratios in these populations range from values typical of C₃ to those typical of C₄ plants. This variation is statistically explained by a combination of leaf anatomical traits linked to the preponderance of bundle sheath tissue. We hypothesize that increased investment in bundle sheath boosts the strength of the intercellular C₄ pump and shifts the balance of carbon acquisition towards the C₄ cycle. Carbon isotope ratios indicating a stronger C₄ pathway are associated with warmer, drier environments, suggesting that incremental anatomical alterations can lead to the emergence of C₄ physiology during local adaptation within metapopulations.

KEYWORDS C₃–C₄ intermediate, C₄ photosynthesis, evolution, population genetics

2.2. Introduction

The majority of plants use the ancestral C₃ photosynthetic pathway, in which atmospheric CO₂ is fixed directly by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) in a reaction that constitutes the entry point of the Calvin–Benson–Bassham cycle. However, Rubisco can also bind O₂, which starts the energetically costly photorespiratory pathway (Bowes et al., 1971; Lorimer and Andrews, 1973). The efficiency of the C₃ type, therefore, decreases in all conditions that increase the partial pressure of O₂ relative to CO₂ within the leaf (Ku et al., 1983; Peixoto et al., 2021; Sage et al., 2014). These conditions include warm, arid and saline habitats in the low-CO₂ atmosphere that prevailed over the last 30 million years (Sage, 2001, 2004). Plants have evolved a number of strategies to minimize photorespiration costs, the most significant of which are CO₂-concentrating mechanisms, such as C₄ photosynthesis (Edwards and Ku, 1987; Hatch, 1971). In C₄ plants, the initial fixation of carbon is mediated by phosphoenolpyruvate carboxylase (PEPC), an enzyme without affinity for O₂ (Burnell and Hatch, 1988; Hatch and Slack, 1966). This initial reaction usually takes place in the mesophyll of C₄ plants, which CO₂ reaches via simple diffusion through the stomata. The resulting C₄ acid is then transported into a different leaf compartment, usually the bundle sheath that surrounds the veins, where Rubisco is segregated in C₄ plants (Hatch, 1971, 1987). CO₂ is released within this compartment isolated from the atmosphere, and the intermediate compounds of the cycle are brought back to the location of PEPC activity and regenerated (Hatch, 1987). Through this mechanism, the C₄ cycle increases the relative partial pressure of CO₂ around Rubisco, almost completely suppressing photorespiration (von Caemmerer and Furbank, 2003).

The C₄ pathway relies on many anatomical and biochemical novelties compared to the C₃ ancestral state (Hatch, 1987). A number of enzymes, including PEPC and those needed to transform and transport the C₄ acid, release CO₂ and regenerate the intermediate compounds, need to be active at high levels in specific leaf compartments (Hatch, 1971; Hatch and Osmond, 1976). In addition, specific leaf properties are needed to support the segregation of reactions and allow for a rapid transfer of metabolites (Dengler et al., 1994; Hattersley, 1984; Lundgren et al., 2014). In particular, mesophyll and bundle sheath cells need to be in close proximity, with no more than one cell separating them in C₄ plants (Hattersley and Watson, 1975). In addition, a large fraction of the leaf must be dedicated to the tissue that houses the Calvin–Benson–Bassham cycle (Brown and Hattersley, 1989; Dengler and Nelson, 1999). Despite this apparent complexity, the C₄ pathway evolved more than 60 times independently from C₃ ancestors (Sage et al., 2011). This remarkable evolutionary phenomenon is explained by the existence of intermediate stages, which bridge the gap between the C₃ and C₄ states (Edwards and Ku, 1987; Heckmann et al., 2013; Kennedy and Laetsch, 1974; Monson et al., 1984; Sayre and Kennedy, 1977; Williams et al., 2013; Yorimitsu et al., 2019). In particular, some plants fix part of their atmospheric carbon via the C₃ cycle and part via the C₄ cycle, in proportions that vary from weak to strong C₄ involvements (Edwards and Ku, 1987; Monson et al., 1986; Stata et al., 2019). Theoretical models predict the existence of a path from C₃ to C₄, where each step increases photosynthetic efficiency (Heckmann et al., 2013; Monson and Moore, 1989; Williams et al., 2013). However, empirical tests of this hypothesis are largely missing and the detailed events allowing the transition from a weak to a strong C₄ pathway remain unexplored.

Investigating photosynthetic systems and how they respond to environmental gradients can be done effectively using the stable carbon isotope ratio ($\delta^{13}\text{C}$; Farquhar et al., 1989; O'Leary, 1981). Plants

contain lower levels of ¹³C than ambient air, and the ¹³C content of plant tissue largely depends on the photosynthetic pathway (Bender, 1968; Troughton, 1979), although additional variation can be observed in relation to environmental conditions (Guy et al., 1980; Winter, 1981). Rubisco discriminates against ¹³C and preferentially fixes ¹²C during photosynthesis (O'Leary, 1981). In contrast, PEPC discriminates less than Rubisco, making the isotopic composition of plant dry matter a useful tool to identify different photosynthetic types (O'Leary, 1981). $\delta^{13}\text{C}$ has been widely used as a proxy for photosynthetic type (Bender, 1968; Brown, 1977; von Caemmerer, 1992; Cerling, 1999; Cerros-Tlatilpa and Columbus, 2009; Gowik et al., 2011; Lundgren et al., 2015; Olofsson et al., 2021; Smith and Brown, 1973; Smith and Epstein, 1971; Stata et al., 2019), with values of -10‰ to -15‰ for C₄ species, and -22‰ to -31‰ for C₃ species (95% confidence intervals from the data compilation of Cerling et al., 1997).

Alloteropsis semialata is the only species known to have both C₃ and C₄ genotypes (Ellis, 1974). In addition, some non-C₄ populations of *A. semialata*, found in the grassy ground layer of the Central Zambezian miombo woodlands, perform a weak C₄ cycle in addition to the direct fixation of CO₂ via the C₃ cycle (Dunning et al., 2017; Lundgren et al., 2016). Comparative analyses have shown that their physiology results from the up-regulation of particular C₄ enzymes (Dunning et al., 2019) and the acquisition of C₄-like anatomical characters (Lundgren et al., 2019) compared to C₃ populations. Importantly, the $\delta^{13}\text{C}$ of these plants range from values typical of a weak or null C₄ involvement to values indicative of a C₄ cycle responsible for more than half of carbon acquisition (i.e., -15‰ to -22‰ ; von Caemmerer, 1992, 2000; Lundgren et al., 2015; Monson et al., 1988; Olofsson et al., 2021; Stata and Sage, 2019). These non-C₄ populations of *A. semialata*, therefore, constitute an attractive system for investigating the changes responsible for the transition from a weak to a strong C₄ cycle.

In this study, we compare the leaf anatomy of wild populations of non-C₄ *A. semialata* representing a range of carbon isotope values, to quantify the importance of anatomical changes during the transition from weak to strong C₄ cycles. We first use $\delta^{13}\text{C}$, which reflects the proportion of inorganic carbon fixed by Rubisco as opposed to PEPC, to estimate the strength of the C₄ cycle in each individual. We further correlate the average $\delta^{13}\text{C}$ values of each population with the local climate to test for links between C₄ activity and environmental conditions. We then use genomic data to establish the history of non-C₄ populations spread across the Central Zambezian miombo woodlands, and we quantitatively describe the leaf anatomy of multiple non-C₄ individuals per population. These data are used to assess the variation for functionally important C₄ leaf characters and test for relationships with the strength of the C₄ cycle across the phylogenetic tree for this species. Our work sheds new light on the role of ecological variation in driving leaf anatomical changes responsible for the transition from weak to strong C₄ cycles.

2.3. Materials and Methods

2.3.1. Plant sampling and carbon isotopes

Our sampling focused on the Central Zambezian woodlands, where non-C₄ *A. semialata* are known to occur (Bianconi et al., 2020; Lundgren et al., 2016). Populations were collected irrespective of their photosynthetic type in Tanzania and Zambia between 2014 and 2019 (Olofsson et al., 2021), and all 28 populations that included individuals with a non-C₄ carbon isotope ratio ($\delta^{13}\text{C} < -17\text{‰}$) were selected for analyses in this study. The carbon isotope threshold was chosen based on previous observations of photosynthetic physiology in comparison with $\delta^{13}\text{C}$ in this species (Lundgren et al., 2016) and in other groups (Stata and Sage, 2019). This existing sampling was augmented during a field trip to Zambia in January 2020. All populations of *A. semialata*, localized during random walk-and-search stops, were again collected irrespective of their photosynthetic type. Leaf samples

of multiple individuals were placed in silica gel and in 70% ethanol, and Global Positioning System (GPS) coordinates were recorded for each locality (Supporting Information: Table 2.S.1). In addition, herbarium collections were used to increase geographical coverage. Four previously analysed herbarium samples (Lundgren et al., 2015; Olofsson et al., 2016) and eight new ones from Burundi and the Democratic Republic of Congo (DRC) were included here (Supporting Information: Table 2.S1). The GPS coordinates were used to retrieve for each locality the values of 18 bioclimatic variables from the WorldClim database (Fick and Hijmans, 2017), using the Raster package (Hijmans, 2021) from R (R Core Team, 2020). These data represent the average climate for the region from 1970 to 2000 (Fick and Hijmans, 2017).

Values of $\delta^{13}\text{C}$ were retrieved from previous studies or generated here (Supporting Information: Table 2.S1; Bianconi et al., 2020; Lundgren et al., 2015; Olofsson et al., 2021). For the new samples, 1–2 mg of dried tissue of leaves were used to measure the $\delta^{13}\text{C}$ with an ANCAGSL preparation module that is joined to a 20–20 stable isotope analyser (PDZ Europa). The $\delta^{13}\text{C}$ was expressed relative to the standard Pee Dee Belemnite, and all samples with a $\delta^{13}\text{C} < -17\text{‰}$ were considered as non-C₄. When sufficient material was available, the $\delta^{13}\text{C}$ of non-C₄ individuals that were selected for anatomical analyses were measured three times independently. The median of these three technical replicates was used in subsequent analyses (Supporting Information: Table 2.S2). For the 16 samples where there was insufficient material to replicate the measurements, the use of an unreplicated value reduces the precision of $\delta^{13}\text{C}$ estimates. However, the exclusion of these values from the regression analyses did not qualitatively change the results.

To help interpret the $\delta^{13}\text{C}$ estimates, we carried out a sensitivity analysis using simple models of carbon isotope discrimination (von Caemmerer, 1992; Farquhar et al., 1982). The analysis

particularly emphasized the effects of variation in water-use efficiency via the ratio of intercellular to atmospheric CO₂ partial pressures (p_i/p_a), and bundle sheath leakiness (ϕ), in plants with a weak C₄ cycle (i.e. ‘type II’ C₃–C₄ intermediates; Monson et al., 1986).

2.3.2. Genome scan and phylogenetic analyses

The phylogenetic relationships among all individuals were inferred by combining sequence data obtained with different approaches. The reduced representation sequencing approach of Olofsson, Dunning et al. (2019) was used to scan the genomes of samples stored in silica gel. Genomic DNA was extracted with the DNeasy Plant Mini Kit (Qiagen). Two restriction enzymes (*EcoRI* and *MseI*) were used to digest the extracted DNA, and a barcode and a common adapter were ligated. Standard Illumina primers were then added through PCR, with 16 cycles. Each sample was individually barcoded and pools of up to 96 libraries were size selected (target of 300–600 bp) and sequenced as 125 paired-end reads on an Illumina HiSeq. 2500. The new data were combined with the subset of those from Olofsson et al. (2021) corresponding to non-C₄ individuals. The whole genomes of the new herbarium samples were sequenced as 150 bp paired-end Illumina reads at low coverage, using the approach of Olofsson et al. (2016). Existing whole-genome sequence data for other *Alloteropsis* samples representing the multiple lineages of *A. semialata* as well as the other species in the genus were added to the data set (Supporting Information: Table 2.S3; Bianconi et al., 2020; Lundgren et al., 2015; Olofsson et al., 2016).

A nuclear phylogenetic tree was inferred using the method of Bianconi et al. (2020). For each of the 7408 putative single-copy genes in Panicoideae grasses identified by Bianconi et al. (2020), orthologous sequences of *A. semialata* retrieved from the species chromosome-level genome assembly (reference AUS1; Dunning et al., 2019) were used as a reference to map the paired-end

reads from all sequence data sets using Bowtie2 v. 2.3.5 (Langmead and Salzberg, 2012) with default parameters. Using a bash-scripted pipeline that implements the mpileup function of Samtools v.1.9 (Li et al., 2009) for variant calling, gene sequences were reconstructed by incorporating variant sites into a consensus sequence as in Olofsson, Cantera et al. (2019). All polymorphic sites were called as ambiguous bases. This method produces sequences that are already aligned to the reference. Gene alignments were trimmed with TrimAl v.1.4 (Gutiérrez-Rodríguez et al., 2009) to remove sites with missing data in more than 50% of individuals. For each gene alignment, individuals with sequences shorter than 100 bp after trimming were removed. In the end, individuals with more than 95% missing data across gene alignments were discarded. The remaining trimmed gene alignments were concatenated, resulting in a 223 023 bp alignment encompassing 252 samples. A maximum likelihood phylogenetic tree was inferred with RAxML v.8.2.12 (Stamatakis, 2014), using the GTRCAT substitution model and 100 bootstrap pseudoreplicates.

2.3.3. Microscopy and leaf anatomical measurements

Leaf material collected in the field was fixed and dehydrated in ethanol before embedding. For a minimum of three randomly sampled individuals per population (or all individuals for populations where fewer than three non-C₄ individuals were located in the field), leaf fragments of 5–7 mm in length were embedded using the Technovit Kit (Technovit 7100; Heraeus Kulzer GmbH). In all cases, these samples for anatomy were taken from the same plant at the same time as the $\delta^{13}\text{C}$ samples. For herbarium samples or individuals only available as silica gel dried material, the leaves were first rehydrated in 1% KOH solution for 24 h in the refrigerator at 4°C. After embedding, 11- μm -thick transverse sections were obtained with a rotary microtome (Leica Biosystems) and were stained with 1% Toluidine Blue O for 1.5 min (Sigma-Aldrich). The slides were photographed

using an Olympus BX51 microscope with a mounted camera (Olympus). To recreate large stretches of leaf tissue in a cross-section, sequences of images taken along a single leaf were stitched together using the Hugin software (Hugin Development Team, 2015).

All measurements of leaf anatomical properties were made using ImageJ v.1.53f (Schneider et al., 2012). For each leaf, a segment connecting the vertical mid-axis of two consecutive secondary veins (recognized by the presence of metaxylem) was selected, avoiding the midrib and leaf edges (Supporting Information: Figure 2.S1a). For each segment, the total areas in the cross-section of mesophyll (including airspace), outer bundle sheath, inner bundle sheath, and vascular tissue were measured. The proportion of the photosynthetic part of the leaf dedicated to the re-fixation of carbon acquired via the C₄ cycle, which is the inner bundle sheath also known as the mestome sheath in *A. semialata* (Hattersley et al., 1977), was calculated relative to the mesophyll as the area of inner sheath divided by the sum of the areas of mesophyll and inner bundle sheath (i.e., inner bundle sheath fraction; IBSF).

The number of veins was recorded separately for tertiary veins, which are associated with extraxylary fibres and epidermis thinning, and lower-order veins, which are smaller than tertiary veins and lack extraxylary fibres (hereafter referred to as ‘minor veins’). The segment length was measured along a line connecting the centres of secondary and tertiary veins (referred to as ‘major veins’) and used to calculate the average interveinal distance based on secondary and tertiary veins. The average minimum distance between the outside of the outer sheath of consecutive bundles (including minor veins) was measured as a proxy for the maximum distance between mesophyll and inner bundle sheath cells (i.e., bundle sheath distance, BSD). The size of individual inner sheath

cells was measured as the average width of the most equatorial cells of each tertiary vein within the segment (i.e., inner bundle sheath width, IBSW) (Supporting Information: Figure 2.S1).

2.3.4. Statistical analyses

All analyses were done in R version 3.6.3. (R Core Team, 2018). The climatic variation among populations was summarized using a principal component analysis with the *prcomp* function in R (R Core Team, 2018). The first two principal components were extracted and used as explanatory variables in multiple linear models. Multiple regression analyses were first used to test for the effects of these climatic variables on photosynthetic diversity, as inferred from the median of $\delta^{13}\text{C}$ per population. Photosynthetic diversity within populations was not considered in this analysis since the climate as measured here does not vary within populations. The least significant variable was successively removed until all remaining variables were significant.

Second, we also used multiple regression analysis to test for relationships between $\delta^{13}\text{C}$ and anatomical traits, accounting for lineage as a fixed effect. The least significant variable was successively removed until all remaining variables were significant.

Finally, to take phylogenetic relationships into account, phylogenetic generalized least-squares (PGLS) analysis was also carried out in R using the *caper* package (Orme et al., 2022) using the phylogenetic tree and anatomical traits as explanatory variables. We ran the model with the least significant variable removed until only significant variables ($p < 0.05$) remained. The phylogeny described relationships among the sampled populations, but did not include the relationships among individuals within each population. To run this analysis, we therefore only included the individual from each population that was used to construct the phylogeny.

2.4. Results

2.4.1. Plant sampling captures a range of carbon isotope ratios

Across all 842 *A. semialata* individuals analysed previously (Bianconi et al., 2020; Lundgren et al., 2015; Olofsson et al., 2021) or here, the value of $\delta^{13}\text{C}$ ranges from -9.1‰ to -29.0‰ (Figure 2.1 and Supporting Information: Table 2.S1). Within the Central Zambezian region, both C₄ individuals with $\delta^{13}\text{C}$ above -17‰ and non-C₄ individuals with $\delta^{13}\text{C}$ below -17‰ are found (Figure 2.1), as previously reported (Lundgren et al., 2015, 2016; Olofsson et al., 2016, 2021). In addition to the five populations previously identified (Olofsson et al., 2021), C₄ and non-C₄ individuals were found in sympatry in three new populations (Figure 2.1). In total, 38 populations from the Central Zambezian region contained non-C₄ individuals based on $\delta^{13}\text{C}$ values, which is consistent with previous reports from this region (Bianconi et al., 2020; Dunning et al., 2017; Lundgren et al., 2016). The $\delta^{13}\text{C}$ of the 231 non-C₄ collected from these populations range from -27.2‰ to -18.2‰ (Figure 2.1b), consistent with variation among individuals in the proportion of carbon fixed by C₄ photosynthesis. In addition to the variation among populations, the $\delta^{13}\text{C}$ also varied within populations (Supporting Information: Table 2.S1 and Figure 2.S2).

The $\delta^{13}\text{C}$ of the 109 non-C₄ individuals selected for anatomical analyses covered the extreme values from this range (Figure 2.1b and Supporting Information: Table 2.S5). Of these 109 individuals, 55 had $\delta^{13}\text{C}$ above -23‰ , which indicates a significant involvement of the C₄ cycle in atmospheric CO₂ capture (Stata and Sage, 2019). The selected samples, therefore, represent a panel of non-C₄ accessions with a diversity of C₄ cycle strength (Figure 2.1b).

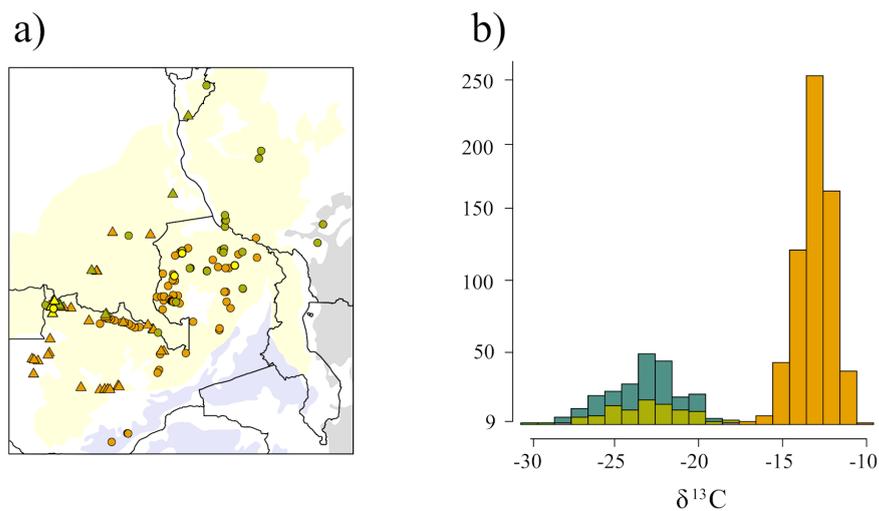


Figure 2.1

Photosynthetic diversity of *Alloteropsis semialata* in Central Zambebian miombo woodlands. (a) Sampled populations are indicated, in orange for those containing only C₄ individuals, in olive green for those containing only non-C₄ individuals, and in yellow for those containing both C₄ and non-C₄ individuals. Previously published populations are shown with circles and newly published ones with triangles. The approximate extent of miombo woodland biogeographic regions is shown in the background, in light yellow for the Central Zambebian woodlands, in light grey for the Eastern Zambebian woodlands and in light blue for Southern Zambebian woodlands (distribution based on Maquia et al., 2019). (b) Distribution of stable carbon isotope ratios for all individuals of *A. semialata* sampled from this region is represented with a histogram (see Supporting Information: Table 2.S1 for details). Values indicative of a C₄ type are in orange, while those indicative of a non-C₄ type are in olive green for individuals sampled for leaf anatomy and in pine green for other non-C₄ individuals.

2.4.2. Ecology affects the strength of the C₄ cycle

The first axis of the principal component analysis on climatic variables explains 63.1% of the total variation, while the second component explains 25.8% (Supporting Information: Figure 2.S3). The first principal component captures a combination of temperature and precipitation variables, with positive values corresponding to wetter and colder regions (Supporting Information: Table 2.S6 and Figure 2.S3). The second principal component is mainly correlated with precipitation in the coldest quarter, with positive values corresponding to drier regions during this quarter. The δ¹³C was statistically associated with the climatic first principal component, with populations inhabiting the drier and warmer habitats (negative values in the principal component analysis) having higher δ¹³C (Table 2.1a and Figure 2.S2a). These results are consistent with the hypothesis that a larger proportion of carbon is fixed via the C₄ pathway in habitats favouring C₄ plants.

Summary of multiple regression analyses.

Table 2.1:

Summary of (a) multiple regression analysis testing the effects of climate, (b) multiple regression testing the effects of climate and anatomy, and (c) phylogenetic generalised least squares (PGLS) analysis testing the effects of anatomy only. For (b) and (c), the full model tested all three anatomical variables as predictors of $\delta^{13}\text{C}$, and the reduced model was re-run after removing the least significant effect (BSD). The statistics shown are t (d.f.) p-value for each variable, with bold indicating significance, and the adjusted R² for each model.

a)

Response variable	PC1	PC2	R ²
$\delta^{13}\text{C}$	-64.03 (2,34) < 0.0001	NS	0.27

b)

Response variable	IBSF	IBSW	BSD	R ²
$\delta^{13}\text{C}$	3.34 (9,99) 0.001	4.71(9,99) < 0.0001	-2.28 (9,99) 0.02	0.57

c)

Model	IBSF	IBSW	BSD	R ²
Full PGLS	1.56 (3,26) 0.13	3.68 (3,26) 0.001	-1.24 (3,26) 0.22	0.25
Reduced PGLS	3.80 (2,27) 0.004	3.13 (2,27) 0.0007	-	0.23

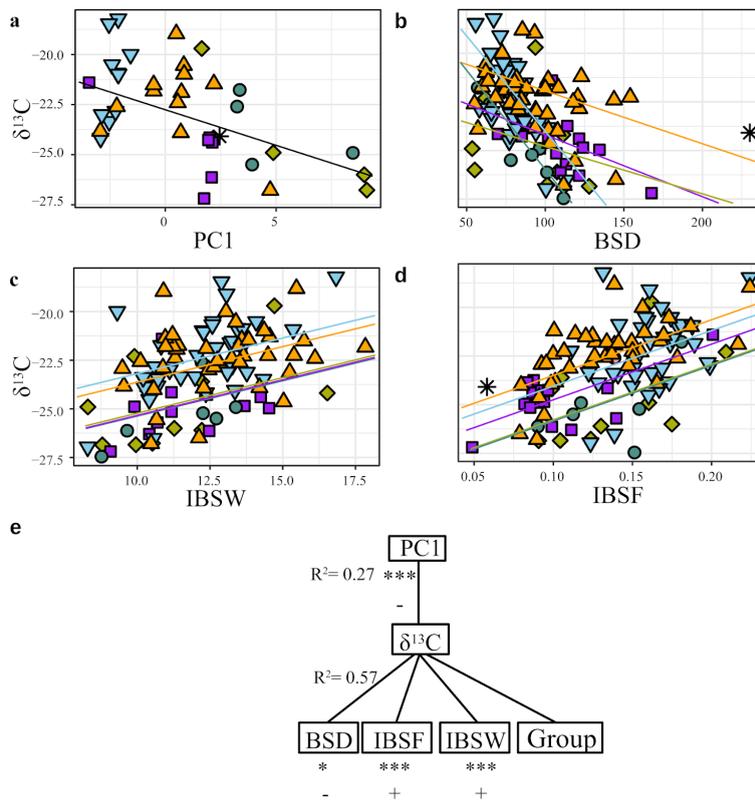


Figure 2.2

Relationships between ecology, anatomy and photosynthetic type. For the sampled non-C₄ individuals of *Alloteropsis semialata*, plots show (a) the averaged carbon isotope ratio ($\delta^{13}\text{C}$) for each population as a function of the position along the climatic first principal component (PC1), (b) the carbon isotope ratio ($\delta^{13}\text{C}$) as a function of bundle sheath distance (BSD), (c) the carbon isotope ratio ($\delta^{13}\text{C}$) as a function of inner bundle sheath width (IBSW) and (d) the carbon isotope ratio ($\delta^{13}\text{C}$) as a function of inner bundle sheath fraction (IBSF). In each case, colours and shapes indicate the main phylogenetic groups. (e) The hierarchical relationships are indicated, with in each case an indication of the directionality of the relationship (+ for positive and - for negative), its significance level (* < 0.05, *** < 0.001) and the model R².

2.4.3. Phylogenetic lineages of non-C₄ *A. semialata* occupy different regions across the Central Zambezan miombo woodlands

Phylogenetic analyses based on data extracted from the different data sets confirmed that all non-C₄ from the Central Zambezan region form a monophyletic group (corresponding to clade II of Olofsson et al., 2016, 2021), distinct from the non-C₄ accessions from southern Africa and C₄ individuals (Figure 2.3 and Supporting Information: Figure 2.S4). Within this non-C₄ group, well-supported subclades correspond to distinct geographic regions, which mostly represent

subdivisions of the two groups recognized by Olofsson et al. (2021) based on a smaller sampling (groups IIa and IIb). An accession from DRC (DRC11; new lineage IIc) is sister to all others, and a small clade containing accessions from Burundi and the middle and southern western regions of Tanzania is then sister to the rest (lineage IIb.1; Figure 2.3 and Supporting Information: Figure 2.S4). The remaining accessions are separated into four, well-supported groups. The first of these contains accessions from the south of Tanzania and the north-east of Zambia (lineage IIb.2), and the second (lineage IIb.3) is located slightly west of it, expanding into DRC (Figure 2.3 and Supporting Information: Figure 2.S4). The third of these groups (lineage IIa.1) includes individuals from the west of Zambia and adjacent areas from DRC, and the fourth (lineage IIa.2) is composed of Zambian accessions that are spread in the east of Zambia and neighbours the other groups (Figure 2.3 and Supporting Information: Figure 2.S4). Despite their geographic separation, the six groups largely overlap in the climatic space (Supporting Information: Table 2.S5 and Figure 2.3).

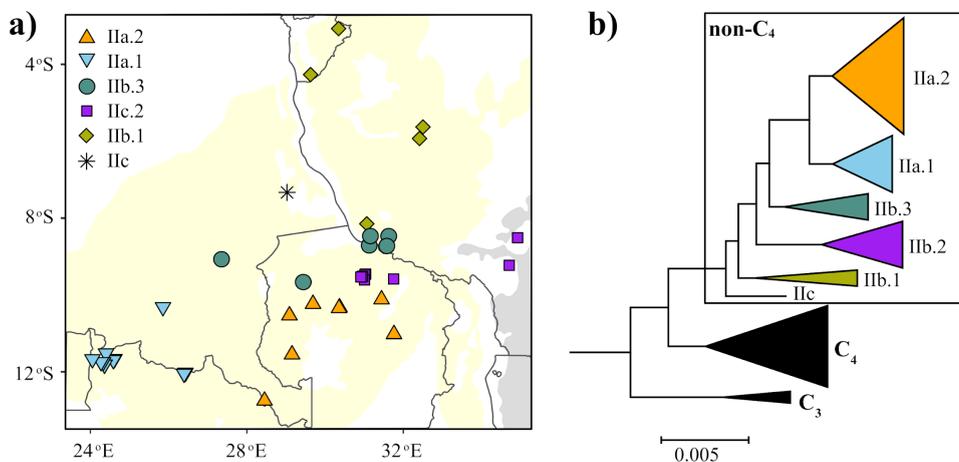


Figure 2.3

Distribution of non-C₄ accessions of *Alloteropsis semialata*. (a) The sampled populations are shown on a geographical map, with shapes and colours indicating the phylogenetic groups, as in Figure 2.2. The extent of the miombo woodlands is shown in the background, as in Figure 2.1. (b) Simplified phylogenetic tree, where the main non-C₄ lineages are collapsed and coloured as in panel a. The full phylogenetic tree is available in Supporting Information: Figure 2.S4.

2.4.4. Variation in leaf anatomy explains carbon isotope ratios

All sampled accessions had starch staining in the inner bundle sheaths, supporting the hypothesis based on $\delta^{13}\text{C}$ that this tissue is used for the Calvin–Benson–Bassham cycle in these individuals, although the mesophyll in some plants also stained for starch. While most non-C₄ individuals had only major veins (secondary and tertiary), the presence of a few minor veins was observed in some non-C₄ individuals (Supporting Information: Figure 2.S5), but these never reached the large numbers observed in C₄ accessions of *A. semialata* (Lundgren et al., 2019). Important quantitative variation was observed among the non-C₄ individuals. In particular, IBSW varied from 8.3 to 17.8 μm (Supporting Information: Table 2.S5), BSD varied from 53.5 to 230.1 μm , and the IBSF varied from 0.05 to 0.22 μm (Supporting Information: Table 2.S5).

The variation in $\delta^{13}\text{C}$ is partially explained by anatomical variation (Figure 2.2 and Table 2.1b). Multiple regression analysis of the whole data set showed that the three variables IBSF, BSD, and IBSW together explain 57% of the variation in $\delta^{13}\text{C}$. The relationships between $\delta^{13}\text{C}$ and both IBSF and IBSW are positive, so that individuals with greater amounts of bundle sheath tissues and larger BSDs have a stronger C₄ pathway. Individuals with greater amounts of bundle sheath tissues and larger bundle sheath cells, therefore, have a stronger C₄ pathway. Conversely, the relationship between $\delta^{13}\text{C}$ and BSD is negative, so that individuals with a smaller distance between consecutive bundle sheaths have a stronger C₄ cycle.

In the multiple regression model, there is a significant interaction between BSD and lineage, indicating that the intercepts and slopes change among lineages (Table 2.1 and Figure 2.2b). There is an additional effect of the phylogenetic lineage for IBSF, but the interaction between IBSF and lineage is not significant, indicating that only the intercepts change among lineages (Figure 2.2d and

Table 2.1). PGLS analysis using a subset of the data confirmed the multiple regression results, showing that IBSF and IBSW together explain 23% of the variation in $\delta^{13}\text{C}$, whereas the relationship with BSD is weaker and NS in the full model (Table 2.1c). The relationships between $\delta^{13}\text{C}$ and both IBSF and IBSW were positive and highly significant.

These results indicate that, across the studied region, individuals with anatomical traits usually associated with C₄ plants acquire a greater proportion of their atmospheric carbon via the C₄ cycle. Importantly, the same relationship is observed within a single population for which more individuals were sampled (Figure 2.4 and Supporting Information: Table 2.S7). This indicates that intraspecific variation diversity provides a substrate for natural selection, and demonstrates that there is considerable variation in both $\delta^{13}\text{C}$ and anatomy at a single location experiencing the same climatic conditions.

Since $\delta^{13}\text{C}$ may also be influenced by environmental effects on water-use efficiency and bundle sheath leakiness, we explored these alternatives using a model sensitivity analysis. The analysis indicated that some of the observed variations in $\delta^{13}\text{C}$ among individuals could arise from differences in their water-use efficiency or bundle sheath leakiness (Supporting Information: Table 2.S4).

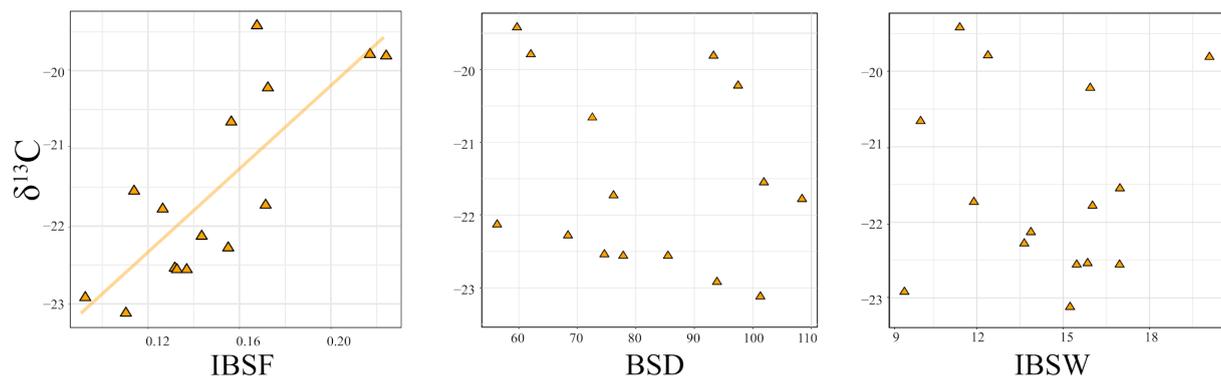


Figure 2.4

Intrapopulation variation in leaf anatomy and carbon isotope ratios. The relationship between the carbon isotope ratio ($\delta^{13}\text{C}$) and the fraction of inner bundle sheath tissue (IBSF) is shown for 15 non- C_4 individuals sampled within a single population (population ZAM1715). The phylogenetic generalized least-squares (PGLS) result is indicated ($p < 0.001$, $R^2 = 0.24$), but it is not significant for other anatomical traits (bundle sheath cell [BSD] and inner bundle sheath width [IBSW]).

2.5. Discussion

2.5.1. Anatomical variation is related to the strength of the C_4 cycle in non- C_4 plants

The distribution of carbon isotopes among non- C_4 samples of *A. semialata* collected in the Central Zambebian miombo woodlands supports this region as an important centre of variation for photosynthetic types within the group (Figure 2.1b). The most negative values could be associated with a CO_2 fixation pathway either based on the C_3 cycle or complemented by a photorespiratory pump (also called the ‘ C_2 cycle’; von Caemmerer, 2000; Khoshravesh et al., 2016; Sage et al., 2012). The few individuals with low $\delta^{13}\text{C}$ values whose physiology was previously characterized had CO_2 -compensation points that were not compatible with a pure C_3 type (e.g., from population TAN2; Lundgren et al., 2016, 2019). Although this does not necessarily imply that all non- C_4 from the region have C_4 cycle activity, we can safely conclude that the C_4 cycle in these populations with a low carbon isotope ratio is at best very weak. Conversely, the numerous non- C_4 individuals with carbon isotope values greater than -23‰ unambiguously acquired a large fraction of their CO_2 via the C_4 cycle, while still using the C_3 cycle (potentially involving a photorespiratory pump) for part

of their carbon acquisition (Brown and Hattersley, 1989; von Caemmerer, 1992; Farquhar et al., 1989; Stata and Sage, 2019). This intraspecific variation would provide a fertile ground for genome-wide association studies aiming to elucidate the genetic determinism of C₄ characters (Simpson et al., 2021), and already allows investigations about the local-scale drivers of C₄ activity.

Sensitivity analysis using a model of carbon isotope discrimination in C₃–C₄ intermediate plants indicated that some of the observed variation in $\delta^{13}\text{C}$ among individuals could arise from environmental effects on water-use efficiency in these field-sampled leaves. However, three lines of evidence argue against an environmental effect being the dominant cause of $\delta^{13}\text{C}$ variation. First, previous work with this species has shown that the anatomical characteristics observed here are associated with physiological values of CO₂-compensation point consistent with the operation of a weak C₄ cycle (Lundgren et al., 2016). Second, the same study also showed that differences in $\delta^{13}\text{C}$ among genotypes in the field are preserved in a common environment, showing that they are fixed rather than plastic (Lundgren et al., 2016). Recent experiments have supported this result, by showing that the $\delta^{13}\text{C}$ values of multiple C₃–C₄ intermediate genotypes are unaffected by temperature under controlled environmental conditions (Alenazi, unpublished data). Finally, the relationships between $\delta^{13}\text{C}$ and anatomical traits are observed within a single population at a site experiencing the same climatic conditions, as well as among populations. In combination, these findings imply that, although some of the observed variations in $\delta^{13}\text{C}$ may be environmental, the primary cause is genetic differences in leaf anatomy. However, we cannot discount the possibility that some of the differences in $\delta^{13}\text{C}$ among individuals arise from genetic variation in water-use efficiency rather than genetic variation in the strength of the C₄ cycle.

The characteristics of the inner bundle sheath, which is used to segregate the part of the C₄ cycle of *A. semialata* responsible for the release of CO₂ (Hattersley et al., 1977; Lundgren et al., 2019), vary among non-C₄ *A. semialata*, and this variation statistically explains a large part of the variation in $\delta^{13}\text{C}$ (Table 2.1 and Figure 2.2). These patterns indicate that plants with a combination of a higher fraction of the leaf dedicated to the inner bundle sheath, larger bundle sheath cells and shorter distances between consecutive bundle sheaths increase the fraction of CO₂ initially fixed by PEPC (which starts the C₄ cycle) as opposed to Rubisco (which starts the C₃ cycle, but also a potential photorespiratory pump). Both enzymes are present in the mesophyll cells of *A. semialata* (Lundgren et al., 2016; Ueno and Sentoku, 2006), although it is not established these are both active. The increased CO₂ fixation by PEPC might result from enzymatic changes, but PEPC gene expression does not vary substantially among non-C₄ individuals of *A. semialata* (Dunning et al., 2019). Even if small changes to enzyme expression and activity cannot be excluded, our results suggest that the strengthening of the C₄ biochemical cycle implied by $\delta^{13}\text{C}$ is at least partially driven by anatomical changes (Figure 2.2), which contradicts the widespread assumption that the transition from weak to strong C₄ cycle is mainly driven by the upregulation of C₄ enzymes (Heckmann et al., 2013; Sage, 2004). Our data suggest that quantitative changes in leaf anatomy are associated with the strength of the C₄ cycle in non-C₄ plants of *A. semialata*.

We hypothesize that leaf structural properties can shift the balance of carbon fixation towards the C₄ cycle. If both Rubisco and PEPC are active in the mesophyll, the enzymes will compete for CO₂ fixation. While the product of Rubisco can be directly processed in the mesophyll, the product of PEPC needs to be transported to the bundle sheath to be decarboxylated. Importantly, a high rate of mesophyll-to-bundle sheath transport relies on a concentration differential, with low C₄ acid and high C₃ acid concentrations in the bundle sheath cells compared to the mesophyll (Arrivault et al.,

2017; Schlüter et al., 2017). A long distance between mesophyll and bundle sheath cells will hamper the diffusion of metabolites, while a small bundle sheath area will be insufficient to process the large quantities of C₄ acids produced in a large area of mesophyll, weakening the biochemical pull and leading to the accumulation of C₄ acids in the mesophyll (Bräutigam et al., 2018). These products will inhibit PEPC activity (Chollet et al., 1996), thereby favouring CO₂ fixation by Rubisco. Any increase of the bundle sheath area or decrease of the distance between bundle sheaths would conversely increase the strength of the C₄ pump, which would in turn favour CO₂ fixation by PEPC over Rubisco in the mesophyll (Bräutigam et al., 2016; von Caemmerer and Furbank, 1999). Our interpretation of results is that such a process increased the strength of the C₄ cycle in some *A. semialata* populations, providing a path from weak to strong C₄ activity via incremental anatomical changes (Figures 2d and 5). Importantly, our comparative analyses show that three different leaf properties are independently associated with carbon isotope ratios in *A. semialata* (Table 2.1 and Figure 2.2). This result implies that different anatomical changes are correlated with similar strengthening of the C₄ pathway in this species, possibly providing multiple targets for natural selection.

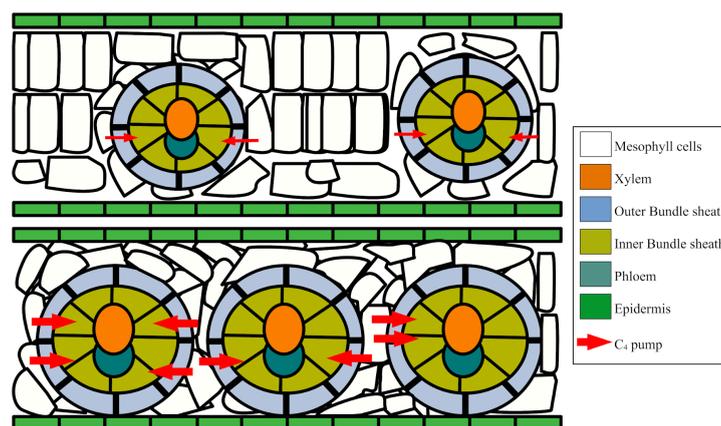


Figure 2.5

Hypothetical effect of anatomy on the strength of the C₄ cycle. Cross-sections are represented for two hypothetical individuals, with the bottom ones having larger bundle sheaths and shorter distances between them that increase the strength of the C₄ pump (represented with red arrows).

We recognize that in conducting the cross-sectional analysis of leaf samples via microscopy, it is crucial to consider the potential variation in anatomical features along the length of the leaf. The position from which a section is taken may influence the observed structures, as different regions of the leaf can exhibit distinct morphological characteristics. This consideration is particularly pertinent in studies involving herbarium specimens, where the precise origin of the sectioned material is often indeterminate.

2.5.2. Transition to stronger C₄ is associated with high temperatures

The higher carbon isotope ratios, which can be found in multiple phylogenetic lineages, are correlated with more negative values along the first principal component of the analysis of climatic data (Figure 2.2a). These values correspond to warmer and drier environments (Figure 2.2a and Table 2.1), which are known to favour C₄ over C₃ plants (Ehleringer, 1978; Ehleringer and Björkman, 1977; Hatch, 1987; Sage, 2001; Teeri and Stowe, 1976). Previous comparative work across the whole grass family has indicated that evolutionary transitions from C₃ to C₄ photosynthesis occurred in hot environments (Edwards and Smith, 2010), have been faster in tropical than temperate climates (Watcharamongkol et al., 2018), and coincided with shifts into drier regions (Edwards and Smith, 2010). Phylogeographic analyses (Bianconi et al., 2020), grounded in previous characterizations of these populations (Lundgren et al., 2015; 2016; Olofsson et al., 2016), show that the non-C₄ group of *A. semialata* likely emerged within the Central Zambezian miombo woodlands, and has since remained tightly associated with this biome. Our new results suggest that climatic variation within this region can drive the evolution of more C₄-like types, in a process of local adaptation. Microhabitat variation linked to solar radiation and surface temperature (e.g., associated with tree cover) or heterogeneity in soil moisture (e.g., associated with

soil depth or texture) might have equivalent selective effects within populations growing in the same climate. However, we do not have the fine-grained spatial data required to test this hypothesis.

The strengthening of the C₄ cycle decreases photorespiration (von Caemmerer and Furbank, 2003), which is elevated in warmer and drier areas (Ehleringer et al., 1991; Sage et al., 2012), providing the selective impetus for the observed relationships. We conclude that, as plants colonized habitats associated with increased photorespiration, anatomical changes were gradually selected to fix more CO₂ via the C₄ cycle. Over time, this led to non-C₄ populations with a strong C₄ cycle in some regions. A similar evolutionary path might have further led to plants acquiring almost all of their CO₂ via the C₄ cycle, as the C₄ group of *A. semialata* differs from the non-C₄ mainly by the presence of minor veins, which increase further the fraction of bundle sheath tissue (Lundgren et al., 2019), and the upregulation of a few genes (Dunning et al., 2019). Interestingly, the presence of sparse minor veins is detected as a rare polymorphism in some non-C₄ individuals (6.4% of the 109 individuals analysed here), which could be a result of standing genetic variation and, in a few cases, of introgression from C₄ populations (Olofsson et al., 2016). Overall, our investigations are consistent with the hypothesis that the transition from weak C₄ activity within non-C₄ individuals to fully C₄ plants can be mediated by climatic selection for anatomical changes that gradually shift the balance towards CO₂ fixation by PEPC.

Despite the overall strong relationship between anatomy and carbon isotopes, there is an additional effect of the genetic groups (Figure 2.2e). In particular, for a given fraction of bundle sheath, groups IIa.1e and IIa.2, and to a lesser extent IIb.2, have more positive carbon isotope ratios than the other groups (Figure 2.2d). While biochemical investigations are required to test this hypothesis, it is likely that this effect results from slight differences in the expression of some C₄-related enzymes.

Such properties might have evolved a limited number of times and have then been mostly retained within each group. If correct, this would indicate that biochemical changes happen infrequently and therefore lead to punctuated transitions, while anatomical tuning provides a rapid path to adaptation during the evolution of C₄ photosynthesis.

2.6. Conclusion

Our intraspecific analyses of non-C₄ *A. semialata* show that a strengthening of the C₄ cycle is statistically associated with changes in multiple leaf anatomical traits. The data are therefore consistent with the hypothesis that these changes improve the biochemical pumping of C₄ acids from mesophyll to bundle sheath cells, which shifts the fixation of CO₂ in the mesophyll towards PEPC. Importantly, stronger C₄ pathways, as detected via higher carbon isotope ratios, are correlated with warmer and drier habitats, pointing to local adaptation. Overall, these patterns suggest that, during the spread of non-C₄ *A. semialata*, habitats promoting photorespiration have selected for increased C₄ involvement. While some enzymatic changes might have happened in a punctuated manner, potentially explaining slight variation among phylogenetic lineages, our data imply that quantitative anatomical changes provided rapid evolutionary paths to physiological adaptation. Over time, such a process is likely to have led to the C₄ populations of *A. semialata* that are now found around the world.

Acknowledgements

We thank Jovin Manalayil, Louis Migerwa, Ben Vaughan and Daniel Twigg for their help in generating leaf cross-section images. The work was funded by the European Research Council (Grant No.: ERC-2014-STG-638333) and the Natural Environment Research Council (Grant No.: NE/M00208X/1). Matheus E. Bianconi was supported by the Royal Society (Grant No.: RGF\EA\181050), Pascal-Antoine Christin was funded by a Royal Society University Research Fellowship (Grant No.: URF\R\180022) and Luke T. Dunning was supported by a Natural Environment Research Council Independent Research Fellowship (Grant No.: NE/T011025/1).

2.7. References

- Arrivault, S., Obata, T., Szecówka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A. R., and Stitt, M. (2017). Metabolite pools and carbon flow during C₄ photosynthesis in maize: ¹³CO₂ labeling kinetics and cell type fractionation. *Journal of Experimental Botany*, 68, 283–298
- Bender, M.M. (1968). Mass spectrometric studies of carbon 13 variations in corn and other grasses. *Radiocarbon*, 10, 468–472.
- Bianconi, M. E., Dunning, L. T., Curran, E. V., Hidalgo, O., Powell, R. F., Mian, S., Leitch, I. J., Lundgren, M. R., Manzi, S., Vorontsova, M. S., Besnard, G., Osborne, C. P., Olofsson, J. K., and Christin, P. A. (2020). Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B. Biological sciences*, 287, 20201960.
- Bowes, G., Ogren, W.L. and Hageman, R.H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications*, 45, 716–722.
- Brautigam, C.A., Zhao, H., Vargas, C., Keller, S. and Schuck, P. (2016). Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nature Protocols*, 11, 882–894.
- Bräutigam, A., Schlüter, U., Lundgren, M.R., Flachbart, S., Ebenhöf, O. and Schönknecht, G., Christin, P. A, Bleuler, S., Droz, J.M, Osborne, C. P., Weber, P., Gowik, U. (2018). Biochemical mechanisms driving rapid fluxes in C₃–C₄ photosynthesis. *BioRxiv*. [Preprint] Available from: <https://doi.org/10.1101/387431> [Accessed 9th August 2018].
- Brown, R.H. and Hattersley, P.W. (1989). Leaf anatomy of C₃–C₄ species as related to evolution of C₄ photosynthesis. *Plant Physiology*, 91, 1543–1550.
- Brown, W.V. (1977). The Kranz syndrome and its subtypes in grass systematics. *Memoirs of the Torrey Botanical Club*, 23, 1–97.
- Burnell, J.N. and Hatch, M.D. (1988). Photosynthesis in phosphoenolpyruvate carboxykinase-type C₄ plants: pathways of C₄ acid decarboxylation in bundle sheath cells of *Urochloa panicoides*. *Archives of Biochemistry and Biophysics*, 260, 187–199.
- Cerling, T.E. (1999). Paleo Records of C₄ plants and ecosystems. In: R.F. Sage and R.K. Monson (Eds.). *C₄ plant biology*. San Diego: Academic Press, pp. 445–469.

- Cerling, T. E., Harris, J. M., MacFadden, B. J., Leakey, M. G., Quade, J., Eisenmann, V., and Ehleringer, J. R. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature*, 389, 153–158.
- Cerros-Tlatilpa, R. and Columbus, J.T. (2009). C₃ photosynthesis in *Aristida longifolia*: implication for photosynthetic diversification in Aristidoideae (Poaceae). *American Journal of Botany*, 96, 1379–1387.
- Chollet, R., Vidal, J. and O'Leary, M.H. (1996). Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 47, 273–298.
- Christin, P. A., Wallace, M. J., Clayton, H., Edwards, E. J., Furbank, R. T., Hattersley, P. W., Sage, R. F., Macfarlane and Ludwig, M. (2012). Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae. *Journal of Experimental Botany*, 63, 6297–6308.
- Dengler, N. (1994). Quantitative leaf anatomy of C₃ and C₄ grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Annals of Botany*, 73, 241–255.
- Dengler, N.G. and Nelson, T. (1999). Leaf structure and development in C₄ plants. In: Sage, R.F. and Monson, R.K. (Eds.). *C₄ plant biology*. San Diego: Academic Press, pp. 133–172.
- Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P., and Christin, P.-A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C₄ photosynthesis among close relatives. *Evolution*, 71, 1541–1555.
- Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J. K., Mapaura, A., Grundy, I. M., Kayombo, C. J., Dunning, L. A., Kentatchime, F., Ariyaratne, M., Yakandawala, D., Besnard, G., Quick, W. P., Bräutigam, A., Osborne, C. P., and Christin, P. A. (2019). Key changes in gene expression identified for different stages of C₄ evolution in *Alloteropsis semialata*. *Journal of Experimental Botany*, 70, 3255–3268.
- Edwards, E.J. and Smith, S.A. (2010). Phylogenetic analyses reveal the shady history of C₄ grasses. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2532–2537.
- Edwards, G.E. and Ku, M.S. (1987). Biochemistry of C₃–C₄ intermediates. In: M.D. Hatch and Boardman (Eds.). *The biochemistry of plants: a comprehensive treatise*, vol. 10. New York: Academic Press, pp. 275–325.
- Ehleringer, J. and Björkman, O. (1977). Quantum yields for CO₂ uptake in C₃ and C₄ plants: dependence on temperature, CO₂, and O₂ concentration. *Plant Physiology*, 59, 86–90.
- Ehleringer, J.R. (1978). Implications of quantum yield differences on the distributions of C₃ and C₄ grasses. *Oecologia*, 31, 255–267.

- Ehleringer, J.R., Sage, R.F., Flanagan, L.B. and Pearcy, R.W. (1991). Climate change and the evolution of C₄ photosynthesis. *Trends in Ecology and Evolution*, 6, 95–99.
- Ellis, R.P. (1974). Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades. *Bothalia*, 11, 273–275.
- Farquhar, G.D., Ehleringer, J.R. and Hubick, K.T. (1989). Carbon isotope discrimination and photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, 40, 503–537.
- Farquhar, G.D., O'Leary, M.H. and Berry, J.A. (1982). On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Australian Journal of Plant Physiology*, 9, 121–137.
- Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.
- Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M. and 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.
- Gutiérrez-Rodríguez, A., Latasa, M., Mourre, B. and Laws, E. (2009). Coupling between phytoplankton growth and microzooplankton grazing in dilution experiments: potential artefacts. *Marine Ecology Progress Series*, 383, 1–9.
- Guy, R.D., Reid, D.M. and Krouse, H.R. (1980). Shifts in carbon isotope ratios of two C₃ halophytes under natural and artificial conditions. *Oecologia*, 44, 241–247.
- Hatch, M. and Slack, C. (1966). Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochemical Journal*, 101, 103–111.
- Hatch, M.D. (1971). The C₄ pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C₄ dicarboxylic acid. *Biochemical Journal*, 125, 425–432.
- Hatch, M.D. (1987). C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta/Bioenergetics*, 895, 81–106.
- Hatch, M. D., and Osmond, C. B. (1976). Compartmentation and transport in C₄ photosynthesis. In Lüttge, U. and Pitman, M. G. (Eds.), *Transport in plants III: intracellular interactions and transport processes* (pp. 144–184). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hattersley, P.W. (1984). Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Annals of Botany*, 53, 163–180.
- Hattersley, P.W. and Watson, L. (1975). Anatomical parameters for predicting photosynthetic pathways of grass leaves: the “maximum lateral cell count” and the “maximum cells distant count”. *Phytomorphology*, 25, 325–333.

- Hattersley, P.W., Watson, L. and Osmond, C.B. (1977). In situ immunofluorescent labelling of Ribulose-1,5-bisphosphate carboxylase in leaves of C₃ and C₄ plants. *Australian Journal of Plant Physiology*, 4, 523–539.
- Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P., and Lercher, M. J. (2013). Predicting C₄ photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell*, 153, 1579–1588.
- Hijmans, R.J. (2021). Geographic data analysis and modeling R package raster version 3.4-10. Accessed March 16, 2023. <https://cran.r-project.org/web/packages/raster/index.html>
- Hugin Development Team. (2015). Hugin - Panorama photo stitcher. <https://hugin.sourceforge.io>
- Kennedy, R.A. and Laetsch, W.M. (1974). Plant species intermediate for C₃, C₄ photosynthesis. *Science*, 184, 1087–1089.
- Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., and Sage, T. L. (2016). C₃-C₄ intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C₂ photosynthesis. *Journal of Experimental Botany*, 67, 3065–3078.
- Ku, M.S.B., Monson, R.K., Littlejohn Jr., R.O., Nakamoto, H., Fisher, D.B. and Edwards, G.E. (1983). Photosynthetic characteristics of C₃-C₄ intermediate Flaveria species: I. Leaf anatomy, photosynthetic responses to O₂ and CO₂, and activities of key enzymes in the C₃ and C₄ pathways. *Plant Physiology*, 71, 944–948.
- Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- Lorimer, G.H. and Andrews, T.J. (1973). Plant photorespiration—an inevitable consequence of the existence of atmospheric oxygen. *Nature*, 243, 359–360.
- Lundgren, M. R., Besnard, G., Ripley, B. S., Lehmann, C. E., Chatelet, D. S., Kynast, R. G., Namaganda, M., Vorontsova, M. S., Hall, R. C., Elia, J., Osborne, C. P., and Christin, P. A. (2015). Photosynthetic innovation broadens the niche within a single species. *Ecology Letters*, 18, 1021–1029.
- Lundgren, M. R., Christin, P. A., Escobar, E. G., Ripley, B. S., Besnard, G., Long, C. M., Hattersley, P. W., Ellis, R. P., Leegood, R. C., and Osborne, C. P. (2016). Evolutionary implications of C₃-C₄ intermediates in the grass *Alloteropsis semialata*. *Plant, Cell and Environment*, 39, 1874–1885.

- Lundgren, M. R., Dunning, L. T., Olofsson, J. K., Moreno-Villena, J. J., Bouvier, J. W., Sage, T. L., Khoshravesh, R., Sultmanis, S., Stata, M., Ripley, B. S., Vorontsova, M. S., Besnard, G., Adams, C., Cuff, N., Mapaura, A., Bianconi, M. E., Long, C. M., Christin, P. A., and Osborne, C. P. (2019). C₄ anatomy can evolve via a single developmental change. *Ecology Letters*, 22, 302–312.
- Lundgren, M.R., Osborne, C.P. and Christin, P.-A. (2014). Deconstructing Kranz anatomy to understand C₄ evolution. *Journal of Experimental Botany*, 65, 3357–3369.
- Maquia, I., Catarino, S., Pena, A. R., Brito, D. R. A., Ribeiro, N. S., Romeiras, M. M., and Ribeiro-Barros, A. I. (2019). Diversification of African Tree Legumes in Miombo-Mopane Woodlands. *Plants (Basel, Switzerland)*, 8, 182.
- Monson, R.K. and Moore, B. (1989). On the significance of C₃–C₄ intermediate photosynthesis to the evolution of C₄ photosynthesis. *Plant, Cell and Environment*, 12, 689–699.
- Monson, R.K., Moore, B., KU, M.S.B. and Edwards, G.E. (1986). Co-function of C₃ and C₄ photosynthetic pathways in C₃, C₄ and C₃-C₄ intermediate Flaveria species. *Planta*, 168, 493–502.
- Monson, R.K., Teeri, J.A., Ku, M.S.B., Gurevitch, J., Mets, L.J. and Dudley, S. (1988). Carbon-isotope discrimination by leaves of Flaveria species exhibiting different amounts of C₃- and C₄-cycle co-function. *Planta*, 174, 145–151.
- O'Leary, M.H. (1981). Carbon isotope fractionation in plants. *Phytochemistry*, 20, 553–567.
- Olofsson, J. K., Bianconi, M., Besnard, G., Dunning, L. T., Lundgren, M. R., Holota, H., Vorontsova, M. S., Hidalgo, O., Leitch, I. J., Nosil, P., Osborne, C. P., and Christin, P. A. (2016). Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology*, 25, 6107–6123.
- Olofsson, J. K., Cantera, I., Van de Paer, C., Hong-Wa, C., Zedane, L., Dunning, L. T., Alberti, A., Christin, P. A., and Besnard, G. (2019). Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. *Molecular Ecology Resources*, 19, 877–892.
- Olofsson, J. K., Curran, E. V., Nyirenda, F., Bianconi, M. E., Dunning, L. T., Milenkovic, V., Sotelo, G., Hidalgo, O., Powell, R. F., Lundgren, M. R., Leitch, I. J., Nosil, P., Osborne, C. P., and Christin, P. (2021). Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology*, 30, 2116–2130.
- Olofsson, J. K., Dunning, L. T., Lundgren, M. R., Barton, H. J., Thompson, J., Cuff, N., Ariyaratne, M., Yakandawala, D., Sotelo, G., Zeng, K., Osborne, C. P., Nosil, P., and Christin, P.-A. (2019). Population-Specific Selection on Standing Variation Generated by Lateral Gene Transfers in a Grass. *Current Biology*, 29, 3921–3927.e5.

- Orme, David., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, Nick and Pearse, W. (2012). CAPER: Comparative Analyses of Phylogenetics and Evolution in R.
- Peixoto, M.M., Sage, T.L., Busch, F.A., Pacheco, H.D.N., Moraes, M.G., Portes, T.A., Almeida, R. A. Graciano-Ribeiro, D., Sage, R.F. (2021). Elevated efficiency of C₃ photosynthesis in bamboo grasses: a possible consequence of enhanced refixation of photorespired CO₂. *GCB Bioenergy*, 13, 941–954.
- R Core Team. (2018). R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. Available from: <https://www.R-project.org/>
- R Core Team. (2020). The R Project for Statistical Computing. <https://www.r-project.org>
- Sage, R.F. (2001). Environmental and evolutionary preconditions for the origin and diversification of the C₄ photosynthetic syndrome. *Plant Biology*, 3, 202–213.
- Sage, R.F. (2004). The evolution of C₄ photosynthesis. *New Phytologist*, 161, 341–370.
- Sage, R.F., Christin, P.-A. and Edwards, E.J. (2011). The C₄ plant lineages of planet Earth. *Journal of Experimental Botany*, 62, 3155–3169.
- Sage, R.F., Khoshravesh, R. and Sage, T.L. (2014). From proto-Kranz to C₄ Kranz: building the bridge to C₄ photosynthesis. *Journal of Experimental Botany*, 65, 3341–3356.
- Sage, R.F., Sage, T.L. and Kocacinar, F. (2012). Photorespiration and the evolution of C₄ photosynthesis. *Annual Review of Plant Biology*, 63, 19–47.
- Sayre, R.T. and Kennedy, R.A. (1977). Ecotypic differences in the C₃ and C₄ photosynthetic activity in *Mollugo verticillata*, a C₃-C₄ intermediate. *Planta*, 134, 257–262.
- Schlüter, U., Bräutigam, A., Gowik, U., Melzer, M., Christin, P. A., Kurz, S., Mettler-Altmann, T., and Weber, A. P. (2017). Photosynthesis in C₃-C₄ intermediate *Moricandia* species. *Journal of Experimental Botany*, 68, 191–206.
- Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9, 671–675.
- Simpson, C.J.C., Reeves, G., Tripathi, A., Singh, P. and Hibberd, J.M. (2021). Using breeding and quantitative genetics to understand the C₄ pathway. *Journal of Experimental Botany*, 73, 3072–3084.
- Smith, B.N. and Brown, W.V. (1973). The Kranz syndrome in the Gramineae as indicated by carbon isotopic ratios. *American Journal of Botany*, 60, 505–513.
- Smith, B.N. and Epstein, S. (1971). Two categories of ¹³C/¹²C ratios for higher plants. *Plant Physiology*, 47, 380–384.

- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
- Stata, M., Sage, T.L. and Sage, R.F. (2019). Mind the gap: the evolutionary engagement of the C₄ metabolic cycle in support of net carbon assimilation. *Current Opinion in Plant Biology*, 49, 27–34.
- Teeri, J.A. and Stowe, L.G. (1976). Climatic patterns and the distribution of C₄ grasses in North America. *Oecologia*, 23, 1–12.
- Troughton, J.H. (1979). $\delta^{13}\text{C}$ as an indicator of carboxylation reactions. In: M. Gibbs and E. Latzko (Eds.) *Photosynthesis II. Encyclopedia of Plant Physiology*. Berlin: Springer.
- Ueno, O. and Sentoku, N. (2006). Comparison of leaf structure and photosynthetic characteristics of C₃ and C₄ *Alloteropsis semialata* subspecies. *Plant, Cell and Environment*, 29, 257–268.
- von Caemmerer, S. (1992). Carbon isotope discrimination in C₃–C₄ intermediates. *Plant, Cell and Environment*, 15, 1063–1072.
- von Caemmerer, S. (2000). *Biochemical models of leaf photosynthesis*. Collingwood: CSIRO.
- von Caemmerer, S. and Furbank, R.T. (1999). The modelling of C₄ photosynthesis. In: R.F. Sage and R.K. Monson (Eds.) *C₄ plant biology*. San Diego: Academic Press, pp. 173–211.
- von Caemmerer, S. and Furbank, R.T. (2003). The C₄ pathway: an efficient CO₂ pump. *Photosynthesis Research*, 77, 191–207.
- Watcharamongkol, T., Christin, P.-A. and Osborne, C.P. (2018). C₄ photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology Letters*, 21, 376–383.
- Williams, B.P., Johnston, I.G., Covshoff, S. and Hibberd, J.M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C₄ photosynthesis. *eLife*, 2, e00961.
- Winter, K. (1981). CO₂ and Water Vapour Exchange, Malate Content and $\delta^{13}\text{C}$ Value in *Cicer arietinum*: Grown under Two Water Regimes. *Zeitschrift für Pflanzenphysiologie*, 10, 421–430.
- Yorimitsu, Y., Kadosono, A., Hatakeyama, Y., Yabiku, T. and Ueno, O. (2019). Transition from C₃ to proto-Kranz to C₃–C₄ intermediate type in the genus *Chenopodium* Chenopodiaceae. *Journal of Plant Research*, 132, 839–855.

Supplementary Figures

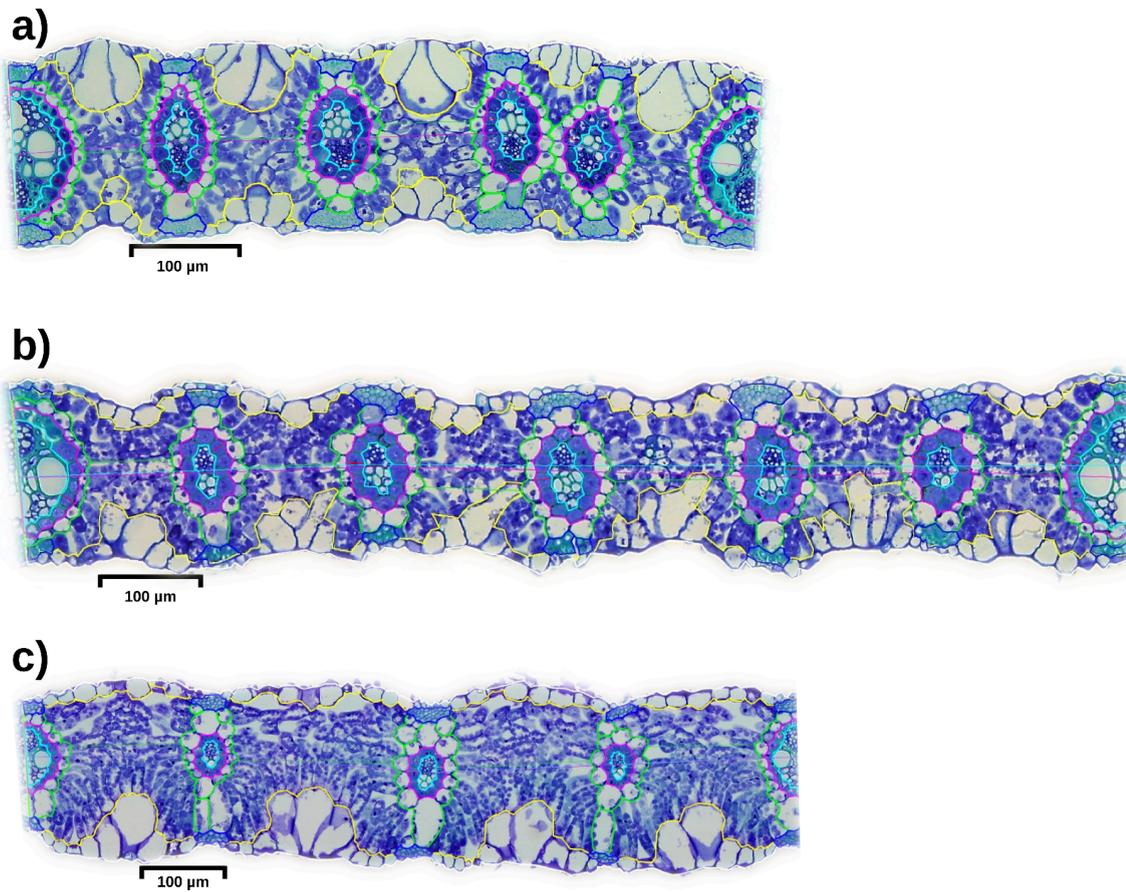


Figure S1: Examples of leaf sections.

Leaf cross sections are shown for three individuals that represent the extreme values of anatomical traits observed among non-C₄ individuals; a) individual ZAM2090-03 ($\delta^{13}\text{C} = -18.5\%$, BSD = 56 μm , IBSW = 12.9 μm and IBSF = 0.22), b) ZAM20103-03 ($\delta^{13}\text{C} = -23\%$, BSD = 66 μm , IBSW = 11.1 μm and IBSF = 0.20), and c) individual ZAM1940-01 ($\delta^{13}\text{C} = -27.2\%$, BSD = 168 μm , IBSW = 9.1 μm and IBSF = 0.05). The measured areas are shown for individual ZAM1940-01, with the inside of the epidermis delimited in yellow, the fibers delimited in dark blue, the outside of the outer sheath bundle in green, the outside of the inner bundle sheath in magenta, and the outside of the vascular bundle in cyan. The green lines show the distances between consecutive bundle sheaths and the magenta segments connect the centers of veins. The horizontal red line inside the inner cell is indicating to the width of inner bundle sheath.

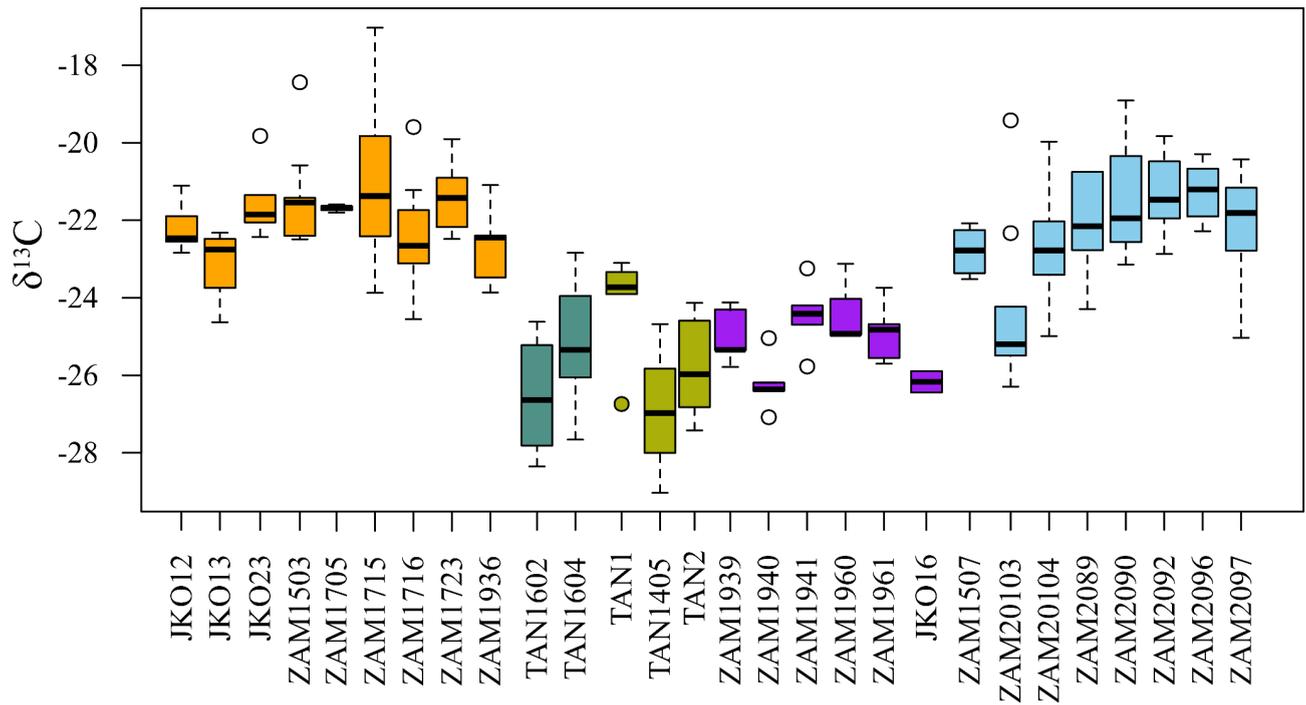


Figure S2: Variation in carbon isotope ratios within each population.

Vertical extending lines denote adjacent values, and horizontal black lines within each box represent median values; dots indicate observations outside the range of adjacent values. In each case, shapes and colours indicate the phylogenetic groups, as in Figure 2.2.

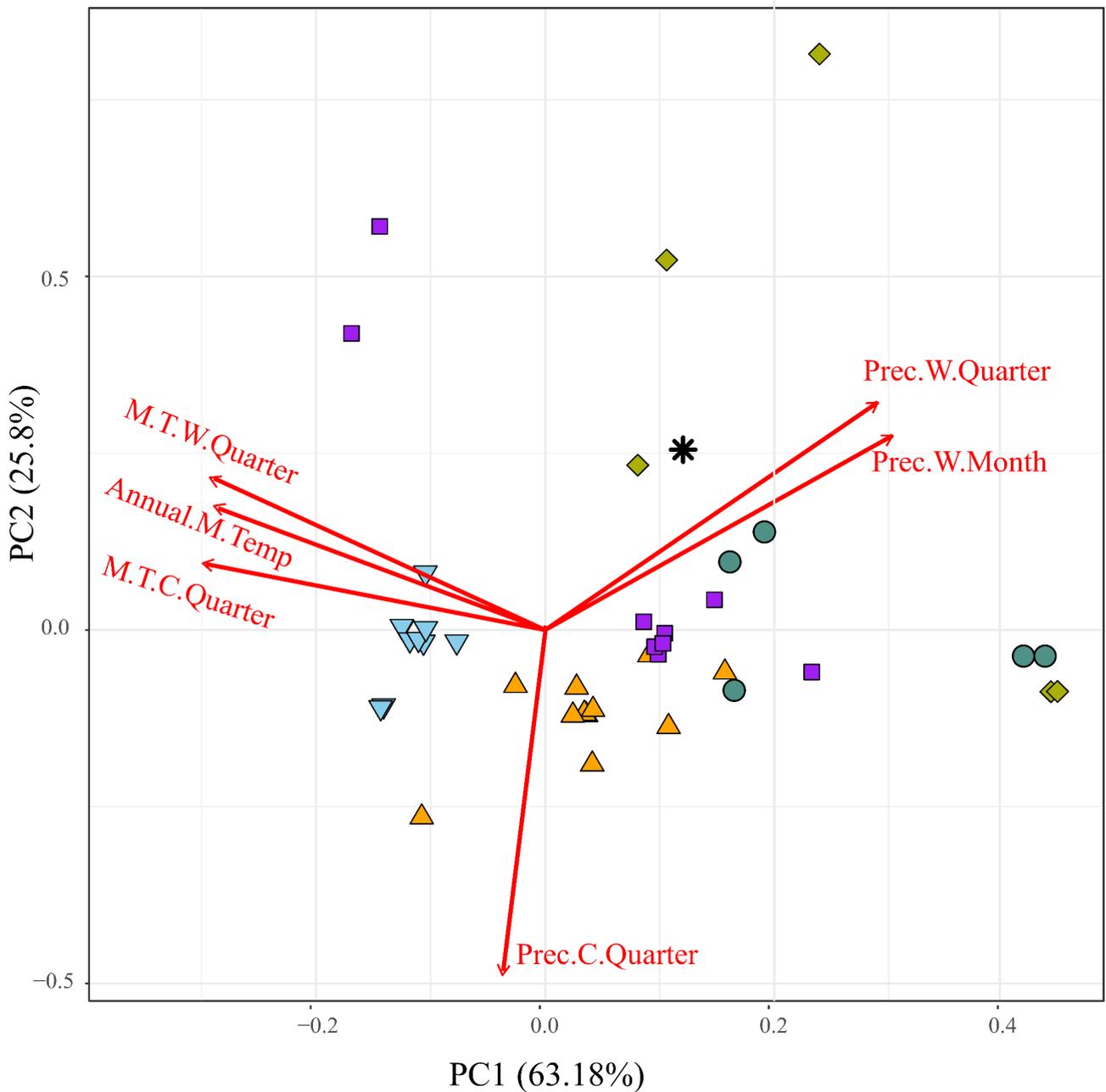


Figure S3: Principal component analysis of climatic variables.

The populations are plotted in the space created by the first two principal components and are indicated with shapes and colours that match the phylogenetic groups, as in Figure 2.2. The climatic variables are plotted in the same space and indicated with arrows. M.T.W.Quarter = Mean temperature of wettest quarter, M.T.C.Quarter = Mean temperature of coldest quarter, Annual.M.Temp = Annual mean temperature, Prec.W.Month = Precipitation of wettest month, Prec.W.Quarter = precipitation of wettest quarter, Prec.C.Quarter = Precipitation of coldest quarter and M.T.W.Quarter = Mean temperature of wettest quarter.

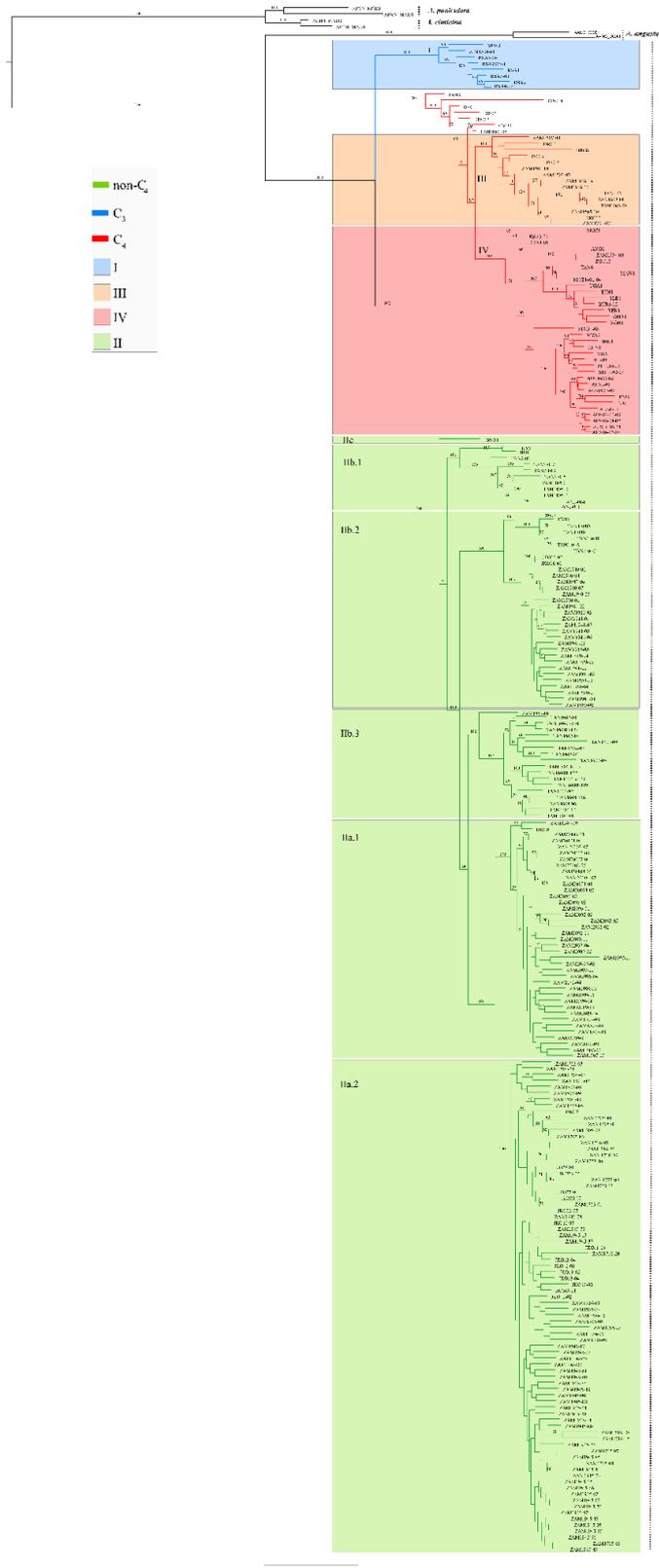


Figure S3:

The full resolution phylogeny tree is found in <https://doi.org/10.1111/pce.14607>

Supplementary Tables

Supplementary tables of this chapter are available at:

<https://doi.org/10.1111/pce.14607>

Chapter 3. Identifying genomic regions associated with C₄ photosynthetic activity and leaf anatomy in *Alloteropsis semialata*

Chapter 3. Identifying genomic regions associated with C₄ photosynthetic activity and leaf anatomy in *Alloteropsis semialata*

Ahmed S Alenazi^{1,2*}, Lara Pereira^{2*}, Pascal-Antoine Christin², Colin P Osborne³, Luke T Dunning^{2**}

Affiliations:

¹Department of Biological Sciences, College of Science, Northern Border University, Arar 91431, Saudi Arabia.

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

³Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

*These authors contributed equally to the work

**Corresponding author: Luke T. Dunning; Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom; +44 (0) 1142220027; l.dunning@sheffield.ac.uk

Data availability

All *A. semialata* genomic data was previously published, and the additional phenotype data generated here is available in the SI.

This work is published in **New Phytologist**. Volume243, Issue5, September 2024.

Author contributions

ASA, LP, PAC, CPO and LTD designed the study. ASA conducted the experimental work and generated the phenotype data. ASA, LP and LTD analysed the data. All authors interpreted the results and helped write the manuscript.

Personal contribution: I generated cross section data, measured the anatomical traits and performed statistical analysis. I also performed GWAS, LD and selection test analyses with guidance from Dr. Lara Pereira and Dr. Luke T Dunning and then wrote the manuscript with help from co-authors.

3.1. Abstract

C₄ photosynthesis is a complex trait requiring numerous developmental and metabolic alterations to increase productivity in hot, high light and dry environments. Despite this complexity, it has convergently evolved over 60 times. Comparisons between species with different photosynthetic types have identified the core C₄ enzymes and numerous related genes. However, understanding the individual steps in the transition to C₄ is complicated by the fact that this phenotypic variation is often segregated between species that have been independently evolving for millions of years. Here we leverage the photosynthetic diversity within the grass *Alloteropsis semialata*, the only known species to have C₃, intermediate C₃+C₄, and C₄ genotypes. Using a genome wide association study (GWAS) we identify genomic regions associated with the strength of the C₄ cycle, measured using the $\delta^{13}\text{C}$ carbon isotope ratio. These regions include regulators of C₄ decarboxylation enzymes (*RIPK*) and partitioning of the C₄ cycle (*SLP1*). We then conducted a GWAS for leaf morphological traits correlated with $\delta^{13}\text{C}$ in the C₃+C₄ intermediates, focusing on those intermediate for the expansion of the C₄ bundle sheath tissue. Associated genomic regions include leaf anatomy regulators, such as genes associated with leaf vein patterning (*GSL8*) and meristem determinacy (*GRF1*). Both the strengthening of the C₄ cycle and leaf anatomy modifications appear to have a relatively simple genetic architecture in *A. semialata*, supporting a gradual stepwise transition. The association of these candidate genes with C₄ photosynthesis will require further functional verification, which may prove to be relevant for engineering a C₄ leaf anatomy in C₃ species.

3.2. Introduction

Oxygenic photosynthesis originated over 2 billion years ago and is the ultimate source of nearly all energy used by living organisms. Almost 90% of plants fix carbon using the ancestral C₃ cycle, but this process is inefficient in hot environments (Sage and Monson, 1999). This is because the key enzyme responsible for the initial fixation of atmospheric CO₂ (Ribulose-1,5-bisphosphate carboxylase/oxygenase, Rubisco) is less able to discriminate CO₂ from O₂ at higher temperatures, and as a result, energy is lost through photorespiration (Farquhar et al., 1982). To reduce photorespiration, plants have evolved C₄ photosynthesis, wherein atmospheric CO₂ is initially assimilated into a 4-carbon organic acid by phosphoenolpyruvate carboxylase (PEPC) in the mesophyll cells, before shuttling the acid to the neighboring bundle sheath cells where it is decarboxylated and the CO₂ recaptured by Rubisco (Hatch, 1971; Edwards and Ku, 1987). This compartmentalization of Rubisco effectively prevents photorespiration. C₄ photosynthesis is a complex trait that relies on both changes to the leaf anatomy and the coordinated regulation of multiple metabolic enzymes (Hatch, 1987). In order to understand the sequence of events that led to C₄ evolution, comprehensive genomic and phenotypic datasets have been generated in many systems, such as *Flaveria* (Adachi et al., 2023) and *Alloteropsis* (Pereira et al., 2023). These existing data sets can potentially be mined for quantitative genetics approaches to identify novel genetic factors involved in the evolution of C₄ (Simpson et al., 2021).

By comparing species with different photosynthetic types, the core C₄ enzymes, multiple accessory genes, and loci associated with C₄ leaf anatomy (often termed ‘Kranz’ anatomy) have been identified (Langdale et al., 1987, 1988; Slewinski et al., 2012; Cui et al., 2014). However, decomposing the individual steps during the transition to C₄ is confounded by the fact that variation in photosynthetic type is usually segregated between distinct species that have been independently

evolving for millions of years, meaning that they differ in many aspects besides those linked to the photosynthetic pathway (Heyduk et al., 2019). The interspecific segregation of variation in photosynthetic type makes it challenging to apply quantitative genetics methods, such as quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS), since these rely on traits varying within a species, or the ability to hybridize species with divergent phenotypes. GWAS has been used to investigate the variation of C₄ traits within C₄ species, such as photosynthetic performance during chilling in maize (Strigens et al., 2013), and to identify genes associated with stomatal conductance and water use efficiency in sorghum (Ferguson et al., 2021; Pignon et al., 2021). However, to date there has been no QTL region identified for differences in C₄ carbon fixation or Kranz anatomy (Simpson et al., 2021).

In grasses, the proportion of carbon that is fixed through the C₄ cycle can be measured using the stable carbon isotope ratio ($\delta^{13}\text{C}$) (O'Leary, 1981; Farquhar et al., 1989). Both ¹²C and ¹³C occur naturally in the atmosphere, and in C₃ plants, Rubisco preferentially fixes ¹²C during photosynthesis (O'Leary, 1981). Conversely, in C₄ plants, carbon is initially fixed by CA and PEPC, and this coupled enzyme system discriminates less than Rubisco between the two isotopes (O'Leary, 1981). The rate of CO₂ release in the bundle sheath is coordinated with the rate of CO₂ fixation by Rubisco, which reduces the fractionation effect of this enzyme. $\delta^{13}\text{C}$ is therefore commonly used as a proxy for photosynthetic type and the relative strength of the C₄ cycle (Bender, 1968; Smith and Epstein, 1971; Smith and Brown, 1973; Von Caemmerer, 1992; Cerros-Tlatilpa and Columbus, 2009; Gowik et al., 2011; Lundgren et al., 2015; Stata et al., 2019; Olofsson et al., 2021). While there is intraspecific variation in $\delta^{13}\text{C}$ for C₄ species such as maize and *Gynandropsis* (Voznesenskaya et al., 2007), we do not know whether this variation arises from differences in anatomy or biochemistry (Simpson et al., 2021). In addition, some of the observed variation in $\delta^{13}\text{C}$

could also be due to environmental effects on water use efficiency (Farquhar and Richards, 1984), particularly if the phenotypic data comes from individuals sampled in the field. However, differences in the $\delta^{13}\text{C}$ between accessions of some species are maintained in a common environment (Lundgren et al., 2016), indicating that the $\delta^{13}\text{C}$ ratio likely has a genetic component. Intraspecific, heritable variation in $\delta^{13}\text{C}$ offers an excellent opportunity for using quantitative genetic approaches to discover C₄ QTLs.

The grass *Alloteropsis semialata* has long been used as a model to study C₄ evolution, since it is the only species known to have C₃, C₄, and intermediate genotypes that diverged relatively recently and can be crossed, allowing gene flow among them (reviewed by Pereira et al., 2023). The common ancestor of this species is thought to be an intermediate with some chloroplasts in its bundle sheath and performing a very weak C₄ cycle, with the C₃ being a reversal from this intermediate state as that lineage colonized cooler environments in southern Africa (Dunning et al., 2017). The intermediate populations are found in the grassy ground layer of the Central Zambezian miombo forests that we refer to as ‘C₃+C₄’ because they perform a weak C₄ cycle in addition to directly fixing CO₂ through the C₃ cycle (Lundgren et al., 2016; Dunning et al., 2017). Comparative studies have shown that the transition to a purely C₄ physiology in *A. semialata* is caused by the overexpression of relatively few core C₄ enzymes (Dunning et al., 2019a) and the acquisition of C₄-like morphological traits, notably the presence of minor veins (Lundgren et al., 2019). The $\delta^{13}\text{C}$ of the C₃+C₄ plants ranges from values characteristic of a weak (or absent) C₄ cycle to values that show that the C₄ cycle accounts for more than half of the carbon acquisition (Von Caemmerer, 1992; Lundgren et al., 2015; Stata et al., 2019; Olofsson et al., 2021). Furthermore, the strengthening of the C₄ cycle in the C₃+C₄ intermediates (measured using $\delta^{13}\text{C}$) is significantly associated with alterations in a number of leaf anatomical traits related to the preponderance of inner bundle sheath

(IBS) tissue, the cellular location of the C₄ cycle in this species (Alenazi et al., 2023), including the distance between consecutive bundle sheaths, the width of IBS cells, and the proportion of bundle sheath tissue in the leaf (Alenazi et al., 2023).

Alloteropsis semialata therefore represents an ideal system to identify the genes correlated with the strengthening of the C₄ cycle. Here, we first conducted a global analysis to identify candidate genes associated with the strength of the C₄ cycle ($\delta^{13}\text{C}$) using genomic data from 420 individuals representing C₃, C₃+C₄, and C₄ phenotypes. We then focused specifically on the C₃+C₄ intermediates, to identify candidate genes associated with the relative expansion of bundle sheath tissue during the transition from a weak to a strong C₄ cycle. The high level of interspecific variation in *A. semialata* permits a fine-scale understanding of the genetic basis of C₄ evolution, including the intermediate steps involved in the assembly of this complex trait. This is crucially important to identify the initial changes required for the emergence of this trait, something that may ultimately have applications in the engineering of C₄ photosynthesis in C₃ crops such as rice.

3.3. Materials and Methods

3.3.1. Genome data, $\delta^{13}\text{C}$ values, and population genetic analyses

For the genomic analyses, we compiled previously published double digest restriction-site associated DNA sequencing (ddRADSeq) data sets for *Alloteropsis semialata* (R. Br.) Hitchc. individuals that also had known $\delta^{13}\text{C}$ values from field-collected leaves measured using mass spectrometry (Lundgren et al., 2015, 2016; Bianconi et al., 2020; Olofsson et al., 2021; Alenazi et al., 2023). Depending on the source of the $\delta^{13}\text{C}$ values, these were either single measures (Lundgren et al., 2015, 2016; Bianconi et al., 2020), replicated if the $\delta^{13}\text{C}$ values did not match other individuals of the population and genomic group (Olofsson et al., 2021), or medians of triplicate

technical replicates if sufficient material was available (Alenazi et al., 2023). In total, the data set comprised 420 individuals collected from 87 populations across Africa and Asia (Supporting Information Table 3.S1), representing the full range of photosynthetic types found in *A. semialata* ($45 \times C_3$; $132 \times C_3+C_4$; $243 \times C_4$).

The ddRADseq data were downloaded from NCBI Sequence Read Archive and cleaned using TRIMMOMATIC v.0.38 (Bolger et al., 2014) to remove adapter contamination (ILLUMINACLIP option in palindrome mode) and low-quality bases ($Q < 3$ from both 5' and 3' ends; $Q < 15$ for all bases in four-base sliding window). The cleaned ddRADseq data were then mapped to a chromosomal scale *A. semialata* reference genome for a C_4 Australian individual (Dunning et al., 2019b) using BOWTIE2 v.2.2.3 with default parameters (Langmead and Salzberg, 2012). We called single-nucleotide polymorphisms (SNPs) from these alignments using the GATK v.3.8 (McKenna et al., 2010) pipeline with default parameters. We generated individual variant files (gVCF) with HAPLOTYPECALLER and then combined them into a single multi-sample VCF file with Genotype GVCFs. Biallelic SNPs were extracted from this file using SELECTVARIANTS, and high-quality SNPs retained using VARIANTFILTRATION ($MQ > 40$, $QD > 5$, $FS < 60$, $MQRankSum > -12.5$ $ReadPosRankSum > -8$). Finally, we used VCFTOOLS to filter the remaining SNPs to remove those with $> 30\%$ missing data and/or a minor allele frequency < 0.05 (Danecek et al., 2011).

The evolutionary relationship among individuals was inferred using a maximum likelihood phylogenetic tree. We used VCF2PHYLIP v.2.8 (Ortiz, 2019) to generate a nucleotide alignment from the filtered VCF file. To reduce the effect of linked SNPs on phylogenetic reconstruction, we thinned the data set so that SNPs were at least 1 kb apart (starting from the first SNP on each chromosome). The phylogenetic tree was inferred using RAXML v.8.2.12 (Stamatakis, 2014) with

the GTRCAT model and 100 bootstrap replicates (Dataset S1). Finally, to verify previous phylogenetic groupings (Alenazi et al., 2023), we determined the population structure of the C_3+C_4 accessions using ADMIXTURE v.1.3.0 (Alexander et al., 2009). We ran the analysis with multiple values of k (range: 2–7), with 10 replicate runs for each value. The optimal k was inferred using Admixture's cross-validation error method. We also used PLINK-v.1.9 to perform a principal component analysis to quantify population structure and to generate a pairwise kinship matrix (Purcell et al., 2007).

3.3.2 Leaf anatomical traits of C_3+C_4 *A. semialata*

Leaf anatomy data for all 132 C_3+C_4 individuals were either extracted from a previous study ($n = 100$; Alenazi et al., 2023) or generated here using the same method from field-preserved samples ($n = 32$; Table 3.S2). The measurements themselves were taken from leaf cross-sections that were prepared from silica-dried leaf material following the method described by Alenazi *et al.* (2023). The slide images were captured using a mounted camera on an Olympus BX51 microscope (Olympus, Hamburg, Germany), and images from the same leaf were stitched together with Hugin's software (Hugin Development Team, 2015). All measurements of leaf anatomical characteristics were made using IMAGEJ v.1.53f (Schneider et al., 2012), avoiding the midrib and leaf margins. For each individual, the anatomical measurements are based on the mean of at least five technical replicates, each measured between independent pairs of secondary veins in the same cross-section using.

We recorded the total cross-sectional areas between secondary veins (i.e. veins accompanied by extraxylary fibers and epidermal thinning) for mesophyll (including airspaces; MS) and IBS tissues (Figure 3.1). We used these values to then calculate the inner bundle sheath fraction

(IBSF = IBS/[MS + IBS]), which is the portion of the photosynthetic part of the leaf that can be responsible for refixing carbon obtained through the C₄ cycle. Finally, we also measured the bundle sheath distance (BSD) and the inner bundle sheath width (IBSW) using the mean widths of equatorial cells.

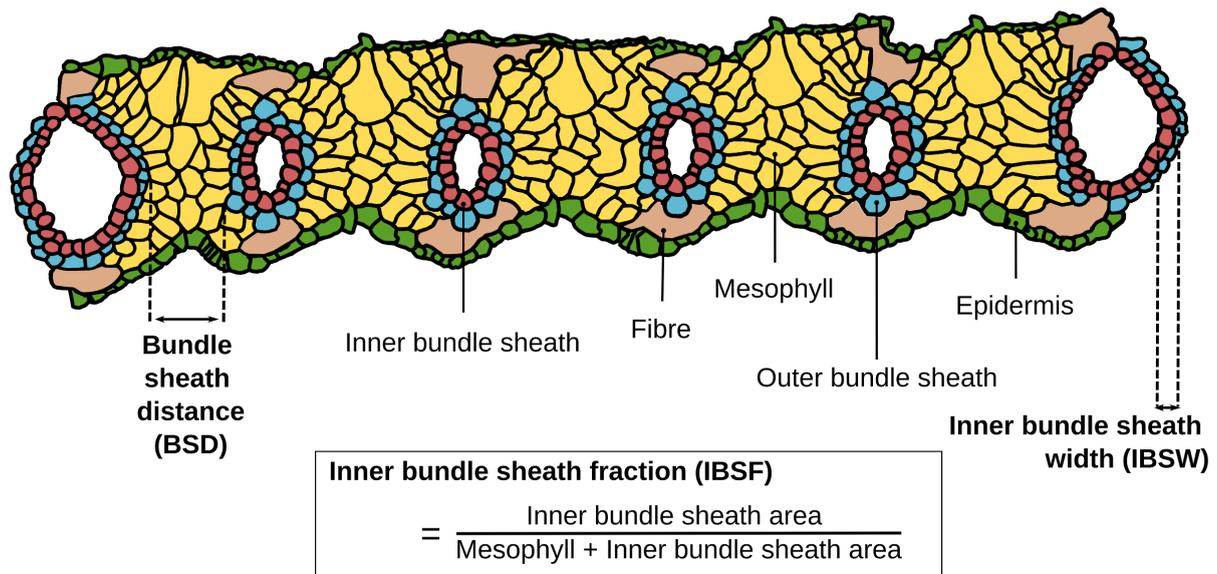


Figure 3.1: A cross-section of a typical C₄ *Alloteropsis semialata* leaf.

The schematic is traced from a cross-section of accession JKO23-03_16, with tissue types labelled. The anatomical measurements used for the genome-wide association study (GWAS) are indicated in bold.

3.3.3. Estimating trait heritability

To estimate the proportion of phenotypic variation explained by underlying genetic differences, we calculated the heritability of $\delta^{13}\text{C}$ (complete and restricted C₃+C₄ datasets) and the leaf anatomical traits (C₃+C₄ dataset) using Genome-wide Complex Trait Analysis (GCTA) v.1.94.1 (Yang et al., 2011). A genetic relationship matrix was inferred from the previously generated SNP calls and combined with the phenotype values in GCTA. Heritability was then estimated for each trait using the restricted maximum likelihood method.

3.3.4. Genome-wide association study of photosynthetic traits

We performed a GWAS for the strength of the C₄ cycle measured using $\delta^{13}\text{C}$, and leaf anatomy traits previously correlated with the strength of the C₄ cycle (IBSF, BSD, and IBSW; Figure 3.1; Alenazi et al., 2023), with the objective of ultimately proposing some candidate genes underpinning the phenotype. We used the variation in photosynthetic type which exists across *A. semialata* as a whole, before focusing on anatomical variation in the C₃+C₄ individuals that have been associated with the strength of the C₄ cycle (Alenazi et al., 2023). We defined our associated regions of the genome as the linkage block containing a significant SNP from the GWAS. We then identified the gene models located within the correlated region as candidate genes and assessed their functional relevance, gene expression pattern, and selective forces they have been evolving under.

The GWAS itself was performed using the `RMVP` package (Yin et al., 2021) in `RSTUDIO` v.4.3, with the `MVP.Data` function and default parameters used for single-locus GWAS analysis for each phenotypic trait with the fixed and random model circulating probability unification (FarmCPU) approach (Yin et al., 2021). Population structure and genetic relatedness can confound a GWAS and result in false associations (Chen et al., 2016). We therefore included the previously generated pairwise kinship matrix so that the relationships among individuals could be accounted for. The phenotypic data for each trait were normalized (if required), and a Bonferroni corrected SNP significance threshold of $P \leq 0.05$ was used.

3.3.5. Linkage disequilibrium

Linkage blocks are regions of the genome that are likely to be co-inherited, and the association of the significant SNPs identified from the GWAS could be caused by any gene within this region. To determine the linkage block encapsulating each SNP, we used `HAPLOVIEW` v.4.1 (Barrett et al., 2005).

The input map and binary files were processed using PLINK v.1.9 (Purcell et al., 2007), and we used a solid spine of linkage disequilibrium (LD) with default parameters to infer linkage block size (Kim et al., 2018). This approach requires the first and last SNPs in a block to be in strong LD with all intermediate markers (normalized deviation (D') ≥ 0.8), but the intermediate markers do not necessarily need to be in LD with each other. Identifying linkage blocks is heavily impacted by the distribution of SNPs across the genome, something that is accentuated by reduced sequencing methods such as ddRADSeq. We therefore used the genome-wide mean linkage block size if the analysis failed to place a significant SNP in a block of its own (Figure 3.S1). To do this, we positioned the significant SNP at the center of the artificial linkage block and if necessary truncated it to avoid incorporating unlinked SNPs up and/or downstream from this marker.

3.3.6 Identification of candidate genes

The linkage blocks associated with the phenotype of interest contain the causal gene(s) in addition to those that happen to be in close physical linkage (hitchhiking). To try and identify plausible candidate genes in each region, we compared their functional annotations, expression patterns, and the selective pressures they are evolving under.

ORTHOFINDER v.2.5.4 (Emms and Kelly, 2015) was used to identify orthologous genes to the loci in the associated regions. To do this, we combined the *A. semialata* protein sequences with nine other plant species (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Hordeum vulgare*, *Oryza sativa*, *Physcomitrium patens*, *Solanum lycopersicum*, *Triticum aestivum*, and *Zea mays*) downloaded from Phytozome v.13 (Goodstein et al., 2012). Orthogroup phylogenies are presented in Dataset S2. We then used publicly available databases (e.g. TAIR (Berardini et al., 2015), RAP-DB (Sakai et al.,

2013), and maizeGDB (Monaco et al., 2013)) and literature searches to extrapolate the functions of each orthogroup containing a gene from a correlated linkage block identified from the GWAS.

Gene expression data for the candidate genes was extracted from a Dunning et al. (2019a). The gene expression data come from mature leaf tissue grown under controlled conditions (60% relative humidity, day/night temperatures of 25/20°C), sampled in the middle of the photoperiod (Dunning et al., 2019a). To test for differential expression between the photosynthetic types, we used two-tailed *t*-tests, with *P*-values Bonferroni corrected to account for multiple testing.

Finally, we used whole-genome resequencing data (Bianconi et al., 2020) for 45 *A. semialata* individuals to determine whether the genes in the GWAS regions were evolving under positive selection. In short, the datasets were downloaded from NCBI sequence read archive and mapped to the reference genome using BOWTIE2, and consensus sequences generated using previously developed methods (Olofsson et al., 2016; Dunning et al., 2022), and a maximum likelihood phylogeny tree for each gene was inferred using RAxML (Stamatakis, 2014) with 100 bootstrap. We then inferred the selective pressure each gene was evolving under by running the M0 model in CODEML v.4.9 h.

3.4. Results

3.4.1. Population structure

The broadscale phylogenetic (Figure 3.2a) and population genetic (Figure 3.2b) analyses recovered those previously inferred by earlier studies, with the different photosynthetic types (C₃, C₃+C₄, and C₄) belonging to separate clades (Olofsson et al., 2016, 2021; Bianconi et al., 2020). Within the C₃+C₄ intermediates, individuals are separated into five populations geographically spread across

the Central Zambesian miombo woodlands (Figure 3.2). This reconfirms the phylogenetic groupings previously demarcated (Alenazi et al., 2023), although the earliest diverging sixth lineage is absent in this study because it is only represented by a single herbarium individual from the Democratic Republic of the Congo and lacks ddRADSeq data. The population structure analysis (Figure 3.2c) largely concurs with the phylogenetic groupings, although it indicates gene flow between populations. The distribution of the C_3+C_4 groups has a pattern largely matching a scenario of isolation by distance along an east–west axis through Zambia and Tanzania (Figure 3.2d). $Q-Q$ plots show no sign of P -value inflation for the subsequent GWAS results, indicating that population structure was sufficiently corrected for using the pairwise kinship matrix (Figure 3.S2).

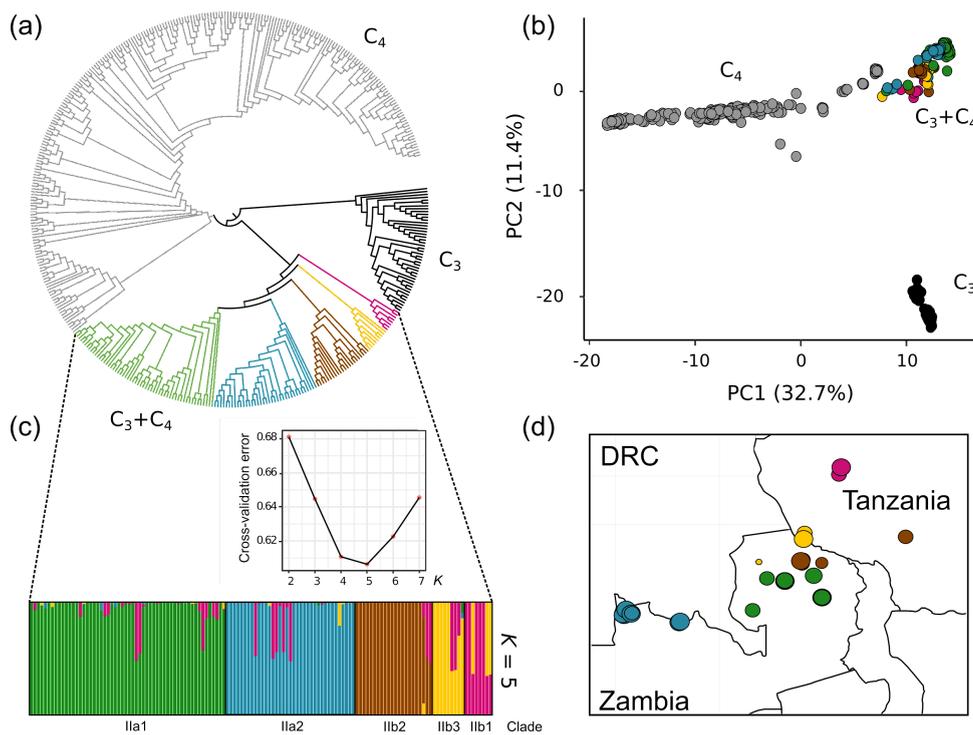


Figure 3.2: Population genomics of the *Alloteropsis semialata* accessions.

(a) A cladogram of the maximum likelihood phylogenetic tree, with individual clades recovered within the C_3+C_4 lineage coloured (same colours used in all panels). (b) A principal component analysis of the genotypes, showing the first two axes. (c) Admixture results for the C_3+C_4 *A. semialata* accessions for $K = 5$, the optimal number of population clusters based on the cross-validation error. (d) location of the C_3+C_4 populations used in this study, with the size of the point proportional to the number of samples (range 1- 20 samples per population).

3.4.2. Identifying regions of the genome correlated with the strength of the C₄ cycle

We used the $\delta^{13}\text{C}$ values as a proxy for the strength of the C₄ cycle for all 420 *A. semialata* samples used in this study with 8,082 SNPs analysed across the genome (Table 3.S3). As expected, the $\delta^{13}\text{C}$ values supported the demarcation of the main nuclear clades into the C₃, C₃+C₄ and C₄ phenotypes (Figure 3.3a). For C₃ and C₄ accessions, we found $\delta^{13}\text{C}$ average values of -26.67 and -12.63 with little dispersion within each group, whereas for C₃+C₄ accessions, we found substantial variation ranging from -28.35 to -18.47 with an average of -23.87. The heritability estimate, which represents the proportion of phenotypic variation due to genetic variation in the population, was high for $\delta^{13}\text{C}$ when considering all photosynthetic types ($h^2 = 0.75$; SE = 0.06; n = 420), and three-fold lower when just considering the C₃+C₄ intermediates ($h^2 = 0.25$; SE = 0.00; n = 132).

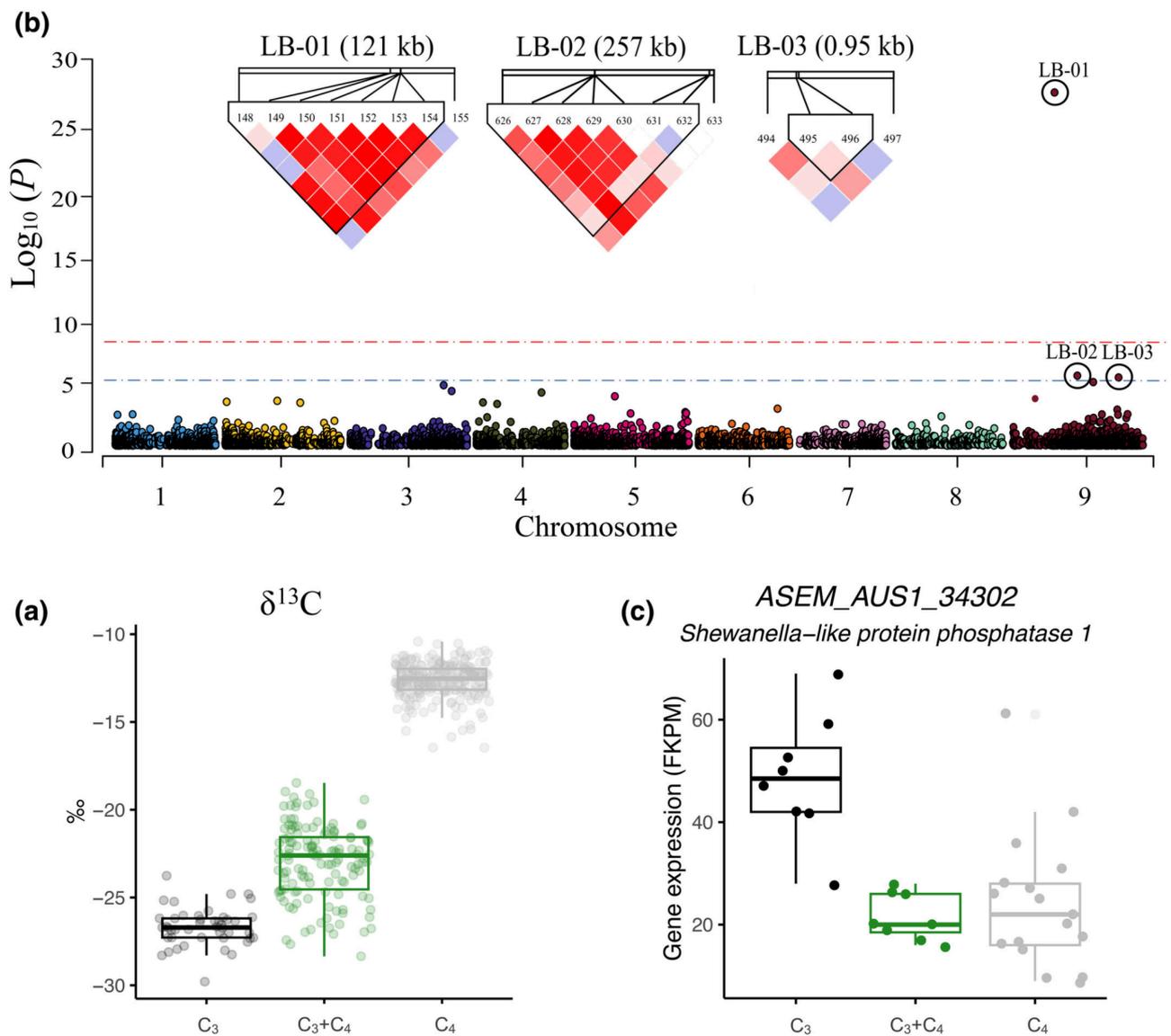


Figure 3.3.

Genetic variation associated with the strength of the C_4 cycle in *Alloteropsis semialata*. (a) The stable carbon isotope ratio ($\delta^{13}C$) was used to infer the strength of the C_4 cycle, with values measured for each of the photosynthetic types shown. The boxes show the median value and the interquartile range, and the whiskers represent $1.5\times$ the interquartile range. (b) Manhattan plot showing the results of a genome-wide association study (GWAS) for $\delta^{13}C$ using all samples. The blue and red dotted lines indicate Bonferroni corrected P -values of 0.05 and 0.001, respectively. Significant single-nucleotide polymorphisms (SNPs) are labelled with a block ID. Heatmap of pairwise linkage disequilibrium (LD) between markers surrounding each significant SNP, ranging from white indicating low LD ($LOD < 2$ and $D' < 1$) to bright red indicating strong LD ($LOD \geq 2$ and $D' = 1$). (c) Boxplot of gene expression for the candidate gene that is significantly differentially expressed. The boxes show the median value and the interquartile range, and the whiskers represent $1.5\times$ the interquartile range.

We conducted a combined GWAS using all individuals (Figure 3.3b), as well as various partitions by photosynthetic type (Figure 3.S3). When considering all individuals, the GWAS identified three significant SNPs on chromosome 9, which all corresponded to relatively narrow regions based on the LD (Figure 3.3b). The region with the highest association with $\delta^{13}\text{C}$ (LB-01) is a 121-kb region at 32.2 Mb (Tables 1, S3; Figure 3.3b). The same region was also significant when repeating the GWAS within the C₃+C₄, and when combining the C₃+C₄ with either the C₃ or C₄ individuals (Table S3; Figure 3.S2), but not when excluding the C₃+C₄ individuals. These results imply that the underlying causative gene segregates only within the C₃+C₄ group. There were six predicted protein-coding genes in the LB-01 region, and all were expressed in the leaf tissue of at least one *A. semialata* individual (Table S4). One of these genes (*Shewanella-like protein phosphatase 1*, *SLP1* (*ASEM_AUS1_34305*)) was significantly more highly expressed in the C₃ than in the other photosynthetic types (C₃ vs C₄ Bonferroni-adjusted *P*-value = 0.073; C₃ vs C₃+C₄ Bonferroni-adjusted *P*-value = 0.015; Figure 3.3c; Table 3.S4), although there is no consistent differential expression between photosynthetic types when individual populations are compared separately (Dunning et al., 2019a). None of the six genes were found to be strictly evolving under positive selection with a dN/dS ratio (ω) > 1 (Table S4), although those with the highest values may be seeing a relaxation of purifying selection (e.g. ω = 0.83 for *ASEM_AUS1_34303*). The annotated genes in the LB-01 region have a variety of functions (Table S4), including loci associated with the regulation of the Calvin cycle (*SLP1* (*ASEM_AUS1_34302*) (Kutuzov and Andreeva, 2012; Johnson et al., 2020)) and the activation of NADP-malic enzyme 2 (NADP-ME2), a C₄ decarboxylation enzyme (*RPM1-Induced Protein Kinase*, *RIPK* (*ASEM_AUS1_34305*) (Wu et al., 2022)).

Table 3.1: Significantly correlated regions of the genome identified in the genome wide association studies.

* This SNP was not located in a linkage block in our analyses, we therefore defined the region using the genome wide median block size

** This SNP was not located in a linkage block in our analyses, we therefore defined the region using the genome wide median block size that was truncated if there was a closely located unlinked SNP up or downstream.

Phenotype	Chromosome	SNP position	LD block (kb)	$-\log_{10} P$	Bonferroni adjusted P -value	Number of genes
$\delta^{13}C$	9	32191256	121	29.18	5.33E-26	6
	9	49592498	0.95**	5.73	1.51E-02	1
	9	81051792	63	5.58	2.12E-02	1
IBSF	9	58663539	362	6.88	3.58E-04	33
	2	634463	6**	4.83	4.95E-02	2
	5	23291361	59*	5.89	3.53E-03	4
	4	63231217	13	6.71	5.34E-04	2
	8	23357684	275	16.47	9.25E-14	24
BSD	9	10612628	11	5.1	2.17E-02	1
	9	48749591	0.14**	5.59	7.02E-03	0

The two other regions identified in the $\delta^{13}\text{C}$ GWAS using all individuals (LB-02 and LB-03; Figure 3.3b) were not significant when partitioning the data by photosynthetic type (Table S3). Both these regions are delimited by LD blocks narrow in size and that contain one annotated gene each. The candidate gene in LB-02 (*ASEM_AUSI_29467*) was not expressed at all in any *A. semialata* mature leaves, while the one in LB-03 (*ASEM_AUSI_14480*) was expressed in all individuals, but was not differentially expressed between photosynthetic types. In addition, both genes do not seem to have been under positive selection (Table S4). One of these genes (*ASEM_AUSI_29467*) encodes a SCARECROW-LIKE protein 9 (SCL9) protein belonging to the GRAS gene family, a group of transcription factors shown to play a key role in C₄ leaf anatomy and photosynthetic development in maize (Slewinski et al., 2012; Hughes and Langdale, 2020). The lack of expression (or differential expression) of these candidate genes in transcriptomes generated from mature leaf tissues is likely explained by the involvement of these genes, such as *SCL9*, in leaf development. The other gene encodes a protein associated with the suppression of nonphotochemical quenching and maintaining the efficiency of light harvesting (*suppressor of quenching 1, SOQ1* (*ASEM_AUSI_14480*) (Brooks et al., 2013; Duan et al., 2023)).

The $\delta^{13}\text{C}$ GWAS analyses were repeated with a subset of photosynthetic types (Figure 3.S3; Table 3.S4), and these also identified potentially interesting candidate genes, particularly those related to leaf vein patterning. WIP C2H2 zinc finger protein (*WIP2* (*ASEM_AUSI_03361*); LB-20; C₄ and C₃+C₄ individuals) is paralogous to the WIP6 transcription factor TOO MANY LATERALS that specifies vein rank in maize and rice (Vlad et al., 2024). Defectively Organized Tributaries 4 (*DOT4* (*ASEM_AUSI_05127*); LB-23; C₃+C₄ individuals) is orthologous to a vein patterning gene in *Arabidopsis thaliana* (Petricka et al., 2008).

3.4.3. Identifying regions of the genome associated with C₄ leaf anatomy in the C₃+C₄ intermediates

We studied the genetic basis of three leaf anatomical traits previously associated with the strength of the C₄ cycle ($\delta^{13}\text{C}$) using the 132 C₃+C₄ individuals (Figure 3.1; Alenazi et al., 2023) with 2,754 SNPs analysed across the genome (Table 3.S3). The heritability estimates for the three leaf anatomical traits in the C₃+C₄ intermediates ranged from roughly equivalent to the value for $\delta^{13}\text{C}$ value to much lower (IBSF $h^2 = 0.22$ (SE = 0.04); BSD $h^2 = 0.12$ (SE = 0.06); IBSW $h^2 = 0.06$ (SE = 0.06); $n = 132$). No significantly correlated genomic region was detected for IBSW (Figure 3.S4), the trait with the lowest heritability. However, we did detect SNPs significantly associated with BSD and IBSF.

i. Bundle sheath distance

The distance between consecutive bundle sheaths plays a significant role in determining the rate and efficiency of photosynthesis in plants, with smaller distances being significantly correlated with higher $\delta^{13}\text{C}$ (more C₄-like) values (Alenazi et al., 2023). The C₃+C₄ intermediate individuals showed a range of BSDs from 55.14 to 178.36 μm , with variation between subclades (Figure 3.4a). The GWAS identified two significant regions associated with BSD, both on chromosome 9 (Tables 1, S3; Figure 3.4). Only one annotated gene was identified in the correlated genomic regions associated with BSD, the function of which is associated with leaf development (*Glucan Synthase-Like 8*, *GSL8* (*ASEM_AUSI_16831*); Table 3.S4 (Linh and Scarpella, 2022)).

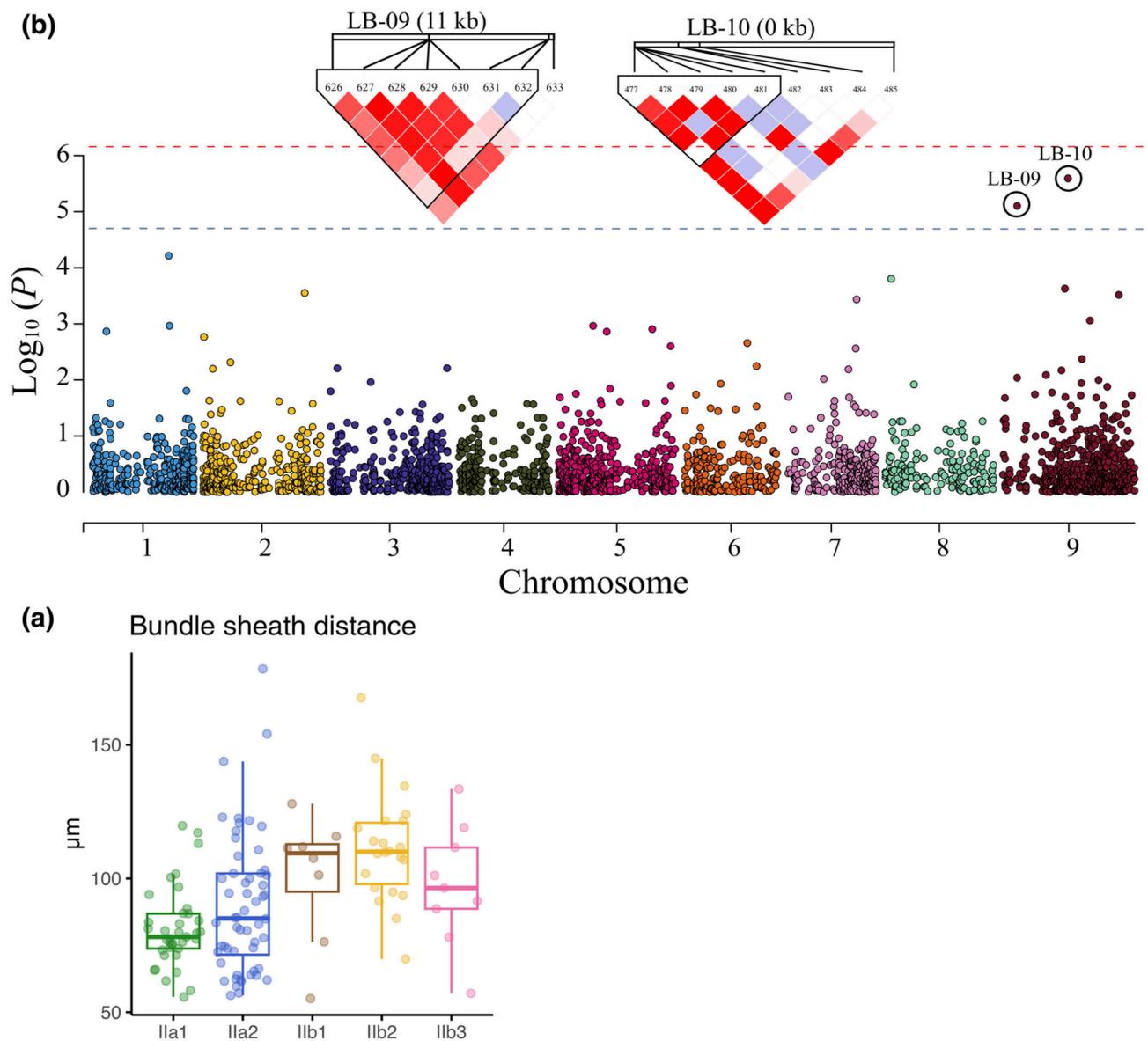


Figure 3.4.

Genetic variation associated with bundle sheath distance (BSD) in the C_3+C_4 *Alloteropsis semialata*. a) The boxplot shows the BSD variation for each of the C_3+C_4 subclades. The box indicates the median value and the interquartile range, and the whiskers represent $1.5\times$ the interquartile range. b) A Manhattan plot showing the results of a Genome-Wide Association Study (GWAS) for BSD. The blue and red dotted lines indicate Bonferroni corrected P-values of 0.05 and 0.001, respectively. Significant SNPs are labelled with a block ID. Heatmap of pairwise linkage disequilibrium (LD) between markers surrounding each significant SNP, ranging from white indicating low LD ($\text{LOD} < 2$ and $D' < 1$) to bright red indicating strong LD ($\text{LOD} \geq 2$ and $D' = 1$).

ii. Inner bundle sheath fraction (IBSF)

Inner bundle sheath fraction represents the portion of the leaf that can be used for C_4 photosynthesis (Figure 3.1). A higher IBSF in the C_3+C_4 *A. semialata* has been significantly correlated with a

higher $\delta^{13}\text{C}$ (more C₄ like; Alenazi et al., 2023). In the C₃+C₄ populations, there is a range from 0.05 to 0.29, with variation between subclades (Figure 3.5a). We identified five regions of the genome correlated with IBSF, each on a different chromosome (Tables 1, S3; Figure 3.5). Expression was detected in mature leaves for 62% of the 65 genes located in the five regions, with no consistent differential expression between photosynthetic types in mature leaves (Dunning et al., 2019a), although two are on average more highly expressed in the C₃ vs C₄ accessions (*ASEM_AUSI_21119* Bonferroni-adjusted *P*-value = 0.073; *ASEM_AUSI_17094* Bonferroni-adjusted *P*-value = <0.001). Five out of the 65 genes were also evolving under strong positive selection with a dN/dS ratio (ω) > 1 using the one-ratio model (Table S4). The annotated genes in the correlated regions of the genome have a variety of functions (Tables 1, S4), including loci directly connected to the response to light stress (*Ferulate 5-Hydroxylase 1*, *FAH1* (*ASEM_AUSI_36251*) (Maruta et al., 2014)) and leaf development (*GATA transcription factor 19*, *GAT19* (*ASEM_AUSI_21136*), *CCR4-NOT transcription complex subunit 11*, *CNOT11* (*ASEM_AUSI_25789*) and *GRF1-interacting factor 1*, *GIF1* (*ASEM_AUSI_21151*)) (Sarowar et al., 2007; Zhang et al., 2018; An et al., 2020).

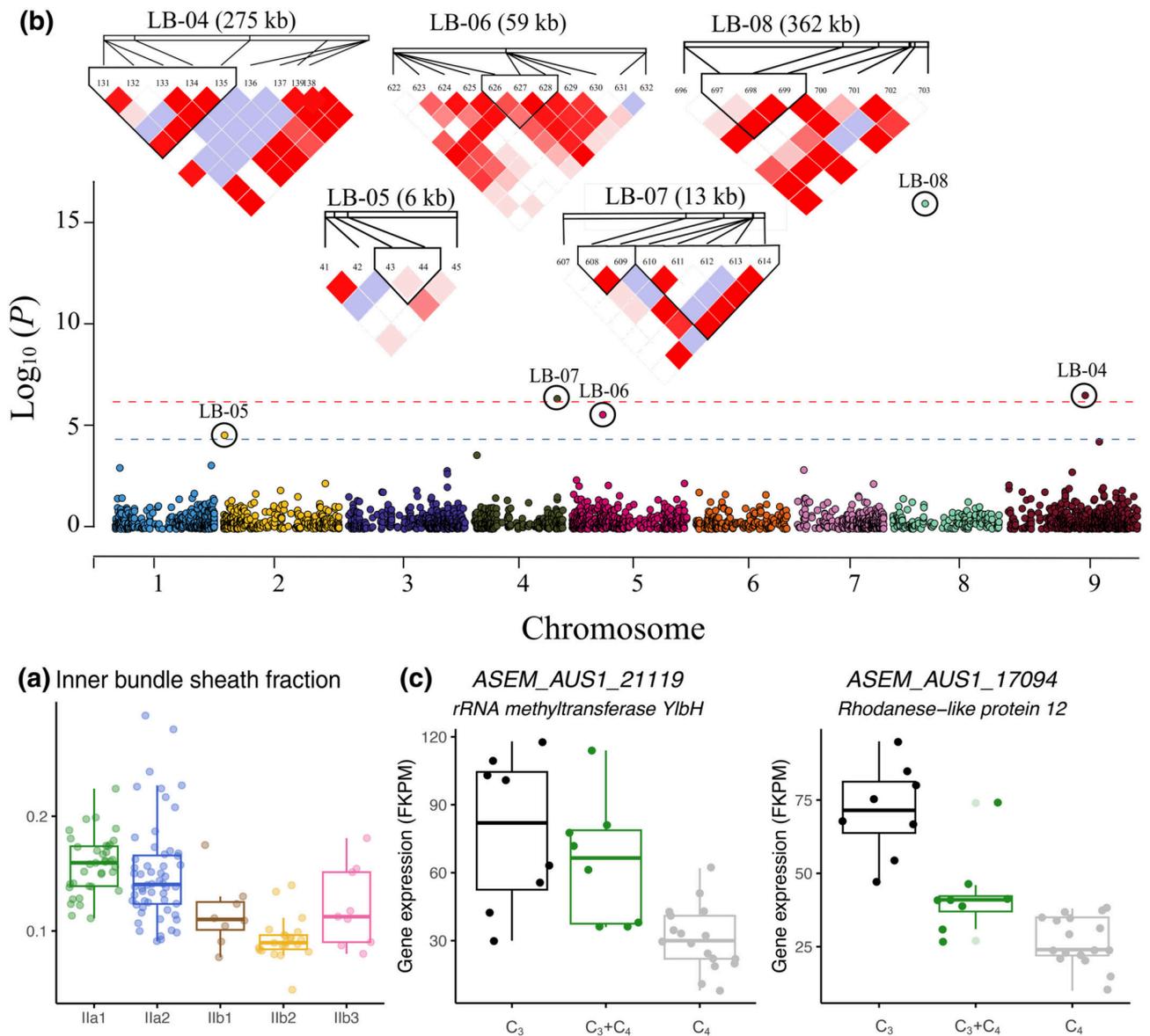


Figure 3.5.

Genetic variation associated with inner bundle sheath fraction (IBSF) in the C_3+C_4 *Alloteropsis semialata*. a) The boxplot shows the IBSF variation for each of the C_3+C_4 subclades. The box indicates the median value and the interquartile range, and the whiskers represent $1.5\times$ the interquartile range. b) a Manhattan plot showing the results of a Genome-Wide Association Study (GWAS) for IBSF. The blue and red dotted lines indicate Bonferroni corrected P-values of 0.05 and 0.001, respectively. Significant SNPs are labelled with a block ID. Heatmap of pairwise linkage disequilibrium (LD) between markers surrounding each significant SNP, ranging from white indicating low LD ($\text{LOD} < 2$ and $D' < 1$) to bright red indicating strong LD ($\text{LOD} \geq 2$ and $D' = 1$). c) Boxplot of gene expression for the candidate genes that are significantly differentially expressed. The boxes show the median value and the interquartile range, and the whiskers represent $1.5\times$ the interquartile range.

3.5. Discussion

Alloteropsis semialata has C₃, C₃+C₄, and C₄ genotypes that recently diverged, and it is therefore a useful model to study the initial steps leading to the establishment of the C₄ phenotype since these modifications are not conflated with other changes that accumulate over time (Pereira et al., 2023), and its emergence in this species provided an immediate demographic advantage (Sotelo et al., 2024). Here, we estimate the heritability and identify regions of the genome correlated with variation in both the stable carbon isotope ratio ($\delta^{13}\text{C}$) and leaf anatomical traits known to influence $\delta^{13}\text{C}$ from field-based measurements (Alenazi et al., 2023). Despite a relatively modest sample size ($n = 420$ for $\delta^{13}\text{C}$; $n = 132$ for leaf anatomy), we identified regions of the genome significantly associated with these traits, which may indicate that the genetic architecture of C₄ evolution in *A. semialata* is relatively simple, although a broader study may identify additional loci. At present, functional validation of the candidate loci is not possible in *Alloteropsis*, as no proven stable transformation system has been established (Pereira et al., 2023). This is something that would greatly advance the utility of *Alloteropsis semialata* as a model system to study C₄ evolution in the future.

3.5.1. Genetic basis of the carbon isotope ratio ($\delta^{13}\text{C}$) in *Alloteropsis semialata*

Using linked phenotype and genotype information for 420 *A. semialata* individuals, we identified three associated regions of the genome, containing eight protein-coding genes (Figure 3.3). The underlying differences in the $\delta^{13}\text{C}$ between photosynthetic types are driven by C₄ plants evolving to fix carbon with the PEPC enzyme rather than Rubisco. However, genes encoding PEPC were not detected in the associated regions identified in our GWAS. This absence could be due to variation in the specific PEPC gene copy used for C₄ in the different individuals masking the signal, with up to five different versions known to be used by different *A. semialata* populations (Dunning et al.,

2017). Among these five copies, three were laterally acquired (Christin et al., 2012), complicating the matter further as they appear as large structural variants inserted randomly into the genome (Dunning et al., 2019b) and are only present in a subset of individuals (Raimondeau et al., 2023). However, based on the annotations of the genes in the associated regions, we did identify candidate genes with functions potentially associated with the $\delta^{13}\text{C}$, the most promising of which include those co-expressed with Rubisco (*SLP1* (*ASEM_AUS1_34302*)), the activation of the NADP-ME C₄ decarboxylating enzyme (*RIPK* (*ASEM_AUS1_34305*)), the development of C₄ ‘Kranz’ anatomy (*SCL9* (*ASEM_AUS1_29467*)), and the suppression of nonphotochemical quenching (*SOQ1* (*ASEM_AUS1_14480*)).

SLP1 encodes a Shewanella-like protein phosphatase 1, an ancient chloroplast phosphatase that is generally more highly expressed in photosynthetic tissue (Kutuzov and Andreeva, 2012; Johnson et al., 2020). In *Arabidopsis thaliana*, it is co-expressed with a number of photosynthetic genes (including all of the Calvin cycle enzymes and Rubisco activase) and it is predicted to play a role in the light-dependent regulation of chloroplast function (Kutuzov and Andreeva, 2012). In *A. semialata*, *SLP1* is significantly more highly expressed in the C₃ individuals compared with the other photosynthetic types. This greater expression in C₃ individuals could indicate a higher Calvin cycle activity at the whole leaf level; meanwhile, in the C₃+C₄ and C₄ individuals, its expression would be increasingly restricted to the IBS tissue. Subdivision of the light signaling networks is one of the key steps in the partitioning of photosynthesis across tissue types in C₄ species (Hendron and Kelly, 2020), and *SLP1* is potentially one of the regulators of this key innovation in *A. semialata*. *RIPK* is an enzyme that plays a role in disease resistance and plant immunity (Liu et al., 2011), but has pleiotropic effects. In *A. thaliana*, *RIPK* directly phosphorylates NADP-ME2 (*AT5G11670*) to enhance its activity and increase cytosolic NADPH concentrations (Wu et al., 2022). In C₄ species,

CO₂ is initially fixed in the mesophyll by CA and PEPC before being transported to an internal leaf compartment and released for Rubisco to assimilate through the Calvin cycle. Preliminary studies in *A. semialata* concluded that NADP-ME was the predominant decarboxylating enzyme, although its activity varied with temperature (Freen et al., 1983). Subsequent transcriptome work showed that NADP-ME expression (specifically the *nadpme-1P4* gene that is a many-to-many ortholog of *AT5G11670*) has a mean expression level four times higher in C₄ and C₃+C₄ individuals (mean = 300 RPKM; SD = 235) than in C₃ plants (mean = 75 RPKM; SD = 32), although this difference is not always consistent between populations (Dunning et al., 2019a). The other decarboxylating enzyme commonly used by C₄ *Alloteropsis* is phosphoenolpyruvate carboxykinase (PCK), but like PEPC, a C₄ copy of PCK was also laterally acquired (Christin et al., 2012), complicating its identification in a GWAS analysis because it is absent in the C₃ (Dunning et al., 2019b).

SCL9 belongs to the GRAS gene family of transcription factors that regulate plant development (Hirsch and Oldroyd, 2009). This multigene family includes two known C₄ Kranz anatomy regulators identified in maize, *SHORTROOT* (Slewinski et al., 2014) and *SCARECROW* (Slewinski et al., 2012). Orthologous *SCARECROW* (*SCR*) genes have divergent functions, being recruited for distinct roles in leaf development within maize, rice, and *A. thaliana* (Hughes and Langdale, 2022). In addition to its influence on leaf anatomy, *SCR* is also required for maintaining photosynthetic capacity in maize (Hughes and Langdale, 2020). The correlation of the *SCARECROW-LIKE SCL9* gene with the strength of the C₄ cycle in *A. semialata* may indicate that convergence in C₄ phenotypes are a result of the parallel recruitment of GRAS transcription factors between species, although there is divergence in the specific loci recruited for this purpose.

SOQ1 is a chloroplast-localized thylakoid membrane protein that regulates nonphotochemical

quenching in *A. thaliana* (Brooks et al., 2013; Duan et al., 2023). In full sunlight, plants absorb more light energy than they can process, which can ultimately result in the generation of free radicals that damage the photosynthetic apparatus (Müller et al., 2001). To overcome this, plants have evolved nonphotochemical quenching, which enables them to dissipate the excess energy as heat. This problem is potentially exacerbated in C₄ species, which typically grow in high-light conditions compared with their C₃ counterparts (Sage and Monson, 1999). Preliminary evidence indicates that C₄ species exhibit a significantly faster and greater nonphotochemical quenching relaxation than their C₃ relatives, including between photosynthetic types in *A. semialata* (Acre Cubas, 2023). *SOQ1* may therefore play a direct role in regulating differences in the nonphotochemical quenching responses among *A. semialata* photosynthetic types, and it may represent a good candidate gene to target for reduced photoinhibition associated with fluctuating light conditions in crops (Long et al., 1994).

3.5.2. The genetic basis of C₄ leaf anatomy

In *A. semialata*, the IBS is the site of C₄ photosynthesis, and three leaf anatomical variables linked to the proliferation of this tissue explain the strength of the C₄ cycle ($\delta^{13}\text{C}$) in the C₃+C₄ intermediate individuals: IBSW, BSD, and IBSF (Alenazi et al., 2023). IBSW has the lowest heritability ($h^2 = 0.06$ (SE = 0.06)), and we failed to identify any significant SNPs correlated with this phenotype in our GWAS. This absence of significant genetic factors contributing to the trait may indicate that IBSW has a complex genetic architecture or high phenotypic plasticity. In *A. semialata*, $\delta^{13}\text{C}$ is largely genetically based as it is highly heritable after population structure has been accounted for ($h^2 = 0.75$ (SE = 0.06)), and field-based differences are preserved in a common environment (Lundgren et al., 2016). However, slight variation in $\delta^{13}\text{C}$ can still be caused by environmental effects on water use efficiency (Farquhar and Richards, 1984). The previously

observed correlation of field-based IBSW measurements with $\delta^{13}\text{C}$ may potentially arise from such environmental-induced plasticity (Alenazi et al., 2023). For example, bundle sheath cells in wheat have a larger diameter (more C₄-like) under drought conditions (David et al., 2017).

3.5.3. Plasmodesmata and reduced distance between bundle sheaths

We identified two regions of the genome associated with BSD that contain a single protein-coding gene. This gene is *GSL8* (*ASEM_AUS1_16831*), a member of the GSL family that encodes enzymes synthesizing callose. *GSL8* plays an important role in tissue-level organization (Chen et al., 2009), including stomatal (Guseman et al., 2010) and leaf vein patterning (Linh and Scarpella, 2022). Mutants of *GSL8* in *A. thaliana* formed networks of fewer veins in their leaves (Linh and Scarpella, 2022). This change in venation is mediated by the aperture of plasmodesmata, channels through cell walls that connect neighboring cells (Paterlini, 2020; Band, 2021), which is regulated by *GSL8* (Saatian et al., 2018; Linh and Scarpella, 2022). Normal vein patterning is reliant on an auxin hormone signal travelling through these plasmodesmata, and any interference of this signal disrupts leaf vein development (Linh and Scarpella, 2022). *GSL8* might play a role in strengthening the C₄ cycle in *A. semialata* by reducing the distance between bundle sheaths through modulation of the auxin signal. The transition to being fully C₄ in *A. semialata* is also correlated with the presence of minor veins, which reduces both the number of mesophyll cells and the distance between bundle sheaths in C₃+C₄ in comparison with C₃ populations (Lundgren et al., 2019). Therefore, *GSL8* may play a pleiotropic role in the strengthening of C₄ photosynthesis in *A. semialata* by increasing both the proportion of bundle sheath tissue in the leaf, and the connectivity between the two distinct cell types required to complete the cycle. The $\delta^{13}\text{C}$ GWAS using the C₄ and C₃+C₄ (which differ in the presence of minor veins; Lundgren et al., 2019) interestingly identified a paralog of a gene recently shown to specify vein rank in maize (Vlad et al., 2024), and potentially, *WIP2* has been co-opted for

a similar function in *A. semialata*.

3.5.4. The genetic basis of the inner bundle sheath fraction in *Alloteropsis semialata*

The inner bundle sheath fraction has the highest heritability of all the three leaf anatomy measures used ($h^2 = 0.22$ (SE = 0.04)). Since it is a composite trait, it is more likely to be influenced by multiple developmental processes. Our GWAS identified five regions of the genome significantly associated with IBSF, containing 65 predicted protein-coding genes. Interestingly, we found a number of genes associated with leaf development that could play a role in the development of C₄ leaf architecture. These include homologs of genes that alter leaf area and vascular development (*GATI9* (*ASEM_AUSI_21136*)) (An et al., 2020), leaf thickness (*CNOT11* (*ASEM_AUSI_25789*)) (Sarowar et al., 2007), and leaf width by regulating meristem determinacy (*GIF1* (*ASEM_AUSI_21151*)) (Zhang et al., 2018). *GIF1* (also called *ANGUSTIFOLIA3*) is perhaps the most interesting of these genes, since it is expressed in the mesophyll cells of leaf primordium and can influence the proliferation of other clonally independent leaf cells (e.g. epidermal cells (Kawade et al., 2013)). The numerous regulators of leaf development identified in the GWAS point to an interacting balance of growth regulators to increase the proportion of bundle sheath tissue within the leaf for C₄ photosynthesis.

There are other genes in these regions with a diverse set of functions, although it is unclear how they could modulate IBSF, including genes associated with light stress and lignin biosynthesis. *FAH1* encodes ferulic acid 5-hydroxylase (F5H) 1, a cytochrome P450 protein that, when disrupted, reduces anthocyanin accumulation under photooxidative stress (Maruta et al., 2014) and is more highly expressed in the C₃ (mean RPKM = 7.38; SD = 5.29) than other photosynthetic types (mean RPKM = 1.00; SD = 2.10; Table 3.S4). These loci could also play a role in C₄ photosynthesis,

although most likely they might just be in close physical linkage.

3.6. Conclusion

C₄ photosynthesis is a complex trait that requires the rewiring of metabolic gene networks and alterations to the internal leaf anatomy. We investigated the genetic basis of these key innovations in *Alloteropsis semialata*, which has recently diverged C₃, C₃+C₄ intermediate, and C₄ phenotypes. We performed a GWAS that identified regulators of C₄ decarboxylation enzymes (*RIPK*), nonphotochemical quenching (*SOQ1*), and several genes involved in tissue-level organization and leaf development (e.g. *SCL9*, *GSL8*, and *GIFI*). Interestingly, these tend to come from the same gene families as the previously identified C₄ leaf anatomy regulators in other species. This parallel recruitment appears to mirror the pattern observed in the core metabolic enzymes, with the paralog recruited for the C₄ function depending on its ancestral expression pattern and catalytic properties (Wang et al., 2009; Hibberd and Covshoff, 2010; Christin et al., 2013, 2015; Aubry et al., 2014; Emms et al., 2016; Moreno-Villena et al., 2018). Thus, the easiest path to C₄ leaf anatomy would be context-dependent, which likely has implications for engineering C₄ anatomy in C₃ species.

Acknowledgments

ASA is supported by a PhD scholarship from the Northern Border University in Saudi Arabia, LP is supported by a Natural Environment Research Council grant NE/V000012/1, PAC was funded by a Royal Society University Research Fellowship (grant URF\R\180022), and LTD is funded by a NERC fellowship (grant NE/T011025/1).

3.7. References

- Alenazi, A. S., Bianconi, M. E., Middlemiss, E., Milenkovic, V., Curran, E. V., Sotelo, G., Lundgren, M. R., Nyirenda, F., Pereira, L., Christin, P.-A., Dunning, L. T., Osborne, C. P. (2023). Leaf anatomy explains the strength of C₄ activity within the grass species *Alloteropsis semialata*. *Plant, Cell and Environment*, 8, 2310–2322.
- Alexander, D. H., Novembre, J., Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 9, 1655–1664.
- An, Y., Zhou, Y., Han, X., Shen, C., Wang, S., Liu, C., Yin, W., Xia, X. (2020). The GATA transcription factor GNC plays an important role in photosynthesis and growth in poplar. *Journal of Experimental Botany*, 71, 1969–1984.
- Acre Cubas, L. (2023). A comparative analysis of C₃ and C₄ photosynthesis under dynamic light conditions. PhD Thesis, Cambridge University, UK.
- Band, L. R. (2021). Auxin fluxes through plasmodesmata. *New Phytologist*, 231, 1686-1692.
- Barrett, J. C., Fry, B., Maller, J., Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 2, 263–265.
- Bellasio, C., Griffiths, H. (2014). Acclimation of C₄ metabolism to low light in mature maize leaves could limit energetic losses during progressive shading in a crop canopy. *Journal of Experimental botany*, 65, 3725-3736.
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, 53, 474–485.
- Bianconi, M. E., Dunning, L. T., Curran, E. V., Hidalgo, O., Powell, R. F., Mian, S., Leitch, I. J., Lundgren, M. R., Manzi, S., Vorontsova, M. S., Besnard, G., Osborne, C. P., Olofsson, J. K., and Christin, P. A. (2020). Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B. Biological Sciences*, 287, 20201960.
- Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 15, 2114–2120.
- Chen, X. Y., Liu, L., Lee, E., Han, X., Rim, Y., Chu, H., Kim, S. W., Sack, F., Kim, J. Y. (2009). The Arabidopsis callose synthase gene *GSL8* is required for cytokinesis and cell patterning. *Plant Physiology*, 150, 105–113.
- Chen, L., Ge, B., Casale, F. P., Vasquez, L., Kwan, T., Garrido-Martín, D., Ecker, S. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell*, 5, 1398–1414.e24.

- Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., Hartwell, J and Osborne, C. P. (2012). Adaptive evolution of C₄ photosynthesis through recurrent lateral gene transfer. *Current Biology*, 22, 445-449.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 15, 2156–2158.
- 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.
- Duan, S., Dong, B., Chen, Z., Hong, L., Zhang, P., Yang, Z., Wang, H-B., Jin, H-L. (2023). HHL1 and SOQ1 synergistically regulate nonphotochemical quenching in *Arabidopsis*. *Journal of Biological Chemistry*, 299, 104670.
- Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P., and Christin, P.-A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C₄ photosynthesis among close relatives. *Evolution*, 71, 1541– 1555.
- Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J. K., Mapaura, A., Grundy, I. M., Kayombo, C. J., Dunning, L. A., Kentatchime, F., Ariyaratne, M., Yakandawala, D., Besnard, G., Quick, W. P., Bräutigam, A., Osborne, C. P., and Christin, P. A. (2019a). Key changes in gene expression identified for different stages of C₄ evolution in *Alloteropsis semialata*. *Journal of Experimental Botany*, 70, 3255– 3268.
- Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., Dionora, J., Quick, W. P., Park, M., Bennetzen, J. L., Besnard, G., Nosil, P., Osborne, C. P., and Christin, P.-A. (2019b). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences of the United States of America*, 10, 4416–4425.
- Dunning, L. T., Olofsson, J. K., Papadopulos, A. S. T., Hibdige, S. G. S., Hidalgo, O., Leitch, I. J., Baleeiro, P. C., Ntshangase, S., Barker, N., and Jobson, R. W. (2022). Hybridisation and chloroplast capture between distinct *Themeda triandra* lineages in Australia. *Molecular Ecology*, 31, 5846-5860.
- Edwards, G. E., Ku, M. S. (1987). Biochemistry of C₃–C₄ intermediates. In: M.D. Hatch and Boardman (Eds.) *The biochemistry of plants: a comprehensive treatise*, vol. 10. New York: Academic Press, pp. 275– 325.
- Emms, D. M., Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 1, 157–157.

- Farquhar, G., O'Leary, M., Berry, J. (1982). On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Australian Journal of Plant Physiology*, 2, 121–137.
- Farquhar, G. D., Richards, R. A. (1984). Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Functional Plant Biology*, 11, 539-552.
- Ferguson, J. N., Fernandes, S. B., Monier, B., Miller, N. D., Allen, D., Dmitrieva, A., Schmuker, P., Lozano, R., Valluru, R., Buckler, E. S., Gore, M. A., Brown, P. J., Spalding, E. P., and Leakey, A. D. B. (2021). Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiology*, 187, 1481–1500.
- Frean, M. L., Ariovich, D., Cresswell, C. F. (1983). C₃ and C₄ Photosynthetic and anatomical forms of *Alloteropsis semialata* (R. Br.) Hitchcock: 2. A comparative investigation of leaf ultrastructure and distribution of chlorenchyma in the two forms. *Annals of Botany*, 51, 811-821.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178–D1186.
- Guseman, J. M., Lee, J. S., Bogenschutz, N. L., Peterson, K. M., Virata, R. E., Xie, B., Kanaoka, M. M., Hong, Z., and Torii, K. U. (2010). Dysregulation of cell-to-cell connectivity and stomatal patterning by loss-of-function mutation in *Arabidopsis thaliana* (glucan synthase-like 8). *Development*, 137, 1731-1741.
- Hatch, M. D. (1987). C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 895, 81–106.
- Hatch, M. D. (1971). The C₄ pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C₄ dicarboxylic acid. *Biochemical Journal*, 125, 425– 432.
- Hendron, R. W., Kelly, S. (2020). Subdivision of light signaling networks contributes to partitioning of C₄ photosynthesis. *Plant Physiology*, 182, 1297-1309.
- Heyduk, K., Moreno-Villena, J. J., Gilman, I., Christin, P. A., Edwards, E. J. (2019). The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics*, 20, 485-493.
- Hirsch, S., Oldroyd, G. E. (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signaling and Behavior*, 4, 698-700.
- Hugin Development Team. (2015). Hugin - Panorama photo stitcher. <https://hugin.sourceforge.io>
- Hughes, T. E., Langdale, J. A. (2020). SCARECROW gene function is required for photosynthetic development in maize. *Plant Direct*, 4, e00264.

- Hughes, T. E., Langdale, J. A. (2022). SCARECROW is deployed in distinct contexts during rice and maize leaf development. *Development*, 149, dev200410.
- Johnson, J. J., White-Gloria, C., Toth, R., Labandera, A. M., Uhrig, R. G., Moorhead, G. B. (2020). SLP1 and SLP2: ancient chloroplast and mitochondrial protein phosphatases. *Protein Phosphatases and Stress Management in Plants: Functional Genomic Perspective*, 1-9.
- Kawade, K., Horiguchi, G., Usami, T., Hirai, M. Y., Tsukaya, H. (2013). ANGUSTIFOLIA3 signaling coordinates proliferation between clonally distinct cells in leaves. *Current Biology*, 23, 788-792.
- Kutuzov, M. A., Andreeva, A. V. (2012). Prediction of biological functions of Shewanella-like protein phosphatases (Shelphs) across different domains of life. *Functional and Integrative Genomics*, 12, 11–23.
- Langdale, J. A., Metzler, M. C., Nelson, T. (1987). The argentia mutation delays normal development of photosynthetic cell-types in *Zea mays*. *Developmental Biology*, 122, 243–255.
- Langdale, J. A., Rothermel, B. A., Nelson, T. (1988). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Genes and Development*, 1, 106–115.
- Linh, N. M., Scarpella, E. (2022). Leaf vein patterning is regulated by the aperture of plasmodesmata intercellular channels. *PLoS Biology*, 20, e3001781.
- Liu, J., Elmore, J. M., Lin, Z. J. D., Coaker, G. (2011). A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell host and microbe*, 9, 137-146.
- Long, S. P., Humphries, S., Falkowski, P. G. (1994). Photoinhibition of photosynthesis in nature. *Annual Review of Plant Biology*, 45, 633-662.
- Lundgren, M. R., Besnard, G., Ripley, B. S., Lehmann, C. E. R., Chatelet, D. S., Kynast, R. G., Namaganda, M., Vorontsova, M. S., Hall, R. C., Elia, J., Osborne, C. P., Christin, P., and Pannell, J. (2015). Photosynthetic innovation broadens the niche within a single species. *Ecology Letters*, 18, 1021–1029.
- Lundgren, M. R., Christin, P., Escobar, E. G., Ripley, B. S., Besnard, G., Long, C. M., Hattersley, P. W., Ellis, R. P., Leegood, R. C., and Osborne, C. P. (2016). Evolutionary implications of C₃–C₄ intermediates in the grass *Alloteropsis semialata*. *Plant, Cell and Environment*, 39, 1874–1885.
- Lundgren, M. R., Dunning, L. T., Olofsson, J. K., Moreno-Villena, J. J., Bouvier, J. W., Sage, T. L., Khoshravesh, R., Sultmanis, S., Stata, M., Ripley, B. S., Vorontsova, M. S., Besnard, G., Adams, C., Cuff, N., Mapaura, A., Bianconi, M. E., Long, C. M., Christin, P., Osborne, C. P., and Pannell, J. (2019). C₄ anatomy can evolve via a single developmental change. *Ecology Letters*, 22, 302– 312.

- Maruta, T., Noshi, M., Nakamura, M., Matsuda, S., Tamoi, M., Ishikawa, T., Shigeoka, S. (2014). Ferulic acid 5-hydroxylase 1 is essential for expression of anthocyanin biosynthesis-associated genes and anthocyanin accumulation under photooxidative stress in Arabidopsis. *Plant Science*, 219-220, 61–68.
- Monaco, M. K., Sen, T. Z., Dharmawardhana, P. D., Ren, L., Schaeffer, M., Naithani, S., Amarasinghe, V., Thomason, J., Harper, L., Gardiner, J., Cannon, E. K. S., Lawrence, C. J., Ware, D., and Jaiswal, P. (2013). Maize Metabolic Network Construction and Transcriptome Analysis. *The Plant Genome*, 1, 1–12.
- Müller, P., Li, X. P., Niyogi, K. K. (2001). Non-photochemical quenching. A response to excess light energy. *Plant Physiology*, 125, 1558-1566.
- Olofsson, J. K., Curran, E. V., Nyirenda, F., Bianconi, M. E., Dunning, L. T., Milenkovic, V., Sotelo, G., Hidalgo, O., Powell, R. F., Lundgren, M. R., Leitch, I. J., Nosil, P., Osborne, C. P., and Christin, P. (2021). Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology*, 9, 2116–2130.
- Olofsson, J. K., Bianconi, M., Besnard, G., Dunning, L. T., Lundgren, M. R., Holota, H., Vorontsova, M. S., Hidalgo, O., Leitch, I. J., Nosil, P., Osborne, C. P., and Christin, P.-A. (2016). Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology*, 25, 6107– 6123.
- Ortiz, E. M. (2019). vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis.
- Paterlini, A. (2020). Uncharted routes: exploring the relevance of auxin movement via plasmodesmata. *Biology Open*, 9, 11.
- Pereira, L., Bianconi, M. E., Osborne, C. P., Christin, P.-A., Dunning, L. T. (2023). *Alloteropsis semialata* as a study system for C₄ evolution in grasses. *Annals of Botany*, 132, 365-382.
- Pignon, C. P., Leakey, A. D., Long, S. P., Kromdijk, J. (2021). Drivers of natural variation in water-use efficiency under fluctuating light are promising targets for improvement in Sorghum. *Frontiers in Plant Science*, 12, 627432.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 3, 559–575.
- Raimondeau, P., Bianconi, M. E., Pereira, L., Parisod, C., Christin, P. A., Dunning, L. T. (2023). Lateral gene transfer generates accessory genes that accumulate at different rates within a grass lineage. *New Phytologist*, 240, 2072-2084.
- Sage, R. F., Monson, R. K. (1999). C₄ plant biology. San Diego: Academic Press.

- Saatian, B., Austin, R. S., Tian, G., Chen, C., Nguyen, V., Kohalmi, S. E., Geelen, D., and Cui, Y. (2018). Analysis of a novel mutant allele of *GSL8* reveals its key roles in cytokinesis and symplastic trafficking in *Arabidopsis*. *BMC Plant Biology*, 18, 1-17.
- Sarowar, S., Oh, H. W., Cho, H. S., Baek, K.-H., Seong, E. S., Joung, Y. H., Choi, G. J., Lee, S., and Choi, D. (2007). *Capsicum annuum* CCR4-associated factor CaCAF1 is necessary for plant development and defence response. *The Plant Journal*, 51, 792-802.
- Schneider, C. A., Rasband, W. S., Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9, 671– 675.
- Simpson, C. J. C., Reeves, G., Tripathi, A., Singh, P., Hibberd, J. M. (2021). Using breeding and quantitative genetics to understand the C_4 pathway. *Journal of Experimental Botany*, 73, 3072– 3084.
- Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C. C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., and Itoh, T. (2013). Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant and Cell Physiology*, 54, e6.
- Slewinski, T. L., Anderson, A. A., Zhang, C., Turgeon, R. (2012). Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant and Cell Physiology*, 53, 2030-2037.
- Slewinski, T. L., Anderson, A. A., Price, S., Withee, J. R., 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.
- Sotelo, G., Gamboa, S., Dunning, L. T., Christin, P.-A., Varela, S. (2024). C_4 photosynthesis provided an immediate demographic advantage to populations of the grass *Alloteropsis semialata*. *New Phytologist*, 242.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312– 1313.
- Stata, M., Sage, T. L., Sage, R. F. (2019). Mind the gap: the evolutionary engagement of the C_4 metabolic cycle in support of net carbon assimilation. *Current Opinion in Plant Biology*, 49, 27– 34.
- Strigens, A., Schipprack, W., Reif, J. C., Melchinger, A. E. (2013). Unlocking the Genetic Diversity of Maize Landraces with Doubled Haploids Opens New Avenues for Breeding. *PloS One*, 2, e57234–e57234.
- Von Caemmerer, S. (1992). Carbon isotope discrimination in C_3 – C_4 intermediates. *Plant, Cell and Environment*, 15, 1063–1072.

- Voznesenskaya, E. V., Koteyeva, N. K., Chuong, S. D. X., Ivanova, A. N., Barroca, J., Craven, L. A., and Edwards, G. E. (2007). Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). *Functional Plant Biology : FPB*, 34, 247–267.
- Wu, Z., Burns, J. K. (2003). A morphological study on the responses of mature citrus leaves to high light. *Journal of the American Society for Horticultural Science*, 128, 590–596.
- Wang Y, Bräutigam A, Weber AP, Zhu XG. (2014). Three distinct biochemical subtypes of C₄ photosynthesis? A modelling analysis. *Journal of experimental botany* 65: 3567-3578.
- Yang J, Lee SH, Goddard ME, Visscher PM. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88, 76-82.
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Liu X. 2021. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics and Bioinformatics*, 4, 619–628.
- Zhou JM, Zhang Y. (2020). Plant immunity: danger perception and signaling. *Cell*, 181, 978-989.
- Zhang, D., Sun, W., Singh, R., Zheng, Y., Cao, Z., Li, M., Lunde, C., Hake, S., and Zhang, Z. (2018). GRF-interacting factor1 regulates shoot architecture and meristem determinacy in maize. *The Plant Cell*, 30, 360-374.

Supplementary Figures

Identifying genomic regions associated with C₄ photosynthetic activity and leaf anatomy in *Alloteropsis semialata*

Ahmed S Alenazi, Lara Pereira, Pascal-Antoine Christin, Colin P Osborne, Luke T Dunning

Supplementary Figures: 3

Supplementary Figures

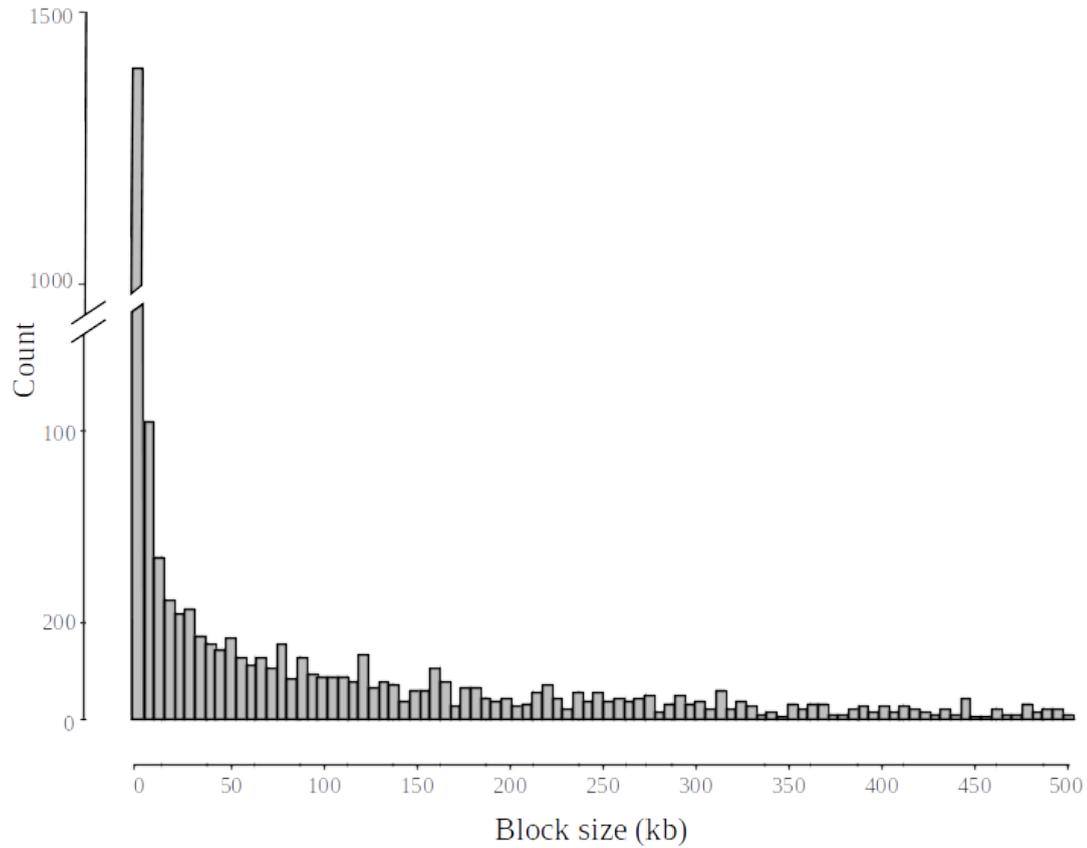


Figure 3.S1: Distribution of linkage block sizes in the *Alloteropsis semialata* genome. The block sizes were inferred using a solid spine of linkage disequilibrium (SSLD) analysis. 2,601 linkage blocks were identified, with a mean length of 58,996 kb (SD = 109,378 kb).

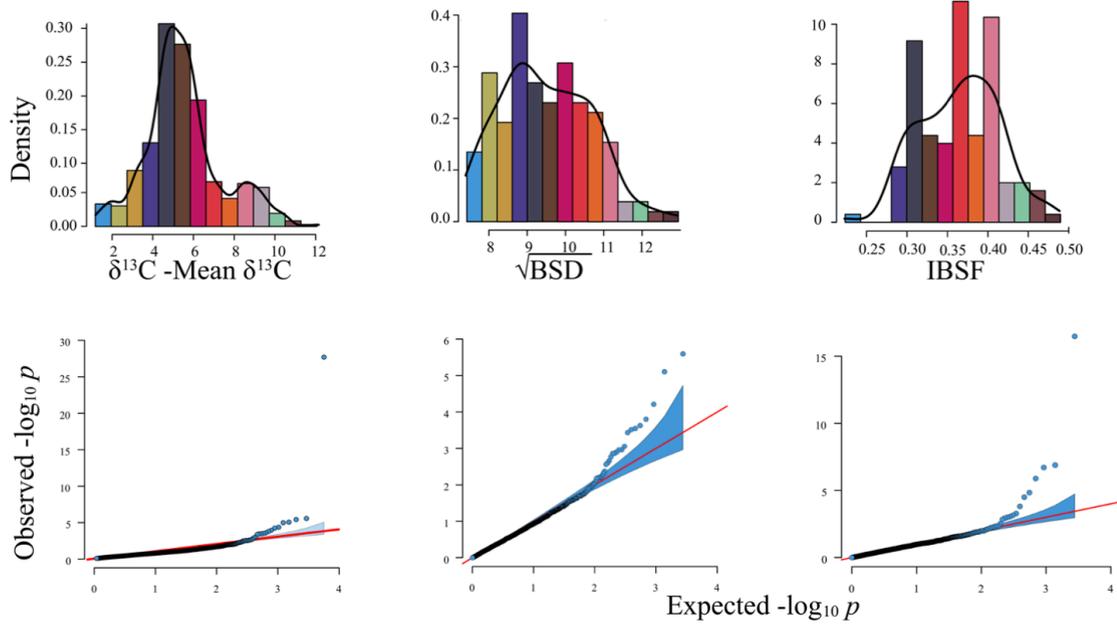
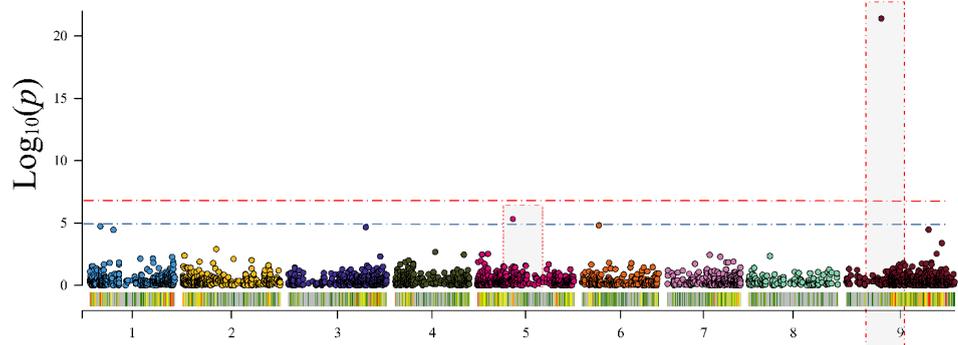
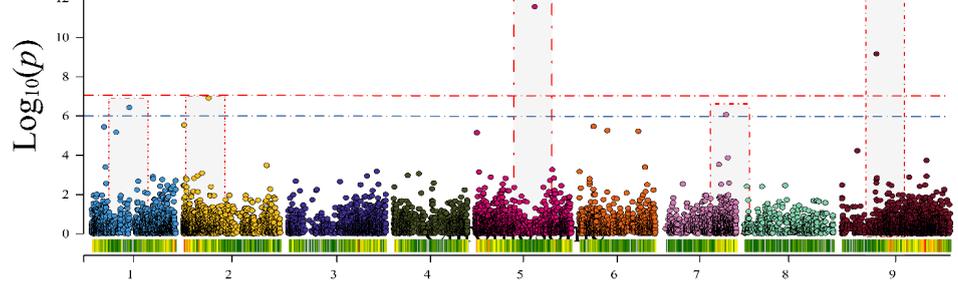


Figure 3.S2: Density and Q-Q plots for the studied traits. The blue shaded areas in the Q-Q plots represent the 95% confidence bands around the p-values under the null hypothesis for the GWAS results presented in the main text.

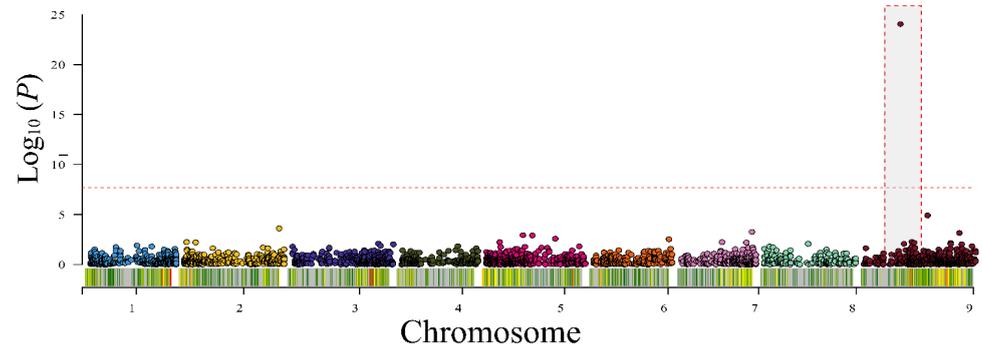
a) $C_3 + C_4$ vs C_3



b) $C_3 + C_4$ vs C_4



c) $C_3 + C_4$



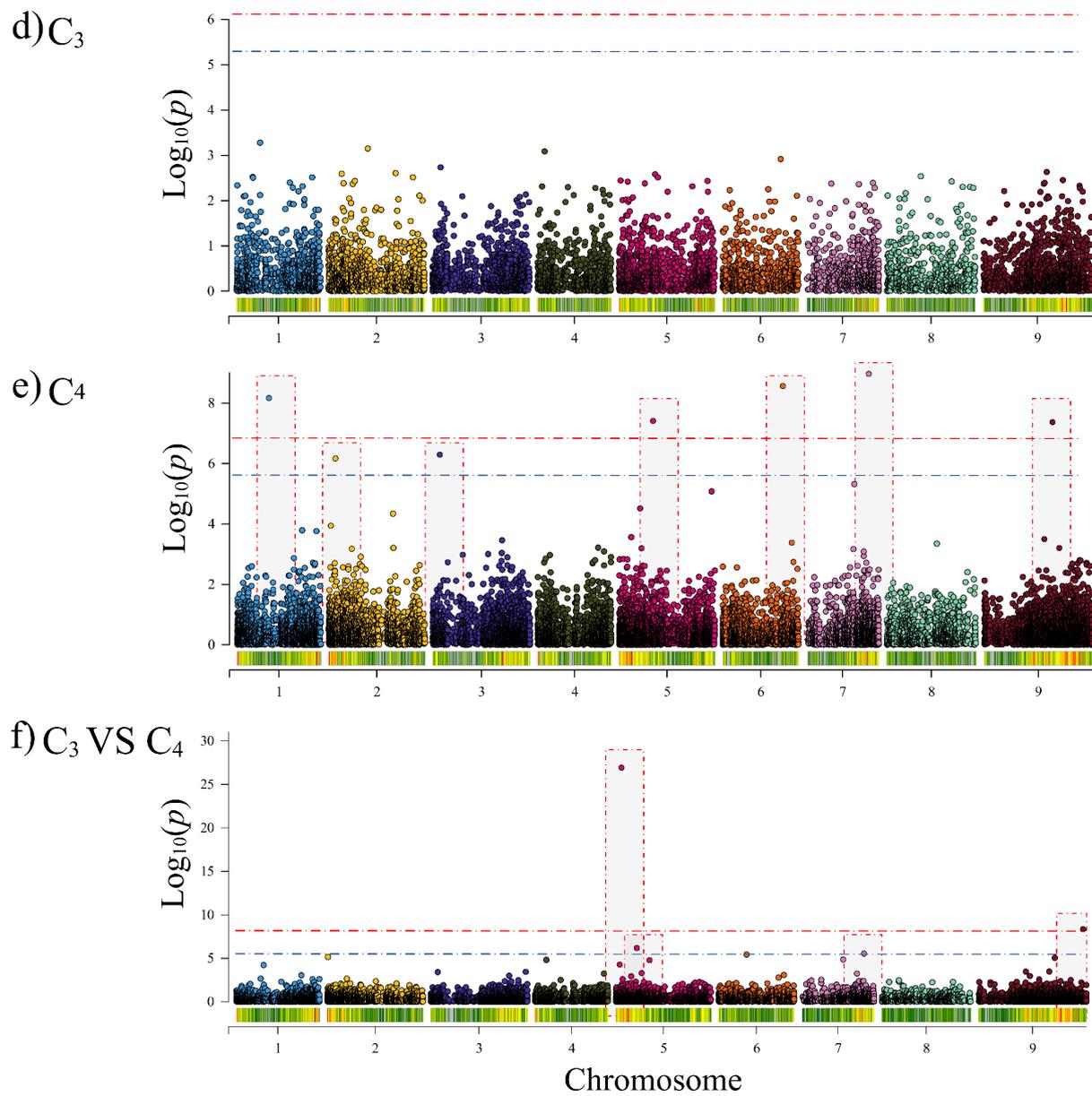


Figure 3.S3: Manhattan plot showing the results of a Genome-Wide Association Study (GWAS) for $\delta^{13}C$ using various subdivisions of samples based on photosynthetic type. The blue and red dotted lines indicate Bonferroni corrected P-values of 0.05 and 0.001, respectively. Significant SNPs are labelled with a block ID.

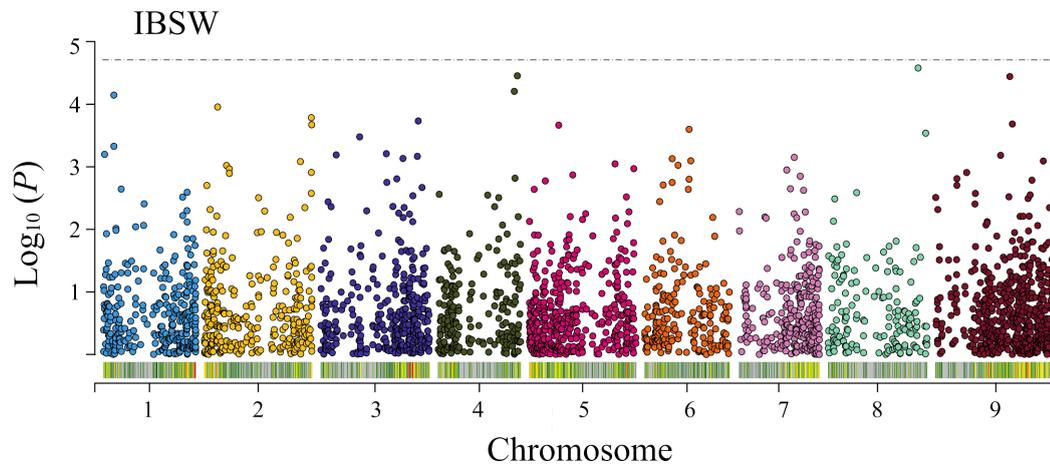


Figure 3.S4: Manhattan plot showing the results of a Genome-Wide Association Study (GWAS) for inner bundle sheath width (IBSW). The blue dotted line indicates a cutoff significance value of a Bonferroni corrected P-value of 0.05.

Supplementary Tables

All tables available as individual tabs in separate excel file

Table 3.S1: Details of samples used in the genome wide association study. This includes sample collection information, $\delta^{13}\text{C}$ and NCBI Sequence Read Archive accession numbers.

Table 3.S2: Leaf anatomy measurements for the C₃+C₄ *Alloteropsis semialata*.

Table 3.S3: Summary of significant regions detected. This includes the SNP location and linkage block size.

Table 3.S4: Summary of genes in significant regions. This includes gene annotation, expression levels and estimates of omega (dN/dS ratio).

Supplementary tables for this chapter are available at <https://doi.org/10.1111/nph.19933>

Supplementary datasets

Data set 1: Maximum likelihood phylogeny of all accessions used in this study.

https://drive.google.com/file/d/10FJ_fB51-nHo8vg0PCUG-ko2HvTR537c/view?usp=sharing

Data set 2: Orthogroup phylogenies

<https://drive.google.com/file/d/1A5LRHA6TJWU2bD2pDvRioRdHxwGn9Xqq/view?usp=sharing>

Chapter 4. Gene duplication and the repeated origin of C₄ photosynthesis in PACMAD grasses

Chapter 4. Gene duplication and the repeated origin of C₄ photosynthesis in PACMAD grasses

Lara Pereira^{1*}, Ahmed S Alenazi^{1,2*}, Sahr Mian³, Ilia J Leitch³, Pascal-Antoine Christin¹, Colin P Osborne⁴, Luke T Dunning^{1**}

Affiliations:

¹ Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

² Department of Biological Sciences, Northern Border University, Saudi Arabia

³ Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom

⁴ Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

*These authors contributed equally to the work

**Corresponding author: Luke T. Dunning; Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom; +44 (0) 1142220027; l.dunning@sheffield.ac.uk

Personal contribution: I conducted an OrthoFinder analysis to determine orthologous genes and duplication events across all genomic data, and Dr. Lara Pereira has generated two reference genomes. Following this, I performed blast searches to align genetic data and identify gene functions using various biological databases. Finally, I wrote this chapter with the assistance of my co-authors.

4.1. Abstract

C₄ photosynthesis is a complex trait that requires many biochemical and anatomical components to be coordinated. It has evolved convergently over 60 times, yet these origins are phylogenetically clustered. For example grasses can be divided into two main clades (BOP and PACMAD), and the greater than 20 C₄ origins are only found in the PACMAD clade. Here, we examined the genetic precursors that might have enabled the evolution of C₄ in this clade. Using the new Aristidoideae genomes we assembled representing the Aristidoideae sub-clade, we could identify genes that were duplicated specifically in the base of the PACMAD clade, which represents the last common ancestor of all the C₄ grass origins in this family. The genes duplicated in this ancestral clade were involved in the regulation of water use efficiency (WUE) as part of the response to environmental stress via cell wall modifications (e.g. Beta-1,2-xylosyltransferase XYXT1) and stomatal opening (e.g. Trehalose 6-phosphate phosphatase I (*AtTPPI*)). Anatomical modifications related to WUE regulation or CO₂ conductance may therefore have acted as precursors for the convergent evolution of Kranz anatomy. This chapter highlights the profound genetic and anatomical complexities underlying the convergent evolution of C₄ photosynthesis in grasses, with the evolutionary trajectory of a species dictated by its evolutionary history.

4.2. Introduction

Complex or polygenic traits are composed of multiple components working in synergy to increase an organism's fitness. When comparing different species, we can observe how these complex traits manifest into impressive phenotypic innovations and novel adaptations. However, the evolutionary dynamics behind their progressive emergence are often unknown, especially when their function relies on multiple components working in synchrony. The evolution of complex traits is typically envisaged as a stepwise progression of multiple intermediate stages, each offering an incremental fitness advantage that can be selected for through a process of descent with modification (Darwin, 1859). The order that the individual components are assembled is of great interest (Meléndez-Hevia et al., 1996; Lenski et al., 2003), especially when there is potentially a key innovation that is foundational to the emergence of a particular trait (e.g. Ourisson and Nakatani, 1994). The presence of such evolutionary precursors may be one explanation for the non-random phylogenetic clustering of certain convergent traits (Christin et al., 2015).

Gene duplication is a major driving force in evolution, providing new genetic material that can be recruited for specialised or novel gene function (Magadum et al., 2013). Gene duplication has contributed to the developmental programme of various organisms, including the HOX genes that are responsible for mammalian body form (Wagner et al., 2003), and developmental and regulatory loci that underpin the diversity of flowering plants (Angiosperms) (Moore and Purugganan, 2005). Duplicated genes can originate through whole genome duplication (polyploidization) or small-scale duplications, either through unequal crossing-over events or mediated by transposable elements. The fate of these duplicates can be gene loss / pseudogenization; subfunctionalization, where each copy randomly loses subfunctions of the original gene at either protein activity or gene expression levels; or neofunctionalization, where one duplicate retains the ancestral function while its paralog

gains a novel function (Panchy et al., 2016). Gene duplication and neofunctionalization are likely to be very important for the emergence of novel complex traits.

The C₄ photosynthetic pathway is a complex trait that increases the efficiency of carbon fixation in hot and dry environments when compared to the ancestral C₃ pathway (Sage et al., 2012; Atkinson et al., 2016). C₄ photosynthesis requires the coordinated modification of numerous anatomical and biochemical components to concentrate CO₂ around the carbon fixing enzyme Rubisco (Hatch, 1987). Gene duplication and subsequent subfunctionalization has been proposed as a prerequisite to C₄ evolution (Monson, 2003). Most core enzymes in the C₄ pathway are encoded by large multigene families, with different copies recruited depending on their ancestral expression patterns and catalytic properties (Wang et al., 2009; Hibberd and Covshoff, 2010; Aubry et al., 2011; Christin et al., 2013a, 2015; Emms et al., 2016; Moreno-Villena et al., 2018). Gene duplication may also play a more immediate role in the emergence of the C₄ cycle, with a whole genome duplication facilitating the evolution of C₄ anatomy and biochemistry in the eudicot *Gynandropsis gynandra* (Huang et al., 2021), and a single photorespiratory gene duplication underpinning the development of the photorespiratory CO₂ pump in C₃–C₄ intermediate *Flaveria* species (Schulze et al., 2013). To date, most research has focused on genes known to be directly involved in C₄ metabolism, with the genetic basis of leaf anatomical enablers largely unknown (Christin et al., 2013b). A global analysis using comparative genomics as a tool to identify gene families that have been duplicated and retained in two independent C₄ origins unveiled 41 orthologous genes with heterogeneous functions, including transcription factors, transporters, and genes involved in hormone metabolism, among others (Emms et al., 2016).

Despite its apparent complexity, C₄ photosynthesis has evolved over 60 times in land plants, representing a remarkable example of convergent evolution (Heyduk et al., 2019). The origins of C₄ photosynthesis are phylogenetically clustered (Christin et al., 2015), with over 20 C₄ origins in the grasses alone (Grass Phylogeny Working Group II, 2012). Even within the grass family (Poaceae), the multiple origins of C₄ photosynthesis are structured, only occurring in the **PACMAD** clade (six subfamilies: **Panicoideae**, **Aristidoideae**, **Chloridoideae**, **Micrairoideae**, **Arundinoideae**, and **Danthonioideae**) which is sister to the **BOP** clade (three subfamilies: **Bambusoideae**, **Oryzoideae**, and **Pooideae**) where C₄ species are conspicuously absent (Grass Phylogeny Working Group II, 2012).

The absence of C₄ photosynthesis in BOP grasses suggests that species from the PACMAD clade carry certain precursors that make them prone to evolve C₄. These preadaptations are likely related with leaf anatomy rather than with biochemical pathways (Christin et al., 2013b). The genetic basis controlling leaf development has only been partially elucidated (Lewis and Hake, 2016), preventing a comprehensive inspection of specific genes that might be responsible for these preadaptations. Global comparative genomics studies are therefore needed in order to elucidate the genetic factors that may contribute to C₄ evolvability. High-quality reference genomes have been published for more than a hundred grass species, offering a great opportunity to further investigate the role of gene duplication in C₄ evolution in a broad perspective (Figure 4.1). However, a reference genome for any species belonging to the earliest diverging PACMAD subfamily (Aristidoideae) was lacking, hampering an analysis of the specific duplications that occurred in the last common ancestor of this clade.

In this study, we aimed to identify genes that may contribute to preadaptations for C₄ photosynthesis present in the PACMAD clade. To achieve this aim, we generated reference genomes for two members of the Aristidoideae family, *Aristida adscensionis* and *Stipagrostis hirtigluma*. We then used these genomes together with publicly available genomes for other grass species to identify duplication events at the base of the PACMAD clade that were retained in all C₄ species. We investigated their predicted gene function and found that half of the duplicated and retained genes were involved in cell wall biogenesis. We hypothesise that cell wall structure may modify the permeability of bundle sheath and/or mesophyll cells and affect water and CO₂ diffusion. These changes may have facilitated the convergent evolution of C₄ photosynthesis in the PACMAD clade and be related to the key anatomical enablers (Christin et al., 2013b).

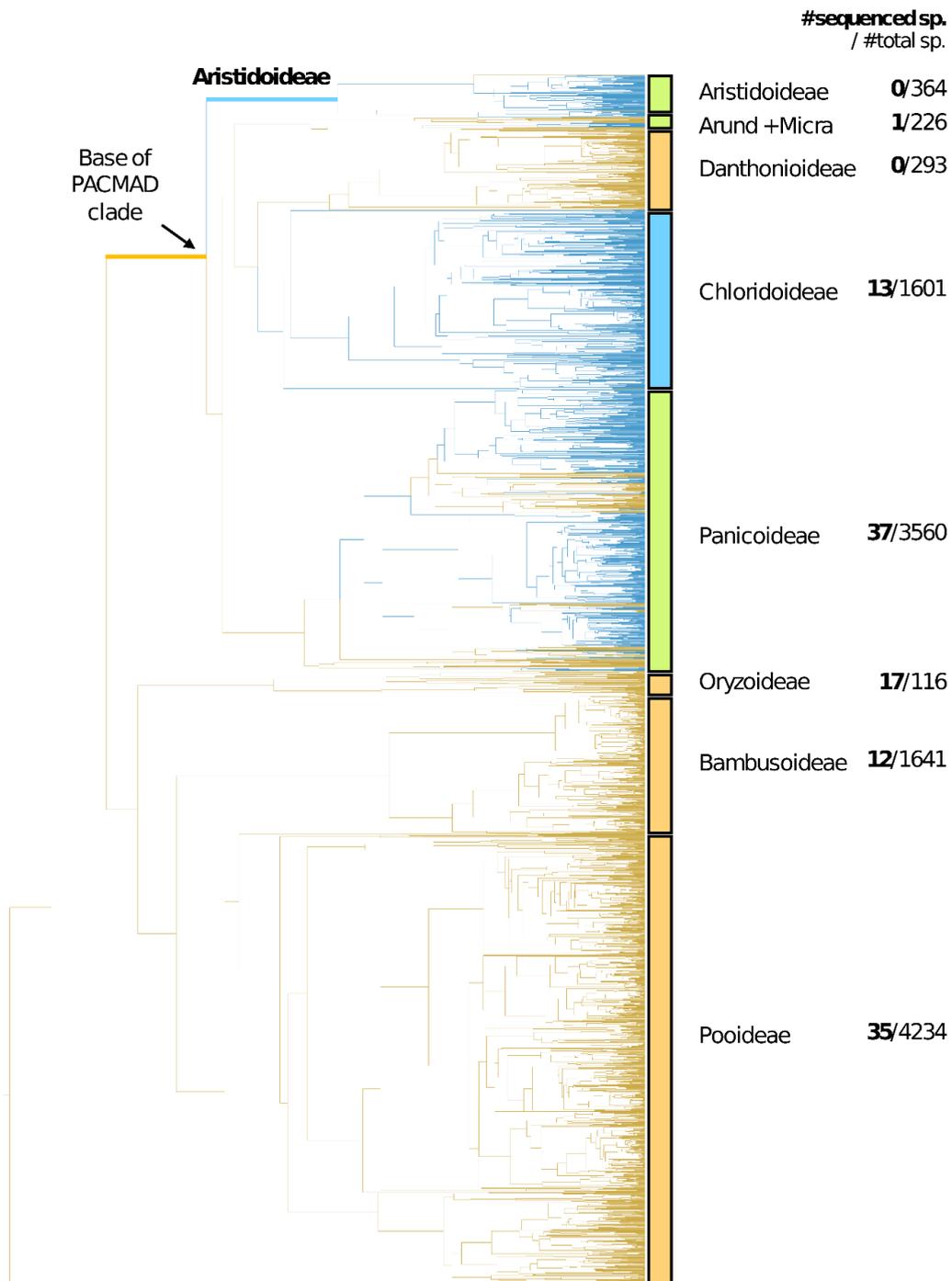


Figure 4.1.

Phylogeny, C₄ origins and sequenced genomes in grasses. The tree has been redrawn from (Spriggs et al., 2014), C₃ species in yellow and C₄ species in blue. The number of species per family was extracted from (Soreng et al., 2015) and the number of sequenced genomes per family was determined by searching in the NCBI database. The photos of *Aristida adscensionis* and *Stipagrostis hirtigluma* were extracted from the Kew - Royal Botanical Gardens website (<https://powo.science.kew.org/>).

4.3. Materials and methods

4.3.1. Tissue collection and DNA extractions

Aristida adscensionis seed was obtained from the USDA Germplasm Resources Information Network (GRIN) (accession number = PI 269867, collection location = Pakistan), and *Stipagrostis hirtigluma* var. *patula* seeds from Kew Millennium Seed Bank (accession number = 74902, collection location = Botswana). Seeds were germinated and grown at the Arthur Willis Environment Centre, University of Sheffield (UK) under semi-controlled conditions (12 h daylight, 25/20 °C day/night temperature). The 1C genome size for each species was estimated by flow cytometry using the one-step protocol (Doležel et al., 2007) with minor modifications (Clark et al., 2016).

High molecular weight genomic DNA was extracted for each species from flash frozen leaf tissue stored at -80 °C using the NucleoBond® HMW DNA kit (Macherey-Nagel), following manufacturer's protocol, and purified with AMPure XP Beads (Beckman Coulter Life Sciences). DNA quality, purity and concentration were estimated with a Femto Pulse (Agilent), NanoDrop (Thermo Fisher Scientific), and Qubit (Thermo Fisher Scientific), respectively. PacBio HiFi DNA libraries were prepared and run on Sequel II SMRT Cells in CCS run mode at the NERC Environmental Omics Facility (University of Liverpool, UK). In total, we ran two SMRT cells for *A. adscensionis* and three SMRT cells for *S. hirtigluma*.

4.3.2. Organelle assembly

To remove organelle data prior to nuclear genome assembly, we first generated reference chloroplast and mitochondrial organelle genomes.

Chloroplast sequences were identified by aligning the raw HiFi reads with BLASR v5.3.5-0 (Bussotti et al., 2011) to published chloroplast genome assemblies for each species downloaded from NCBI GenBank (*A. adscensionis* = MZ373986.1; *S. pennata* = MZ373985.1). To be considered for downstream analysis, reads were filtered by alignment length (≥ 5 kb) and alignment similarity ($\geq 99\%$). These reads were then *de novo* aligned using hifiasm v0.16.1 (Cheng et al., 2021) with default parameters. Primary contigs were inspected using circlator clean v1.5.5 (Hunt et al., 2015), discarding shorter contigs that were nested within larger sequences. The remaining contigs were annotated with the GeSeq v2.03 (Tillich et al., 2017) web portal. The annotations were used to identify overlapping contigs that were then manually trimmed, merged and circularised in Geneious Prime (v2023.0.1, Build 2022-11-28) to obtain the complete chloroplast genome, before being re-annotated with GeSeq.

Assembling mitochondrial genomes can be problematic for plants as they are not necessarily circular and can exist in multiple conformations within a cell (Sloan et al., 2012). To assemble the mitochondrial genome we used MitoHiFi v3.2 (Uliano-Silva et al., 2023) with a minimum percentage of each HiFi read sequence in the blast match to a closely related mitochondrial genome (-p) set to 30%. As expected, a fully resolved mitochondrial genome was not obtained. To select unique, long contigs, we used circlator clean, and annotated the remaining contigs with GeSeq. Only contigs containing at least one gene model were included in the draft mitochondrial contigs.

4.3.3. Nuclear assembly

The estimated coverage and ploidy for each species was calculated using the k-mer spectrum generated with GenomeScope 2.0 (Ranallo-Benavidez et al., 2020). Prior to nuclear genome assembly, we removed reads that belong to the organelles by mapping the raw HiFi reads to the

chloroplast and mitochondrial genomes with minimap2 v2-2.24 (Li, 2018), discarding those where $\geq 80\%$ of the read aligned. The organelle-filtered HiFi reads were then *de novo* assembled using hifiasm, using default parameters except for the number of haplotypes in *S. hirtigluma* being set to four as it was inferred to be tetraploid. The assembly completeness was checked using BUSCO v5.4.2 (Simão et al., 2015) with the poales_odb10 database, and Inspector v1.2 (Chen et al., 2021) was used to estimate assembly quality values (QV).

4.3.4. Genome annotation

Extensive *de-novo* TE Annotator v1.9.4 (Ou et al., 2019) was used to annotate transposable elements (TEs), and RepeatMasker v4.1.2-p1 to annotate simple repeats. The identified TEs and simple repeats were then soft-masked in each draft genome using BEDtools maskfasta (Quinlan and Hall, 2010). To annotate the softmasked reference genome we used both RNA and protein evidence separately, running BRAKER2 v2.1.6 (Brůna et al., 2021) with each independently before combining both annotation files with TSEBRA v1.0.3 (Gabriel et al., 2021). For the RNA annotation we used previously generated RNA-seq data for both species that was downloaded from the NCBI Sequence Read Archive (SRA) (Accession numbers = SRR13316956, SRR15729218, SRR6221622, SRR6221624, SRR6221636, SRR6221637). These data were then cleaned with trimmomatic (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:30 MINLEN:50), aligned to the draft genome using hisat2 (Kim et al., 2019), before using the resulting sorted BAM file in the BRAKER2 pipeline. For the protein informed annotation, we used the Plant dataset Odb10 from BUSCO. To prepare the protein dataset we used ProtHint to generate a genome annotation file that was then incorporated into the BRAKER2 annotation pipeline.

4.3.5. Identifying orthologous genes and identifying duplication events

To identify orthologous genes between the two reference genomes produced here and 18 other grasses with published genomes from across the Poaceae (PACMAD and BOP) and representing multiple C₄ origins (Figure 4.2; Table 4.S1), we used Orthofinder v.2.5.4 (Emms and Kelly, 2019). Orthologs were identified using the coding sequences (CDS), aligned using MUSCLE v3.2 (Edgar, 2004), and a maximum likelihood tree inferred using FastTree v2 (Price et al., 2010). Orthofinder reconciles gene duplications against the species tree it infers to determine the branch on which the duplication occurred. Of particular interest are the orthogroups inferred to have been duplicated at the base of the PACMAD clade, the branch representing the last common ancestor of all grass C₄ origins.

The genes duplicated on the PACMAD branch were further filtered to keep only those that had been duplicated and retained in all C₄ species included in the analysis, as these are potentially the most important loci for the convergent evolution of C₄. The function of these orthogroups was then inferred by annotating the maize gene against the SwissProt database, the MaizeGDB (Woodhouse et al., 2021) and the function of the *Arabidopsis thaliana* ortholog using TAIR (Berardini et al., 2015).

4.3.6. Gene ontology

The annotations for each orthogroup were derived from the annotations of *Oryza sativa* genes in Gene Ontology (Ashburner et al., 2000; Aleksander et al., 2023). The genes that were duplicated in the base of PACMAD were used in the PANTHER Overrepresentation Test in panther resource v.18 (Thomas et al., 2021; Mi et al., 2013) to determine if specific gene ontology (GO) terms were overrepresented in these genes compared to a set of rice genes that is released from Orthofinder. We

used Fisher's exact test to determine which GO terms were significantly more prevalent (FDR < 0.05) in the selected genes and then the multiple testings were corrected using the Bonferroni correction ($p < 0.05$).

4.4. Results

4.4.1. Assembly of two early diverging PACMAD grasses

We assembled reference genomes for two divergent Aristidoideae species, *Aristida adscensionis* and *Stipagrostis hirtigluma*. For *A. adscensionis*, we generated 31.0 Gb of HiFi data, which is approximately 53x coverage of the nuclear genome based on the flow cytometry estimated 1C of 0.58 Gb (Table 4.1). The analysis of the k-mer spectrum suggested that the *A. adscensionis* individual sequenced was diploid, and the final assembly was 451 Mb (77.8% of the flow cytometry estimated genome size), relatively contiguous (186 contigs and N50 = 5.6 Mb) and 95.8% of BUSCOs were detected (Table 4.1). Genome annotation showed that 53.0% of the genome is made up of TEs, and we were able to identify 31,208 protein coding genes (Table 4.1).

We generated 67.7 Gb of HiFi data for *S. hirtigluma* (Table 4.1). The analysis of the k-mer spectrum suggested it was a tetraploid, therefore the haploid genome estimated by flow cytometry would be 0.48 Gb (141x coverage). The final assembly was 613 Mb, which is higher than the haploid flow cytometry estimate, probably because divergent sequences from different subgenomes are assembled separately. The assembly was more fragmented (458 contigs, N50 = 12.1 Mb) than the *A. adscensionis* assembly, despite the higher coverage and larger N50 obtained, most likely due to the added complexity of assembling a polyploid. In terms of completeness, 96.9% of BUSCOs were detected (Table 4.1), and 29.7% were duplicated, supporting the partial redundancy of the assembly.

Genome annotation showed that 48.72% of the genome was made up of TEs, and we were able to identify 30,617 protein coding genes (Table 4.1).

Table 4.1.

Assembly statistics for the two *de novo* genomes.

	<i>Aristida adscensionis</i>	<i>Stipagrostis hirtigluma</i>
<u>Sequencing</u>		
Number of reads	1889280	4507740
Total sequence (Gb)	30.97	67.68
Average read length (kb)	16.39	15.01
<u>Assembly</u>		
Ploidy	Diploid	Tetraploid
Estimated genome size (1C/Gb)	0.58	0.48*
Total assembly length (Mb)	451.51	613.38
Number of contigs	186	458
Contig N50 (Mb)	5.6	12.14
Longest contig (Mb)	17.85	31
QV	63.55	25.06
BUSCO Complete (Duplicated)	94.1% (2.7%)	95.7% (29.7%)
BUSCO Fragmented	1.7%	1.2%
BUSCO Missing	4.2%	3.1%
<u>Annotation</u>		
% Transposable Element	53.0%	48.7%
Number of gene models	31208	30617

4.4.2. Duplication events at the base of PACMAD clade

The PACMAD clade has evolved C₄ photosynthesis has evolved at least 20 times in the PACMAD clade of grass whereas the sister BOP clade only contains C₃ species (Figure 4.1). This is potentially driven by genetic enablers that evolved in the last common ancestor of the PACMAD clade, facilitating the recurrent emergence of C₄ in this clade. We identified 407 duplication events represented by 369 orthogroups that were duplicated at the base of PACMAD clade (Table 4.S2). There was no difference between the proportion of orthogroups retained in the C₃ versus C₄ species (two tailed t-test p-value = 0.737), or in the total number of orthologs within each of these 369 orthogroups (two tailed t-test p-value = 0.292 [0.707 when correcting for ploidy level]). There were eight orthogroups duplicated on the PACMAD branch and retained in all C₄ species (Figure 4.2). Some of these orthogroups were subsequently further duplicated, with as many as 13 *Digitaria exilis* orthologs in orthogroup OG0000383 (Figure 4.2). This redundancy could have an impact in the evolvability of the C₄ cycle, as well as other traits. We further investigated the function of these C₄ conserved PACMAD duplications.

Four of the eight PACMAD duplicated orthogroups retained in the C₄ descendants are involved in cell wall biogenesis, with genes encoding: Beta-1,2-xylosyltransferase *XYXT1* (*XYXT1*) involved in the arabinosylation of xylan in grasses (Anders et al., 2012; Chiniqy et al., 2012); Peroxidase 2 (*PER2_5*) involved in lignin biosynthesis and xylem development (Fernández-Pérez et al., 2015; Hoffmann et al., 2020); a Walls are Thin 1 related (*WAT1*-related) protein belonging to a plant-specific gene family essential for secondary wall formation in *Arabidopsis* (Ranocha et al., 2010, 2013); and probable xyloglucan endotransglucosylase/hydrolase protein 12 [*XTH12*], that is expressed in the vascular bundles of elongating leaves and contributes to xyloglucan modifications in rice (Maris et al., 2009; Hara et al., 2014).

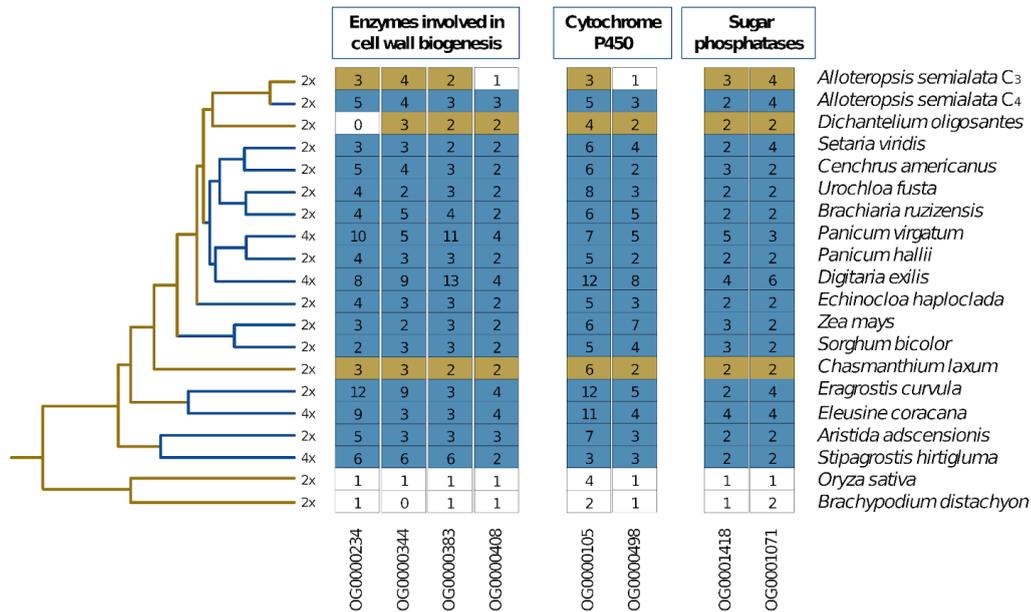


Figure 4.2.

Species tree containing all the species used in this study, including their ploidy, and the copy number found in each species for the orthogroups duplicated at the base of PACMAD. Yellow is used for C₃ species and blue for C₄ species, and white means no duplication.

Genes encoding two primary metabolic enzymes are also duplicated at the base of PACMAD and retained in all C₄ lineages: (DL)-glycerol-3-phosphatase and trehalose 6-phosphate phosphatase (TPPJ). (DL)-glycerol-3-phosphatase (GPP) is involved with core glycerol metabolism, and TPPJ participates in the trehalose biosynthetic process. Trehalose is a disaccharide that serves as a carbon source, a structural component and a stress protectant (Van Houtte et al., 2013). In plants, trehalose 6-phosphate also acts as a cue for important developmental processes such as flowering, embryogenesis and shoot branching (Fichtner and Lunn, 2021).

The last two enzymes duplicated at the base of PACMAD and retained are both cytochrome P450's, a protein superfamily that plays a key role in plant development and defence (Xu et al., 2015).

The gene ontology terms for the 369 duplicated genes in the base of PACMAD has revealed 36 above the threshold with p value 2.8×10^{-2} . exhibit functional diversity, encompassing role in cellular anatomical entity, cell modification, chloroplast stroma, and potassium transporter.

4.5. Discussion

C₄ photosynthesis is a remarkable example of convergent evolution, yet its occurrence is restricted to relatively few lineages (Christin et al., 2015). In grasses, 99% of species are found in two roughly equally sized clades (BOP = 5,753 species; PACMAD = 5,726 species; Soreng et al., 2017), and the >20 origins of C₄ in this family are solely restricted to the PACMAD grasses (Grass Phylogeny Working Group II, 2012). Here we assembled two genomes for the Aristidoideae grasses (*Aristida adscensionis* and *Stipagrostis hirtigluma*), a subfamily containing over 350 species. We then used comparative genomics to determine the genes that were duplicated at the base of PACMAD as these potentially represent genetic enablers underlying the repeated evolution of C₄ in this clade.

4.5.1. Duplication of a photorespiratory enzyme and a C₂ pump

One of the driving forces behind C₄ evolution is the reduction of photorespiration, an energetically costly process that increases at high temperatures (Sage et al., 2012). Indeed, the evolutionary trajectory from C₃ to C₄ predicts that the establishment of a photorespiratory CO₂ pump that operates across different leaf compartments is one of the key steps (Bräutigam and Gowik, 2016). The Asteraceae *Flaveria* genus contains several C₃-C₄ intermediates, and the emergence of these from their C₃ ancestor is thought to be partially driven by gene duplication and subfunctionalization (Schulze et al., 2013). The duplicated gene encoding the P-protein of the glycine decarboxylation complex became preferentially expressed in the bundle sheath cells, helping to establish the photorespiratory CO₂ pump across leaf compartments. Interestingly, we also identified a duplicate

copy of (DL)-glycerol-3-phosphatase 2 (GPP) that encodes a protein with homology to 2-phosphoglycolate phosphatase (PGP), the entry enzyme to the photorespiratory pathway (Schwarte and Bauwe, 2007).

4.5.2. Cell wall modifications may limit CO₂ diffusion and enhance water use efficiency

Half of the genes that were found to be duplicated at the base of the PACMAD clade and subsequently retained in the descendant C₄ species were associated with cell wall biogenesis (Figure 4.2 and Table 4.2). This indicates that compositional modifications to the cell wall were potentially a precursor to the origin of C₄ photosynthesis. In C₄ plants, CO₂ is initially incorporated into four-carbon acids either oxaloacetate or malate in the mesophyll cells, then the C₄ acid is transported into the bundle sheath cells where they are decarboxylated and the CO₂ that is released is subsequently fixed by Rubisco. A low CO₂ conductance across the cell walls from the bundle sheath to the mesophyll is an essential feature of the C₄ pathway (Von Caemmerer and Furbank, 2003), otherwise there is a risk that much of the CO₂ concentrated in the bundle sheaths would diffuse out. Cell wall composition and thickness are the main determinants of foliar gas exchange, and cell wall modifications are likely to impact photosynthetic efficiency (Carriquí et al., 2020; Flexas et al., 2021). Cell walls in grasses are composed of cellulose fibres encased in the hemicellulose polymer glucuronoarabinoxylan (Vogel, 2008). The PACMAD duplicated genes are involved in (1) hemicellulose biosynthesis, particularly in the arabinosylation of xylans (*XYXT1*; Anders et al., 2012), (2) hemicellulose modifications, such as splitting and reconnection of the xyloglucan crosslinks in the cell wall (*XTH12*; Hara et al., 2014), (3) secondary cell wall formation, including an upstream regulator (*WAT1-related*; Ranocha et al., 2010) and a peroxidase likely involved in lignification (*PER2_5*; Fernández-Pérez et al., 2015). Duplication of these genes may

have facilitated their subfunctionalization and specialisation to reduce CO₂ conductance and/or facilitate subsequent anatomical modification.

Table 4.2.

Gene functional annotation of the orthogroups duplicated at the base of PACMAD and retained in all C₄ species.

Orthogroup	Gene description	Swissprot top hit [species]	e-value	Arabidopsis ortholog(s)
OG0000234	Glycosyltransferase family 61 protein	<u>Beta-1,2-xylosyltransferase XYXT1</u> [<i>Oryza sativa Japonica Group</i>]	2.0E-104	AT3G18180, AT3G18170
OG0000344	Peroxidase (EC1.11.1.7)	<u>Peroxidase 2</u> [<i>Zea mays</i>]	2.0E-85	AT4G33420
OG0000383	WAT1-related protein	<u>WAT1-related protein</u> [<i>Arabidopsis thaliana</i>]	2.0E-46	AT3G45870, AT4G19185, AT5G45370
OG0000408	xyloglucan endotransglucosylase (EC2.4.1.207)	<u>Probable xyloglucan endotransglucosylase/hydrolase protein 12</u> [<i>Arabidopsis thaliana</i>]	4.0E-90	AT4G28850
OG0000105	Cytochrome P450	<u>Cytochrome P450 81Q32</u> [<i>Catharanthus roseus</i>]	3.0E-147	AT3G25180, AT2G25160
OG0000498	Cytochrome P450	<u>Cytochrome P450 CYP72A616</u> [<i>Paris polyphylla</i>]	0	
OG0001071	(DL)-glycerol-3-phosphatase 2	<u>(DL)-glycerol-3-phosphatase 1</u> [<i>Arabidopsis thaliana</i>]	4.0E-88	AT4G25840, AT5G57440
OG0001418	Trehalose 6-phosphate phosphatase (EC3.1.3.12)	<u>Trehalose 6-phosphate phosphatase RA3</u> [<i>Zea mays</i>]	0	AT1G35910

Furthermore, cytochrome P450 enzymes are responsible for vascular differentiation and the cell division and expansion has been duplicated was the base of the PACMAD clade and then retained at all 18 C₄ species. Cytochrome P450 enzymes prevent plants from becoming dehydrated during drought stress (Minerdi et al., 2023) which can play an important role in environment conditions that promote a photorespiration and they are also associated with the lignin biosynthesis process (Gou et al., 2018). Moreover, several P450 enzymes are involved in the synthesis of structural components of the plant cell wall (Grausem et al., 2014) and vein density (Rizal et al., 2015). The *AtTPPI* gene that was discovered here to be duplicated at the base of the PACMAD clade and retained in all species (OG0001418 in Figure 4.2) encodes one of trehalose-6-phosphate phosphatase (TPP family) that has been shown in Arabidopsis to regulate stomatal aperture during drought stress thereby facilitating the enhanced water use efficiency (Lin et al., 2020). Also, *AtTPPI* was found to participate in increasing the length of the primary root, and linked to reduced transpiration rate suggesting an enhancement of water use efficiency (Lin et al., 2020). Moreover, the overexpression of *AtTPPI* might influence the abundance of the auxin transporter (*PIN3*) (Lin et al., 2020) which is found in the leaf midrib cells (Li et al., 2024) and has been linked with drought stress adaptation (Krugman et al., 2011). Auxin moves between cells via plasmodesmata and is known to be involved in leaf development and vein patterning (Linh and Scarpella, 2022). Also, auxin is involved in the regulator of cell elongation, which has been linked with adaptation to fluctuations of blue light levels during light stress (Ma and Li, 2019; Li et al., 2024). Moreover, cell elongation is reduced during drought stress thereby influencing overall plant water use efficiency (Boutraa et al., 2011).

In addition to the aforementioned duplicated genes that were retained in all C₄ PACMAD descendants, the GO analysis also revealed that some duplicated genes already present at the base of

the PACMAD clade have functions that are related to cell modification and chloroplast stroma.

Also, the GO analysis showed some functions related to potassium ion transport which is a crucial component of mechanisms that control stomatal guard cells and the opening and closing of stomata (Sawhney and Zelitch, 1969; Osakabe et al., 2013).

4.6. Conclusion

The decoding, assembly and annotation of two new genomes from the Aristidoideae allowed us to investigate the genetic changes that happened in the PACMAD common ancestor that may have enabled the repeated evolution of C₄ photosynthesis in this group. We found eight duplication events that were retained in all subsequent, daughter C₄ lineages, including genes with functions related to cell wall composition, metabolism and stress responses.

4.7. References

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25, 25–29.
- Anders, N., Wilkinson, M. D., Lovegrove, A., Freeman, J., Tryfona, T., Pellny, T. K., Weimar, T., Mortimer, J. C., Stott, K., Baker, J. M., Defoin-Platel, M., Shewry, P. R., Dupree, P., and Mitchell, R. A. (2012). Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 989–993.
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. *Genesis*, 53, 474-485.
- Bianconi, M. E., Dunning, L. T., Moreno-Villena, J. J., Osborne, C. P., and Christin, P. A. (2018). Gene duplication and dosage effects during the early emergence of C₄ photosynthesis in the grass genus *Alloteropsis*. *Journal of Experimental Botany*, 69, 1967-1980.
- Boutraa, T., Akhkha, A., Al-Shoaibi, A. A., and Alhejeli, A. M. (2010). Effect of water stress on growth and water use efficiency (WUE) of some wheat cultivars (*Triticum durum*) grown in Saudi Arabia. *Journal of Taibah University for Science*, 3, 39-48.
- Bräutigam, A., and Gowik, U. (2016). Photorespiration connects C₃ and C₄ photosynthesis. *Journal of Experimental Botany*, 67, 2953–2962.
- Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3, lqaa108.
- Brown, A. D. (1978). Compatible solutes and extreme water stress in eukaryotic micro-organisms. *Advances in Microbial Physiology*, 17, 181-242.
- Bussotti, G., Raineri, E., Erb, I., Zytnicki, M., Wilm, A., Beaudoin, E., Bucher, P., and Notredame, C. (2011). BlastR--fast and accurate database searches for non-coding RNAs. *Nucleic Acids Research*, 39, 6886–6895.

- Carriqui, M., Nadal, M., Clemente-Moreno, M. J., Gago, J., Miedes, E., and Flexas, J. (2020). Cell wall composition strongly influences mesophyll conductance in gymnosperms. *The Plant Journal : for cell and molecular biology*, 103, 1372–1385.
- Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., and Chong, Z. (2021). Accurate long-read de novo assembly evaluation with Inspector. *Genome Biology*, 22, 312.
- Chiniquy, D., Sharma, V., Schultink, A., Baidoo, E. E., Rautengarten, C., Cheng, K., Carroll, A., Ulvskov, P., Harholt, J., Keasling, J. D., Pauly, M., Scheller, H. V., and Ronald, P. C. (2012). XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 17117–17122.
- Christin, P. A., and Besnard, G. (2009). Two independent C₄ origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *American Journal of Botany*, 96, 2234–2239.
- Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013). Parallel recruitment of multiple genes into C₄ photosynthesis. *Genome Biology and Evolution*, 5, 2174–2187.
- Christin, P. A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., Garrison, L. M., Vorontsova, M. S., and Edwards, E. J. (2013b). Anatomical enablers and the evolution of C₄ photosynthesis in grasses. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 1381–1386.
- Clark, J., Hidalgo, O., Pellicer, J., Liu, H., Marquardt, J., Robert, Y., Christenhusz, M., Zhang, S., Gibby, M., Leitch, I. J., and Schneider, H. (2016). Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist*, 210, 1072–1082.
- Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, 2, 2233–2244.
- Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J., Mapaura, A., Grundy, I., Kayombo, C., Dunning, L., Kentatchime, F., Ariyaratne, M., Yakandawala, D., Besnard, G., Quick, P., Bräutigam, A., Osborne, C. and Christin, P. A. (2019). Key changes in gene expression identified for different stages of C₄ evolution in *Alloteropsis semialata*. *Journal of Experimental Botany*, 70, 3255–3268.
- Emms, D. M., Covshoff, S., Hibberd, J. M., and Kelly, S. (2016). Independent and parallel evolution of new genes by gene duplication in two origins of C₄ photosynthesis provides new insight into the mechanism of phloem loading in C₄ species. *Molecular Biology and Evolution*, 33, 1796–1806.

- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 1-14.
- Evans, J. R., Kaldenhoff, R., Genty, B., and Terashima, I. (2009). Resistances along the CO₂ diffusion pathway inside leaves. *Journal of Experimental Botany*, 60, 2235-2248.
- Fernández-Pérez, F., Pomar, F., Pedreño, M. A., and Novo-Uzal, E. (2015). Suppression of Arabidopsis peroxidase 72 alters cell wall and phenylpropanoid metabolism. *Plant Science*, 239, 192–199.
- Fichtner, F., and Lunn, J. E. (2021). The role of trehalose 6-phosphate (Tre6P) in plant metabolism and development. *Annual Review of Plant Biology*, 72, 737-760.
- Flexas, J., Clemente-Moreno, M. J., Bota, J., Brodribb, T. J., Gago, J., Mizokami, Y., Nadal, M., Perera-Castro, A. V., Roig-Oliver, M., Sugiura, D., Xiong, D., and Carriquí, M. (2021). Cell wall thickness and composition are involved in photosynthetic limitation. *Journal of Experimental Botany*, 72, 3971–3986.
- Fujita, Y., Fujita, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). ABA-mediated transcriptional regulation in response to osmotic stress in plants. *Journal of Plant Research*, 124, 509-525.
- Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., and Stanke, M. (2021). TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*, 22, 1-12.
- Grass Phylogeny Working Group II. (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C₄ origins. *New Phytologist*, 193, 304-312.
- Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224, iyad031.
- Hara, Y., Yokoyama, R., Osakabe, K., Toki, S., and Nishitani, K. (2014). Function of xyloglucan endotransglucosylase/hydrolases in rice. *Annals of Botany*, 114, 1309-1318.
- Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P. A., and Edwards, E. J. (2019). The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics*, 20, 485-493.
- Hoffmann, N., Benske, A., Betz, H., Schuetz, M., and Samuels, A. L. (2020). Laccases and peroxidases co-localize in lignified secondary cell walls throughout stem development. *Plant Physiology*, 184, 806-822.
- (<https://github.com/Maggi-Chen/Inspector>) (v1.0.2).

- Huang, C. F., Liu, W. Y., Lu, M. Y. J., Chen, Y. H., Ku, M. S., and Li, W. H. (2021). Whole-genome duplication facilitated the evolution of C₄ photosynthesis in *Gynandropsis gynandra*. *Molecular Biology and Evolution*, 38, 4715-4731.
- Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 16, 294.
- Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., and Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, 8, 1-6.
- Jun, X. U., WANG, X. Y., and GUO, W. Z. (2015). The cytochrome P450 superfamily: Key players in plant development and defense. *Journal of Integrative Agriculture*, 14, 1673-1686.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37, 907-915.
- Krugman, T., Peleg, Z., Quansah, L., Chagué, V., Korol, A. B., Nevo, E., Saranga, Y., Fait, A., Chalhoub, B., and Fahima, T. (2011). Alteration in expression of hormone-related genes in wild emmer wheat roots associated with drought adaptation mechanisms. *Functional and Integrative Genomics*, 11, 565-583.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9, 357-359.
- Lewis, M. W., and Hake, S. (2016). Keep on growing: building and patterning leaves in the grasses. *Current Opinion in Plant Biology*, 29, 80-86.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094-3100.
- Li, Y., Han, D., Hu, G., Dauvillee, D., Sommerfeld, M., Ball, S., and Hu, Q. (2010). *Chlamydomonas* starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol. *Metabolic Engineering*, 12, 387-391.
- Li, J., Yang, J., Gao, Y., Zhang, Z., Gao, C., Chen, S., and Liesche, J. (2023). Parallel auxin transport via PINs and plasmodesmata during the *Arabidopsis* leaf hyponasty response. *Plant Cell Reports*, 43, 4.
- Ma, L., and Li, G. (2019). Auxin-dependent cell elongation during the shade avoidance response. *Frontiers in Plant Science*, 10, 465042.
- Linh NM, Scarpella E. 2022. Leaf vein patterning is regulated by the aperture of plasmodesmata intercellular channels. *PLoS Biology* 20: e3001781.
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, 92, 155-161.

- Maris, A., Suslov, D., Fry, S. C., Verbelen, J. P., and Vissenberg, K. (2009). Enzymic characterization of two recombinant xyloglucan endotransglucosylase/hydrolase (XTH) proteins of *Arabidopsis* and their effect on root growth and cell wall extension. *Journal of Experimental Botany*, 60, 3959-3972.
- Mi, H., Muruganujan, A., and Thomas, P. D. (2012). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41, D377-D386.
- Moore, R. C., and Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, 8, 122-128.
- Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P., and Christin, P. A. (2018). Highly expressed genes are preferentially co-opted for C₄ photosynthesis. *Molecular Biology and Evolution*, 35, 94-106.
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., and Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, 20, 275.
- Osakabe, Y., Arinaga, N., Umezawa, T., Katsura, S., Nagamachi, K., Tanaka, H., Ohiraki, H., Yamada, K., Seo, S. U., Abo, M., Yoshimura, E., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2013). Osmotic stress responses and plant growth controlled by potassium transporters in *Arabidopsis*. *The Plant Cell*, 25, 609–624.
- Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, 171, 2294-2316.
- Pereira, L., Bianconi, M. E., Osborne, C. P., Christin, P. A., and Dunning, L. T. (2023). *Alloteropsis semialata* as a study system for C₄ evolution in grasses. *Annals of Botany*, 132, 365-382.
- Pfalz, M., Mikkelsen, M. D., Bednarek, P., Olsen, C. E., Halkier, B. A., and Kroymann, J. (2011). Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification. *The Plant Cell*, 23, 716-729.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS One*, 5, e9490.
- Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11, 1432.
- Ranocha, P., Denancé, N., Vanholme, R., Freyrier, A., Martinez, Y., Hoffmann, L., Köhler, L., Pouzet, C., Renou, J. P., Sundberg, B., Boerjan, W., and Goffner, D. (2010). Walls are thin 1 (WAT1), an *Arabidopsis* homolog of

- Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant Journal*, 63, 469–483.
- Rizal, G., Thakur, V., Dionora, J., Karki, S., Wanchana, S., Acebron, K., Larazo, N., Garcia, R., Mabilangan, A., Montecillo, F., Danila, F., Mogul, R., Pablico, P., Leung, H., Langdale, J. A., Sheehy, J., Kelly, S., and Quick, W. P. (2015). Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *The Plant Journal*, 84, 257–266.
- Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature*, 441, 227–230.
- Sawhney, B. L., and Zelitch, I. (1969). Direct determination of potassium ion accumulation in guard cells in relation to stomatal opening in light. *Plant physiology*, 44, 1350–1354
- Schlüter, U., Bräutigam, A., Droz, J. M., Schwender, J., and Weber, A. P. (2019). The role of alanine and aspartate aminotransferases in C₄ photosynthesis. *Plant Biology*, 21, 64-76.
- Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., Gowik, U., and Westhoff, P. (2013). Evolution of C₄ photosynthesis in the genus *flaveria*: establishment of a photorespiratory CO₂ pump. *The Plant cell*, 25, 2522–2535.
- Schwarte, S., and Bauwe, H. (2007). Identification of the photorespiratory 2-phosphoglycolate phosphatase, PGLP1, in *Arabidopsis*. *Plant Physiology*, 144, 1580-1586.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210-3212.
- Sloan, D. B., Alverson, A. J., Chuckalovcak, J. P., Wu, M., McCauley, D. E., Palmer, J. D., and Taylor, D. R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology*, 10, e1001241.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., Filgueira, T., Davis, J., and Morrone, O. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53, 117-137.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., Barberá, P., Gillespie, L. J., and Zuloaga, F. O. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution*, 55, 259–290.

- Spriggs, E. L., Christin, P. A., and Edwards, E. J. (2014). C₄ photosynthesis promoted species diversification during the Miocene grassland expansion. *PloS One*, 9, e97722.
- Tholen, D., and Zhu, X. G. (2011). The mechanistic basis of internal conductance: a theoretical analysis of mesophyll cell photosynthesis and CO₂ diffusion. *Plant Physiology*, 156, 90–105.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45, W6–W11.
- Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L., and Mi, H. (2021). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science*, 31, 8–22.
- Thornton, P. K., Jones, P. G., Alagarswamy, G., Andresen, J., and Herrero, M. (2010). Adapting to climate change: agricultural system and household impacts in East Africa. *Agricultural Systems*, 103, 73-82.
- Uliano-Silva, M., Ferreira, J. G. R. N., Krashennnikova, K., Darwin Tree of Life Consortium, Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., and McCarthy, S. A. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*, 24, 288.
- von Caemmerer, S., and Furbank, R. T. (2003). The C₄ pathway: an efficient CO₂ pump. *Photosynthesis Research*, 77, 191–207.
- Vogel J. (2008). Unique aspects of the grass cell wall. *Current Opinion in Plant Biology*, 11, 301–307.
- Van Houtte, H., Vandesteene, L., López-Galvis, L., Lemmens, L., Kissel, E., Carpentier, S., Feil, R., Avonce, N., Beeckman, T., Lunn, J. E., and Van Dijck, P. (2013). Overexpression of the trehalase gene AtTRE1 leads to increased drought stress tolerance in Arabidopsis and is involved in abscisic acid-induced stomatal closure. *Plant Physiology*, 161, 1158–1171
- Voznesenskaya, E. V., Chuong, S. D., Kiirats, O., Franceschi, V. R., and Edwards, G. E. (2005). Evidence that C₄ species in genus *Stipagrostis*, family Poaceae, are NADP-malic enzyme subtype with nonclassical type of Kranz anatomy (Stipagrostoid). *Plant Science*, 168, 731-739.
- Wagner, G. P., Amemiya, C., and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proceedings of the National Academy of Sciences, USA* 100, 14603-14606.
- Wang, Z. T., Ullrich, N., Joo, S., Waffenschmidt, S., and Goodenough, U. (2009). Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. *Eukaryotic Cell*, 8, 1856-1868.

Woodhouse, M. R., Cannon, E. K., Portwood, J. L., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., and Andorf, C. M. (2021). A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biology*, 21, 1-10.

Supplementary Tables of this chapter are available at:

<https://drive.google.com/file/d/19W33WDiJAhKFTNVGAhmDg8Npbkyzhi7R/view?usp=sharing>

https://drive.google.com/file/d/1_39W3YU9pYt_gkBbDF5U4RifZyjvuU6q/view?usp=sharing

Chapter 5: General Discussion

5. General discussion

The overarching aim of this thesis was to investigate the evolutionary steps required to assemble the C₄ cycle, a complex trait requiring the intimate coordination of numerous anatomical and biochemical components. To do this, I first focused on the remarkable photosynthetic variation in *Alloteropsis semialata*, the only species that has both C₄ and non-C₄ photosynthetic genotypes. Using this system I conducted a comparative analysis to identify leaf traits associated with the proportion of carbon fixed using the C₄ cycle, showing that the proportion of bundle sheath tissue increases in plants with higher C₄ activity (Chapter 2). I then followed up the results in the next data chapter, using a genome wide association study (GWAS) to identify several candidate genes related to the strength of the C₄ cycle and an increasing proportion of bundle sheath tissue (Chapter 3). Finally, in the last data chapter, I took a broader perspective to investigate the genetic precursors that may have enabled the repeated evolution of C₄ photosynthesis in the PACMAD grasses. Genes associated with lignification and the regulation of stomatal aperture were discovered to have been duplicated at the base of the PACMAD clade and subsequently retained in all C₄ origins, potentially facilitating the repeated convergent evolution of Kranz anatomy (Chapter 4). All three data chapters highlight the importance of anatomical modifications during the emergence of the C₄ cycle.

5.1. Understanding the evolutionary steps to C₄ carbon fixation

The current best theory proposes that the C₄ cycle evolved via a series of gradual modifications to biochemical and anatomical traits and numerous intermediate phases during the evolutionary progression from C₃ to C₄ (Sage, 2004; Sage et al., 2012). These stages have been represented within an adaptive fitness landscape, which connects various biological properties (Figure 5.1; Edwards, 2014). However, it has remained unclear how individual C₄ trait components were

acquired. The results presented in Chapter 2 suggested that the C_3 - C_4 intermediate *A. semialata* accessions with carbon isotope values less negative than -23‰ acquire a significant fraction of their CO_2 via the C_4 cycle while still operating the C_3 cycle (potentially involving a photorespiratory pump) for part of their carbon acquisition (Brown and Hattersley, 1989; von Caemmerer, 1992; Farquhar et al., 1982; Stata and Sage, 2019). Increasing C_4 activity in *A. semialata* was correlated with leaf anatomical modifications, specifically increasing the inner bundle sheath volume, and decreasing the distance between veins. The strength of the C_4 cycle was also found to be correlated with the local environmental conditions, suggesting that each population has locally adapted its photosynthetic rate.

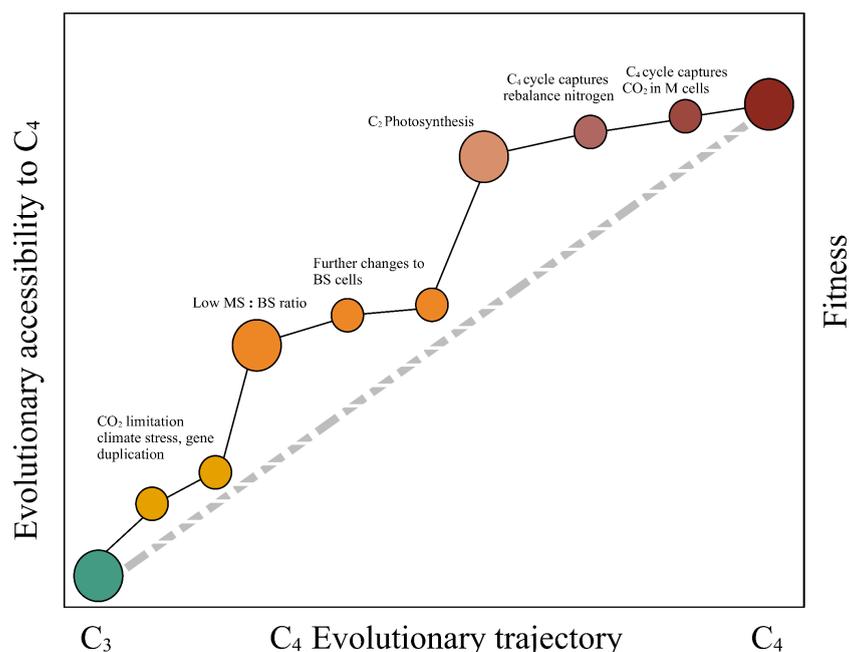


Figure 5.1.

The model suggests the main C_4 evolution stages and C_4 trajectory with significant incremental steps for each stage. Only two major intermediate stages are shown, the first is increasing the volume of bundle sheath and the other is C_2 photosynthesis that precedes the full C_4 photosynthesis. Adapted from Edwards (2014).

Leaf cellular has been proposed as a key anatomy driver of C_4 evolution, but to date only a couple of genes associated with the C_4 leaf anatomy (Kranz anatomy) have been identified and

characterised functionally (Simpson et al., 2021). This is because most variation in photosynthetic types is discretely segregated between species, rather than within, making quantitative genetics approaches difficult. Here, this challenge was overcome using *A. semialata* as a study system, and through conducting the first GWAS for the strength of the C₄ cycle and Kranz anatomy. The GWAS results identified multiple candidate genes that could function in the establishment of identified Kranz anatomy, with functions associated with leaf developments including vein patterning and meristem determinacy (Chapter 3). Interestingly, the identified genes include known paralogs of previously identified Kranz anatomy regulators, for example *SHORTROOT* and *SCARECROW*, indicating that the evolution of C₄ leaf anatomy may repeatedly co-opt loci from the same gene family.

Finally, thanks to the two newly assembled Aristidoideae genomes generated here, we are able for the first time to identify genes specifically duplicated at the base of the PACMAD clade, representing the last common ancestor of all C₄ origins in grasses. We found that genes associated with lignification and the regulation of stomatal aperture were duplicated at the base of the PACMAD clade, potentially facilitating the repeated convergent evolution of Kranz anatomy within this clade by increasing the water use efficiency (WUE) (Chapter 4).

In summary, this research has revealed a quantitative basis for the anatomical modifications that offered rapid evolutionary paths to physiological adaptation (Chapter 2), that the convergent origins of C₄-related anatomy traits may be associated with different members of the same gene family (Chapter 3), and that genes known to regulate WUE through anatomical changes were duplicated at the base of PACMAD clade (Chapter 4). Overall the findings support the idea that constructing the

appropriate leaf anatomy where C_4 photosynthesis can operate may be the foundational step in its evolution. Once the anatomy is in place, then the metabolic enzymes can be recruited.

5.2. Will these results be useful for engineering C_4 photosynthesis in C_3 species?

Improving agricultural productivity by engineering the C_4 photosynthetic pathway into C_3 crops such as rice has been proposed as a method to enhance rice productivity and support global food security (Kajala et al., 2011; Dash et al., 2022; Sahoo et al., 2024; Talukder et al., 2024). There has been a lot of effort to install C_4 photosynthesis in C_3 rice, for example the C_4 Rice Project (reviewed in Furbank et al., 2023). A higher CO_2 assimilation efficiency as a result of C_4 photosynthesis would produce faster growing crops with potentially higher yields. However, there are some challenges, mostly due to the fact that the C_4 pathway involves multiple anatomical and biochemical changes (Dengler et al., 1994; Hattersley, 1984; Lundgren et al., 2014). Since we do not know the levels of core C_4 enzymes needed, researchers have used the level of C_4 enzyme expression in maize as a target (reviewed in Furbank et al., 2023). Transferring and expressing these traits in C_3 plants is a complex task. Also, C_4 plants have distinct regulatory mechanisms for photosynthesis, which need to be integrated into C_3 plants once the anatomy has been established (Christin et al., 2013; Furbank et al., 2023; Talukder et al., 2024). The C_4 rice engineering goal is to incorporate the complete C_4 photosynthetic pathway, along with its anatomical specialisations, into rice. Some progress had been made in expressing several maize photosynthetic genes in rice leaves to achieve C_4 photosynthesis within a single cell. This approach aimed to enable carboxylation by PEPC and decarboxylation by NADP-malic enzyme (NADP-ME) to occur in mesophyll cells (Miyao et al. 2011; Miyao 2003; Ermacova et al., 2023; reviewed in Furbank et al., 2023). The expression of essential photosynthetic enzymes, and specific protein expression in the bundle sheath or mesophyll cells are needed to build a prototype. In short, there are many obstacles in developing tools to

achieve high expression levels of C_4 pathway enzymes and precise cell-specific expression in the mesophyll or bundle sheath compartments for constructing a rice C_4 system. To date, the C_4 rice project has had mixed fairly limited success despite a significant amount of funding.

The intraspecific variation in *A. semialata* allows a fine-scale understanding of the genetic basis of C_4 evolution (Chapter 3), including the intermediate steps involved in the assembly of this complex trait. This is crucially important to identify the initial modifications required to support the emergence of this trait, something that may ultimately have applications in the engineering of C_4 photosynthesis in C_3 crops such as rice. Hence, the results from Chapter 2, which revealed the significant variations in anatomical traits that explain C_4 strengthening, can help guide the C_4 engineering processes by focusing on those traits (Chapter 2). The results in Chapter 3 assist in this by showing that we have identified the right gene families of leaf anatomy regulators that are common across C_4 lineages. However, these are large multigene families and different paralogs can be recruited, potentially meaning the tailoring of a C_4 leaf anatomy will require the use of specific loci.

5.3. Anatomical measurements of C_3 and C_4 would enrich the analysis

The sample size we had was enough to detect C_4 variation among the C_3 - C_4 *A. semialata* in (Chapter 2) using carbon isotope ratio. Adding anatomical measurements for C_3 and C_4 populations would enrich the study and provide a broader understanding of the evolutionary dynamics within *A. semialata*. Studying anatomical traits across these populations might reveal the specific structural modifications that underlie the transition from C_3 to C_4 and help to define the intermediate steps forms that may have led from C_3 to C_4 photosynthesis. Hence, both populations provide the potential to reveal the genetic basis underlying the anatomical differences leading to morphological

and physiological adaptations. However, that was not always possible as there were no cross-sections for these accessions and, even if there were cross sections, anatomical measurements take a prohibitively long time to accomplish within a PhD.

It would also be nice to look at variation in these traits within populations at greater and spatial resolutions. This would help validate hypotheses regarding the selective pressures driving the evolution of the C_4 pathway in *A. semialata*, and as a result would significantly enhance the depth and breadth of the GWAS analysis in Chapter 3. Furthermore, it would be beneficial to analyse the anatomical and carbon isotope ratio variation within another C_3+C_4 *A. semialata* population or at least increase the number of individuals of the ZAM1715 population that were already analysed. This population consists of C_3+C_4 intermediates and C_4 individuals, which might limit our conclusions. Hence, it helps identify local adaptations that might be unique to a specific environment by isolating the genetically fixed traits from the plastic ones. This can provide clues about how different anatomical traits contribute to the fitness of plants in specific ecological niches.

5.4. What is the role of phenotypic plasticity in C_4 leaf anatomy?

We have shown that wild populations of C_3+C_4 *A. semialata* had different $\delta^{13}C$ values that were correlated with local environmental conditions (Chapter 2). It would have been valuable to conduct a plasticity experiment to test whether these differences had genetic basis by growing a panel of C_3+C_4 *A. semialata* populations in several different environments where variables including temperature, light and soil are varied, and then study the leaf cellular anatomy. The goal would be to determine whether the variation in anatomy has a purely genetic basis rather than being a result of phenotypic plasticity. This would provide useful understanding of the extent of phenotypic plasticity within these populations and how environmental factors influence their growth,

development, and overall fitness of the individual accessions. Understanding these adaptive mechanisms could further elucidate the evolutionary pathways and ecological strategies of *C₄ A. semialata*. Nevertheless, a planned plasticity experiment was hindered by various limitations. We had living material of only four *C₃+ C₄ A. semialata* accessions from Zambia (ZAM1503-03) and Tanzania (TAN1602-03, TAN1-01A and TAN1-04B). These accessions are maintained as part of a living collection in the glasshouse at the Arthur Willis Environment Centre (AWEC) at the University of Sheffield, UK. Having a small number of accessions limited the generalizability and robustness of the findings.

Whilst plasticity may play a role in trait expression, the GWAS results revealed that these traits were largely heritable, apart from the inner bundle sheath width. The inner bundle sheath width may be a plastic component of the *C₄* leaf anatomy, and could be a result of differences in WUE under drought conditions (Lynch et al., 2012).

5.5. Would increasing the density of genome wide markers identify new *C₄* loci?

QTL analysis is a powerful tool to elucidate the genetic basis of traits of interest, and could be a promising strategy to map genomic regions associated with traits of *C₃-C₄* intermediate hybrids. Hybrid F2's would have a recombined genetic basis of the *C₄* traits allowing for finer mapping of genes associated with our traits of interest. Whilst hybrids between different photosynthetic types in *A. semialata* have been generated, these have not progressed beyond the F1 stage (Bianconi et al., 2020). We therefore adopted a GWAS approach.

We used dd-RADseq data to successfully identify regions of the genome correlated with our traits. However, the density of markers was relatively low, meaning that some regions of the genome may

have been missed. Using whole-genome sequencing would allow for the identification of genetic variants at a greater resolution across the entire genome. Also, it could be used to potentially detect structural variants such as insertions, deletions, and copy number variations, which may play a role in trait variation and would have been missed with the dd-RADseq data.

5.6. Could structural variation play a role in C₄ leaf anatomy evolution?

It would be nice to verify if any of the outlying regions identified in the GWAS analysis are associated with structural variants. GWAS results for several beneficial agronomic traits (e.g. yield and stress tolerance) in other plant species have been linked to structural variants (Tao et al., 2016). Structural variants can contribute to genetic diversity and adaptation by influencing the genetic basis underlying the anatomical and biochemical adaptations necessary for C₄ and C₃ (Lappalainen et al., 2013). To do this we would require long read sequencing for multiple *A. semialata* accessions, something that may be possible in the near future as sequencing costs continue to plummet. Indeed, we had planned to use the five available *A. semialata* reference genomes (Raimondeau et al., 2024), but these were generated using a hybrid approach (Illumina and PacBio pre-hifi) and the contiguity was not sufficient to generate structural variant graphs.

5.7. Verifying the function of our candidate genes in the evolution of C₄ leaf anatomy

It would be beneficial to conduct knock-out experiments on the genes identified in Chapter 3 that are associated with the most significant C₄-related anatomical trait. One of the candidate genes that could be promising was *GLS8* which may control the bundle sheath distance either by increasing venation or expanding the number of channels between cells to allow auxin signals travelling

throughout these cells. Bundle sheath distance plays an important role in how C₄ metabolites are moved between the photosynthetic cells, and reducing this distance can help a plant to shift to fixing more of its atmospheric CO₂ via the C₄ cycle. Ideally, the number of mesophyll cells between bundle sheath cells should be determined. Disabling the *gls8* gene would be a real step to understand its function in conservation or divergence. An alternative *GLS8* to knock out approaches would be single cell or spatial transcriptomics which could be used and complementary approach to verify whether or not that the candidate genes were expressed during leaf development. To elucidate this, leaf samples would be collected at specific developmental stages (i.e. individual leaf primordia) to see if the identified genes from GWAS are expressed when the C₄ anatomy is developing. Both of these approaches were beyond the scope of this thesis due to time constraints and the lack of a stable transformation method for *A.semialata*, but are potentially profitable avenues for future research.

5.8. Conclusions

The findings presented here support the conclusion that leaf anatomy is paramount for C₄ emergence in grasses. Here, leaf anatomy was found to correlate with the strength of the C₄ cycle and candidate genes responsible for the emergence of a C₄ leaf anatomy were identified using GWAS. Even when in *A.semialata* attention was shifted towards the elucidation of the impact of gene duplication events on the C₄ trajectory using the entire PACMAD clade, most of the duplication events we identified involved genes linked to anatomical traits, particularly those contributing to plant cell wall biosynthetic pathway. This work has built upon an advanced understanding of C₄ evolution and the evolutionary forces shaping the emergence and refinement of C₄ photosynthetic pathways.

5.9. References

- Bianconi ME, Dunning LT, Curran EV, Hidalgo O, Powell RF, Mian S, et al. 2020. Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B: Biological Sciences* 287: 20201960.
- Christin, P. A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., Garrison, L. M., Vorontsova, M. S., and Edwards, E. J. (2013). Anatomical enablers and the evolution of C₄ photosynthesis in grasses. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 1381–1386
- Dash, S. S. S., Lenka, D., Sahoo, J. P., Tripathy, S. K., Samal, K. C., Lenka, D., and Panda, R. K. (2022). Biochemical characterization of maize (*Zea mays L.*) hybrids under excessive soil moisture stress. *Cereal Research Communications*, 50, 875-884.
- Dengler, N. (1994). Quantitative leaf anatomy of C₃ and C₄ grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Annals of Botany*, 73, 241–255.
- Edwards, E. J. (2014). The inevitability of C₄ photosynthesis. *elife*, 3, e03702.
- Furbank, R., Kelly, S., and von Caemmerer, S. (2023). Photosynthesis and food security: the evolving story of C₄ rice. *Photosynthesis Research*, 158, 121-130.
- Hattersley, P.W. (1984). Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Annals of Botany*, 53, 163–180.
- Hornstein, E. D., and Sederoff, H. (2024). Back to the future: re-engineering the evolutionarily lost arbuscular mycorrhiza host trait to improve climate resilience for agriculture. *Critical Reviews in Plant Sciences*, 43, 1-33.
- Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B. J., Dionora, M. J., Mogul, R. T., Mabilangan, A. E., Danila, F. R., Hibberd, J. M., and Quick, W. P. (2011). Strategies for engineering a two-celled C₄ photosynthetic pathway into rice. *Journal of Experimental Botany*, 62, 3001–3010.
- Lundgren, M. R., Osborne, C. P., and Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C₄ evolution. *Journal of Experimental Botany*, 65, 3357–3369.
- Lynch, D. J., McInerney, F. A., Kouwenberg, L. L., and Gonzalez-Meler, M. A. (2012). Plasticity in bundle sheath extensions of heterobaric leaves. *American Journal of Botany*, 99, 1197-1206.
- Miyao, M. (2003). Molecular evolution and genetic engineering of C₄ photosynthetic enzymes. *Journal of Experimental Botany*, 54, 179-189.

- Miyao, M., Masumoto, C., Miyazawa, S. I., and Fukayama, H. (2011). Lessons from engineering a single-cell C₄ photosynthetic pathway into rice. *Journal of Experimental Botany*, 62, 3021-3029.
- Sage, R. F. (2004). The evolution of C₄ photosynthesis. *New Phytologist*, 161, 341-370.
- Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C₄ photosynthesis. *Annual Review of Plant Biology*, 63, 19-47.
- Sahoo, J. P., Mahapatra, D., Mahapatra, M., Dweh, T. J., Kayastha, S., Pradhan, P., Tripathy, S., Samal, K., Mishra, A., Dash, M., and Nanda, S. (2024). Understanding the biochemical, physiological, molecular, and synthetic biology approaches towards the development of C₄ rice (*Oryza sativa L.*). *Cereal Research Communications*, 1-13.
- Simpson CJC, Reeves G, Tripathi A, Singh P, Hibberd JM. 2021. Using breeding and quantitative genetics to understand the C₄ pathway. *Journal of Experimental Botany* 73: 3072– 3084.
- Talukder, P., Sinha, B., Biswas, S., Ghosh, A., Banerjee, A., and Paul, S. (2024). A study on the prospect of converting C₃ Plants into C₄ plants. *Biocatalysis and Agricultural Biotechnology*, 103191.
- Tao, Y., Zhao, X., Mace, E., Henry, R., and Jordan, D. (2019). Exploring and exploiting pan-genomics for crop improvement. *Molecular plant*, 12, 156-169.