The University Of Sheffield.

Phylogenomic investigations of the photosynthetic and genomic diversification of Molluginaceae

by

Lamiaa Adnan Munshi

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

The University of Sheffield
Faculty of Sciences
Department of Animal and Plant Sciences

January 2021

List of contents

List of tables

List of figures

Acknowledgements

My sincere appreciation goes to my principal supervisor Dr. Pascal-Antoine Christin for his generous support, guidance, invaluable advice throughout this project. I have been extremely lucky to have a supervisor who cared so much about my work. His active supervision and unflinching excitement during this work made the research process very enjoyable. Not to mention his unsurpassed knowledge of evolution, phylogenetics, and bioinformatics.

I am thankful to Prof. Colin Osborne, very appreciative of all the valuable discussions and helpful comments that have dedicated to my research.

I would like to express my deep thank to Dr. Jill Olofsson, Dr. Luke Dunning, Dr. Matheus Bianconi, Dr. Daniel Wood, and Dr. Emma Curran for their unflinching encouragement and generously assistance in laboratory work, bioinformatics, and statistical analysis. I would like to thank also Dr. Jose J. Moreno-Villena for advising me and sharing data and ideas on transcriptome work. Thanks are also offered to Dr. Marjorie Lundgren and Dr. Emanuela Samaritani for their valuable help and assistance in (Li-cor) photosynthesis measurements and anatomical preparation.

I am grateful to all members of Christin, Osborne, and Nadeau labs for sharing their great perspectives on ecology, physiology, and genomics in our weekly lab meetings. I would also like to thank my colleague Sam hibdige, look forward to seeing him go on to success in his PhD.

I am particularly grateful to the Ministry of Higher Education, Saudi Arabia for generous funding this research, and special thanks go to the University of Sheffield for hosting me throughout, and to the staff of the Department of Animal and Plant Sciences for their kindly support.

I am greatly indebted and appreciate very much to my beloved father, brother, and all my family, for their support, huge sacrifices throughout the study especially after the death of my mother, and to pursue encouraging me to get this project done.

Last but not least, I wish to express my sincere thanks to all those who have in one way or another helped me in making this study a success.

Abstract

During evolution, organisms acquired many traits that allowed them to colonize almost all possible environments. Some of these tools result from the coordinated action of numerous cells and enzymes and are consequently referred to as complex traits. Due to their apparent complexity, the evolutionary paths leading to the emergence of complex traits remain incompletely understood. C_4 photosynthesis is a derived physiology that boosts productivity in warm and dry habitats. It results from the coordinated action of multiple anatomical and biochemical components and therefore is an excellent example of a complex trait. The C_4 trait has evolved more 62 times independently, offering natural replicates to understand the processes leading to the emergence of complex traits. In particular, the Molluginaceae encompasses multiple C_4 species, species using the ancestral C_3 type, and C_3-C_4 species with intermediate characters. In this work, this study system is used to evaluate the evolutionary dynamics of the leaf anatomy, genes for C_4 enzymes, and chloroplast genomes during the diversification of photosynthetic types. Using a phylogenetic framework, I first compared the leaf anatomies of Molluginaceae species with distinct photosynthetic types, showing that the emergence of C_4 leaf anatomy in Molluginaceae resulted from the constraint combination of characters that individually existed in C_3 ancestors. I then used comparative transcriptomes to show that the C_4 biochemistry emerged from the upregulation of multiple genes, specifically in the C_4 lineages. Finally, I compare the chloroplast genomes of members of the family, showing that a large subgroup of Molluginaceae containing C_4, C_3-C_4 and C_3 species is characterized by sustained increased rates of chloroplast evolution. Overall, this work brings new insights into the events that led to C_4 emergence of Molluginaceae. C_4 leaf anatomical components evolved early in the history of the group and were later recurrently co-opted for transitions to C_4 photosynthesis through gene upregulation and positive selection adapting the encoding enzymes. C_3-C_4 species only occasionally bridge the gap to C_4 anatomy, but in the case of Molluginaceae, they do not decrease the distance to a C_4 biochemistry. The function of C_3-C_4 intermediates as evolutionary facilitators therefore depends on the details of their photosynthetic machinery and their eco-physiological strategies. Transcriptome data moreover suggest that hybridization might have contributed to independent transitions to C_4 photosynthesis in the group. In the future, analyses of complete nuclear genomes will be needed to precisely assess how such processes affected the evolutionary dynamics in the group. My work forms a solid foundation on which to build such efforts.

General introduction

1.1 General introduction

1.1.1 Complex trait evolution.

Since the first evolutionary theories, the processes underlying the origins of traits of apparent complexity have intrigued biologists, including Darwin and Galton (Mayr, 1982). Indeed, some phenotypes, such as the camera eye, some colour patterns and the ability to fly, are genetically complex, gaining their function only when multiple morphological and/or biochemical components act together (Gatesy and Dial, 1996; Taub, 2000; Kozmik et al., 2008). The origins of such traits have been studied by comparing individuals, populations, and species (Darwin, 1859; Haldane, 1915; Fernald, 2006; Barret and Schlüter, 2008), but also in experimental evolution settings.

Without a full understanding of the genetic origins of complex traits, different studies have shown that most novel complex traits evolve via the co-option of pre-existing structures and/or enzymes, through the gradual modification of the constituents of the trait (Panganiban, 1997; True and Carroll, 2002; McLennan, 2008; Visscher et al., 2008). Understanding complex trait origins therefore requires reconstructing the order of character acquisition and genetic modifications that cumulatively lead to the emergence of these novel adaptations (Stapley et al., 2010; Ekblom and Galindo, 2010; Christin et al., 2011). The alterations of components for complex traits have thus been studied at several levels, from genes to biochemical pathways, physiology and external phenotypes (Marazzi et al., 2012; Edwards and Donoghue, 2013; Huang et al., 2016; Moreno-Villena et al., 2018). Our understanding of many complex traits remains, however, incomplete as the extinction of previous intermediate states blurs the evolutionary signal and the developmental processes responsible for the adaptive phenotype are often poorly understood (Ray, 1991; Yokoyama and Radlwimmer,

2001; Lenski et al., 2003; Christin et al., 2011; Heckmann et al., 2013). For traits that emerged over large evolutionary scales, the problem of complex trait origins is better addressed using phylogenetic approaches based on an excellent understanding of the relationships among lineages with variants of the complex trait in question.

1.1.2 Using phylogenetics to study complex traits.

Phylogenetic trees were originally based on individual proteins or sequences of DNA, but the advent of throughput sequencing has allowed the inference of phylogenomic trees based on numerous markers spread across both the organelle and nuclear genomes, providing robust phylogenetic hypotheses (Kayal et al., 2013; Nadeau et al., 2013; Zeng et al., 2014;Larridon et al., 2020). Originally developed with the sole goal of establishing the relationship among species, phylogenetic tools have subsequently been widely adopted by the fields of taxonomy and systematics. Because they allow inferring past events, estimating divergence times, and implementing comparative methods, phylogenetics have later become an indispensable tool in evolutionary biology (Yang and Rannala, 2012; Koboldt et al., 2013; van Dijk et al., 2014; Davis et al., 2016) and a crucial component of many studies in ecology (Thornhill et al., 2016; 2017; Baeckens et al., 2017; Allen et al., 2019). Of particular interest for this thesis, phylogenetics have been successfully used to reconstruct the events leading to the origins of complex traits (Zapata and Jimenez, 2012; Christin et al., 2015; Lauterbach et al., 2017).

Comparative studies are typically performed across a set of species that differ in their phenotype. This approach can be extremely powerful to infer events spread over long evolutionary periods, but species comparisons inflate the amount of inferred changes. Indeed,

some changes that accumulated independently of the trait of interest will coincide with the studied transitions and will be erroneously interpreted as having contributed to the emergence of the trait (Heyduk et al., 2019). In addition, some of the differences among species represent adaptations that happened after the trait of interest originally emerged (Heyduk et al., 2019). These problem can be reduced by comparing closely related taxa that differ in the trait of interest (Dunning et al., 2019). In particular, analyses of intraspecific variants provide powerful systems to elucidate the genetic and selective drivers of novel traits (e.g. Roesti et al., 2012; Villoutreix et al., 2020). For traits that emerged across multiple speciation events, groups of closely related species or species complexes that present variation and putative intermediate stages offer outstanding study systems (Whittall and Hidges, 2007; Dunning et al., 2017; Larridon et al., 2018). Such groups are amenable to phylogenetic approaches, potentially helping understand how complex traits evolved.

1.1.3 C_4 photosynthesis as an ecologically relevant complex trait.

Oxygenic photosynthesis is the most frequent autotrophic metabolic pathway on Earth. This pathway has we know it today originated with the evolution of Ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) more than 2.7 billion years ago (Nisbet et al., 2007; Christin and Osborne, 2013). In photosynthesis, light energy is captured during the light-dependent phase and temporarily stored as ATP and NADPH. The long-term storage of this energy as sugars happens during the light-independent phase, in which atmospheric carbon is fixed by Rubisco and used in the energy-dependent reactions that form the Calvin-Benson-Bassham cycle (Hatch, 1987; Nelson, 1989; Firooznia, 2007). This pathway allows inorganic carbon to enter the biosphere **(**Tabita et al., 2007; 2008).

Rubisco evolved with a tendency to confuse the CO_2 and O_2 substrate, which are both feature-less molecules (Tcherkez et al., 2006). Because O_2 was almost absent from the atmosphere when Rubisco evolved this lack of specificity for CO_2 not counter selected (Christin and Osborne, 2013). However, following the great oxygenation of the atmosphere and continued decreased of CO_2 concentrations, this dual affinity led to the fixation by plants of O_2. This oxygenase reaction of Rubisco produces toxic compounds that need to be recycled by a costly pathway named photorespiration (Ogren, 1984). Plants that fix atmospheric CO_2 directly by Rubisco, named C_3 plants, consequently undergo a drastic decrease of efficiency in conditions where CO_2 availability is reduced. Under the low-CO_2 atmosphere that prevailed over the last 30 million years (0.036% CO_2 and 21% O_2 in current atmosphere), photorespiration is indeed estimated to lead to an up to 50% reduction of carbon fixation. (Ogren, 1984; Leegood, et al., 1995; Miyao, 2003; Yadav and Mishra, 2019). CO_2 solubility decreases faster than that of O_2 with increasing temperature, and Rubisco CO_2:O_2 specificity decreases with temperature. In addition, salinity, drought and low humidity promote stomatal closure to limit water losses, which quickly decreases the internal CO_2 concentration around Rubisco. CO_2 availability is therefore especially low in warm and dry environments, leading to photorespiration that strongly limits C_2 plant productivity. C_2, photosynthesis represents the evolutionary solution to these environements (Hatch, 1987; Ehleringer and Monson 1993; Ehleringer et al., 1997; Sage, 2004; Edwards et al., 2010).

The C_4 process results from a coordinated set of anatomical, physiological, and biochemical modifications that together concentrate CO_2 within the leaf before its fixation by Rubisco (Slack and Hatch, 1967; Christin et al., 2009; 2011; Sage et al., 2014). In C_4 plants, the initial fixation of atmospheric CO_2 (in the form of HCO_3^-) is not catalyzed by Rubisco, but by the enzyme phosphoenolpyruvate carboxylase (PEPC) that does not have any affinity for

O_2 (Webster et al., 1975; Iglesias et al., 1986; Edwards et al., 2010). This reaction occurs in a leaf compartment in direct contact with the atmosphere, which generally consist of mesophyll cells. The resulting four-carbon oxaloacetate is then transformed into another four-carbon acid (malate or asapartate), which is then transported into a different leaf compartment. This second compartment is in most C_4 plants the bundle sheath cells that surround the veins. The four-carbon acid is then decarboxylated in these cells to release CO_2 that will feed Rubisco, segregated in this compartment in C_4 plants. Because bundle sheath cells are nested within the leaf and surrounded by mesophyll cells, they are not in direct contact with the atmosphere and its O_2. The constant pumping by the C_4 cycle of CO_2 into these cells will thus dramatically increase the relative CO_2:O_2 concentration around Rubisco and effectively suppress photorespiration (Dengler and Nelson, 1999; Sage, 2001; von Caemmerer and Furbank, 2003). C_4 photosynthesis is therefore a complex trait that boosts carbon assimilation in warm and dry conditions (Christin et al., 2013; Sage, 2015; Atkinson et al., 2016). Despite being used by only 3% of extant plants, C_4 photosynthesis is nowadays responsible for one fifth to one quarter of global terrestrial primary production (Ehleringer et al., 1997; Sage et al., 1999; 2004).

1.1.4 C_4 biochemical components.

All C_4 plants use PEPC, which defines the pathway (Kellogg, 1999). The decarboxylation of the C_4 acid in the bundle sheath cells produce a three carbon compound, which diffuses back into the mesophyll cells (Fig. 1.1; Hacth and Slack, 1966; Hatch, 1987). It is then regenerated into PEP, the substrate of PEPC. This reaction is catalyzed by pyruvate, phosphate dikinase (PPDK), and the enzyme is used with PEPC by all C_4 plants analysed so far (Fig. 1.1; Hatch,

1987; Kanai and Edwards, 1999; Furbank, 2011). The other biochemical components of the C_4 pathway vary among C_4 lineages (Hatch and Slack, 1966; Bowyer and Leegood, 1997; Furbank et al., 2011; Long and Spence, 2013).

The decarboxylation of the C_4 acid in the bundle sheath can be catalysed by one of three decarboxylating enzymes; NAD-malic enzyme (NAD-ME), NADP-malic mnzyme (NADP-ME) or PEP carboxykinase (PCK). The NADP-ME is the most frequent among C_4 plants (Sage et al., 1999). In C_4 species relying on this enzyme, oxalacetate is typically converted in the mesophyll cells to malate by the enzyme NADP-malate dehydrogenase (NADP-MDH; Fig. 1.1). This malate is then transported to the bundle sheath and decarboxylated to produce CO_2 and pyruvate. This pyruvate is transported back to the mesophyll cells, and used by PPDK to regenerate PEP (Fig. 1.1). In C_4 plants relying on NAD-ME, oxalacetate is typically transformed in the mesophyll cells into aspartate by the enzyme aspartate aminotransferase (ASP-AT). This aspartate is then transported to the bundle sheath cells, where it is transformed back into oxaloacetate by ASP-AT. This oxaloacetate is subsequently transformed into malate by the enzyme NAD-malate dehydrogenase (NAD-MDH). Decarboxylation of this malate by NAD-ME again produces CO_2 and pyruvate. Pyruvate is then transformed into alanine by the enzyme alanine aminotransferase (ALA-AT). One transported back to the mesophyll cells, alanine is transformed back into pyruvate, which is again used by PPDK to regenerate PEP. The decarboxylating enzyme PCK is generally used in addition to NAD-ME or NADP-ME. Its substrate is oxaloacetate, which can be provided to the bundle sheath cells via an aspartate shuttle, as described for the NAD-ME cycle. Unlike NAD-ME and NADP-ME, the decarboxylation by PCK produces CO_2 and PEP, which can in theory directly diffuse back to mesophyll cells and be used by PEPC.

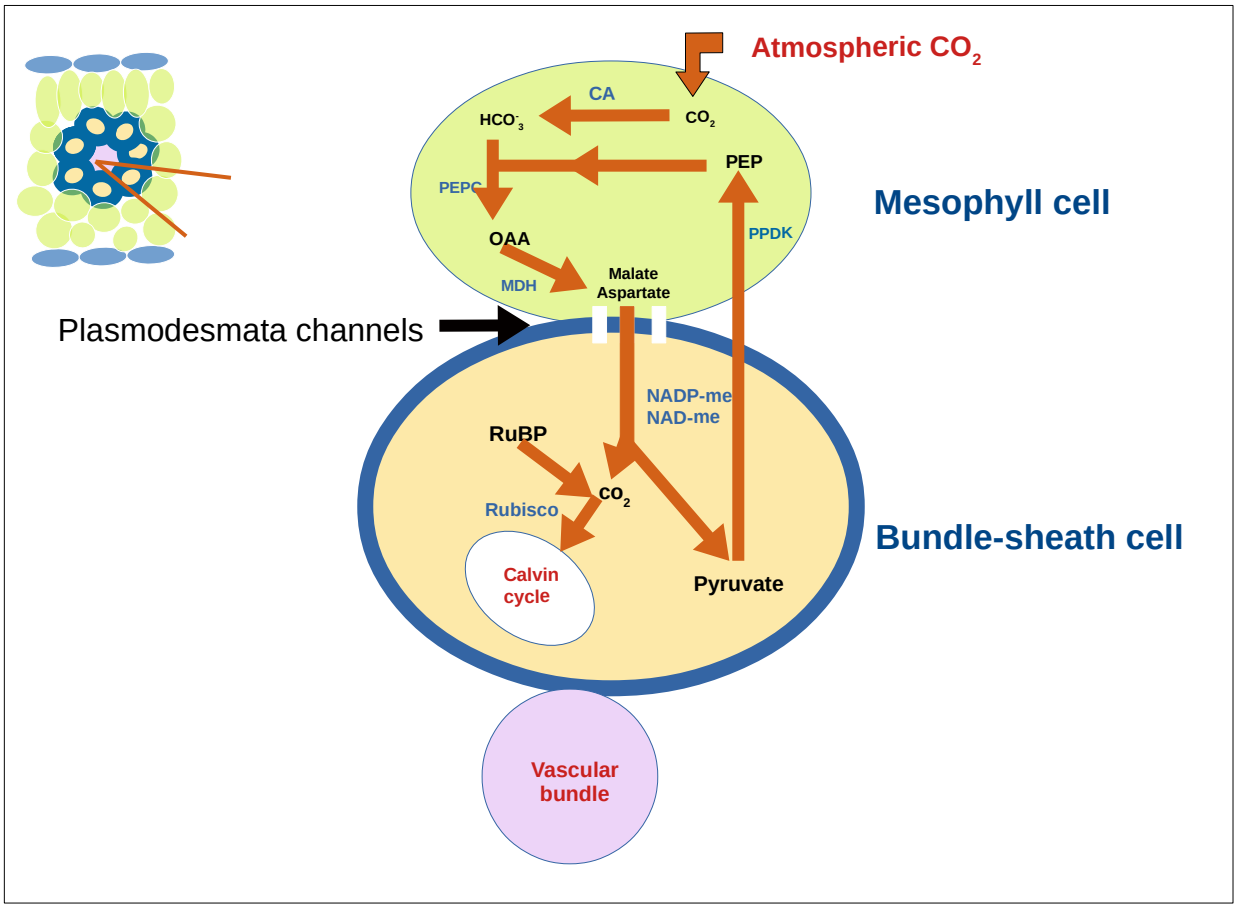

Figure 1.1 C_4 carbon fixation cycle.

Simplified diagram of main biochemical reactions of the C_4 pathway. Two cellular compartments (mesophyll and bundle sheath cells) are connected by plasmodesmata channels (in white). Carbonic anhydrase (CA) converts atmospheric CO_2 into bicarbonate (HCO^-_3). Phosphoenolpyruvate carboxylase (PEPC) catalyzes the fixation of HCO^-_3 on PEP to produce oxaloacetate (OAA). The OAA is the coverted to to malated (by the enzyme malate dehydrogenase - MDH) or aspartate. This acid then diffuses into the inner compartment via plasmodesmata. Therein, it is decarboxylated by either the NADP-malic enzyme (NADP-me) or the NAD-malic enzyme (NAD-me). The released CO_2 is finally fixed by ribulose1,5-bisphosphate carboxylase (Rubisco), which represents the entry point of the Calvin cycle. The remaining pyruvate returns to the outer compartment where PEP is regenerated by pyruvate phosphate dikinase (PPDK).

While the three decarboxylating enzymes were originally seen as defining three distinct C_4 subtypes, it is now accepted that they can occur in in all groups of C_4 plants, in varying proportions (Furbank, 2011). Some C_4 plants rely solely on either the NAD-ME or NADP-ME enzyme, while others use in addition different amounts of PCK (Wang et al., 2014). More recently, transcriptome analyses have suggested that some C_4 plants use both NAD-ME and NADP-ME enzymes (Washburn et al., 2017), while some other C_4 plants might use exclusively PCK (Bräutigam et al., 2018). The C_4 biochemistry can therefore be seen as different combinations of aspartate and malate shuttle associated with distinct set of enzymes. Besides the core C_4 enzymes listed above, carbonic anhydrase is supposed to play in important role in the provision of HCO_3^- to PEPC, although it is not necessary for all C_4 plants (Studer et al., 2014). In addition, many enzymes play accessory or regulator roles in the C_4 cycles, and the transport of metabolites is thought to involve additional proteins, some of them remain to be identified (Jiao and Chollet, 1989; Bräutigam et al., 2011; Pick et al., 2011; Schlüter et al., 2016). The C_4 biochemical pathway therefore results from the complex interaction of numerous enzymes.

1.1.5 C_4 anatomical components.

The C_4 trait relies on the spatial segregation of atmospheric carbon fixation by PEPC and the secondary carbon refixation by Rubisco. In some cases, the spatial segregation is performed within a single cell in which different compartments are created (Voznesenskaya et al., 2001; Edwards et al., 2004). In most cases however, the segregation happens among distinct cells; the mesophyll and bundle sheath cells (Fig. 1.2). These two cell types were already present in the C_3 ancestors, but were responsible for different functions. Indeed, mesophyll cells of C_3

plants are largely responsible for photosynthesis and carbon fixation, while C_3 bundle sheath cells are mostly responsible for the control of exchanges between the mesophyll and the veins (Leegood, 2008; Griffiths et al., 2013). Besides the change of function of these, the C_4 trait requires a set of leaf morphological attributes (Bruhl and Wilson, 2007; Garner et al., 2016; Schüssler et al., 2017), which have collectively been referred to as 'Kranz anatomy' (Hattersley, 1984; Dengler and Nelson, 1999; Voznesenskaya et al., 2001; Lundgren et al., 2014).

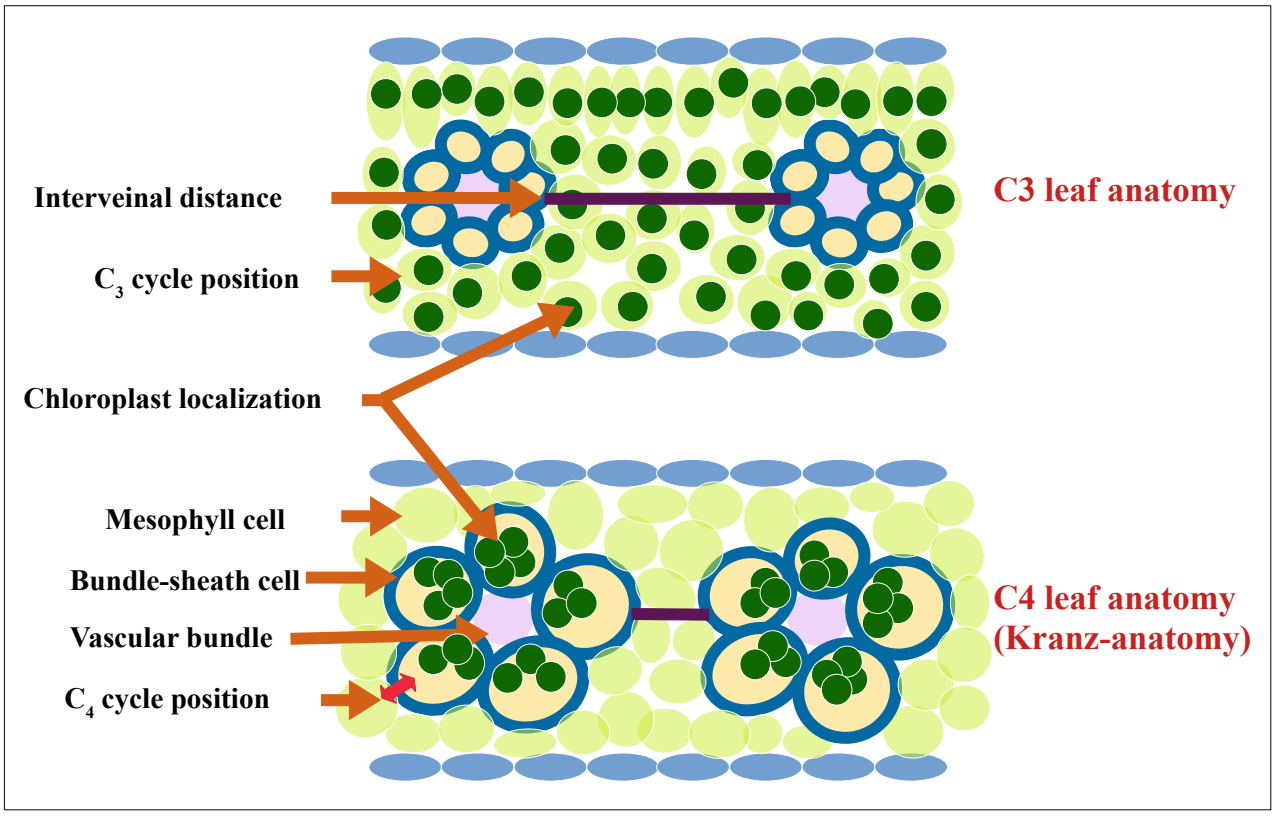

Figure 1.2 Simplified diagrams showing the main anatomical differences between C_3 and C_4 plants.

The leaf arrangement of C_4 plants was first described long before the discovery of C_4 photosynthesis (Haberlandt, 1884), and manifests itself by the appearance of rings (‘Kranz’ in German) in cross sections. These correspond to the ring of bundle sheath cell surrounding the veins that are made dark green by the concentration of chloroplasts, surrounded by rings of lightly coloured mseophyll cells with less chloroplasts. It was later established that these ‘Kranz’ leaves present a number of features, which are essential to the function of the C_4 cycle (Sage 2004; Lundgren et al., 2014; Stata et al., 2014): a) bundle sheath cells are isolated from the atmosphere and contain a high concentration of chloroplasts containing Rubisco, b) mesophyll and bundle sheath cells are separated by a short distance, which facilitates intercellular metabolite diffusion, and c) the bundle sheath tissue represents a larger proportion of the leaf to accommodate a large amount of chloroplasts.

The isolation of bundle sheath cells from the atmosphere results from their position in the middle of the leaf, coupled with limited contact with intercellular airspaces and in some cases cell walls reinforced with a layer of suberin or extra layers of cells around the bundle sheath to limit gas leakage (Hattersley and Browning, 1981; McKown and Dengler, 2007; von Caemmerer et al., 2014). The small distance between mesophyll and bundle sheath cells is best described in mesophyll cells of C_4 plants being a maximum of one cell distant from the bundle sheath or there being a maximum of four mesophyll cells in between consecutive bundle sheaths (Hattersley and Watson, 1975). This can be achieved by a proliferation of veins or a decrease of mesophyll cells in between veins present in the C_3 ancestors (Lundgren et al., 2014). Finally, the increased percentage of the leaf occupied by bundle sheath tissues can results from the proliferation of veins or the increase of individual bundle sheath cells (Hatterlsey, 1984; Soros and Dengler, 2001; McKown and Dengler, 2007; Christin et al., 2013; Lundgren et al., 2014; 2019). The identity of the tissues co-opted for C_4 leaf anatomy

and the alterations in leaf developments involved in the evolution of C_4 photosynthesis varies among C_4 lineages (Dengler et al., 1985; Soros and Dengler, 2001; Kadereit et al., 2003; McKown and Dengler, 2007; Muahidat et al., 2007; Voznesenskaya et al., 2007; Christin et al., 2013). With some exceptions (McKown and Dengler 2007; Dunning et al. 2017), the exact evolutionary path to C_4 leaf evolution is not known because of a lack of detailed studies focusing on closely related plants with distinct photosynthetic types.

1.1.6 C_4 phylogenetic patterns and evolution.

C_4 photosynthesis represents one of the best examples of a complex trait that evolved convergently (Sage 2004; Morris 2006). Indeed, this trait that results from coordinated action of numerous anatomical and biochemical components evolved more than 60 times independently, in various groups spread across the phylogeny of flowering plants including both dicots and monocots (Sage et al., 2011), and even in unicellular algae (Reinfelder et al., 2004). In the phylogenetic trees of several plant families, including Poaceae, Cyperaceae, Amaranthaceae, Chenopodiaceae and Molluginaceae, multiple C_4 lineages are separated by C_3 taxa (Sinha and Kellogg, 1996; Soros and Dengler, 2001; Bruhl and Wilson, 2007; Besnard et al., 2009; Christin et al., 2011; Schüssler et al., 2017). While these patterns could be interpreted as the signature of a few C_4 origins followed by reversals to the ancestral C_3 state, differences in the C_4 anatomy and biochemistry points to independent realizations of the C_4 phenotype (Sinha and Kellogg, 1996; Kellogg, 1999; Kadereit et al., 2003; Muhaidat et al., 2007; Voznesenskaya et al., 2007; Christin et al., 2011).

The repeated origins of C_4 photosynthesis could be explained by the possibility to evolve the derived trait via a single genetic change (Westhoff and Gowik, 2010; Zhu et al., 2010). However, the hypothesis of a C_4 master switch was falsified by the polygenic nature of the trait and its numerous variations. Instead, the multiple biochemical amd morphological changes required to evolve from the ancestral C_3 type to C_4 photosynthesis are likely to have gradually accumulated, in a path involving numerous intermediate stages (Hatch and Slack, 1966; Ku, 1983; Hylton et al., 1988; Voznesenskaya et al., 1999; Christin et al., 2013; Schlüter and Weber, 2016; Lundgren and Christin, 2017; Dunning et al., 2019). The repeated origins of C_4 photosynthesis and their clustering in some groups have even been attributed to the existing of such intermediate stages during the diversification of some groups (Sage, 2001; Christin et al., 2011; 2013). Indeed, an ancestral state intermediate between C_3 and C_4 photosynthesis could then easily be recurrently co-opted by multiple descendants, leading to several C_4 origins in some groups. Of particular interest in this context are extant species presenting characteristics that are intermediate between those typical of C_3 and C_4 plants.

1.1.7 The importance of C_3-C_4 intermediates for C_4 evolution.

The first species reported to have characteristics intermediate between C_3 and C_4 photosynthesis was the eudicot *Mollugo verticillata* (Molluginaceae; Kennedy and Laetsch, 1974). Such species were subsequently referred to as C_3-C_4 intermediates and reported in different lineages of plants (Hattersley et al., 1986; Rajendrudu et al., 1986; McKown et al., 2005; Griffiths et al., 2013; Lundgren et al., 2016). Several properties of the plants can be described as intermediate between C_3 and C_4, and C_3-C_4 plants can be described along a continuum of C_4-like characters. Some species present Kranz-like or C_4-like anatomy without

increased activity of C_4 enzymes (Ku et al., 1983; Moore, 1989). These plants, subsequently referred to as type I intermediates are however less sensitive to photorespiration, thanks to a weak carbon-recycling mechanism (Hunt et al., 1987; Sage et al., 2012). In these C_3-C_4 intermediates, the C_4 cycle occurs in both mesophyll and bundle sheath cells. The product of O_2 fixation in the mesophyll cells are however transported to the bundle sheath cells, where the final step of photorespiration is performed. The CO_2 thereby released is trapped in the bundle sheath cells where it is recaptured by Rubisco. Such type I C_3-C_4 plants have typically reduced CO_2-compensation points (Vogan et al., 2007; Vogan and Sage, 2011). In addition to C_4-like anatomical characters, type II C_3-C_4 intermediates show increased activities of C_4 enzymes, which are responsible for part of the carbon fixation (Moore, 1989). They show further reductions of the CO_2-compensation point, which can approach that of C4 plants (Vogan and Sage, 2011; Lundgren et al., 2016). They do however not exhibit the higher water-use and nitrogen-use efficiencies of C_4 plants (Vogan and Sage, 2011).

The evolutionary status of C_3-C_4 plants is debatable (Kadereit et al., 2017), and it would be difficult to prove that they represent the ancestral state on the road to extant C_4 species (Hancock and Edwards, 2014). However, they are widely considered as proxies for the ancestral intermediate stages on the road from C_3 to C_4 evolution and have consequently been widely used to developed models of C_4 origins (Hylton et al., 1988; Monson and Moore, 1989; Sage et al., 1999, 2012; Vogan et al., 2007; Christin et al., 2013; Lundgren et al., 2014; 2016). Both type I and type II C_3-C_4 have enlarged bundle sheath to mesophyll volume ratios to accommodate more chloroplasts in the bundle sheath. Compared to C_3 plants, they moreover tend to have significantly higher numbers of organelles in the bundle sheaths and higher vein density (Brown and Hattersley, 1989; Sage et al., 1999; Sedelnikova et al., 2018). These anatomical changes are accompanied by a shift of the activity of Rubisco from

mesophyll cells to bundle sheath cells (Stata et al., 2019; Kümpers et al., 2017). These properties are hypothesized to facilitate later transitions to a full C_4 physiology (Sage, 2004), and current models hypothesize a gradual increase of photosynthetic efficiency along an evolutionary path that goes from C_3 plants with some C_4-like characters to type I C_3-C_4 intermediates, type II C_3-C_4 intermediates and finally C_4 plants (Sage, 2004; Heckmann et al., 2013; Mallmann et al., 2014). In many aspects, C_3-C_4 intermediates bridge the gap between the C_3 and C_4 phenotypes, but also their ecology (Ehleringer et al., 1997; Sage et al., 1999; Vogan et al., 2007), and their analysis can therefore inform the origin of the C_4 complex trait.

1.1.8 Genetic changes during C_4 evolution.

The genetic determinants of C_4 leaf anatomy are mostly unknown, but the core enzymes of the C_4 pathway have been known for half a century, since the first description of the C_4 biochemical pathway by Hatch and Slack (1966). All known C_4 enzymes already existed in C_4 ancestors (Aubry et al., 2011), and C_4 evolution therefore involved their co-option for the new C_4 trait. It was initially proposed that the recruitment of these genes followed gene duplication and neofunctionalization (Monson, 1999; Sheen 1999). However, phylogenetic analyses suggest that many C_4 enzymes acquired their C_4 function without gene duplication (Christin et al., 2007; 2009), and the number of genes for enzymes linked to the C_4 pathway is not consistently higher in C_4 than in C_3 genomes (Williams et al., 2012). The importance of gene duplication might therefore vary among C_4 enzymes, as a function of their function in C_3 plants.

Compared to non-C_3 isoforms, forms of the enzymes co-opted for C_4 photosynthesis showed enhanced activity and altered kinetic properties (Monson and Moore, 1989; Nelson and Langdale, 1989; Miyao, 2003; Sage, 2004; Taylor et al., 2010). The recent development of high-throughput sequencing has enabled the comparison of complete transcriptomes of C_3 and C_4 species (Bräutigam et al., 2011; Pick et al., 2011; Schlüter et al., 2016; Dunning et al., 2017; 2019; Moreno-Villena et al., 2018). For each C_4 enzyme, the paralogs ancestrally most abundant in the leaves seem to have been preferentially co-opted (Emms et al., 2016; Moreno-Villena et al., 2018). Genes for several enzymes were reported to increase in expression in C_3-C_4 intermediates of the genus *Flaveria*, potentially to rebalance nitrogen between mesophyll and bundle sheath cells (Mallmann et al., 2014). The development of a complex C_4 physiology then required massive overexpression of C_4-related genes (Bräutigam et al., 2011; Moreno-Villena et al., 2018). The few C_4 enzymes analyzed do not vary significantly in terms of their kinetics between C_3 and C_3-C_4 species and most kinetic changes seem limited to the C_4 state (Engelmann et al., 2003; Gowik et al., 2006; Phansopa et al., 2020). The genetic changes linked to C_4 evolution have however been studied so far in a limited number of plant lineages.

1.1.9. Molluginaceae as a system to study C_4 evolution.

Most C_4 species belongs to the large families of monocots that are grasses (Poaceae) and sedges (Cyperaceae). However, many groups of eudicots include multiple C_4 lineages, in some cases with C_3-C_4 intermediate in addition to C_4 species (Sage et al., 2011). In particular, the order Caryophyllales includes a minimum of 23 species and four families with C_3-C_4 plants (Chenopodiaceae, Molluginaceae, Nyctaginaceae, and Portulacaceae; Sage et al.,

2011). Among those, Molluginaceae (carpet weed family) comprise 13 genera and about 120 species (Mabberely, 2017). It consists of herbs or dwarf shrubs with fleshy or succulent leaves, distributed mainly in the tropical and subtropical regions, but widespread in southern Africa. In the overview of Molluginaceae by Endress and Bittrich, (1993), 13 genera were included in the following order: *Corbichonia* Scop., *Limeum* L., *Macarthuria* Hügel ex Endl., *Psammotropha* Eckl. & Zeyh., *Adenogramma* Rchb., *Glischrothamnus* Pilger, *Mollugo* L., *Glinus* L., *Hypertelis* E.Mey. ex Fenzl, *Pharnaceum* L., *Suessenguthiella* Friedrich, *Coelanthum* E.Mey. The circumscription of the family has however changed following molecular phylogenetic analyses and analyses of pigments, leading to some genera moved to the families Macarthuriaceae, Kewaceae, and Corbichoniaceae (Brockington et al., 2011; Christenhusz et al., 2014; Thulin et al., 2016). In addition, many genera within the family have been renamed based on molecular studies (Christin et al., 2011; Thulin et al., 2016).

The Molluginaceae encompass a diversity of photosynthetic types among relatively few species, including C_3 and C_4 plants and the first described C_3-C_4 intermediates (Sayre and Kennedy, 1977; Christin et al., 2011; Lundgren and Christin, 2017). In addition to *Mollugo verticillata*, which could possess intraspecific variants of photosynthetic types (Sayre and Kennedy, 1979), *Paramollugo nudicaulis* (previously *Mollugo nudicaulis*) and *Hypertelis spergulacea* represent two additional C_3-C_4 lineages (Kennedy et al. 1980; Christin et al., 2011). The three C_4 species have been moved to the *Hypertelis* genus, which they share with the C_3-C_4 *H. spergulacea* (Thulin *et al.,* 2016). Two of them were previously analysed, and the history of their genes for PEPC suggested they evolved C_4 photosynthesis from a common ancestor with some C4-like characters (Christin et al., 2011).

Thanks to the high diversity of photosynthetic types within a small family, Molluginaceae represents an exciting system to understand C_4 evolution. In addition, many of its species are small with a short generation time, and therefore amenable to experimental evolution. However, the group remains poorly studied and has not been the subject of detailed anatomical, biochemical or genomics work.

1.2 Thesis aims and structure

This PhD study aims to use Molluginaceae as a study system to retrace the events leading to C_4 evolution at the anatomical, biochemical (gene expression) and genomic levels. My work is divided into three inter-related chapters, which explore different aspects of the evolutionary origins of the photosynthetic diversity in Molluginaceae. They all rely on a solid phylogenetic framework to understand how the complex C_4 trait could evolve recurrently during the diversification of plant family.

The first chapter evaluates the leaf anatomical diversity that exists within Molluginaceae and the events that led to the emergence of C_4 leaf anatomy. Multiple accessions are sampled for several C_3, C_3-C_4 and C_4 species, and their leaf properties are compared to determine the higher-level characteristics that are specific to the C_4 type and the changes in cell size and number responsible for these characteristics. The traits are then mapped onto a species phylogeny to determine the order in which anatomical changes happened. In particular, I was interested in 1) determining which changes occurred before, during, and after the evolution of C_4 photosynthesis and 2) testing the hypothesis that C_3-C_4 intermediates bridged the gap between C_3 and C_4 leaf anatomies in the group.

The second chapter uses comparative transcriptomics to reconstruct the history of changes in gene expression and coding sequences underlying transitions to a C_4 biochemistry in Molluginaceae. The transcript abundance of all genes encoding enzymes related to the C_4 cycles is quantified in the leaves of multiple accessions of C_3, C_3-C_4 and C_4 species. The data allow the identification of the genes used by each C_4 species and their comparison with C_3 and C_3-C_4 species. In addition, tests of positive selection retrace episodes of adaptive evolution of the coding sequences. Similarly to Chapter 1, the data are used 1) to identify the genes in gene

expression that occurred before, during and after C_4 evolution and 2) to test the hypothesis that C_3-C_4 intermediates bridge the gap between C_3 and C_4 biochemistries in the group.

The third chapter uses phylogenomics to infer the relationships among Molluginaceae species. Using whole sequence data, the chloroplast genomes of multiple accessions with distinct photosynthetic types were assembled and compared. In particular, this work aimed to test whether the patterns previously detected on selected chloroplast markers (phylogenetic relationships and variation in the rate of molecular evolution) could be expanded to the whole chloroplast genomes. Together, my work elucidates the historical processes that led to the diversification of Molluginaceae, including their C_4-related leaf anatomy and biochemistry, and their chloroplast genomes.

1.3 References

Allen, J.M., Germain-Aubrey, C.C., Barve, N., Neubig, K.M., Majure, L.C., Laffan, S.W., Mishler, B.D., Owens, H.L., Smith, S.A., Whitten, W.M., Abbott, J.R., Soltis, D.E., Guralnick, R. and Soltis, P. S. 2019. Spatial phylogenetics of Florida vascular plants: The effects of calibration and uncertainty on diversity estimates. *Iscience.* **11**, pp.57–70.

Atkinson, R.R., Mockford, E.J., Bennett, C., Christin, P.A., Spriggs, E.L., Freckleton, R.P., Thompson, K., Rees, M. and Osborne, C.P. 2016. C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants.* **2**(5), 16038.

Aubry, S., Brown, N.J., and Hibberd, J.M. 2011. The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. J*ournal of Experimental Botany.* **62**(9), pp.3049–3059.

Baeckens, S., Van Damme, R. and Cooper, W. E. 2017. How phylogeny and foraging ecology drive the level of chemosensory exploration in lizards and snakes. *Journal of Evolutionary Biology.* **30**(3), pp.627–640.

Besnard, G., Muasya, A. M., Russier, F., Roalson, E.H., Salamin, N., and Christin, P.A. 2009. Phylogenomics of C_4 photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. *Molecular Biology and Evolution.* **26**(8), pp.1909–1919.

Bowyer, J.R., Leegood, R.C., 1997. Photosynthesis. In: Dey, P.M., Harbone, J.B. (eds). Plant biochemistry. *Academic Press: London.*. pp.49–110.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Weber, A. P. 2011. An mRNA blueprint for C_4 photosynthesis derived from arative transcriptomics of closely related C_3 and C_4 species. *Plant Physiology.* **155**, pp.142–156.

Bräutigam, A., Schlüter, U., Lundgren, M., Flachbart, S., Ebenhöh, O., Schönknecht, G., Christin, P., Bleuler, S., Droz, J., Osborne, C., Weber, A. and Gowik, U. 2018. Biochemical Mechanisms Driving Rapid Fluxes in C_4 Photosynthesis. *BioRxiv* doi: 10.1101/387431

Brockington, S.F., Walker, R.H., Glover, B.J., Soltis, P.S. and Soltis, D.E. 2011. Complex pigment evolution in the Caryophyllales. *New Phytologist*. **190**(4), pp.854–864.

Brown, R.H. and Hattersley, P.W. 1989. Leaf anatomy of C_3-C_4 species as related to evolution of C_4 photosynthesis. *Plant Physiology*. **91**(4), pp.1543–1550.

Bruhl, J. and Wilson, K. 2007. Towards a comprehensive survey of C_3 and C_4 photosynthetic pathways in Cyperaceae. *Aliso.* **23**(1), pp.99–148.

Christenhusz, M. J. M., Brockington, S. F., Christin, P.A. and Sage, R. F. 2014. On the Disintegration of Molluginaceae: A new genus and family (*Kewa*, Kewaceae) segregated from *Hypertelis*, and placement of *Macarthuria* in Macarthuriaceae. *Phytotaxa.* **181**(4), pp.238–242.

Christin, P.A., Arakaki, M., Osborne, C. P. and Edwards, E. J. 2015. Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. *Molecular Biology and Evolution.* **32**(4), pp.846–58.

Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. 2007. C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology.* **17** (14), pp.1241–1247.

Christin, P.A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. 2009. Evolutionary insights on C_4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biology and Evolution.* **1**, pp.221–230.

Christin, P.A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R. and Sage, R.F. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution*. **65**(3), pp.643–660.

Christin, P.A.and Osborne, C.P. 2013. The recurrent assembly of C_4 photosynthesis, an evolutionary tale. *Photosynthesis Research*, **117**(1–3), pp. 163–75.

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 C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences USA*. **110**(4), pp.1381–1386.

Darwin, C. 1859. On the origin of the species by means of natural selection: Or, the preservation of favoured races in the struggle for life (John Murray).

Davis, K.E., Hill, J., Astrop, T.I. and Wills, M.A. 2016. Global cooling as a driver of diversification in a major marine clade. *Nature Communications.***7,** pp.1–8.

Dengler, N.G, Dengler, R.E. and Hattersley, P.W. 1985. Differing ontogenetic origins of PCR ("Kranz") sheaths in leaf blades of C_4 grasses (Poaceae). *American Journal of Botany.* **72**, pp.284–302.

Dengler, N.G., Nelson, T. 1999. Leaf structure and development in C_4 plants. In: Sage, R.F., Monson, R.K. (eds). C_4 plant biology. *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 C_4 photosynthesis among close relatives. *Evolution.* **71** (6), pp.1541–1555.

Dunning, L., Moreno-Villena, J., Lundgren, M., Dionora, J. Salazar, P., Adams, C., Nyirenda, F., Olofsson, J., Mapaura, A., Grundy, I., Kayombo, C., Dunning, L., Kentatchime, F., Ariyarathne, M., Yakandawala, D., Besnard, G., Quick, P., Bräutigam, A., Osborne, Colin. and Christin, P.A. 2019. Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *Journal of Experimental Botany.* **70**(12), pp.3255–3268.

Edwards, E.J., and Donoghue, M..J. 2013. Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany.* **64**(13), pp.4047–4052.

Edwards, G.E., Franceschi, V.R. and Voznesenskaya, E.V. 2004. Single cell C_4 photosynthesis versus the dual cell (Kranz) paradigm. *Annual Review of Plant Biology.* **55**, pp.173–196.

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A. and C_4 Grasses Consortium. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*. **328**, pp.587–591.

Ehleringer, J.R., Cerling, T.E. and Helliker, B.R. 1997. C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*. **112**(3), pp.285–299.

Ehleringer, J.R., and Monson, R.K. 1993. Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology, Evolution, and Systematics.* 24, pp.411–439.

Ekblom, R. and Galindo, J., 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* **107**, pp.1–15.

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 C_4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 Species. *Molecular Biology and Evolution.* **33**(7), pp.1796–806.

Endress, M.E. and Bittrich, V. 1993. Molluginaceae. In: Kubitzki, K., Rohwer, J.G., and Bittrich, V. (eds).The families and genera of vascular plants. vol. 2. *Berlin: Springer Verlag.* pp.419–426.

Fernald, R.D. 2006. Casting a genetic light on the evolution of eyes. *Science.* **313**, pp.1914–1918.

Firooznia, F. 2007. The story of the Calvin cycle: bringing carbon fixation to life. *The American Biolology Teacher.* **69**(6). pp.364–367.

Furbank, R.T. 2011. Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *Journal of Experimental Botany.* **62**(9),pp.3103–8.

Garner, D.M., Mure, C.M., Yerramsetty, P. and Berry, J.O. 2016. Kranz anatomy and the C_4 pathway. *eLS*, (July), pp.1–10.

Gatesy, S.M. and Dial, K.P. 1996. Locomotor modules and the evolution of avian flight. *Evolution*. **50**, pp.331–340.

Gowik, U., Engelmann, S., Bläsing, O.E., Raghavendra, A.S. and Westhoff, P. 2006. Evolution of C_4 phosphoenolpyruvate carboxylase in the genus *Alternanthera:* gene families and the enzymatic characteristics of the C_4 isozyme and its orthologues in C_3 and C_3-C_4 *Alternantheras*. *Planta.* **223**, pp.359–368.

Griffiths, H., Weller, G., Toy, L.F.M., and Dennis, R.J. 2013. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell & Environment.* **36**(2), pp.249–261.

Haberlandt, G. 1884. Physiologische Pflanzenanatomie. *Leipzig: Engelmann.*

Haldane, J.B.S., Sprunt, A.D., and Haldane, N.M. 1915. Reduplication in mice (Preliminary Communication). *Journal of Genetics.* **5**(2), pp.133–135.

Hatch, M. 1987. C_4 Photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta*. **895**(2), pp.81–106.

Hatch, M.D., Slack, C.R. 1966. Photosynthesis by sugar-cane leaves a new carboxylation reaction and pathway of sugar formation. *Biochemical Journal.* **101**, pp.103–111.

Hattersley, P.W. 1984. Characterization of C_4 type leaf anatomy in grasses (Poaceae). Mesophyll:bundle sheath area ratios. *Annals of Botany* **53**, pp.163-179.

Hattersley, P.W. and Browning, A.J. 1981. Occurrence of the suberized lamella in leaves of grasses of different photosynthetic types. I. In parenchymatous bundle sheaths and PCR ("Kranz") sheaths. *Protoplasma.* **109**, pp.371–401.

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**, pp.325–333.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P.M., and Lercher, M.J. 2013. Predicting C_4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell.* **153**(7), pp.1579–1588.

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**, pp.485–493.

Huang, C.H.; Zhang, C.; Liu, M.; Hu, Y.; Gao, T.; Qi, J.; Ma, H. 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Molecular Biology and Evolution.* **33**, pp.2820–2835.

Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A. and Woolhouse, H.W. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. *Planta.* **175**(4), pp.452–459.

Iglesias, A.A., González, D.H. and Andreo, C.S., 1986. The C_4 pathway of photosynthesis and its regulation. *Biochemical Education.* **14**(3), pp.98–102.

Jiao, J.A. and Chollet, R. 1989. Regulatory seryl-phosphorylation of C_4 phosphoenolpyruvate carboxylase by a soluble protein kinase from maize leaves. *Archives of Biochemistry and Biophysics.* **269**(2), pp.526–535.

Kadereit, G., Bohley, K., Lauterbach, M., Tefarikis, D.T. and Kadereit, J.W. 2017. C_3–C_4 intermediates may be of hybrid origin a reminder. *New Phytologist.* **215**, pp.70–76.

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences.* **164**, pp.959–986.

Kanai. R., Edwards, G.E. 1999. The biochemistry of C_4 photosynthesis. In: Sage, R.F., Monson, R.K. (eds). C_4 Plant Biology. *Academic Press.* San Diego, pp.49–87.

Kayal, E., Roure, B., Philippe, H., Collins, A.G. and Lavrov, D. V. 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology.* **13**(1), p.5.

Kellogg, E.A. 1999. Phylogenetic aspects of the evolution of C_4 photosynthesis. In Sage, R. F., Monson. R.K. (eds). C_4 plant biology. *Academic Press: San Diego*. pp.411–444.

Kennedy, R.A., Eastburn, J.L. and Jensen, K.G. 1980. C_3–C_4 photosynthesis in the genus *Mollugo*: structure, physiology and evolution of intermediate characteristics. *American Journal of Botany.* **67**, pp.1207–1217.

Kennedy, R. A. and Laetsch, W. M. 1974. Plant species intermediate for C_3, C_4 photosynthesis. *Science.* **184**(4141), pp.1087–9.

Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K. and Mardis, E. R. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell.* **155**(1), pp.27–38.

Kozmik, Z., Ruzickova, J., Jonasova, K., Matsumoto, Y., Vopalensky, P., Kozmikova, I., Strnad, H., Kawamura, S., Piatigorsky, J. and Paces, V. 2008. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proceedings of the National Academy of Sciences USA*. **105**, pp.8989–8993.

Ku, M., Monson, R., Littlejohn, R., Nakamoto, H., Fisher, D. and Edwards, G. 1983. Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria* species : I. Leaf anatomy, photosynthetic responses to O_2 and CO_2, and activities of key enzymes in the C_3 and C_4 pathways. *Plant Physiology.* **71**(4), pp.944–948.

Kümpers, B.M.C., Burgess, S.J., Reyna-Llorens, I., Smith-Unna, R., Boursnell, C. and Hibberd, J.M. 2017. Shared characteristics underpinning C_4 leaf maturation derived from analysis of multiple C_3 and C_4 Species of *Flaveria*. *Journal of Experimental Botany.* **68**(2), pp.177–189.

Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G.E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A. L., Forest, F. and Baker, W. J. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science.* **10**(January), pp.1–17.

Larridon, I., Walter, H.E., Rosas, M., Vandomme, V. and Guerrero, P. C. 2018. Evolutionary trends in the columnar cactus genus *Eulychnia* (Cactaceae) based on molecular phylogenetics, morphology, distribution, and habitat. *Systematics and Biodiversity.* **16**(7), pp.643–657.

Leegood, R.C. 2008. Roles of the bundle sheath cells in leaves of C_3 Plants. *Journal of Experimental Botany*, **59**(7), pp.1663–1673.

Leegood, R.C., Onek, L.A., Pearson, M. and Lea, P.J. 1995. The isolation and characterization of mutants of the C4 photosynthetic pathway. *Journal of Experimental Botany.* **46**, pp.1363–1376.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. 2017. C_3 cotyledons are followed by C_4 leaves: Intra-individual transcriptome analysis of *Salsola soda* (Chenopodiaceae). *Journal of Experimental Botany*. **68** (2), pp.161–176.

Lenski, R.E., Ofria, C., Pennock, R.T. and Adami, C. 2003. The evolutionary origin of complex features. *Nature*. **423**(6936), pp.139–144.

Long, S.P. and Spence, A.K. 2013. Toward cool C_4 crops. *Annual Review of Plant Biology.* **64**, pp.701–722.

Lundgren, M.R. and Christin, P.A. 2017. Despite phylogenetic effects, C_3–C_4 lineages bridge the ecological gap to C_4 photosynthesis. *Journal of Experimental Botany* . **68**, pp.241–254.

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 C_3-C_4 intermediates in the Grass *Alloteropsis semialata*. *Plant, Cell and Environment.* **39**(9), pp.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. C_4 anatomy can evolve via a single developmental change. *Ecology Letters.* **22**(2), pp.302–312.

Lundgren, M.R., Osborne, C.P. and Christin, P.A. 2014. Deconstructing Kranz anatomy to understand C_4 evolution. *Journal of Experimental Botany.* **65**, pp.3357–3369.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P., Westhoff, P., and Gowik, U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *ELife 3,e02478.*

Mayr, E. 1982. The growth of biological thought: diversity, evolution, and inheritance. *Harvard University Press.*

Marazzi, B., Ane, C., Simon, M. F., Delgado-Salinas, A., Luckow, M. and Sanderson, M. J. 2012. Locating evolutionary precursors on a phylogenetic tree. *Evolution.* **66**(12), pp.3918–3930.

Mckown, A.D. and Dengler, N.G. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany*. **94**, pp.382–399.

McKown, A.D., Moncalvo, J.M. and Dengler, N.G. 2005. Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *American Journal of Botany* . **92,** 1911–1928.

Miyao, M. 2003. Molecular evolution and genetic engineering of C_4 photosynthetic enzymes. *Journal of Experimental Botany.* **54**(381), pp.179–189.

Monson, R.K. 1999. The origins of C_4 genes and evolutionary pattern in the C_4 metabolic phenotype. In: Sage, R.F., Monson, R.K. (eds). C_4 plant biology. *Academic Press: San Diego.* pp. 377–410.

Monson, R.K., Moore, B.D. 1989. On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. *Plant, Cell and Environment.* **12**, pp.689–699.

Moreno-Villena, J.J., Dunning, L.T., Osborne, C.P. and Christin, P.A. 2018. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Molecular Biology and Evolution.* **35**(1), pp.94–106.

Morris, S.C. 2006. Evolutionary convergence. *Current Biology.* **16**(19), pp.826–827.

Muhaidat, R., Sage, R.F. and Dengler, N.G. 2007. Diversity of Kranz anatomy and biochemistry in C_4 eudicots. *American Journal of Botany.* **94**(3), pp.362–381.

Nadeau, N.J., Martin, S.H., Kozak, K.M., Salazar, C., Dasmahapatra, K.K., Davey, J.W,, Baxter, S.W., Blaxter, M.L., Mallet, J., Jiggins, C.D. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology.* ***22***, pp**.** 814–826.

Nelson, T. and Langdale, J.A. 1989. Patterns of leaf development in C_4 Plants. *The Plant cell.* **1**(1), pp.3–13

Nisbet, E.G., Grassineau, N.V., Howe, C.J., Abell, P.I., Regelous, M. and Nisbet, R.E.R. 2007. The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology.* **5**, pp. 311–335.

Ogren, W.L. 1984. Photorespiration: Pathways, regulation, and modification. *Annual Review of Plant Physiology.* **35**, pp.415–442.

Panganiban, G., Irvine, S.M., Lowe, C., Roehl, H., Corley, L.S., Sherbon, B., Grenier, J.K., Fallon, J.F., Kimble, J., Walker, M., Wray, G.A., Swalla, B.J., Martindale, M.Q. and Carroll, S.B. 1997. The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences: USA.* **94**, pp.5162–5166.

Phansopa, C., Dunning, L.T., Reid, J.D. and Christin, P.A. 2020. Lateral gene transfer acts as an evolutionary shortcut to efficient C_4 biochemistry. *Molecular Biology and Evolution*. doi:10.1093/,olbev/msaa143.

Pick, T.R., Bräutigam, A., Schlüter, U., Denton, A. K., Colmsee, C., Scholz, U., Fahnenstich, H., Pieruschka, R., Rascher, U., Sonnewald, U. and Weber, A. M..2011. Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. *Plant Cell.* **23**(12), pp.4208–4220.

Rajendrudu, G., Prasad, J.S., and Das, V.S. 1986. C_3-C_4 intermediate species in *Alternanthera* (Amaranthaceae). leaf anatomy, CO_2 Compensation point, net CO_2 exchange and activities of photosynthetic enzymes. *Plant Physiology.* **80**(2), pp.409–14.

Ray, T. 1991. An approach to the synthesis of life. In: Langton, C.G., Farmer, J. D. and Rasmussen, S. (eds). Artificial Life II. *Addison–Wesley*. pp. 371-408.

Reinfelder, J.R., Milligan, A.J., Morel and F.M.M. 2004. The role of the C_4 pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiology.* **135**: pp.2106–2111.

Roesti, M., Hendry, A. P., Salzburger, W. and Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology.* **21**(12), pp.2852–2862.

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

Sage, R.F. 2004. The evolution of C_4 photosynthesis. *New Phytologist*. **161**, pp.341–370.

Sage, R.F., Christin, P.A., and Edwards, E.J. 2011. The C_4 plant lineages of planet Earth. *Journal of Experimental Botany.* **62**(9), 3155–69.

Sage, R.F., Khoshravesh, R., and Sage, T.L. 2014. From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *Journal of Experimental Botany* . **65**, pp.3341–3356.

Sage, R.F., Sage, T. L., and Kocacinar, F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Rev. Plant Biol.* **63**, pp.19–47.

Sage, R.F. and Stata, M. 2015. Photosynthetic diversity meets biodiversity: The C_4 plant example. *Journal of Plant Physiology.* **172**, pp.104–119.

Sage, R.F., Wedin, D.A., Li, M. 1999. The biogeography of C_4 photosynthesis: patterns and controlling factors. In: Sage, R.F., Monson, R.K. (eds). C_4 Plant Biology. *Academic Press: San Diego*. pp. 313–373.

Sayre, R.T. and Kennedy, R.A. 1977. Ecotypic differences in the C_3 and C_4 photosynthetic activity in *Mollugo verticillata*, a C_3-C_4 intermediate. *Planta.* **134**(3), pp.257–62.

Schlüter, U., Denton, A.K., and Brautigam, A. 2016. Understanding metabolite transport and metabolism in C_4 plants through RNA-seq. *Current Opinion in Plant Biology.* **31**,pp.83–90.

Schlüter, U., and Weber, A. P. 2016. The road to C_4 photosynthesis: evolution of a complex trait via intermediary states. *Plant and Cell Physiology.* **57**(5), pp.881–889.

Schüssler, C., Freitag, H., Koteyeva, N., Schmidt, D., Edwards, G., Voznesenskaya, E. and Kadereit, G. 2017. Molecular phylogeny and forms of photosynthesis in tribe Salsoleae (Chenopodiaceae). *Journal of Experimental Botany.* **68**(2), pp.207–223.

Sedelnikova, O.V., Hughes, T.E. and Langdale, J.A. 2018. Understanding the genetic basis of C_4 Kranz anatomy with a view to engineering C_3 crops. *Annual Review of Genetics*. **52**, pp.249–270

Sheen, J. 1999. C_4 gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology.* **50**, pp. 187–217.

Sinha, N.R. and Kellogg, E.A. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. *American Journal of Botany.* **83**, pp.1458–1470.

Slack, C. and Hatch, M. 1967. Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochemical Journal*. **103**(3), pp.660–665.

Soros, C.L. and Dengler, N.G. 2001. Ontogenetic derivation and cell differentiation in photosynthetic tissues of C_3 and C_4 Cyperaceae. *American Journal of Botany.* **88**, pp.992–1005.

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D, Beckerman, A.P. and Slate, J. 2010. Adaptation genomics: the next generation. *Trends in Ecology and Evolution.* **25**, pp.705–712.

Stata, M., Sage, T.L., Rennie, T.D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., Ludwig, M. and Sage, R.F. 2014. Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. P*lant, Cell and Environment.* **37**, pp.2587–2600.

Stata, M., Sage, T.L. and 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.* **49**, pp.2734.

Studer, R.A., Christin, P.A., Williams, M.A., and Orengo, C.A. 2014. Stability- activity tradeoffs constrain the adaptive evolution of RubisCO. *Proceedings of the National Academy of Sciences: USA.* **111**(6), pp.2223–8.

Tabita, F.R., Hanson, T.E., Li, H.Y., Satagopan, S., Singh, J. and Chan, S. 2007. Function, structure and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiology and Molecular Biology Reviews.* **71,** pp.576-597.

Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E. and Scott, S.S. 2008. Distinct form I, II, III and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of Experimental Botany.* **59**, pp.1515- 1524.

Taub, D.R. 2000. Climate and the U.S. distribution of C_4 grass subfamilies and decarboxylation variants of C_4 photosynthesis. *American Journal of Botany.* **87**, pp.1211–1215.

Taylor, S.H., Hulme, S.P., Rees, M., Ripley, B.S., Ian, Woodward. F. and Osborne, C.P. 2010. Ecophysiological traits in C_3 and C_4 grasses: a phylogenetically controlled screening experiment. *New Phytologist.* **185**, pp.780–791.

Tcherkez, G.G.B., Farquhar, G.D., and Andrews, T.J. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences USA.***103**(19), pp.7246–51.

Thornhill, A.H., Baldwin, B.G., Freyman, W.A., Nosratinia, S., Kling, M.M., Morueta-Holme, N., Madsen, T.P., Ackerly, D.D., and Mishler, B.D. 2017. Spatial phylogenetics of the native California flora. *BMC Biology.* **15**, 96.

Thornhill, A.H., Mishler, B.D., Knerr, N., Gonzalez-Orozco, C.E., Costion, C.M., Crayn, D.M., Laffan, S.W., and Miller, J.T. 2016. Continental-scale spatial phylogenetics of Australian angiosperms provides insights into ecology, evolution and conservation. *Journal of Biogeography.* **43**, pp.2085–2098.

Thulin, M., Moore, A.J., El-seedi, H., Larsson, A., Christin, P.A. and Edwards, E.J. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. *Taxon*. **65**(4), pp.775–793.

True, J.R. and Carroll, S.B. 2002. Gene co–option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol*. **18**, pp.53–80.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics.* **30**(9), pp.418–426.

Villoutreix, R., De Carvalho, C.F., Soria-Carrasco, V., Lindtke, D., De-La-Mora, M., Muschick, M., Feder, J.L., Parchman, T.L., Gompert, Z. and Nosil, P. 2020. Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science.* **369**(6502), pp.460–466.

Visscher, P.M., Hill, W.G. and Wray, N.R. 2008. Heritability in the genomics era-concepts and misconcep-tions. *Nature Reviews Genetics*. **9,** pp. 255–66.

Vogan, P.J., Frohlich, M.W. and Sage, R.F. 2007. The functional significance of C_3-C_4 intermediate traits in *Heliotropium*: gas exchange perspectives. *Plant, Cell and Environment.* **30,** pp.1337-1345.

Vogan, P.J., and Sage, R.F. 2011. Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of *Flaveria* Juss. (Asteraceae). *Plant, Cell and Environment.* **34**, pp.1415–1430.

von Caemmerer, S. and Furbank, R.T. 2003. The C_4 pathway: an efficient CO_2 pump. *Photosynthesis Research.* **77**, pp.191–207.

von Caemmerer, S., Ghannoum, O., Pengelly, J.J. and Cousins, A.B. 2014. Carbon isotope discrimination as a tool to explore C_4 photosynthesis. **J***ournal of Experimental Botany.* **65**, pp.3459–3470.

Voznesenskaya, E.V., Chuong, S.D.X., Koteyeva, N.K., Franceschi, V.R., Freitag, H. and Edward, G.E. 2007. Structural, biochemical, and physiological characterization of C_4 photosynthesis in species having two vastly different types of Kranz anatomy in Genus *Suaeda* (Chenopodiaceae). *Plant Biology.* **9**, pp.745–757.

Voznesenskaya, E.V., Franceschi, V.R., Kiirats, O., Freitag, H. and Edwards, G.E. 2001. Kranz anatomy is not essential for terrestrial C_4 plant photosynthesis. *Nature.* **414**, pp.543–546.

Voznesenskaya, E. V., Franceschi, V.R., Pyankov, V.I. and Edwards, G.E. 1999. Anatomy, chloroplast structure and compartmentation of enzymes relative to photosynthetic mechanisms in leaves and cotyledons of species in the tribe Salsoleae (Chenopodiaceae). *Journal of Experimental Botany*. **50**(341), pp.1779–1795.

Wang, Y., Bräutigam, A., Weber, A.P.M., and Zhu, X.-G. 2014. Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. *Journal of Experimental Botany.* **65**(13), pp.3567–3578.

Washburn, J.D., Schnable, J.C., Conant, G.C., Brutnell, T.P., Shao, Y., Zhang, Y., Ludwig, M., Davidse, G. and Pires, J.C. 2017. Genome–guided phylo–transcriptomic methods and the nuclear phylogenetic tree of the Paniceae grasses. *Scientific Reports.* **7**,13528.

Webster, G.L., Brown, W.V and Smith, B.N. 1975. Systematics of photosynthetic carbon fixation pathways in *Euphorbia*. *Taxon*. **24**, pp.27–33.

Westhoff, P. and Gowik, U. 2010. Evolution of C_4 photosynthesis-Looking for the master switch. *Plant Physiology.* **154**(2), pp.598–601.

Whittall, J.B. and Hodges, S.A. 2007. Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*, **447**(7145), pp.706–709.

Williams, B.P., Aubry, S. and Hibberd, J. M. 2012. Molecular evolution of genes recruited into C_4 photosynthesis. *Trends in Plant Science.* **17**(4), pp.213–220.

Yadav, S. and Mishra, A. 2019. Introgression of C_4 pathway gene(s) in C_3 plants to improve photosynthetic carbon assimilation for crop improvement: A biotechnological approach. *Photosynthesis, Productivity and Environmental Stress.* (i), pp.267–281.

Yang, Z. and Rannala, B. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics.* **13**(5), pp.303–314.

Yokoyama, S. and Radlwimmer, F.B. 2001. The molecular genetics and evolution of red and green colour vision in vertebrates. *Genetics.* **158**, pp.1697–1710.

Zapata, F. and Jiménez, I. 2012. Species delimitation: Inferring gaps in morphology across geography. *Systematic Biology.* **61**(2), pp.179–194.

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications*. **5**, p.4956.

Zhu, X. G., Shan, L., Wang, Y. and Quick, W. P. 2010. C_4 rice - an ideal \rena for systems biology research. *Journal of Integrative Plant Biology.* **52**(8), pp.762–770.

Chapter 1:

Events leading to the evolution of C_4 leaf anatomy in Molluginaceae

Chapter 1: Events leading to the evolution of C_4 leaf anatomy in Molluginaceae

by Lamiaa Munshi[1], Pascal-Antoine Christin[1]

[1] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

Personal contributions: [I generated and analysed the data. I wrote the manuscript with the help of Pascal-Antoine Christin].

2.1 Abstract

Multiple lineages of plants have independently evolved C_4 photosynthesis, a physiology that relies on a specialized leaf anatomy to boost productivity in tropical conditions. The order of events leading to C_4 anatomy remains however unknown, and the evolutionary importance of C_3-C_4 intermediates is still debated. The Molluginaceae family encompasses a variety of photosynthetic types, and previous investigations have suggested that anatomical traits predated C_4 physiology in the group. In this study, we decompose the leaf of Molluginaceae species with various photosynthetic types into their constituents, which we then quantify and analyse in a phylogenetic context. Our analyses show that C_4 species differ from C_3 Molluginaceae in their proportion of different tissues, but the traits responsible for these proportions overlap between photosynthetic types. This indicates that C_4 anatomy in Molluginaceae emerged via the unique combination of widespread traits. Some C_3-C_4 species are anatomically closer to the C_3 than the C_4, which indicates that they do not universally bridge the gap between photosynthetic types. However, the C_3-C_4 species closely related to the C_4 present C_4-like leaf properties, and we suggests that the C_3-C_4 physiology stabilized a gross anatomy that evolved for unrelated reasons, helping its subsequent co-option for C_4 photosynthesis.

Key words: C_4, C_3-C_4, evolution, Kranz anatomy, leaf anatomy, Molluginaceae.

2.2 Introduction

One of the major aims of evolutionary biology is to understand how novel adaptations emerged during the diversification of organisms. This problem requires comparing species that differ in their adaptations and inferring putative ancestral states, which can be performed using phylogeny-based analysis (Dieckmann and Doebeli, 1999; Lenski et al., 2003; Lamb et al., 2007). In the case of traits that evolved multiple times independently, comparative analyses can moreover identify the factors that increase the chance of transitions among states (Brockington et al., 2011; Kawecki et al., 2012; Cacefo et al., 2019), helping to understand why some groups of organisms are more prone to acquire specific adaptations than others (Monteiro and Podlaha, 2009).

In plants, C_4 photosynthesis represents an adaptation over the ancestral C_3 type that boosts productivity in tropical conditions (Ehleringer et al., 1997; Sage, 2001; Sage and Kubien, 2003; Christin et al., 2009; Khoshravesh and Sage, 2014; Schüssler et al., 2017). The C_4 trait results from a coordinated set of anatomical, physiological, and biochemical modifications, that together concentration CO_2 within the leaf before its fixation by Rubisco (Slack and Hatch, 1967; Ehleringer et al., 1997). Rubisco achieves this via a spatial separation of atmospheric CO_2 fixation and CO_2 assimilation. Typically, the former happens in the mesophyll cells, via an enzyme without any affinity for O_2, namely phosphoenolpyruvate carboxylase (PEPC) (Webster et al., 1975; Iglesias et al., 1986; Brown and Hattersley, 1989; Osborne and Sack, 2012). The resulting four-carbon acid, which gave its name to the pathway, is transformed and transported to the bundle sheath cells, a tissue nested within the leaf, where Rubisco is segregated in C_4 plants (Monson et al., 1986; Dengler et al., 1994; Furbank and Taylor, 1995; Busch et al., 2017). CO_2 is released there to feed Rubisco and the CBB cycle

(Hatch and Slack, 1966; Sage, 2004). Besides high, cell-specific enzyme activities, the C_4 trait requires a set of leaf morphological attributes (Hatch, 1987; Christin and Besnard, 2009; Muhaidat et al., 2011; Wang et al., 2017). While these have collectively been referred to as 'Kranz anatomy', they represent a variety of solutions to the same requirements; a) the isolation of the bundle sheath from the atmosphere and the concentration of chloroplasts containing Rubisco there, b) a short distance between mesophyll and bundle sheath cells to allow rapid diffusion of metabolites, and c) a proliferation of the bundle sheath tissue to accommodate a large amount of chloroplast (Sage et al., 2011; Lundgren et al., 2014, 2019; Lauterbach et al., 2019). Despite this apparent complexity, the C_4 trait evolved more than 60 times independently (Sage et al., 2011), with numerous C_4 origins clustered in some parts of the angiosperm phylogeny (Christin and Besnard, 2009; Edwards et al., 2010; Kadereit et al., 2012). This complex trait therefore constitutes an excellent system to understand the factors that promote functional innovations in some plant groups.

Comparative work has shown that the recurrent origins of C_4 photosynthesis were facilitated by the existence of either anatomical and/or biochemical C_4-like traits in some C_3 lineages that could be recurrently co-opting, thereby limiting the evolutionary distance between the C_3 and C_4 phenotypes (Voznesenskaya et al., 1999; Kadereit et al., 2012; Christin et al., 2013; Moreno-Villena et al., 2018). In particular, evolutionarily stable phenotypes exist that are physiologically intermediate between C_3 and C_4 photosynthesis. These so-called C_3-C_4 intermediates encompass plants that use a C_4-like leaf anatomy to perform a photorespiratory bypass and those that fix some, but not all of their CO_2 via a weak C_4 cycle (Christin and Osborne, 2014; Sage, 2016; Schuler et al., 2016; Lundgren and Christin, 2017). These intermediates have been detected in 21 plant lineages, and in many cases are closely related to C_4 plants, which is interpreted as the fingerprint of the gradual transition from C_3 to C_4

photosynthesis. Whether all C_3-C_4 intermediates offer similar paths to C_4 photosynthesis is however debated, and several such lineages are not closely related to any C_4 plants (Sage et al., 2011). This is notably the case of two C_3-C_4 groups within the Molluginaceae family, which have been used as examples of C_3-C_4 lineages that might lack elements critical for further transitions to a C_4 state (Edwards and Donoghue, 2013), a hypothesis that is yet to be formally tested.

With 31 species, Molluginaceae is a small family that encompasses a diversity of photosynthetic types. Besides C_4 individuals recently shown to belong to three distinct species (Thulin et al., 2016), it contains a number of C_3 species and three C_3-C_4 lineages that include the first described C_3-C_4 intermediates (Sayre and Kennedy, 1977; Christin et al., 2011; Lundgren and Christin, 2017). The three C_4 species belong to the genus *Hypertelis* that also includes the C_3-C_4 *H. spergulacea*. Based on the qualitative comparison of leaf anatomies obtained from herbarium samples, the C_4 species were described as representing independent origins that co-opted C_4-like traits present in their common ancestor (Christin et al., 2011). A denser sampling coupled with quantitative measures of the individual anatomical traits are however needed to precisely evaluate the location of changes along the phylogeny of the family, to evaluate the potential of distinct C_3-C_4 lineages to serve as evolutionary intermediates.

In this study, we compare the anatomical traits of Molluginaceae species representing a diversity of photosynthetic types in a phylogenetic context to i) identify the changes involved in the emergence of a C_4 leaf anatomy, ii) test the hypothesis that C_3-C_4 phenotypes acted as evolutionary intermediates facilitating subsequent transitions to C_4 photosynthesis and iii) determine whether the leaf anatomy of distinct C_3-C_4 lineages has the same potential

to facilitate the emergence of C_4 photosynthesis. Our results show that some C_3-C_4 are closer to the anatomical requirements of C_4 plants than others, potentially restricting the possibility of C_4 origins to some groups of such intermediates.

2.3 Materials and Methods

2.3.1 Plant material.

Accessions of Molluginaceae species were sampled during multiple field trips, acquired from seed banks or from herbarium collections (Table 1). The list of species includes widely distributed taxa from the different photosynthetic types, as well as geographically restricted species, and covers the whole Molluginaceae family. For C_4 and C_3-C_4 species, multiple accessions were sampled to cover the potentially existing within each species. When possible, three different leaves were analyzed per accession. For field-collected individuals, leaves were directly preserved in ethanol. For accessions available as seeds, plants were grown in controlled accessions at the University of Sheffield, in a Conviron chamber, with a 14 hours light period. The temperature was maintained at 25°C during the day and 20°C at night. A 2:1 mixture of compost and sand was used, and plants were watered three times a week to keep the soil damp. Mature leaves were then collected and preserved in ethanol. The final dataset consisted of 13 C_4 accessions from three species, nine accessions from three intermediate species, and nine C_3 species (Table 1).

2.3.2 Sequencing and phylogenetic inference.

A chloroplast marker was selected to infer phylogenetic relationships among the sampled accessions. The region encompassing the intron of *trnK* and the full coding sequence of *matK* (*matK-trnK*) was selected, because it is informative and was previously analysed for species covering the whole Molluginaceae family (Christin et al., 2011). All new samples were included, while existing sequences were retrieved for previously analysed accessions (Christin et al., 2011). For new accessions, DNA was extracted from fresh leaves or silica gel dried sample using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA), following the manufacturer's protocol. The *trnK-matK* marker was amplified in one or several overlapping fragments, depending on the quality of the extracted DNA. Previously published primers and PCR protocols were used (Christin et al., 2011). In short, GoTaq polymerase (Promega, Madison, WI) was used with the following program: initial denaturation for 3 minutes followed with 35 cycles of 95°C for 30 s, annealing at 54°C for 30 s and extension at 72°C for 3 minutes, and a final extension at 72°C for 10 minutes. Amplified fragments were Sanger sequenced using standard protocols and raw chromatograms were manually edited in Geneious v. 135 6.8 (Kearse et al., 2012). Sequencing was preformed by Core Genome Facility in the University of Sheffield.

Table 2.1: List of accessions analyzed in this study.

Species	Code	Origin	Lat	Long	Photo	Source	Voucher
Adenogramma galoides	156	South Africa	-	-	C_3	Ogburn& Edwards 2012	Ogburn146 (BRU)
Adenogramma glomerata	146	South Africa	-	-	C_3	Ogburn& Edwards 2012	Ogburn 142 (BRU)
Hypertelis cerviana	1.2	Namibia	-26.748	17.221	C_4	Field collection	JJ Moreno-Villena HYP-7.3-2 (SHD)
Hypertelis cerviana	8	Namibia	-22.034	16.937	C_4	Field collection	JJ Moreno-Villena HYP-8.3-3 (SHD)
Hypertelis cerviana	-	Spain	39.326	-4.8380	C_4	Growth chamber	No voucher
Hypertelis cerviana	-	Zambia	-11.36	29.6	C_4	Growth chamber	No voucher
Hypertelis spergulacea	13	Namibia	-27.86	16.68	C_3-C_4	Field collection	JJ Moreno-Villena HYP-5.3-3 (SHD)
Hypertelis spergulace	6.2	Namibia	-27.829	16.692	C_3-C_4	Field collection	JJ Moreno-Villena HYP-5.3-1 (SHD)
Hypertelis spergulacea	5.3	Namibia	-28.08	16.8912	C_3-C_4	Field collection	JJ Moreno-Villena HYP-5.3-2 (SHD)
Hypertelis umbellata	3	Namibia	-27.44	17.94	C_4	Field collection	JJ Moreno-Villena HYP-3.3-3 (SHD)
Hypertelis umbellata	-	Arizona	34.395	-111.763	C_4	Growth chamber	No voucher
Hypertelis umbellata	-	Mozambique	11.3409	40.362	C_4	Growth chamber	No voucher
Hypertelis walteri	2.2	Namibia	-27	17.897	C_4	Field collection	JJ Moreno-Villena HYP-3.3-1(SHD)
Hypertelis walteri	3.2	Namibia	-27.145	17.687	C_4	Field collection	JJ Moreno-Villena HYP-3.3-2 (SHD)
Hypertelis walteri	5.2	Namibia	-26.657	16.275	C_4	Field collection	JJ Moreno-Villena HYP-7.3-2 (SHD)
Hypertelis walteri	7	Namibia	-26.083	18.153	C_4	Field collection	JJ Moreno-Villena HYP-7.3-4 (SHD)
Hypertelis walteri	9	Namibia	-26.55	18.149	C_4	Field collection	JJ Moreno-Villena HYP-7.3-3 (SHD)
Hypertelis walteri	10	Namibia	-26.657	16.27	C_4	Field collection	JJ Moreno-Villena HYP-7.3-1 (SHD)
Mollugo verticillata	-	Brazil	-22.389	-49.004	C_3-C_4	Field collection	No voucher
Mollugo verticillata	-	Michigan	43.621	-84.682	C_3-C_4	Growth chamber	Sage &Sage 2007(TRT)
Mollugo verticillata	-	Montana	47.375	-109.63	C_3-C_4	Growth chamber	No voucher

Species	Code	Origin	Lat	Long	Photo	Source	Voucher
Mollugo verticillata	-	South Africa	-28.816	24.991	C_3-C_4	Growth chamber	No voucher
Paramollugo nudicaulis	-	India	22.351	78.667	C_3-C_4	Growth chamber	No voucher
Paramollugo nudicaulis	-	Uganda	0.7913	33.03611	C_3-C_4	Growth chamber	Christin 2015-18 (SHD)
Pharnaceum confertum	-	South Africa	-	-	C_3	Ogburn& Edwards 2012	Ogburn 163 (BRU)
Pharnaceum incanum	-	South Africa	-	-	C_3	Ogburn& Edwards 2012	Ogburn 148 (BRU)
Pharnaceum subtile	-	South Africa	-	-	C_3	Ogburn& Edwards 2012	No voucher
Psammotropha obovata	-	South Africa	-	-	C_3	Ogburn& Edwards 2012	No voucher
Psammotropha quadrangularis	-	South Africa	-	-	C_3	Ogburn& Edwards 2012	Ogburn 160 (BRU)
Suessenguthiella scleranthoides	-	Namibia	-26.64	16.233	C_3	Field collection	JJ Moreno-Villena SUE-4.3-1 (SHD)
Trigastrotheca pentaphylla	-	India	22.351	78.667	C_3	Growth chamber	No voucher

The dataset was manually aligned. The final dataset, comprised of 31 taxa, was used for phylogenetic inference as implemented in the software Bayesian Evolutionary Analysis by Sampling Trees (BEAST) (Drummond and Rambaut, 2007) under a GTR substitution model with a gamma-shape parameter (GTR+G), a relaxed log-normal clock, and a Yule process. The monophyly of both the ingroup and the outgroup (identified based on Christin et al., 2011) was enforced to root the tree. The root of the tree was fixed to 46.7 Ma, based on Christin et al. (2011), using a normal distribution with a sd of 0.0001. Two analyses were run for 10,000,000 generations, and convergence was verified using Tracer (Drummond and Rambaut, 2007). The burn-in period was set to 2,000,000 generations, and the median ages across the remaining trees were mapped on the maximum credibility topology.

2.3.3 Microscopy and anatomical measurements.

Three replicates per each accession were examined. Leaf samples stored in ethanol were embedded using the Tecnovit 7100 kit (Technovit 7100, Heraeuk Kulzer GmbH, Wehrhein, Germany). Mid-regions of leaves were fixed in Carnoy's fixative solution (100% EtOH: acetic acid 4:1) and subsequently dehydrated in a series of ethyl alcohol. The dehydrated leaves were pre-infiltrated with Technovit base solution and EtOH 155 (1:1) overnight, which was followed by vacuum infiltration with 100% Technovit1 solution for 1-1.5h. The samples were subsequently embedded in freshly mixed resin of Tecnovit1 and Harder2 (1:15) solutions and mounted on histoblocks (Technovit 7100, Heraeuk Kulzer GmbH, Wehrhein, Germany). Transverse sections10µm thick were obtained with a Leica microtome (Leica RM2145). Dried sections were stained by applying 1% toluidine blue O (100 mM phosphate buffer 160 ph 7.0 with 0.1 g of toluidine blue O) for 2 minutes at 65°C (Sigma-Aldrich, St.Louis, MO, USA).

Pictures were taken from the mounted slides using a digital camera linked to Fluorescence Microscope Filter Sets for the Olympus BX51 (Olympus BX51, Hamburg, Germany). linked to Fluorescence Microscope Filter Sets for the Olympus BX51 (Olympus BX51, Hamburg, Germany).Molluginaceae species with small leaves were fully measured, while the large leaves were measured as segment in the middle of leaves encompasses two to three veins. Captured images were used for measurements, using the ImageJ software (Schneider et al., 2012). A total of 14 anatomical variables were measured for each replicate. These traits were chosen to capture the variation in leaf anatomy, including the characters that alter the main functions associated with C_4 leaf anatomy (Lundgren et al., 2014). Leaf thickness was measured for the layers of mesophyll in the mid-part of the leaf, while leaf width was measured as the distance between edges.

The interveinal distance was calculated as the distance between the centres of consecutive veins. The diameter of veins were calculated for all veins in the small leaves, and for each segment in the large leaves. Total areas of the different tissue types as bundle sheath, mesophyll, and vascular tissue were measured and divided by leaf width to get values comparable among samples. The size of bundle sheath cell was measured by averaging three cells surrounding the main veins. Based on these measurements, the fraction of bundle sheath area was estimated as the ratio of bundle sheath area divided by total area of mesophyll and bundle sheath.

2.3.4 Statistical analysis.

Regression analyses were used to test whether there was a linear relationship between the relatives areas of bundle sheath and mesophyll. The photosynthetic type was added as a cofactor to determine whether the relationship varied among photosynthetic types. In addition, multiple regression analyses were used to evaluate the contribution of individual traits to the emergent properties of the leaves. For each of the relative areas of bundle sheath and mesophyll, models were tested that included as explanatory variables the leaf thickness, interveinal distance, diameter of veins, and size of bundle sheath cells. Non-significant variables were successively removed until remaining variables had a significant effect on the modelled leaf property. All analyses were performed in R (RStudio Team, 2020). The distribution of all anatomical values was assessed by plotting them on the phylogenetic tree, and a maximum likelihood estimation was used to reconstruct ancestral states for all tree nodes in APE package in R (Paradis et al., 2004).

2.4 Results

2.4.1 High fractions of bundle sheath in C_4 species by constraining the relationship among tissues.

As expected, the leaf morphology of Molluginaceae was very variable, with differences linked to photosynthetic types (Fig. 2.1). In particular, the proportion of bundle sheath tissue (%BS) was larger in all three C_4 species than in any of the C_3 samples (0.37-0.49 versus 0.07-028; Fig. 2.2). The range of the intermediates overlapped with both C_3 and C_4 values (0.13-0.45; Fig. 2.2). The wide range of the intermediates results from differences among phylogenetic groups (Fig. 2.3). Indeed, the whole of the *Hypertelis* clade, which includes the three C_4 species as well as the intermediate *H. spergulacea*, is characterized by large bundle sheath fraction (%BS), while intermediates outside of the clade have low %BS (Fig. 2.3). Ancestral state reconstructions indicate that the common ancestor of the whole family had a low %BS (0.19; 95% CI = 0.12-0.31), and an increase is inferred specifically along the branch leading to the *Hypertelis* clade, from 0.20 (95% CI = 0.14-0.29) to 0.31 (95% CI = 0.23-0.43; Fig. 2.3). Within the *Hypertelis* clade, no consequent modifications of %BS are inferred along the branches leading to the individual C_4 groups, but recent changes of %BS are suggested for some populations of the intermediate *H. spergulacea* and the C_4 *H. cerviana* and *H. umbellata* (Fig. 2.3).

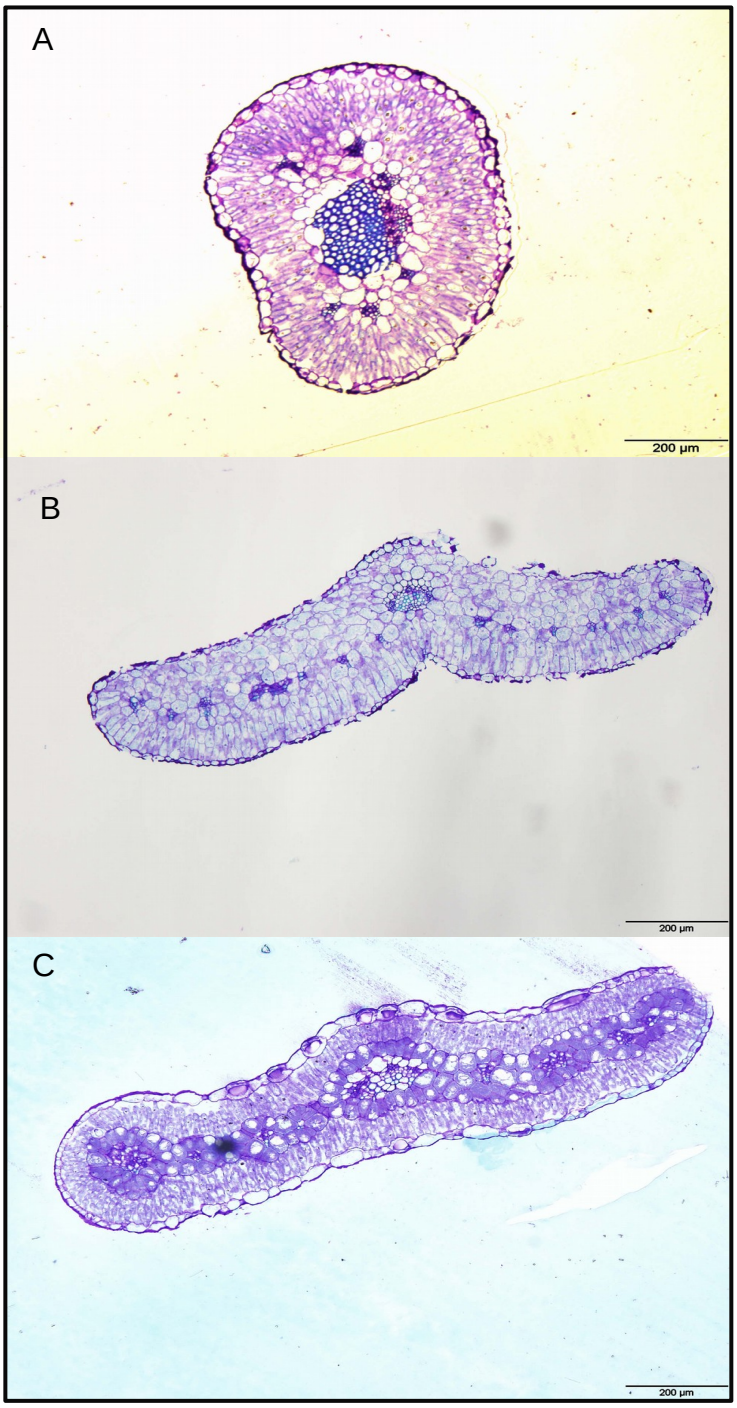

Figure 2.1: Leaf cross sections of Molluginaceae with different photosynthetic types.

(A) C_3 *Suessenguthiella scleranthoides*, (B) C_3-C_4 *Mollugo verticillata*, (C) C_4 *Hypertelis walteri.* The scale bar is shown on the bottom right of each picture.

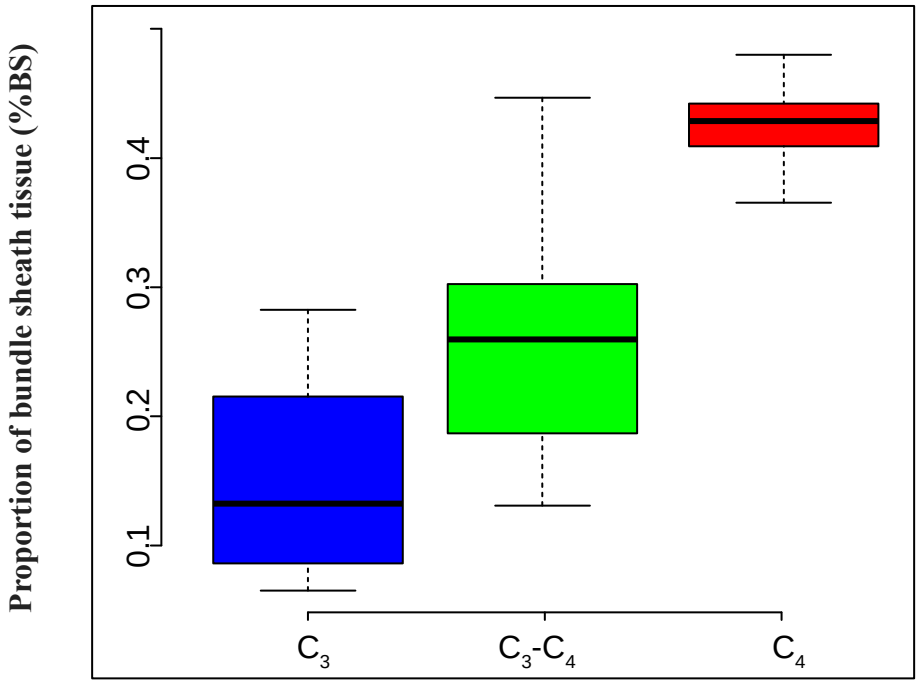

Figure 2.2: Distribution of bundle sheath fraction among photosynthetic types of Molluginaceae.

For each photosynthetic type, the distribution of the proportion of bundle sheath tissue (%BS) is shown with boxplots. Boxes connect the 25th and 75th percentiles, with medians indicated by thick lines. Whiskers connect the maximal values within 1.5 times the interquartile range.

The %BS of a given species is determined both by the amount of mesophyll and the amount of bundle sheath in the leaf. The amounts of each tissue were expressed as area per leaf width, providing a proxy for their relative contribution to the leaf thickness. Both the relative thickness of mesophyll (Mt) and bundle sheath (Bt) varied among species (Fig. 2.4). Individually, each of these variables overlapped between C_3 and C_4 species, but the *Hypertelis* clade is characterized by specific combinations of high Bt per Mt (Fig. 2.4).

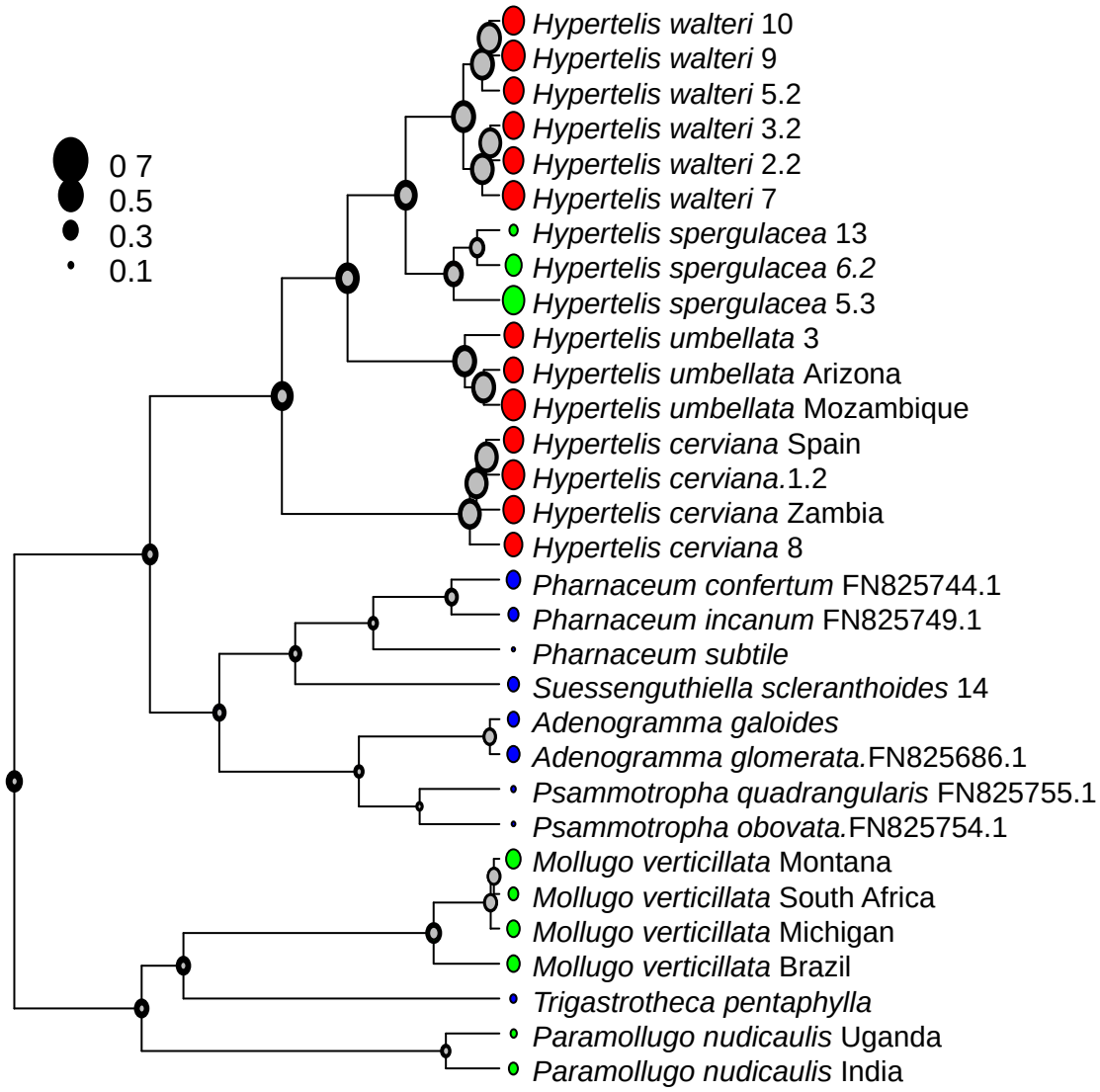

Figure 2.3: Bundle sheath fraction (%BS) mapped on Molluginaceae phylogeny.

The bundle sheath fraction was mapped on the phylogeny inferred from the chloroplast map marker *trnK-matK*. Circles at tips are proportional to the measured %BS and are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Circles at nodes are proportional to the lower (in grey) and upper (in black) bounds of the 95% confidence intervals of inferred ancestral states.

The relationship between Bt and Mt was evaluated while taking into account the photosynthetic types (linear regression with interactions, $R^2 = 0.43$). There was a significant effect of Mt on Bt for the C_4 individuals ($p < 0.003$), but the effect was only marginally significant in the intermediates ($p = 0.08$), and not significant in the C_3 ($p = 0.44$). The consistently high %BS of C_4 species is thus achieved by constraining the relationship between the areas of M and BS.

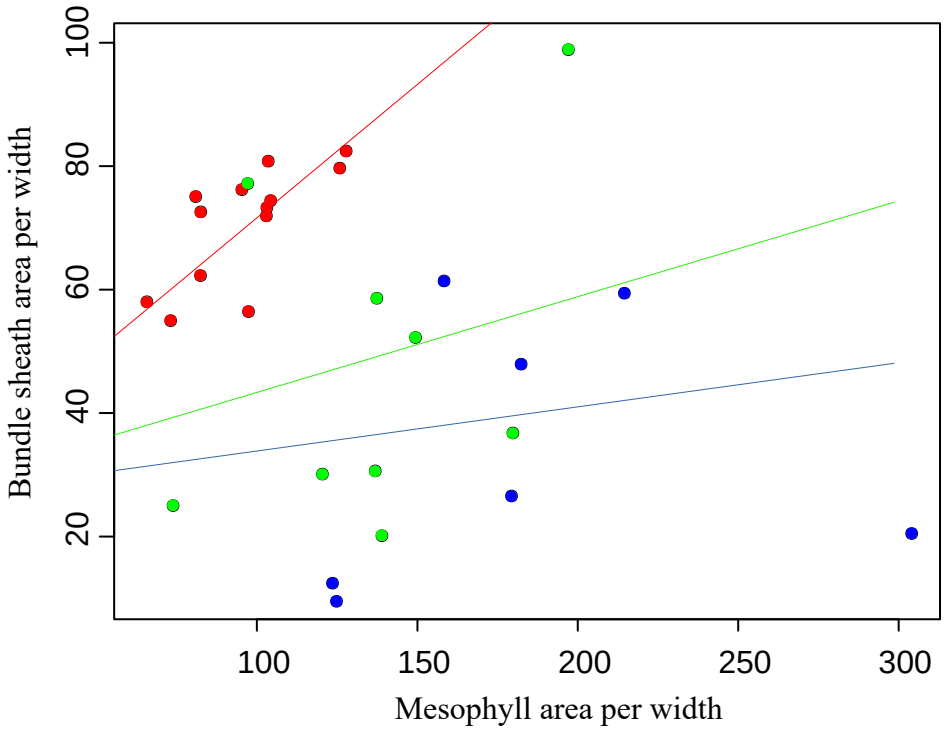

Figure 2.4: Relationship between the relative amounts of bundle sheath and mesophyll tissues.

The amount of bundle sheath and mesophyll tissues were calculated per width. Points are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Regression lines are indicated for each photosynthetic type, with the same colours.

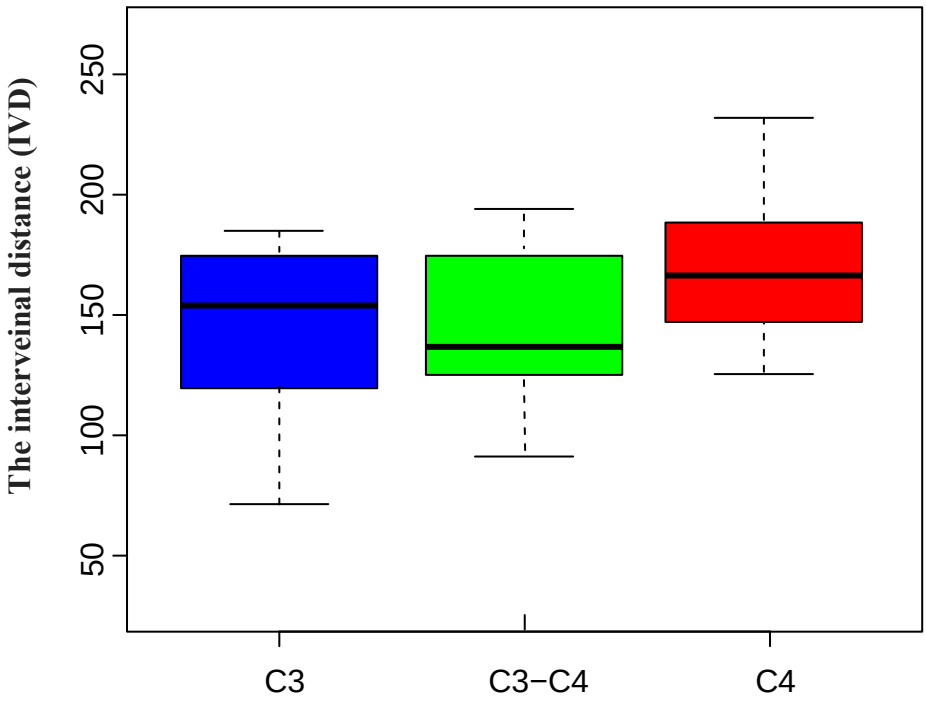

Figure 2.5: Distribution of interveinal distance (IVD) among photosynthetic types of Molluginaceae.

For each photosynthetic type, the distribution of interveinal distance (IVD) is shown with boxplots. Boxes connect the 25th and 75th percentiles, with medians indicated by thick lines. Whiskers connect the maximal values within 1.5 times the interquartile range.

2.4.2 Large bundle sheath fractions via a combination of traits observed in other species.

The overall variation in Bt is explained by a combination of interveinal distance (IVD), size of bundle sheath cells (BS.cell) and mean diameter of the veins (vein.diam; linear model, $R^2 = 0.68$). Each of these three variables individually overlap between C_3 and C_4 species, and with intermediates. IVD in particular is very similar among the three groups (Fig. 2.5), with only the C_3 species *Suessenguthiella scleranthoides* presenting a very large IVD, but this species has a single, very large vein (Fig. 2.6). The C_4 are consistently characterized

by a combination of large vein.diam and large BS.cell (Fig. 2.7), a part of the phenotypic landscape only reached by *S. scleranthoides* among non-C_4 species (but the large IVD of this species leads to a small Bt). Therefore, unique combinations of variables observed in C_3 species lead to important bundle sheath relative thickness in C_4 Molluginaceae. The variation in Mt is explained by a combination of vein.diameter and thickness (linear model, $R^2 = 0.66$). Again, these two variables overlap between C_3 and C_4 species, but C_4 species are characterized by large vein.diam coupled with small thickness (Fig. 2.8).

Variation in IVD (Fig. 2.6) and BS.cell (Fig. 2.9) along the tree is not clearly linked to photosynthetic transitions, and the values inferred for the common ancestor of the *Hypertelis* clade lie in the upper range of the C_3 group, with limited subsequent modifications (Figs 2.8 and 2.9). A relatively high thickness is inferred for this same ancestor, and thin leaves were achieved independently in two of the three C_4 clades (Fig. 2.10). By contrast, vein diameter seems to have increased along the branch leading to the common ancestor of *Hypertelis*, from 126 (95% CI = 78-201) to 174 (95% CI = 115-264), and further increases are observed within some populations of the different C_4 species (Fig. 2.11).

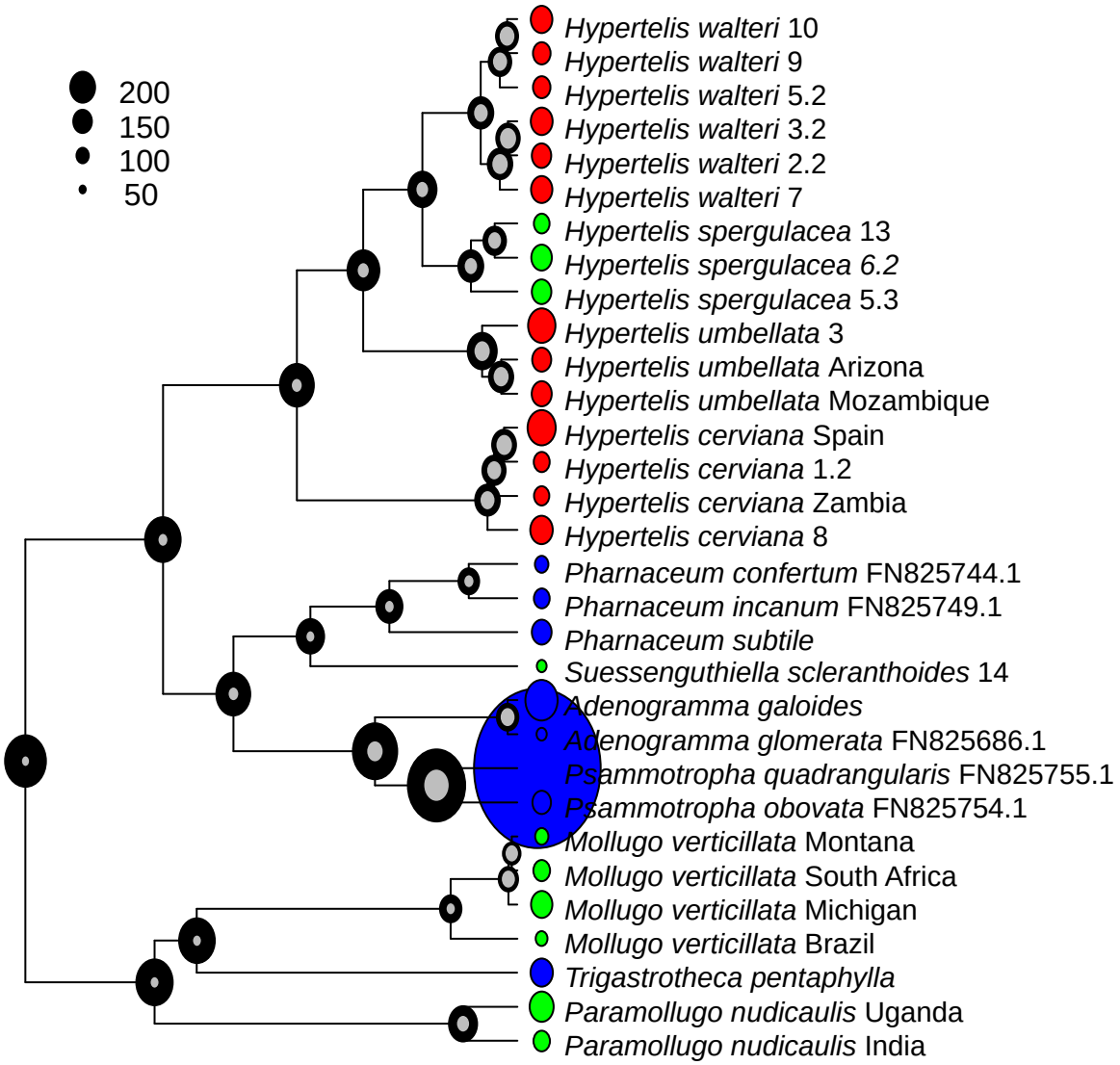

Figure 2.6: Interveinal distance mapped on Molluginaceae phylogeny.

The interveinal distance (IVD) was mapped on the phylogeny inferred from the chloroplast map marker *trnK-matK*. Circles at tips are proportional to the measured IVD and are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Circles at nodes are proportional to the lower (in grey) and upper (in black) bounds of the 95% confidence intervals of inferred ancestral states.

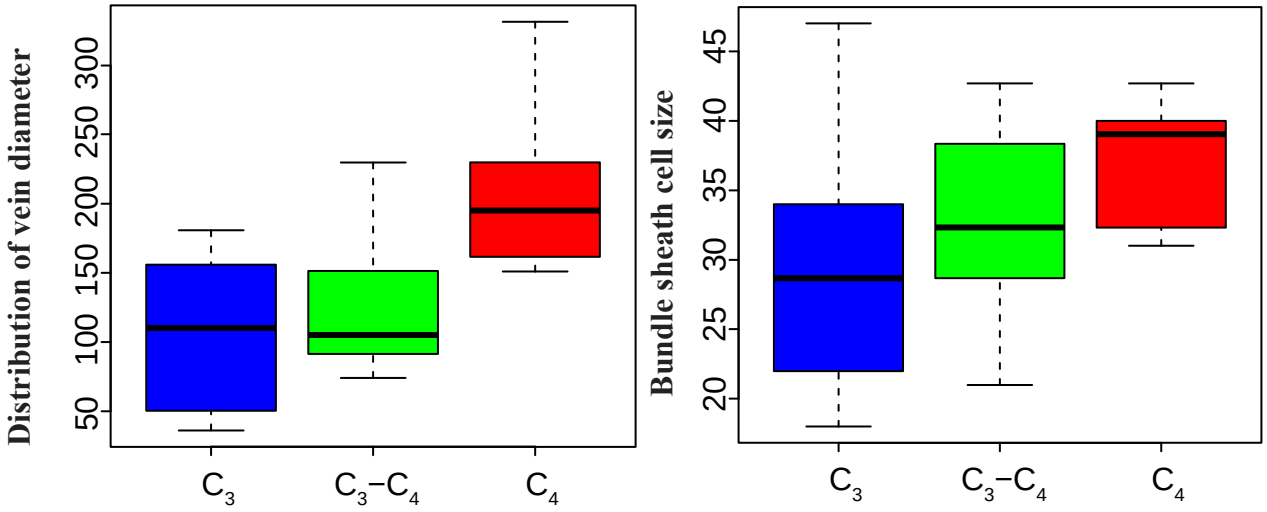

Figure 2.7: Distribution of vein diameter (left) and bundle sheath cell size (right) among photosynthetic types of Molluginaceae.

For each photosynthetic type, the distribution is shown with boxplots. Boxes connect the 25th and 75th percentiles, with medians indicated by thick lines. Whiskers connect the maximal values within 1.5 times the interquartile range

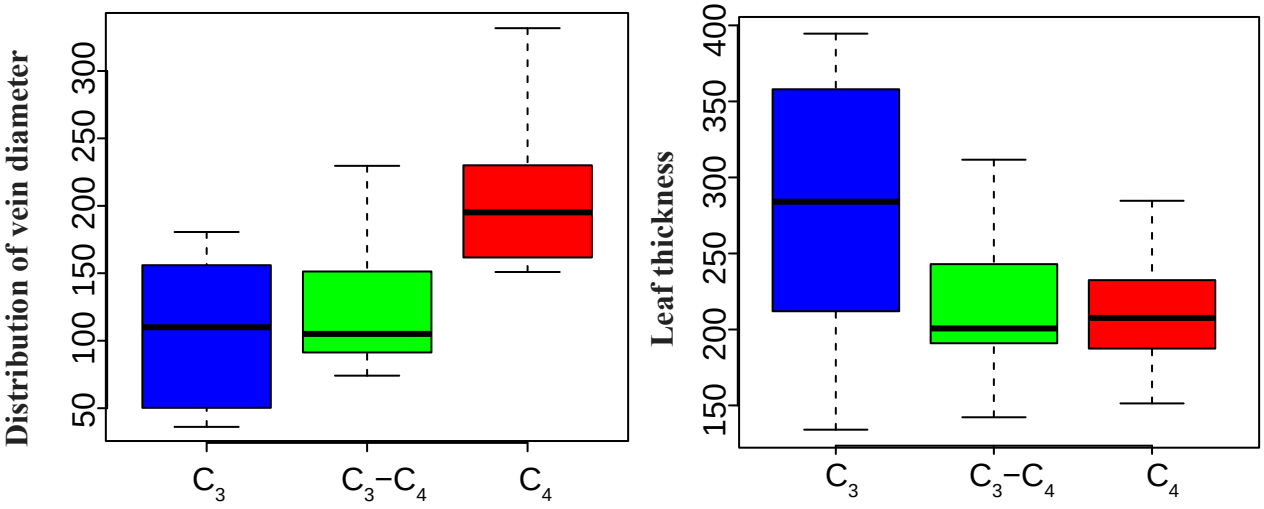

Figure 2.8: Distribution of vein diameter (left) and leaf thickness (right) among photosynthetic types of Molluginaceae.

For each photosynthetic type, the distribution is shown with boxplots. Boxes connect the 25th and 75th percentiles, with medians indicated by thick lines. Whiskers connect the maximal values within 1.5 times the interquartile range.

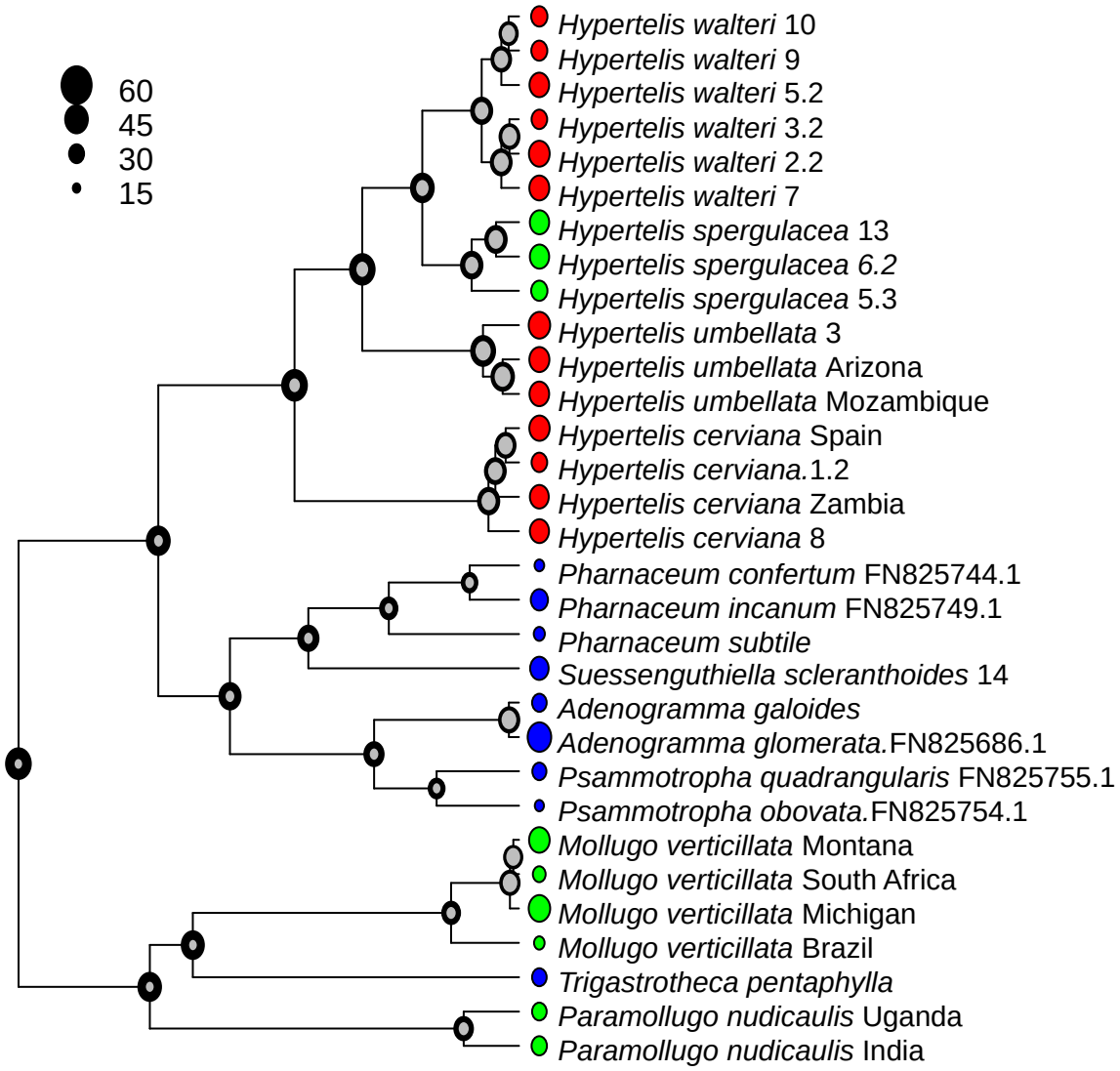

Figure 2.9: Bundle sheath cell size mapped on Molluginaceae phylogeny.

The bundle sheath cell size was mapped on the phylogeny inferred from the chloroplast map marker *trnK-matK*. Circles at tips are proportional to the measured values and are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Circles at nodes are proportional to the lower (in grey) and upper (in black) bounds of the 95% confidence intervals of inferred ancestral states.

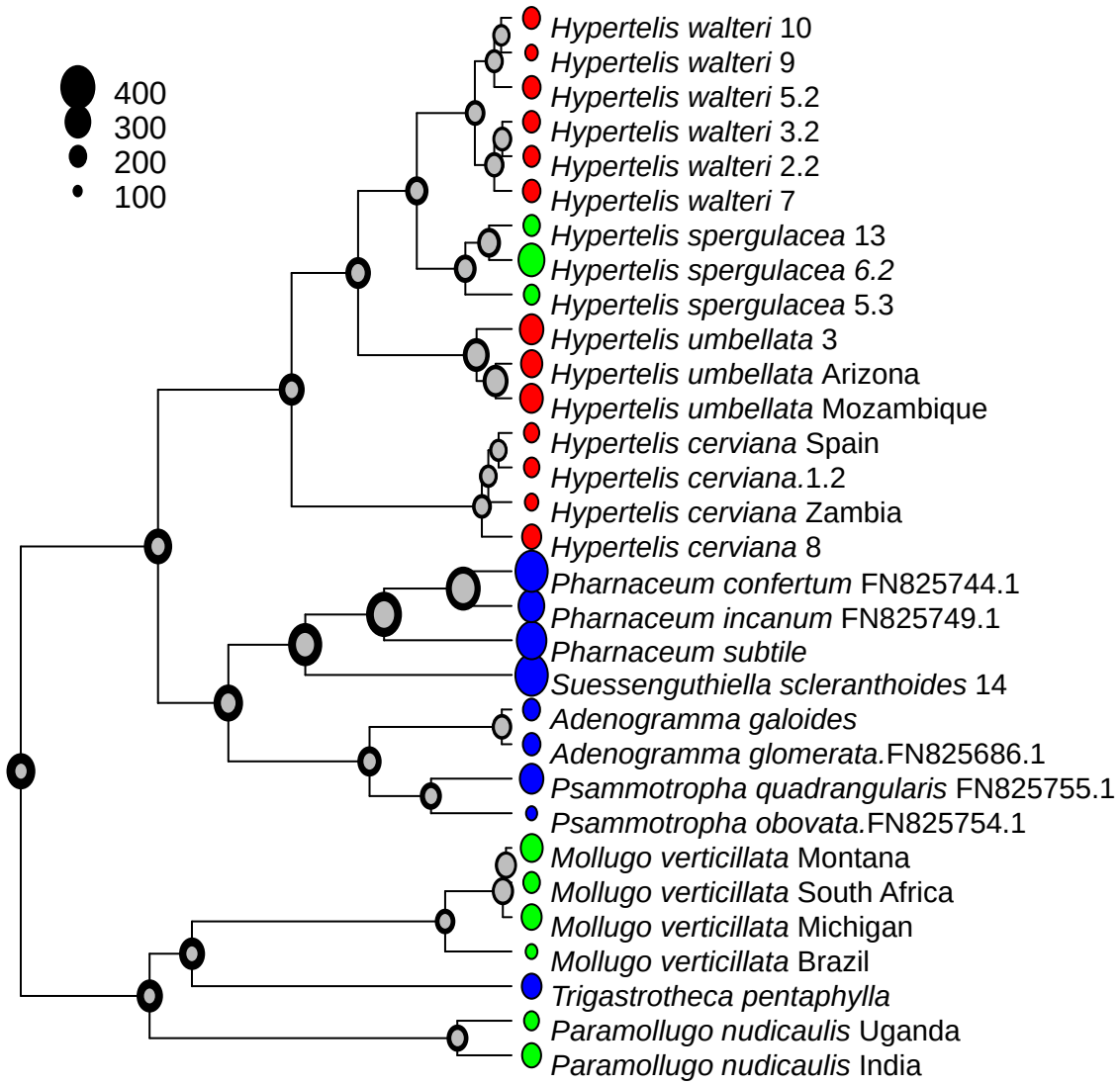

Figure 2.10: Leaf thickness mapped on Molluginaceae phylogeny.

The leaf thickness was mapped on the phylogeny inferred from the chloroplast map marker *trnK-matK*. Circles at tips are proportional to the measured values and are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Circles at nodes are proportional to the lower (in grey) and upper (in black) bounds of the 95% confidence intervals of inferred ancestral states.

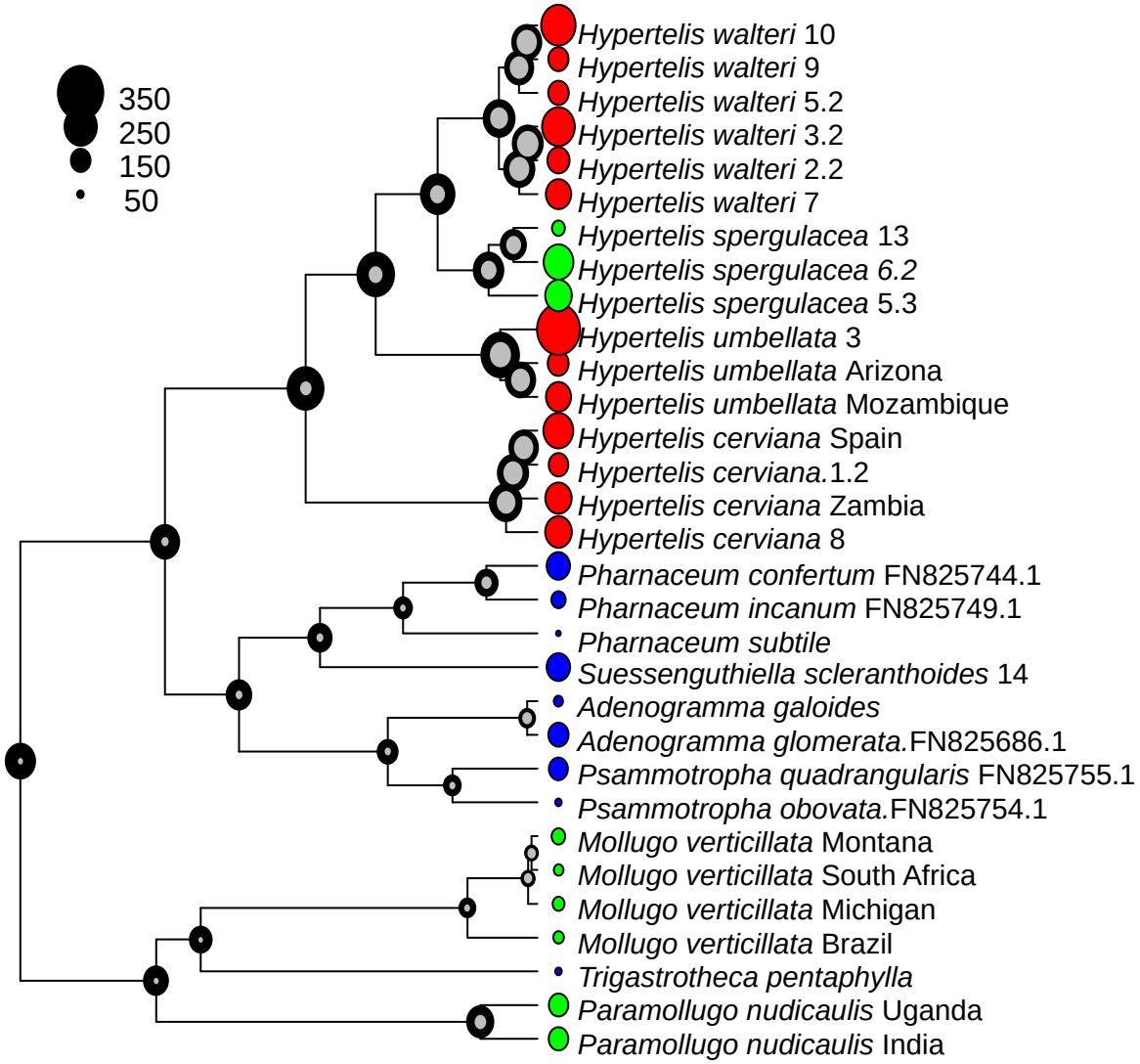

Figure 2.11: Vein diameter mapped on Molluginaceae phylogeny.

The leaf thickness was mapped on the phylogeny inferred from the chloroplast map marker *trnK-matK*. Circles at tips are proportional to the measured values and are colored per photosynthetic type; blue = C_3, green = C_3-C_4, red = C_4. Circles at nodes are proportional to the lower (in grey) and upper (in black) bounds of the 95% confidence intervals of inferred ancestral states.

2.5 Discussion

2.5.1 C_4 anatomy in Molluginaceae evolved via large bundle sheath and large vein diameter.

As expected, C_4 Molluginaceae have a larger bundle sheath per mesophyll area, when compared to C_3 plants from the same family (Fig. 2.1). This proportion is an emergent property that results from the number and sizes of the different tissue types, and none of these tissue-specific characters differs consistently between C_3 and C_4 Molluginaceae (Figs 2.5, 2.7, 2.8). This confirms that, above the cellular level, there is no perfectly diagnostic C_4 anatomical character, as concluded previously for the grass family (Lundgren et al., 2014). Instead, the increased bundle sheath proportion of C_4 Molluginaceae emerges via a unique combination of characters, that are found in isolation in C_3 species but not in combination (Fig. 2.2).

Based on the phylogenetic analyses, the increase of the proportion of bundle sheath happened mainly on the branch leading to the *Hypertelis* clade, which contains both C_4 and intermediate species (Fig. 2.3). Of the multiple variables that together determine this proportion, only the diameter of veins markedly changed along this branch (Fig. 2.11). This indicates that the C_4 anatomy of Molluginaceae emerged mainly via an increase of vein diameters with only small adjustments of the other characters. A previous, qualitative comparison of leaf anatomy among Molluginaceae indicated that bundle sheath cell sizes was modified in branches leading to the *Hypertelis* clade plus its sister group (Christin et al., 2011), a pattern also suggested by our quantitative analyses based on a larger sample size (Fig. 2.7). However, the previous analysis assigned a decrease of interveinal distance to the branch leading to the *Hypertelis* clade (Fig. 2.6). (Christin et al., 2011). Our detailed analyses

confirm changes in the vein architecture at the base of the clade, but indicate that it is not the density of veins per segment that changed, but their diameter (Fig. 2.11).

Increased vein diameter, coupled with bundle sheath cells that are constrained to the upper range of values observed in C_3 Molluginaceae, effectively reduces the distance between consecutive bundle sheaths. Such a short distance is characteristic of the gross C_4 phenotype, and has been reported across a range of taxa (Brown and Hattersley, 1989; Lundgren et al., 2019). However, detailed analyses of the causal properties of the cells indicate that such a phenotype is reached via different mechanisms in distinct groups. In *Alloteropsis semialata*, short distances between consecutive bundle sheaths are achieved via the proliferation of higher order veins (Lundgren et al., 2019), mirroring patterns in the eudicot *Flaveria* (Mckown and Dengler, 2007). By contrast, other grasses achieve short distance between consecutive bundle sheaths by decreasing the number of mesophyll separating them or evolving large bundle sheath cells (Christin et al., 2013; Lundgren et al., 2014). Finally, a number of C_4 Chenopodiaceae evolved continuous bundle sheaths that surround multiple veins (Fisher et al., 1997; Kadereit et al., 2003; Sage, 2004). The mechanism suggested here for the evolution of C_4-specific leaf anatomy in Molluginaceae therefore reinforces the idea that the convergent origins of C_4 photosynthesis used a variety of evolutionary trajectories, which might have facilitated the recurrent emergence of this complex trait in different parts of the angiosperm phylogeny.

2.5.2 The C_3-C_4 ancestral state of Hypertelis constrained the anatomy to C_4-compatible properties.

The distribution of photosynthetic types within *Hypertelis* clade, with the putative C_3-C_4 intermediate *H. spergulacea* nested within three distinct C_4 species (Fig. 2.3), might be interpreted as evidence of a single C_4 origin at the base of the clade followed by a reversal to an intermediate stage in *H. spergulacea*. Establishing the exact order of events leading to the evolutionary diversification of photosynthetic types requires a detailed investigation of individual C_4 component (Christin et al., 2010; Dunning et al., 2017). Analyses of genes encoding the key C_4 enzyme phosphoenolpyruvate carboxylase have established that this gene was adapted for the C_4 function independently in *H. umbellata* and *H. cerviana* (Christin et al. 2011). The genetic determinants of the C_4 pathway of *H. walteri* are yet to be analysed, but the independent adaptation of enzymes for the C_4 context in the two other species suggests that a complete C_4 pathway evolved three times in the *Hypertelis* clade. The common ancestor of the group was likely a C_3-C_4 intermediate, which facilitated recurrent transitions to a full C_4 state in this small group of plants.

Based on the amount of different tissue types, cell sizes and numbers, the four species species of *Hypertelis* are indistinguishable. Indeed, while leaf thickness and size varies both among and within species, individual cells change proportionally so that the ratio of bundle sheath and mesophyll tissues is approximately constant (Figs. 2.2 and 2.4). This suggests that the anatomy observed in extant *Hypertelis* species did not undergo marked changes since the most recent common ancestor of the group, leading to the conclusion that transitions from C_3-C_4 *Hypertelis* involved no important modifications of the gross leaf anatomy besides the redistribution of organelles among cell types. The C_3-C_4 intermediate type, which can involve

solely a photorespiratory pump (also referred to as 'C_2 photosynthesis' (Monson et al., 1986; Hylton et al., 1988; Khoshravesh et al., 2016) or a weak C_4 pathway, has anatomical requirements similar to those of a complete C_4 trait (Sage et al., 2012; Mallmann et al., 2014). The intermediacy at the base of the *Hypertelis* clade therefore constrained leaf anatomy to those combinations of traits that conferred a large proportion of bundle sheath tissue and a small distance between consecutive bundle sheaths. In *Hypertelis*, the C_3-C_4 therefore both bridged the anatomical gap between C_3 and C_4 leaves and reduced the chance of later departures from a C_4 compatible anatomy. The persistence of a C_4-compatible anatomy then facilitated recurrent transitions to full C_4 traits via upregulation of enzymes of the C_4 cycle.

2.5.3 Only some C_3-C_4 lineages act as evolutionary intermediates.

Molluginaceae contain three groups of C_3-C_4 accessions, which are separated in the phylogeny by C_3 species and likely represent independent origins of the intermediate physiology (Christin et al. 2011). The leaf anatomy varies among the three groups, and while *H. spergulacea* has a proportion of bundle sheath equivalent to C_4 species, *M. verticillata* individuals have consistently less bundle sheath than the C_4 species and lie at the upper range of values observed among C_3 Molluginaceae (Fig. 2.3). At the other end of the spectrum, the leaves of *P. nudicaulis* have a low amount of bundle sheath tissue that is typical of C_3 Molluginaceae (Fig. 2.3). The C_4 trait emerged three times within the *Hypertelis* clade, but never from the *Mollugo* and *Paramollugo* groups, although their intermediate types are potentially old (Christin et al. 2011). It has been speculated that these groups might lack C_4 biochemical components (Edwards and Donoghue, 2013), but our investigations show that their anatomy are distant from those observed in C_4 species. We therefore suggest that

variation among C_3-C_4 lineages means that some of them are more likely to act as facilitators of C_4 evolution.

The variation in the degree of leaf specialization among C_3-C_4 lineages of Molluginaceae might depend on the strength of their C_4 cycles. The C_3-C_4 *M. verticillata* expresses a partial C_4 cycle (Sayre and Kennedy, 1979; Hylton et al., 1988; Stata et al., 2019), while *P. nudicaulis* has been described as a C_3-C_4 species without any C_4 activity (Christin and Osborne, 2014; Stata et al., 2019). The biochemical status of *H. spergulacea* is not known, but a stronger C_4 cycle might have favoured the acquisition of more C_4-like leaf characters in this species. If the strength of the C_4 cycle indeed correlates with C_4-like anatomical characters, the question remains of why a stronger C_4 cycle did not evolve in each of the C_3-C_4 groups. Models indeed predict the rapid emergence of a C_4 cycle in C_3-C_4 species to rebalance nitrogen among cells (Mallmann et al. 2013), but the selective pressures might vary across environments.

The two C_3-C_4 species *M. verticillata* and *P. nudicaulis* are weeds that grow in disturbed, open habitats spread across the seasonally warm regions of world (Christin et al., 2011; Lundgren and Christin, 2017; Sage et al., 2018). By contrast, *H. spergulacea* is restricted to semi-deserts from Namibia and South Africa, where extreme temperatures lead to strong photorespiratory stresses. Importantly, the group of species sister to *Hypertelis* are from the same region. We therefore propose that the C_3-C_4 type ancestral to *Hypertelis* emerged in arid regions from southern Africa and allowed the colonization of progressively warmer and more arid ecological niches. Strong selection in these habitats lead to a progressively stronger C_4 cycle, which involved alterations of the leaf anatomy that facilitated later C_4 origins. The success of the two other C_3-C_4 Molluginaceae in wetter and cooler

habitats limited selection for strengthened C_4 cycles, which ultimately limited the accessibility of the C_4 trait. Therefore, the role of C_3-C_4 as evolutionary intermediates depends on the specifics of their phenotypes that is affected by their geographical origins and ecological specialization.

2.6 Conclusions

In this study, we compared in a phylogenetic context the leaf anatomy of Molluginaceae species differing in their photosynthetic type. We show that the large proportion of bundle sheath tissue characterizing the three C_4 species results from a unique combination of individual characters observed among C_3 samples with an increase of the size of veins. One of the C_3-C_4 species is indistinguishable from its close C_4 relatives based on the gross leaf anatomy, which confirms that the C_3-C_4 type can act as an evolutionary intermediate, efficiently bridging the anatomical gap between C_3 and C_4 Molluginaceae. By contrast, the two C_3-C_4 lineages that never gave rise to C_4 descendants harbour a leaf that is more distant to the C_4 requirements, and we suggest that the contrasted ecology of the three C_3-C_4 lineages lead to a variety of selected leaf phenotypes, only one of which was compatible with a full C_4 cycle. The study of this small group of plants with a large photosynthetic diversity indicates that the role of C_3-C_4 intermediates as evolutionary facilitators depends on the implementation of their photosynthetic machinery, which is affected by their history as well as their eco-physiological strategies.

2.7 References

Brockington, S.F., Walker, R.H., Glover, B.J., Soltis, P.S. and Soltis, D.E. 2011. Complex pigment evolution in the Caryophyllales. *New Phytologist*. **190**(4), pp.854–864.

Brown, R.H. and Hattersley, P.W. 1989. Leaf anatomy of C_3-C_4 species as related to evolution of C_4 photosynthesis. *Plant Physiology*. **91**(4), pp.1543–1550.

Busch, F.A., Sage, R.F. and Farquhar, G.D. 2017. Plants increase CO_2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nature Plants*. **4**, pp.46–54.

Cacefo, V., Ribas, A.F., Zilliani, R.R., Neris, D.M., Domingues, D.S., Moro, A.L. and Vieira, L.G.E. 2019. Decarboxylation mechanisms of C_4 photosynthesis in *Saccharum* spp.: Increased PEPCK activity under water-limiting conditions. *BMC Plant Biology*. **19**(1), pp.1–14.

Christin, P.A. and Osborne, C.P. 2014. The evolutionary ecology of C_4 plants. *New Phytologist*. **204**(4), pp.765–781.

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 C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences: USA*. **110**(4), pp.1381–1386.

Christin, P.A. and Besnard, G. 2009. Two independent C_4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *American Journal of Botany*. **96**(12), pp.2234–2239.

Christin, P.A., Freckleton, R.P. and Osborne, C.P. 2010. Can phylogenetics identify C_4 origins and reversals? *Trends in Ecology and Evolution*. **25**(7), pp.403–409.

Christin, P.A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R. and Sage, R.F. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution*. **65**(3), pp.643–660.

Dengler NG, Dengler RE, Donnelly PM, H.P. 1994. Quantitative leaf anatomy of C_3 and C_4 grasses (Poaceae): bundle sheath and mesophyll surface area relationships. *Annals of Botany*. **73**, pp.241–255.

Dieckmann, U. and Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature*. **400**(6742), pp.354–357.

Drummond, A.J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. **7**(214), pp.1–8.

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 C_4 photosynthesis among close relatives. *Evolution*. **71**(6), pp.1541–1555.

Edwards, E.J. and Donoghue, M.J. 2013. Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany*. **64**(13), pp.4047–4052.

Edwards, E.J., Osborne, C.P., Stromberg, C.A.E., Smith, S.A. and C_4 Grasses Consortium. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science*. **328**(5978), pp.587–591.

Ehleringer, J.R., Cerling, T.E. and Helliker, B.R. 1997. C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*. **112**(3), pp.285–299.

Fisher, D.D., Schenk, H.J., Thorsch, J.A. and Ferren, W.R. 1997. Leaf anatomy and subgeneric affiliations of C_3 and C_4 species of *Suaeda* (Chenopodiaceae) in North America. *American Journal of Botany*. **84**(9), pp.1198–1210.

Furbank, R. and Taylor, W. 1995. Regulation of photosynthesis in C_3 and C_4 plants: a molecular approach. *The Plant Cell*. **7**(7), pp.797–807.

Hatch, M. 1987. C_4 Photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta*. **895**(2), pp.81–106.

Hatch, M.D. and Slack, C.R. 1966. Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *The Biochemical Journal*. **101**(1), pp.103–11.

Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A. and Woolhouse, H.W. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. *Planta*. **175**(4), pp.452–459.

Iglesias, A.A., González, D.H. and Andreo, C.S. 1986. The C_4 pathway of photosynthesis and its regulation. *Biochemical Education*. **14**(3), pp.98–102.

Kadereit, G., Ackerly, D. and Pirie, M.D. 2012. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae). *Proceedings of the Royal Society B: Biological Sciences*. **279**(1741), pp.3304–3311.

Kadereit, G., Borsch, T., Weising, K. and Freitag, H. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences*. **164**(6), pp.959–986.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I. and Whitlock, M.C. 2012. Experimental evolution. *Trends in Ecology & Evolution*. **27**(10), pp.547–560.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. **28**(12), pp.1647–1649.

Khoshravesh, R., Stinson, C.R., Stata, M., Busch, F.A., Sage, R.F., Ludwig, M. and Sage, T.L. 2016. C_3-C_4 intermediacy in grasses: Organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C_2 photosynthesis. *Journal of Experimental Botany*. **67**(10), pp.3065–3078.

Lamb, T.D., Collin, S.P. and Pugh, E.N. 2007. Evolution of the vertebrate eye: Opsins, photoreceptors, retina and eye cup. *Nature Reviews Neuroscience*. **8**(12), pp.960–976.

Lauterbach, M., Zimmer, R., Alexa, A.C., Adachi, S., Sage, R., Sage, T., MacFarlane, T., Ludwig, M. and Kadereit, G. 2019. Variation in leaf anatomical traits relates to the evolution of C_4 photosynthesis in Tribuloideae (Zygophyllaceae). *Perspectives in Plant Ecology, Evolution and Systematics*. **39**(May), p.125463.

Lenski, R.E., Ofria, C., Pennock, R.T. and Adami, C. 2003. The evolutionary origin of complex features. *Nature*. **423**(6936), pp.139–144.

Lundgren, M.R. and Christin, P.A. 2017. Despite phylogenetic effects, C_3-C_4 lineages bridge the ecological gap to C_4 photosynthesis. *Journal of Experimental Botany*. **68**(2), pp.241–254.

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. C_4 anatomy can evolve via a single developmental change. *Ecology Letters*. **22**(2), pp.302–312.

Lundgren, M.R., Osborne, C.P. and Christin, P.A. 2014. Deconstructing Kranz anatomy to understand C_4 evolution. *Journal of Experimental Botany*. **65**, pp.3357–3369.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P. and Gowik, U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife*. **2014**(3), pp.2–5.

Mckown, A.D. and Dengler, N.G. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany*. **94**, pp.382–399.

Monson, R.K., Moore, B., Ku, M.S.B. and Edwards, G.E. 1986. Co-function of C_3- and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. *Planta*. **168**, pp.493–502.

Monteiro, A. and Podlaha, O. 2009. Wings, horns, and butterfly eyespots: How do complex traits evolve? *PLoS Biology*. **7**(2), pp.0209–0216.

Moreno-Villena, J.J., Dunning, L.T., Osborne, C.P. and Christin, P.A. 2018. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Molecular Biology and Evolution*. **35**(1), pp.94–106.

Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G. and Sage, R.F. 2011. Characterization of C_3-C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. *Plant, Cell and Environment*. **34**(10), pp.1723–1736.

Ogburn, R.M. and Edwards, E.J., 2013. Repeated origin of three-dimensional leaf venation releases constraints on the evolution of succulence in plants. *Current Biology*, **23**(8), pp.722-726.

Osborne, C.P. and Sack, L. 2012. Evolution of C_4 plants: A new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **367**(1588), pp.583–600.

Paradis, E., Claude, J. and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20**(2), pp.289–290.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Sage, R.F. 2004. The evolution of C_4 photosynthesis. *New Phytologist*. **161**, pp.341–370.

Sage, R.F. 2016. Tracking the evolutionary rise of C_4 metabolism. *Journal of Experimental Botany*. **67**(10), pp.2919–2922.

Sage, R.F., Christin, P.A. and Edwards, E.J. 2011. The C_4 plant lineages of planet Earth. *Journal of Experimental Botany*. **62**(9), pp.3155–3169.

Sage, R.F., Monson, R.K., Ehleringer, J.R., Adachi, S. and Pearcy, R.W. 2018. Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia*. **187**(4), pp.941–966.

Sage, R.F., Sage, T.L. and Kocacinar, F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*. **63**(1), pp.19–47.

Sayre, R.T. and Kennedy, R. A. 1979. Photosynthetic enzyme activities and localization in *Mollugo verticillata* populations differing in the levels of C_3 and C_4 cycle operation. *Plant Physiology*. **64**(2), pp.293–299.

Sayre, R.T. and Kennedy, R.A. 1977. Ecotypic differences in the C_3 and C_4 photosynthetic activity in *Mollugo verticillata*, a C_3-C_4 intermediate. *Planta*. **262**, pp.256–262.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. **9**(7), pp.671–675.

Schuler, M. L., Mantegazza, O., and Weber, A. P. M. (2016). Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. *The Plant Journal*. **87**(1), pp.51–65.

Slack, C. and Hatch, M. 1967. Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochemical Journal*. **103**(3), pp.660–665.

Stata, M., Sage, T.L. and 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**(M), pp.27–34.

Thulin, M., Moore, A.J., El-seedi, H., Larsson, A., Christin, P. and Edwards, E.J. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. *Taxon*. **65**(4), pp.775–793.

Voznesenskaya, E. V., Franceschi, V.R., Pyankov, V.I. and Edwards, G.E. 1999. Anatomy, chloroplast structure and compartmentation of enzymes relative to photosynthetic mechanisms in leaves and cotyledons of species in the tribe Salsoleae (Chenopodiaceae). *Journal of Experimental Botany*. **50**(341), pp.1779–1795.

Wang, S., Tholen, D. and Zhu, X.G. 2017. C_4 photosynthesis in C_3 rice: a theoretical analysis of biochemical and anatomical factors. *Plant Cell and Environment*. **40**, pp.80–94.

Webster, G.L., Brown, W. V and Smith, B.N. 1975. Systematics of photosynthetic carbon fixation pathways in *Euphorbia*. *Taxon*. **24**, pp.27–33.

Appendix for Chapter 1

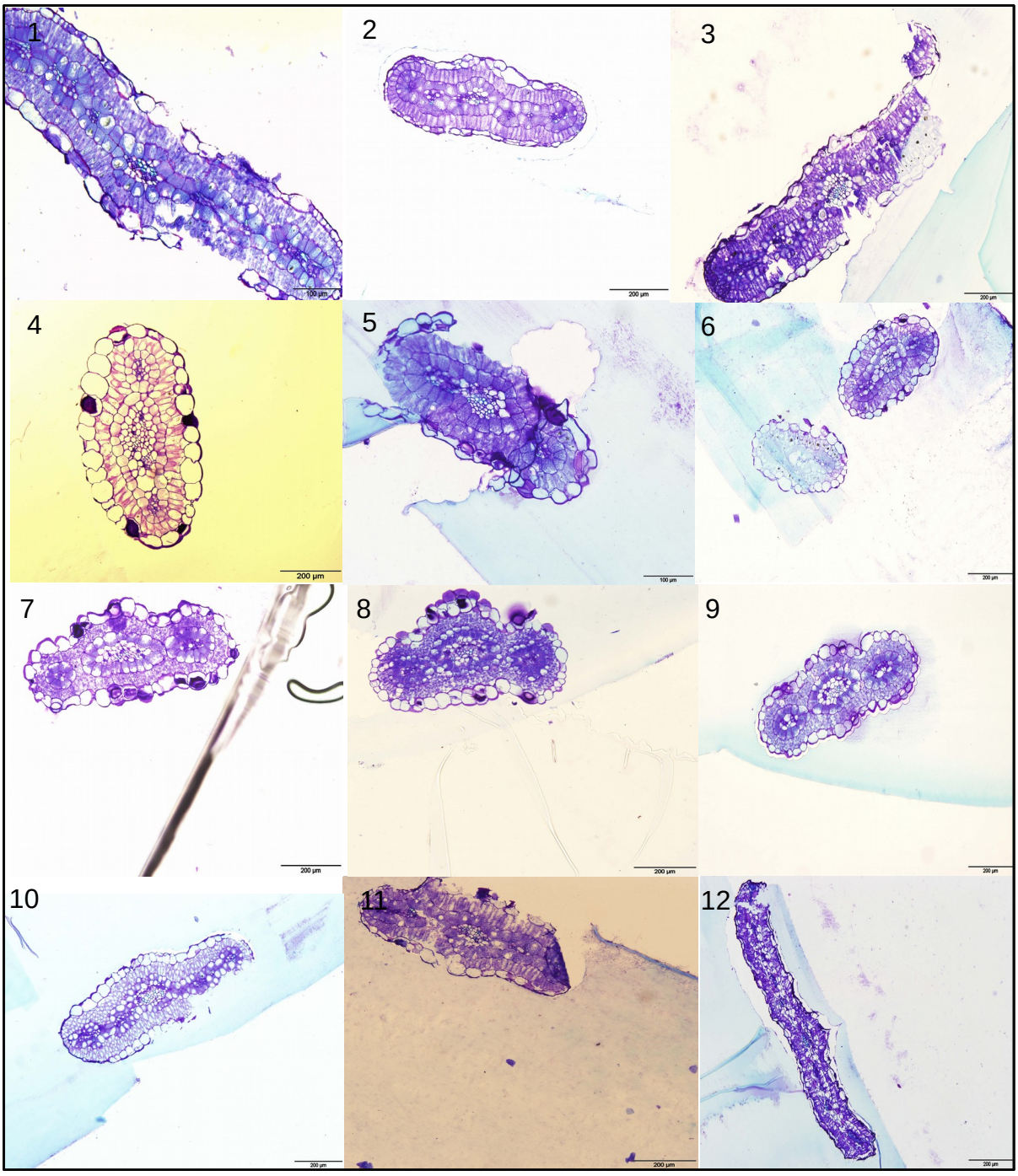

Figure S2.1. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Hypertelis cerviana* 1.2 (C_4), [4-6] *Hypertelis cerviana* 8 (C_4), [7-9] *Hypertelis cerviana* Spain (C_4) and [10-12] *Hypertelis cerviana* Zambia (C_4).

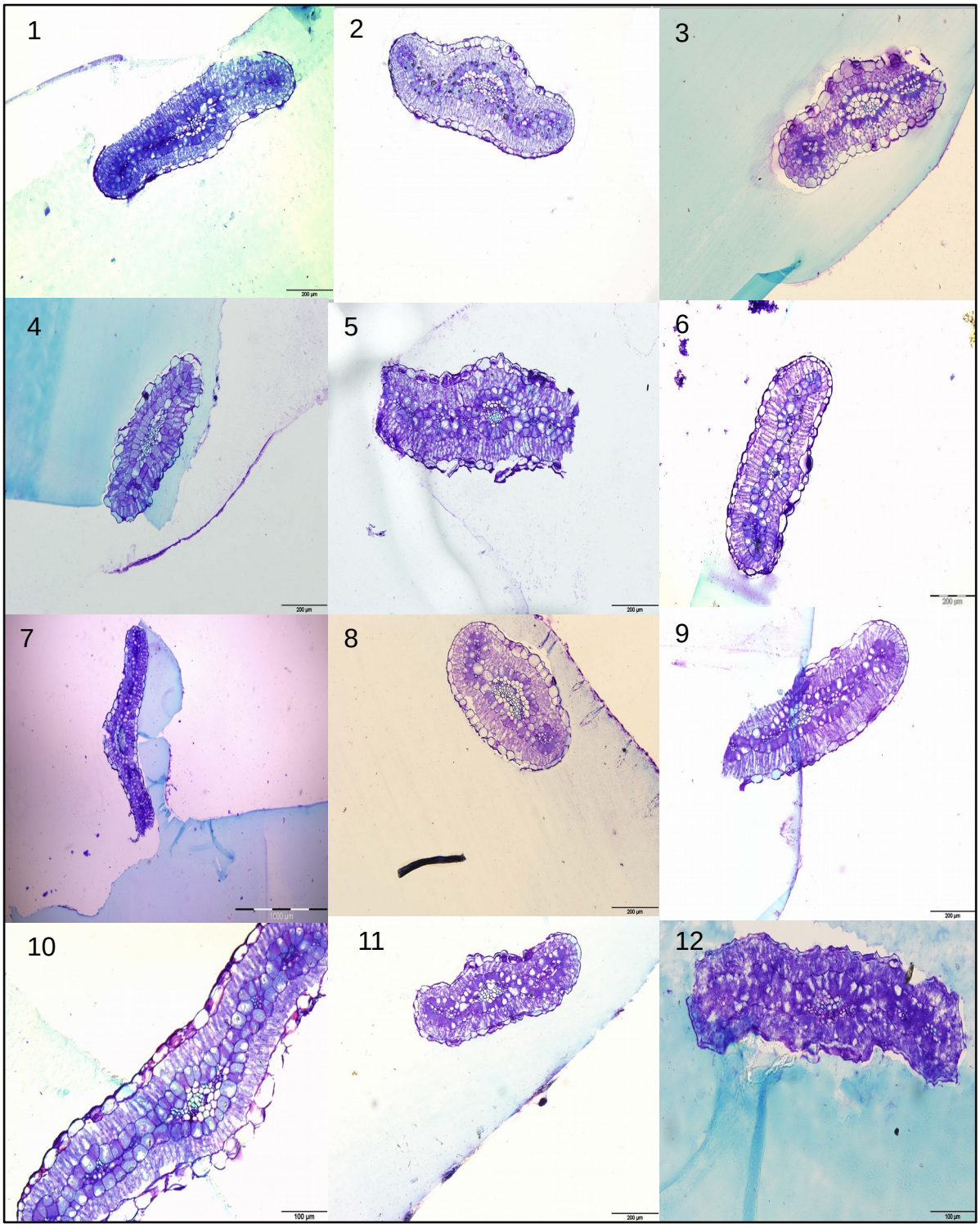

Figure S2.2. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Hypertelis walteri* 2.2 (C_4), [4-6] *Hypertelis walteri* 3.2 (C_4), [7-9] *Hypertelis walteri* 5.2 (C_4) and [10-12] *Hypertelis walteri* 7 (C_4).

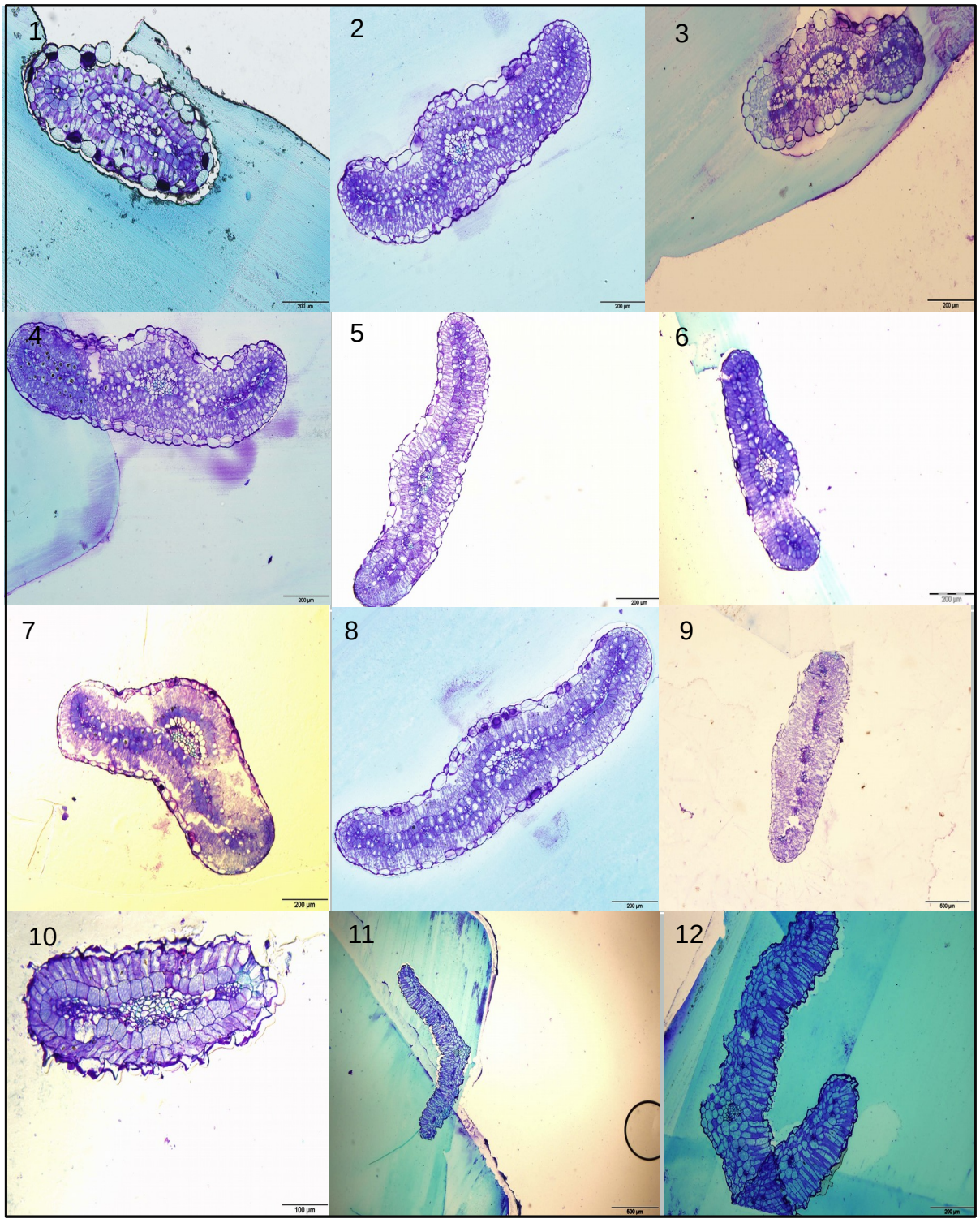

Figure S2.3. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Hypertelis walteri* 10 (C_4), [4-6] *Hypertelis spergulacea* Namibia (C_3-C_4), [7-9] *Hypertelis spergulaceae* 6.2 (C_3-C_4) and [10-12] *Hypertelis spergulaceae* 13 (C_3-C_4).

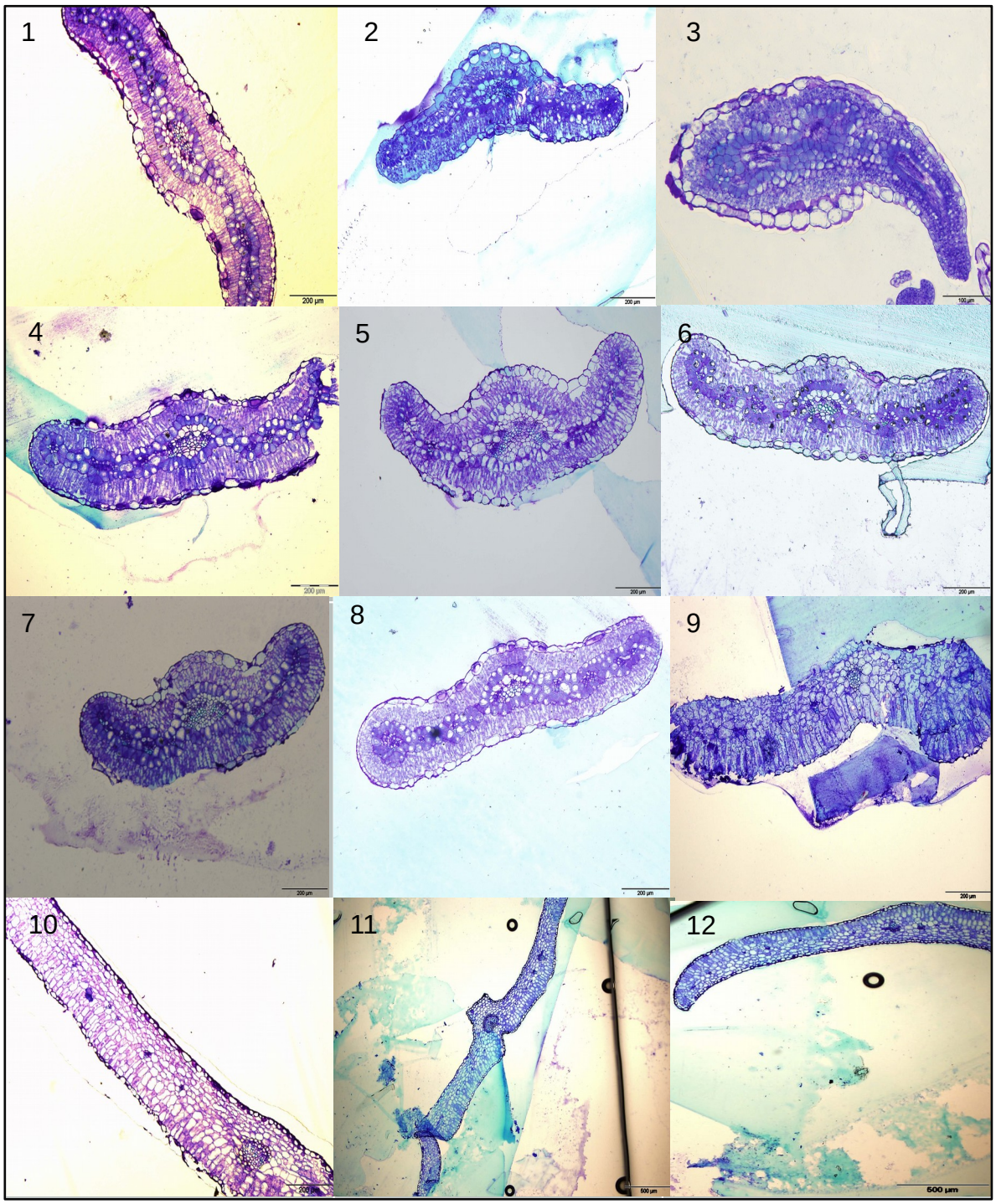

Figure S2.4. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Hypertelis umbellata* Mozambique (C_4), [4-6] *Hypertelis umbellata* 3 (C_4), [7-9] *Hypertelis umbellata* Arizona (C_4) and [10-12] *Trigastrotheca pentaphylla* (C_3).

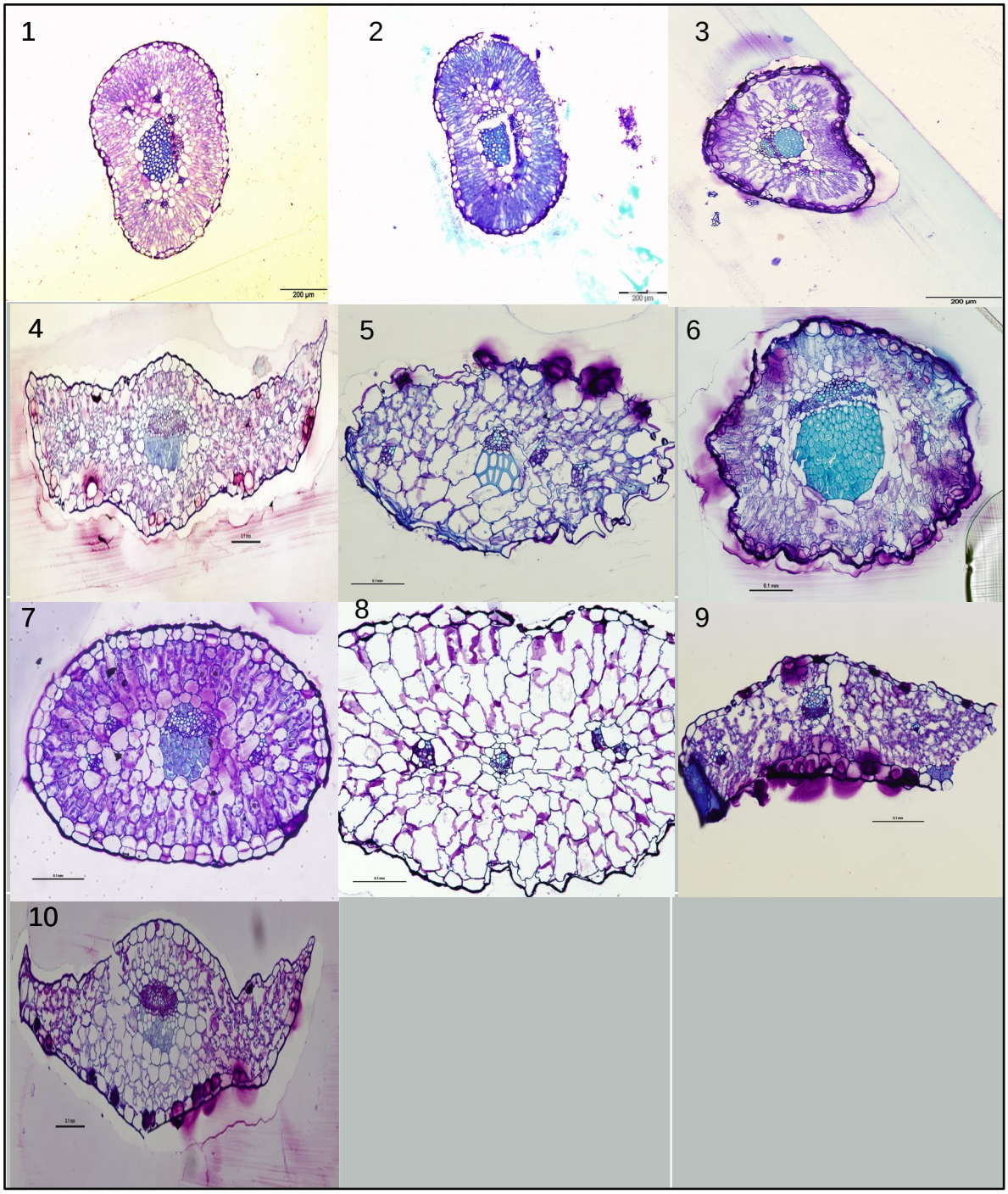

Figure S2.5. Light micrographs of cross sections of *Molluginaceae* species.

For each species one sample each of species is shown; [1-3] *Suessenguthiella scleranthoides* 14(C_3), [4] *Adenogramma galoides* (C_3), [5] *Adenogramma glomerata* (C_3), [6] *Pharnaceum confertum* (C_3); [7] *Pharnaceum incanum* (C_3), [8] *Pharnaceum subtile* (C_3), [9] *Psammotropha obovata* (C_3) and [10] *Psammotropha quadrangularis* (C_3)**.**

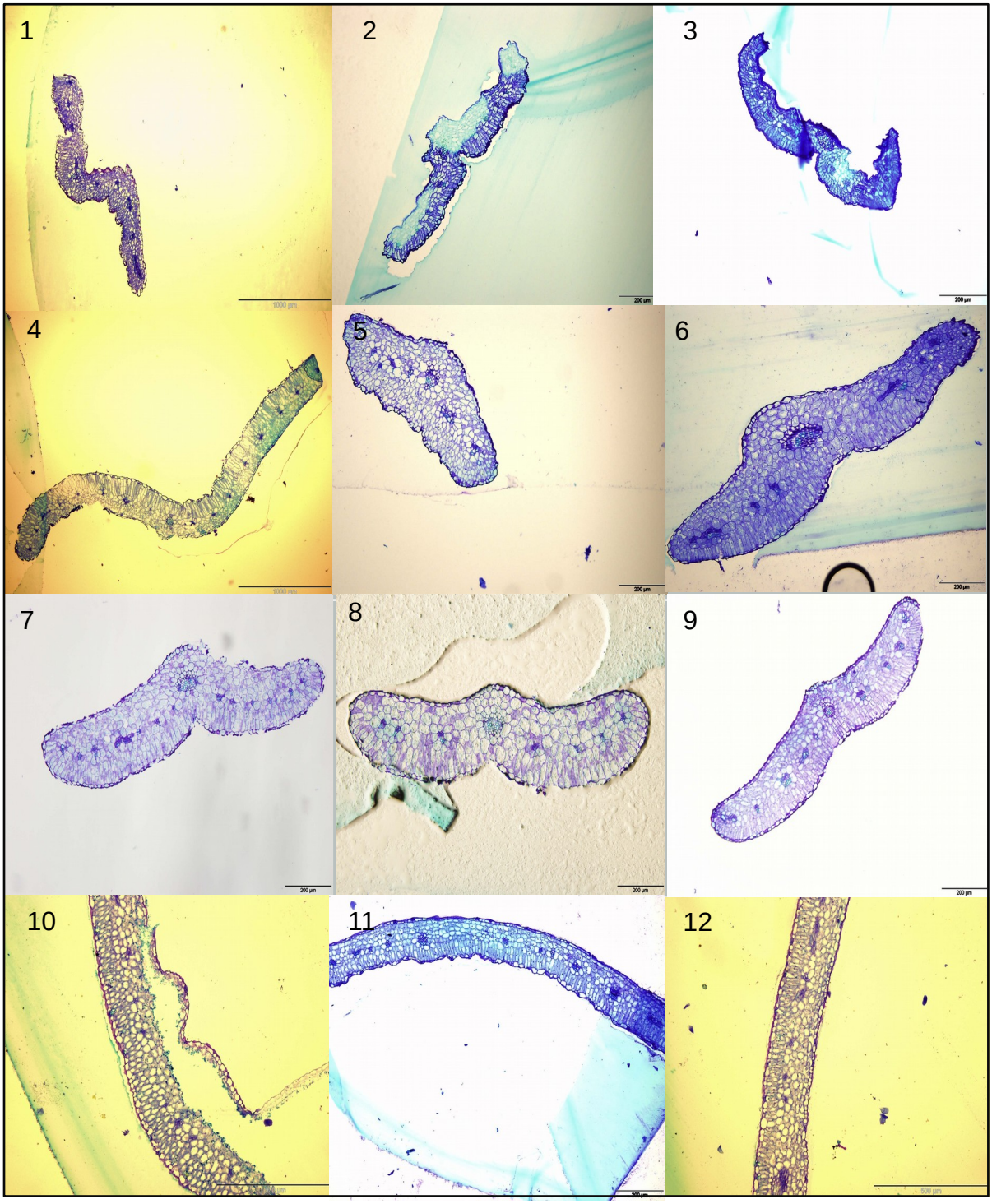

Figure S2.6. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Mollugo verticillata* Brazil (C_3-C_4), [4-6] *Mollugo verticillata* Michigan (C_3-C_4), [7-9] *Mollugo verticillata* Montana(C_3-C_4) and [10-12] Mollugo verticillata South Africa (C_3-C_4).

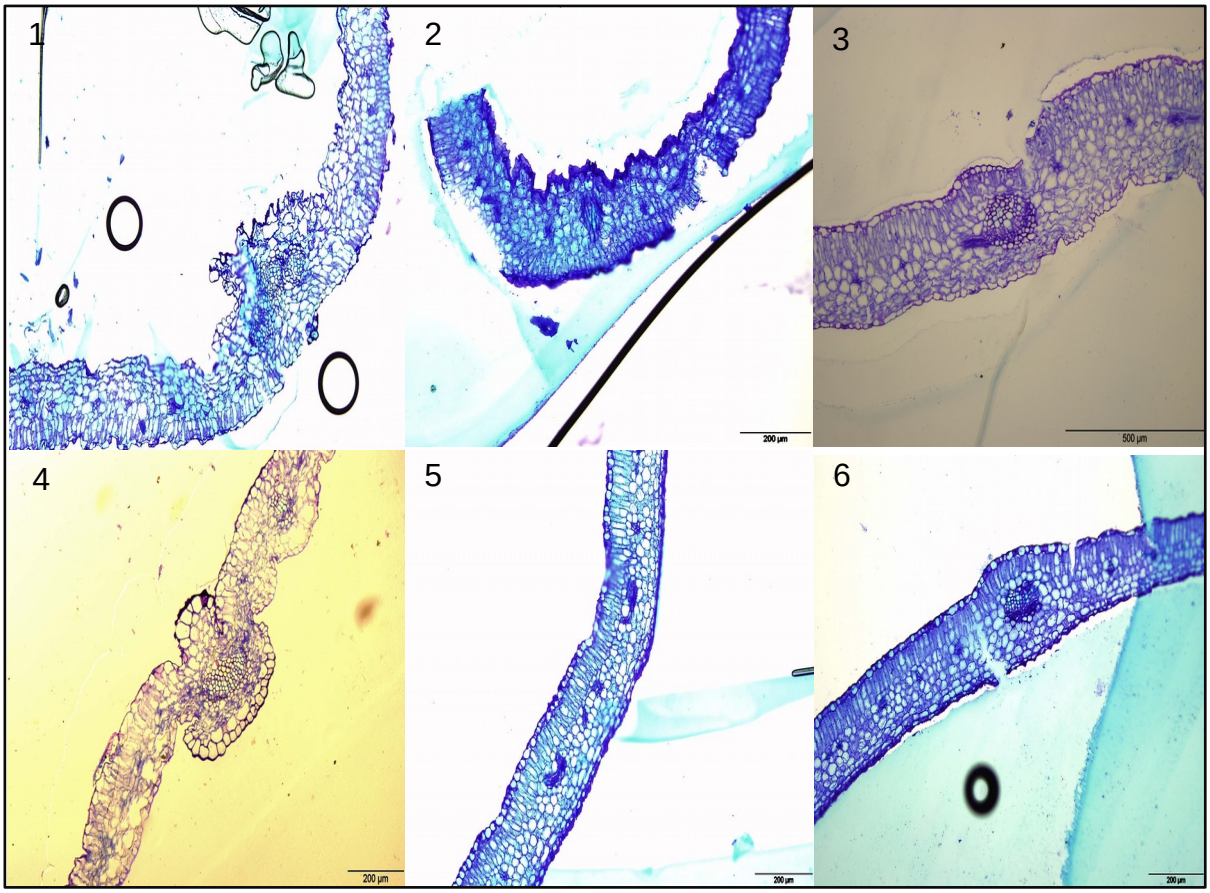

Figure S2.7. Light micrographs of cross sections of *Molluginaceae* species.

For each species, three replicates are shown; [1-3] *Paramollugo nudicaulis* India (C_3-C_4), [4-6] *Paramollugo nudicaulis* Uganda (C_3-C_4).

Chapter 2:
Hybridization might have facilitated the rapid evolution of C_4 photosynthesis in closely related Molluginaceae

Chapter 2: Hybridization might have facilitated the rapid evolution of C_4 photosynthesis in closely related Molluginaceae

by Lamiaa Munshi[1], Jose Moreno-Villena[1], and Pascal-Antoine Christin[1]

[1] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom.

Personal contributions: [I generated the data, with the help of Jose J. Moreno-Villena. Jose Moreno-Villena assembled the transcriptomes and I performed all subsequent analysed. I wrote the manuscript with the help of Pascal-Antoine Christin].

3.1 Abstract

C_4 photosynthesis is a trait that, despite its complexity, has evolved more than 60 times in angiosperms. C_4 origins are especially clustered in some clades, a pattern that has been attributed to the recurrent co-option of anatomical precursors inherited from the common ancestor of these groups. The role of genetic precursors in the addition of a C_4 biochemistry on top of these anatomical traits however remains debated. Here, we analyse the transcriptomes of various Molluginaceae species, including two C_4 species that evolved from a C_3-C_4 intermediate common ancestor. We detect nine C_4 genes that reach high levels in the C_4 species, but these are low in the closely related C_3-C_4 intermediates. We conclude that C_3-C_4 species do not always bridge the gap to C_4 biochemistry, and increased gene expression evolved independently in the two C_4 Molluginaceae. We find evidence of shifts of selective pressures on the co-opted genes, but these are sustained in the descendant branches. These results show that the adaptation of C_4 enzymes starts after their co-option but is then sustained over million of years. Finally, the phylogenetic trees of several of the co-opted genes are compatible with transfers among the C_4 species after the divergence. We suggest that hybridization facilitated rapid C_4 evolution in the group, with a role of the common C_3-C_4 ancestor limited to anatomical characters.

Key words: Biochemical pathway, C_4 photosynthesis, gene expression, Molluginaceae, transcriptomes.

3.2 Introduction

During evolution, plants have adapted to a plethora of living conditions by evolving novel physiological and anatomical traits. The accessibility of new phenotypes varies among lineages, and depends on the ancestral state as well as the evolutionary rate. In particular, the presence of precursors, also called evolutionary enablers, facilitate transitions to derived character. These precursors, which can be inferred along phylogenetic trees (Marazzi et al., 2012), include mutations without direct phenotypic effects (Blount et al., 2012) and visible traits. Because a given precursor can be co-opted by multiple descendants, some novel traits emerge recurrently within some groups of plants. This is especially the case of C_4 photosynthesis, which evolved over the C_3 ancestral condition.

C_4 photosynthesis is a complex trait that results from the coordinated action of multiple anatomical and biochemical components to boost productivity in tropical conditions (Hatch, 1987; Atkinson et al., 2016). It evolved more than 60 times independently in angiosperms, but C_4 origins are unequally spread across the phylogeny (Sage et al., 2011). Indeed, large groups of plants inhabiting tropical regions lack C_4 species, while some families, such as grasses, harbour more than 20 C_4 origins (GPWGII, 2012). These clusters of C_4 origins have been discussed in relation to several factors, include ecology and life history (Kellogg, 1999; Sage, 2001). In grasses, lineages with some C_4-like anatomy are statistically more likely to evolve C_4 photosynthesis (Christin et al., 2013), and close relatives of other C_4 lineages have C_4-like characters (Muhaidat et al., 2011), pointing to anatomical precursors as having important impacts on C_4 evolvability. Other factors might however also play a role. In particular, the presence of genes with suitable expression patterns has been proposed as a likely preconditions to C_4 evolution (Monson, 2003; Moreno-Villena et al., 2018). In addition,

the existence of C_3-C_4 intermediates has been shown to restrict the gap between the C_3 and C_4 phenotypes (Monson et al., 1986; Sage, 2004), both in terms of anatomy (Mckown and Dengler, 2007) and biochemistry (Mallmann et al., 2014). The importance of these different factors has however been studied in only a few groups.

The Molluginaceae family represents an exciting system to study the effects of different factors on facilitating C_4 evolution. It contains a number of C_3 species, but also three C_4 species and three known C_3-C_4 intermediates (Sayre and Kennedy, 1977; Christin et al., 2011; Thulin et al., 2016). Previous phylogenetic investigations have suggested that C_4 photosynthesis evolved twice independently, in the two closely related species that were previously studied (previously treated as the single species *Mollugo cerviana*) that co-opted some anatomical characters inherited from their common C_3-C_4 ancestor (Chapter 1; Christin et al., 2011). However, the genes for C_4 photosynthesis from the family remain largely unstudied, with the exception of those encoding phosphoenopyruvate carboxylase (PEPC) that acquired their C_4 properties independently in each of the two C_4 groups (Christin et al., 2011). It is therefore unknown whether the two C_4 origins inherited both anatomical and biochemical precursors from the common ancestor.

In this study, we used transcriptomes to investigate the expression levels and coding sequences of genes encoding enzymes of the C_4 photosynthetic pathway in various Molluginaceae species. We sampled multiple species of each photosynthetic type and multiple populations per species to test the hypotheses that (1) C_4 evolution in Molluginaceae was facilitated by increased C_4 gene expression in C_3-C_4 species, (2) the co-option of C_4 genes was followed by episodes of positive selection in each origin, and (3) because of the recent origin of C_4 photosynthesis, populations differ in the genetic make-up of their C_4 phenotype. Our

transcriptome analyses instead revealed the importance of gene movements across species boundaries for the recurrent photosynthetic innovation in this small group of plants.

3.3 Material and methods

3.3.1 Plant material and sequencing.

We sampled 14 species of Molluginaceae to represent the phylogenetic diversity in the group and the different C_4 and C_3-C_4 lineages. For widespread species, multiple populations were considered, for a total of 34 accessions (Table 3.1). A species from another family of Caryophyllales was added to the sampling (*Kewa salsoloides*). Plants were grown from seeds at the University of Sheffield. Controlled conditions were maintained using a Conviron chamber, with a temperature of 25/20°C day/night and a 14-hour photoperiod. Plants were grown in individual pots, composed of a 2:1 mixture of M3 compost and sand, and were watered frequently to keep the soil moist.

For each accession, three separate plants were sampled when available to provide biological replicates. In some cases, only one or two plants grew and these were used to represent the population. For each individual, mature and fully expanded leaves were sampled in the middle of the photoperiod (in the time window between 5 and 8 hours after the start of the light). Leaf samples were flash frozen and directly stored at -80 until RNA extractions. RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. The extracted RNA was used to generate cDNA libraries using the Illumina TruSeq mRNA Sample Preparation Kit (Illumina, San Diego, CA), following the provider's instructions. A total of 24 indexed libraries were pooled per lane of flow cell and

sequenced with an Illumina HiSeq 2500 sequencer with 100 cycles in rapid mode, to produce 100bp paired-end reads. Sequencing was performed at the Sheffield Diagnostic Genetics Service, Sheffield Children's hospital, UK.

3.3.2 Transcriptome assembly and phylogenetic annotation.

Raw sequencing reads were cleaned and filtered using the Agalma pipeline version 0.5.0 (Dunn et al., 2012), with default parameters. This pipeline removed low quality reads, reads corresponding to ribosomal RNA, and those containing adaptors. Cleaned reads from all replicates of each population were jointly used for *de novo* transcriptome assembly using Trinity version trinityrnaseq_r20140413p1 (Grabherr et al., 2011).

Transcripts corresponding to genes encoding ten enzymes with a known function in C_4 photosynthesis were identified and annotated using the phylogenetic pipeline from Christin et al. (2015). For each gene family, the dataset covering land plants was retrieved from Christin et al. (2015). These datasets are composed of representatives from various groups with complete genomes, as well as transcripts from Caryophyllales. Homologous transcripts from Molluginaceae were extracted from the transcriptomes generated here. In each case, a BLASTn search was used to identify all transcripts matching any of the sequences from the reference on more than half of their length. All sequences for each gene family were then aligned using Muscle v3.8.31 (Edgar, 2004), and the alignment was manually curated to remove introns and poorly aligned regions on the 5' or 3' ends. A phylogenetic tree was then inferred with PhyML *v. 20120412* (Guindon and Gascuel, 2003), and the longest sequence from each monospecific group of very similar sequences was retained. These Molluginaceae

transcripts were added to the original dataset to constitute the new reference of each gene family. Each contig from each Molluginaceae transcriptome corresponding to a family containing C_4 genes was then identified and annotated, as in Christin et al. (2015). For each transcriptome, all contigs homologous to any species in each of the ten reference datasets were identified using BLASTn searches with an e-value threshold of 0.01 and minimal matching length of 50 bp. The matching region of each of these contigs was successively aligned to the original alignment using Muscle and a phylogenetic tree was inferred using Phyml and a GTR+G substitution model. The phylogenetic tree was automatically inspected, and the contig was assigned to the Caryophyllales-level group of co-orthologs with which it grouped.

To obtain quantitative estimates of transcript abundances, all reads from each individual were mapped back to the respective transcriptome assembly using bowtie2 v. 2.0.5 (Langmead and Salzberg, 2012). For each group of co-orthologs belonging to one of the ten gene families analysed, the number of reads mapped to each contig assigned to it were summed. The transcript abundance of each gene lineage was then computed as number of reads per million of mappable reads (rpm). Co-orthologs reaching > 500 rpm in C_4 accessions were considered as used in the C_4 pathway of these populations, mirroring previous studies (Christin et al. 2015; Moreno-Villena et al. 2018).

Table 3.1. Species sampling and sequencing statistics.

Species	Origin	Photo	Pairs of reads	Average reads	Contigs	Mapped reads %
Hypertelis cerviana	Greece	C_4		2,541,924	59,450	
Replicate 1			1,308,013			95.50%
Replicate 2			3,775,835			94.07%
Hypertelis cerviana	Spain	C_4		5,863,270	75,969	
Replicate 1			5,112,596			96.80%
Replicate 2			5,885,334			96.51%
Replicate 3			6,591,882			96.64%
Hypertelis cerviana	Namibia	C_4		5,116,258	103,167	
Replicate 1			6,084,512			93.69%
Replicate 2			8,461,117			94.13%
Replicate 3			803,146			93.54%
Hypertelis spergulacea	Namibia	C_3-C_4		6,726,489	85,804	
Replicate 1			4,821,422			95.25%
Replicate 2			6,685,254			94.43%
Hypertelis umbellata	Mozambique	C_4	1,286,592	1,286,592	31,557	84.86%
Hypertelis umbellata	Arizona	C_4	2,212,950	4,559,413	111,101	95.10%
Mollugo verticillata	Brazil	C_3-C_4		6,445,136	148,367	
Replicate 1			7,533,667			94.49%
Replicate 2			6,459,415			92.38%
Replicate 3			5,342,327			95.50%
Mollugo verticillata	Michigan	C_3-C_4		4,846,802	157,861	
Replicate 1			3,729,608			94.57%
Replicate 2			4,879,936			93.52%
Replicate 3			5,930,863			94.90%
Mollugo verticillata	Montana	C_3-C_4		2,220,081	80,569	
Replicate 1			3,761,365			91.86%
Replicate 2			678,798			90.96%
Paramollugo nudicaulis	India	C_3-C_4		4,064,535	174,137	
Replicate 1			4,404,843			94.37%
Replicate 2			1,480,877			94.15%
Replicate 3			6,307,885			93.64%
Paramollugo nudicaulis	Madagascar	C_3-C_4		4,892,592	125,878	
Replicate 1			2,437,195			94.05%
Replicate 2			7,240,822			93.94%
Replicate 3			4,999,761			93.89%
Paramollugo nudicaulis	Uganda	C_3-C_4		5,408,701	202,057	
Replicate 1			3,419,896			89.32%
Replicate 2			8,167,995			88.34%
Replicate 3			4,638,214			88.37%
Suessenguthiella	Namibia	C_3		8,199,270	89,336	

Species	Origin	Photo	Pairs of reads	Average reads	Contigs	Mapped reads %
scleranthoides						
Replicate 1			4,133,264			94.07%
Replicate 2			12,265,276			92.86%
Trigastrotheca pentaphylla	India	C_3		6,343,252	134,222	
Replicate 1			5,467,868			94.70%
Replicate 2			4,856,551			95.25%
Replicate 3			8,705,337			94.88%

3.3.3 Tests for positive selection.

A new phylogenetic tree was inferred for each group of co-orthologs deemed to be used for the C_4 pathway of some Molluginaceae based on the transcript abundance. Sequences corresponding to Molluginaceae were extracted from the reference dataset and were realigned as codons. The 3rd positions of codons were extracted from the alignments, as these are less subject to selection. Their alignment was used to infer a phylogenetic tree with Phyml and a GTR+G+I substitution model. Support values were estimated with 100 bootstrap pseudo replicates. The inferred topology was used for positive selection analyses.

For each group of co-orthologs, different codon models were optimized with codeml as implemented in PAML (Yang, 2007). The null model (M1a) assumes that some sites evolve under relaxed selection, while others evolve under purifying selection, with selection constant among all branches. By contrast, branch-site models assume that a third category of sites shift from purifying or relaxed selection to relaxed (in model A) or positive (in model A1) selection in *a priori* defined foreground branches. For each group of co-orthologs, different sets of foreground branches corresponding to different hypotheses were used. First, the hypothesis of a single episode of adaptive evolution was tested by setting the foreground branch to that leading to the most recent common ancestor of the two C_4 clades. Second, the hypothesis of

two independent episodes of adaptive evolution was tested by setting the foreground branches to those leading to each of the two C_4 clades. In each case, the hypothesis of sustained selection was further tested by setting all descending branches as foreground branches. For each gene, the best-fit branch-site model was identified using the Akaike information criterion, after verifying that it was significantly better than the null model M1a using a likelihood ratio test with a p-value corrected for nine genes (equivalent to an AIC difference of 6.59 for model A1 and 4.59 for model A).

2.4 Results

3.4.1 Sequencing, assembly and read mapping.

After cleaning, over 190 million paired-end cleaned reads were retained across 34 samples of Molluginaceae belonging to 14 populations (mean =71329, sd = 56246.10; Table 3.1). One *de novo* transcriptome assembly was generated per population, and these 14 assemblies contained a total of 1,579,475 contigs (mean = 112,820, sd = 47,126; Table 3.1). Reads from each of the 34 samples were mapped on the respective assembly, with an average of 93.67% reads successfully mapped (sd = 2.47%).

The phylogenetic trees inferred from genes encoding C_4 enzymes of Molluginaceae and other plants were congruent with those previously inferred for other Caryophyllales (Christin et al., 2015). Based on the phylogenetic trees, the ten gene families with some members encoding enzymes of the C_4 pathway contain between one and four co-orthologs at the level of Molluginaceae. In no cases were duplications detected at the base of Molluginaceae, but some occurred within sub-clades of Molluginaceae. In most cases, the

multiple contigs from the same species were monophyletic, but one sample of *H. umbellata* (from Arizona) consistently had contigs nested within each of *H. cerviana* and *H. umbellata*. This population, which was grown from seeds obtained from a greenhouse collection, likely results from a hybridization between the two species. For phylogenetic annotation and tests of positive selection, the contig positioned outside of *H. cerviana* was selected.

3.4.2 Expression profiles.

The transcript abundance was reported for all 25 Caryophyllales-level co-orthologs encoding one of the ten enzymes with a known function in C_4 photosynthesis (Table S3.1). A total of 12 genes had consistently low levels (< 500 rpm) in all samples. In two other cases (*nadmdh-1*, *nadhmdh-3*, and *nadpmdh-3*), levels exceeding 500 rpm were reached, but mainly in non-C_4 accessions (Table S3.1). In addition, one gene for carbonic anhydrase (*βca-2E1*) reached high levels in samples of various photosynthetic types (Table S3.1), which is expected as the enzyme is highly transcribed in both C_3 and C_4 plants (Furbank and Taylor, 1995; Sage, 2004). Six genes encoding five distinct enzymes reached high levels only in C_4 accessions, while three more (*nadmdh-2*, *ppc-1E1* and *ppdk-1C1b*) reached levels mainly in C_4 accessions, and in a few non-C_4 accessions (Table S3.1). Based on these patterns, nine different genes for eight enzymes are consequently considered as co-opted for the C_4 pathway of C_4 Molluginaceae (Table 3.2). All these genes are also used by some other C_4 Caryophyllales (Christin et al. 2015; Table 3.2).

Genes for ALAAT (*alaat-1*), NAD-ME (*nadme-2*), PEPC (*ppc-1E1*), and PPDK (*ppdk-1C1b*) reached high levels in all C_4 accessions (Table **3.**3). By contrast, distinct genes

for ASPAT were most upregulated in *H. umbellata* (*aspat-3C1*) and *H. cerviana* (*aspat-1E1*; Table **3.**3). The potential hybrid (*H. umbellata* from Arizona) showed high expression of both. In addition, a gene for NADMDH (*nadmdh-2*) is upregulated exclusively in *H. cerviana* (and the hybrid *H. umbellata*), while the genes *βca-2E3* and *nadpme-1E1* are up-regulated in most but not all C_4 accessions (Table **3.**3). The *H. cerviana* from Greece has levels of most C_4 enzymes below those of other C_4 accessions (Table **3.**3). Despite its close relationship to the C_4 species, the C_3-C_4 *H. spergulacea* does not show elevated levels of genes that were co-opted by C_4 *Hypertelis* (Table 3.3). The levels of most genes co-opted by the C_4 *Hypertelis* are low in the C_3-C_4 *P. nudicaulis* and *M. verticillata*. However, *M. verticillata* from Brazil and some *P. nudicaulis* have relatively high levels (> 500 rpm) of *ppc-1E1*, while *ppdk-1C1b* is above 500 rpm in some other *M. verticillata* (Table 3.3).

Table 3.2. List of genes co-opted for C_4 photosynthesis in Molluginaceae.

Gene	Enzyme	*H. umbellata*[a]	*H. cerviana*[a]	Other Caryophyllales[b]
alaat-1	Alanine aminotransferase	Yes	Yes	Yes
aspat-1E1	Aspartate aminotransferase	Yes	Yes	Some
aspat-3C1	Aspartate aminotransferase	Yes	Some	Some
βca-2E3	Carbonic anhydrases	Yes	Yes	Yes
nadme-2	NAD-malic enzyme	Yes	Yes	Some
nadmdh-2	NAD-malate dehydrogenase	No	Yes	Some
nadpme-1E1	NADP-malic enzyme	No	Some	Yes
ppc-1E1	Phosphoenolpyruvate carboxylase	Yes	Yes	Yes
ppdk-1C1b	Pyruvate orthophosphate dikinase	Yes	Yes	Yes

[a] Genes co-opted for C_4 photosynthesis by each species are indicated with a 'yes'

[b] Whether the gene lineage has been co-opted by other C_4 Caryophyllales is indicated, baed on the results based on Christin et al. (2015).

Table 3.3. Transcript abundance of genes co-opted for C_4 photosynthesis in accessions of Molluginaceae [a].

Species	Photo	*alaat-1*	*aspat-1E1*	*aspat-3C1*	*βca-1E2*	*nadme-2*	*nadmdh-2*	*nadpme-1E1*	*ppc-1E1*	*ppdk-1C1b*
H. cerviana Namibia	C_4	2769	1370	522	1632	2014	909	606	8969	9820
H. cerviana Greece	C_4	961	316	171	395	1134	281	131	1398	4849
H. cerviana Spain	C_4	2895	2542	745	1894	3368	1154	734	9572	14160
H. spergulacea	C_3-C_4	140	66	198	120	99	202	191	291	216
H. umbellata Arizona	C_4	4076	2011	2865	4187	2035	1394	1115	8784	14126
H. umbellata Mozambique	C_4	2295	638	1613	1749	1191	336	368	13547	11734
S. scleranthoides	C_3	0	69	24	377	85	357	125	680	230
T. pentaphylla	C_3	78	29	12	35	87	203	134	372	491
M.verticillata Brazil	C_3-C_4	64	243	61	205	121	415	388	995	267
M. verticillata Montana	C_3-C_4	56	63	77	226	164	232	151	411	472
M.verticillata Michigan	C_3-C_4	62	93	99	257	250	424	387	165	636
P. nudicaulis Madagascar	C_3-C_4	181	134	108	137	80	472	63	389	104
P. nudicaulis Uganda	C_3-C_4	255	187	232	284	137	384	98	693	231
P. nudicaulis India	C_3-C_4	69	66	155	126	120	253	166	393	291

[a] For each accession, the average transcript abundance in reads per million of mappable reads is indicated.

3.4.3 Phylogenetic trees and tests of positive selection.

A chloroplast marker (trn*K-matK*) was selected to infer phylogenetic relationships among the new sampled accessions adding to previously analysed by Christin et al. (2011) (Fig. 3.1). The individual gene trees were largely congruent with those previously inferred from chloroplast markers or nuclear ribosomal DNA (Fig. 3.2; Christin et al., 2011; Thulin et al., 2016). When these species were present, the *Hypertelis* clade was always recovered, and was always sister to *Suessenguthiella*. Within *Hypertelis*, the C_3-C_4 *H. spergulacea* was sister to *H. umbellata*, mirroring chloroplast relationships, in one case (*alaat-1*; Fig. S3.1.a). In four cases however, the two C_4 *H. cerviana* and *H. umbellata* formed a clade that was sister to *H. spergulacea* (*βca-2E3*, *aspat-1E1*, *aspat-3C1* and *ppdk-3C1b*), in most cases with high support (Fig. S3.1). In one other case, the C_3-C_4 *H. spergulacea* was sister to *H. cerviana* (*nadmdh-2*), but with

moderate support (Fig. S3.1). Outside of *Hypertelis*, the C_3-C_4 *M. verticillata* was consistently placed outside of *P. nudicaulis* plus *Suessenguthiella*, but the *Trigastrotheca* jumped from sister to *P. nudicaulis* and sister to *M. verticllata*, the latter placement reflecting the chloroplast relationships (Fig. S3.1).

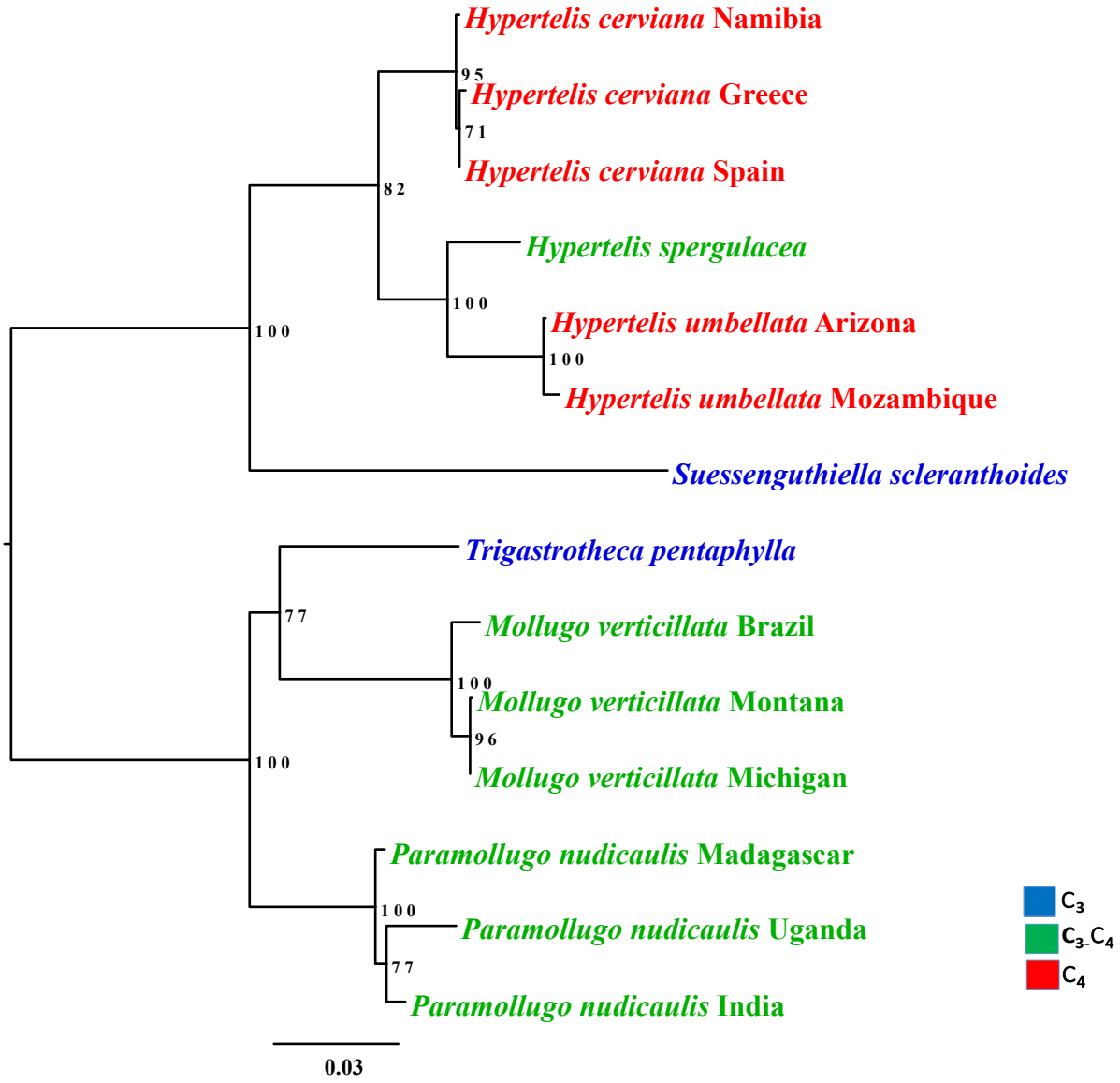

Figure 3.1. Phylogenetic relationships among sampled Molluginace.

This phylogenetic tree was inferred based on the chloroplast marker *trnKmatK*. Bootstrap values are indicated near nodes, and species are coloured based on the photosynthetic type.

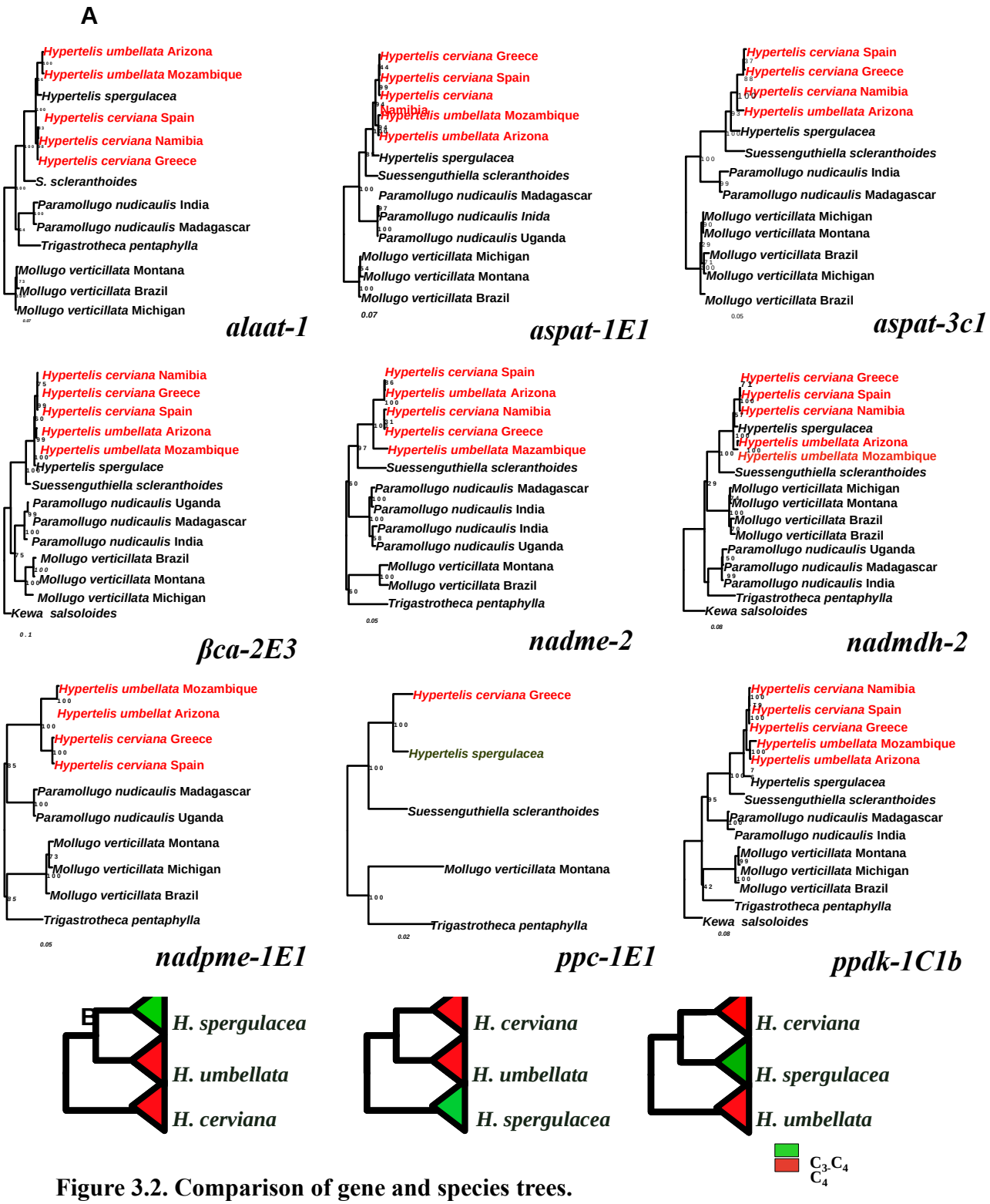

Figure 3.2. Comparison of gene and species trees.

A) Gene trees are shown for the nine genes co-opted by C_4 Molluginaceae. Names of C_4 accessions are in red, and bootstrap support values are shown near branches. Note that some samples are not included in all trees, reflecting a lack of transcript because of low expression or partial assembly. B) The species tree of *Hypertelis* (on the left) is shown next to the two alternative topologies.

For one gene (*ppc-1E1*), the species representation is too incomplete to compare the different branch models (Fig. S3.1). The model assuming no shift of selective pressure is preferred for two genes out of the eight other co-opted for C_4 photosynthesis (*nadmdh-2* and *βca-2E3*). For all other six, assuming a shift of selective pressure in C_4 branches significantly improved the fit of the model (Table 3.4). Assuming a shift solely on the branches at the base of the C_4 groups (either one or two) was never favoured, indicating that in all cases the shift of selective pressures was sustained in the descending branches.

In four cases (*alaat-1*, *aspat-3C1*, *nadme-2* and *ppdk-1C1b*), the shift occurred in the common ancestor of the two C_4 groups. Based on the tree topology, this implies a shift of selection in the ancestor of the intermediate C_3-C_4 *H. spergulacea* in the case of *alaat-1*, but not the other three genes (Fig. 3.2). For two other genes (*aspat-1E1* and *nadpme-1E1*), shifts happened independently on the branches leading to each of the two C_4 groups (Table 3.4). The model assuming positive selection was significantly better than the model assuming relaxed selection on the same branches in only two out of these six genes (Table 3.4).

Table 3.4. Results of positive selection tests[a]

Gene	M1a[b]	One[c]		Two[d]		One, sustained[c]		Two, sustained[d]	
		A[e]	A1[f]	A[e]	A1[f]	A[e]	A1[f]	A[e]	A1[f]
alaat-1	9706	9708	9710	9708	9710	9689*	9690	9708	9710
aspat-1E1	7419	7421	7423	7405	7405	7401	7402	7399*	7400
aspat-3C1	7419	7421	7423	7404	7405	7398*	7400	7400	7402
βca-2E3	4779*	4781	4783	4781	4783	4776	4777	4776	4777
nadme-2	13333	13311	13313	13296	13293	13259	13257*	13293	13286
nadmdh-2	7166*	7168	7170	7168	7170	7168	7170	7168	7170
nadpme-1E1	10405	10407	10409	10403	10405	10406	10408	10398*	10400
ppc-1E1	21079	21036*	21038	21036*	21038	21036*	21038	21036*	21038
ppdk-1C1b	18851	18849	18851	18852	18852	18832	18821*	18841	18831

[a] For each model, the Akaike information criterion is indicated (lower values indicate better models). For each gene, the asterisk indicates the best model, which was identified as the best alternative model if it is significantly better than the null model based on a likelihood ratio test. In the case of *ppc-1E1*, a limited species sampling means that all alternative models are the same.

[b] Null model, without shifts of selective pressures.

[c] Shift of selective pressures along the branch leading to the common ancestor of the two C_4 groups.

[d] Shifts of selective pressures along the branches leading to each of the two C_4 groups.

[e] Shift to relaxed selection.

[f] Shift to positive selection.

3.5 Discussion

3.5.1 C_3-C_4 taxa do not bridge the gap to C_4 biochemistry.

Our transcriptome comparisons revealed strong upregulation of genes for four enzymes in the C_4 Molluginaceae (Fig. 3.3). The set of encoded enzymes matches classical pathways, although the upregulation of two distinct malic enzymes (*nadme-2* and *nadpme-1E1*) differs

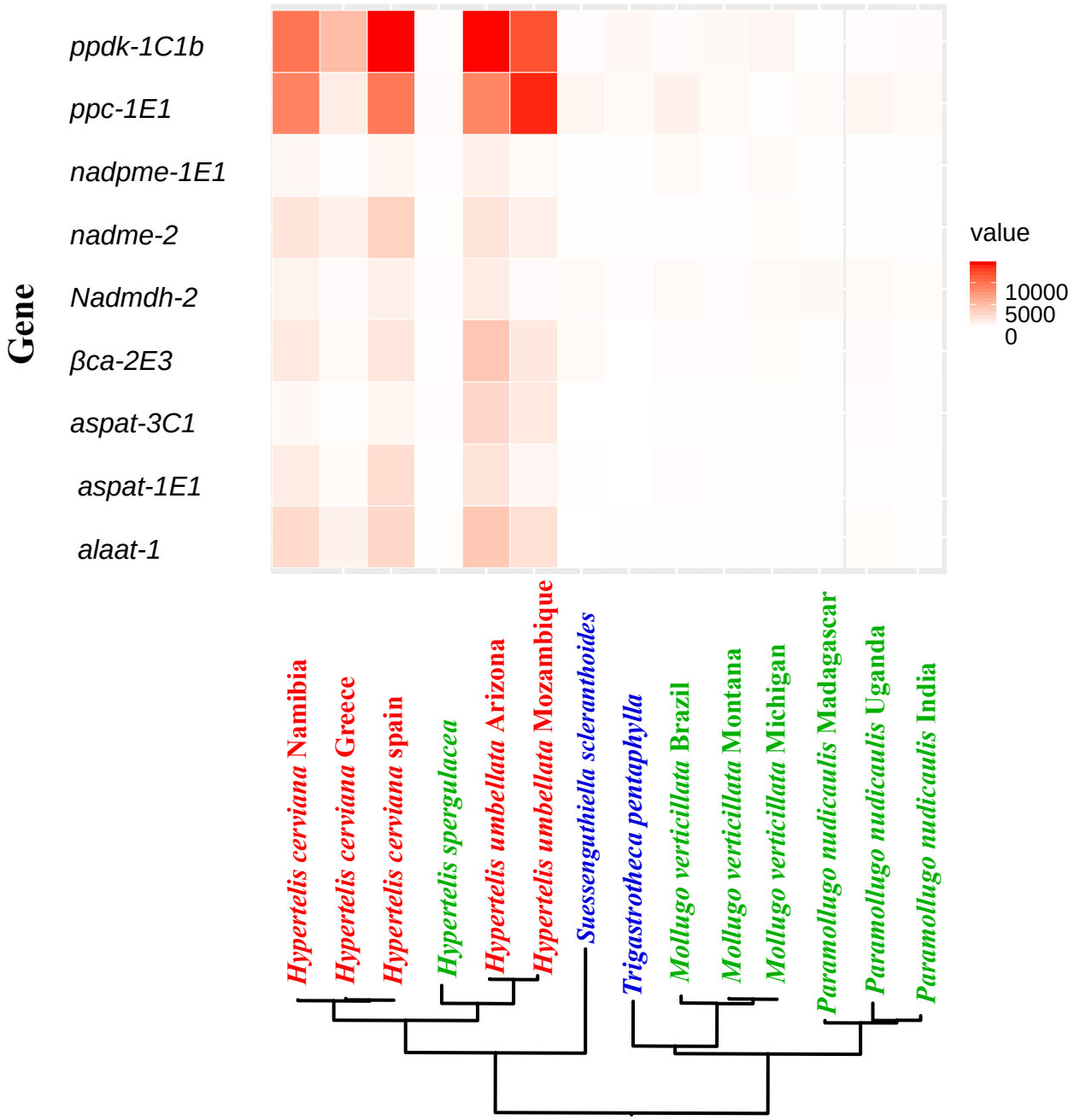

Figure 3.3. Patterns of gene expression among Molluginaceae species.

The heatmap shows the relative abundance of nine genes co-opted by C_4 Molluginaceae in the different species, shown with their phylogenetic tree at the bottom (names in blue for C_3, green for C_3-C_4 and red for C_4 species).

from most other C_4 Caryophyalles analysed so far (Christin et al., 2015; Lauterbach et al., 2017), but has been reported in other groups (Washburn et al., 2015). Importantly, these levels of expression are not shared by the closely related C_3-C_4 species *H. spergulacea* (Fig. 3.3). Our results therefore suggest that the transition from C_3-C_4 to C_4 involved upregulation of multiple enzymes of a magnitude that would be similar to a direct transition from C_3 to C_4 photosynthesis. This conclusion is at odds with studies in *Alloteropsis* and *Flaveria*, where C_3-C_4 species showed an elevated level of C_4 enzymes (Mallmann et al., 2014; Dunning et al., 2019). The intermediates in these genera correspond to the type II, where some of the carbon is fixed by the C_4 cycle (Westhoff and Gowik, 2004; Lundgren et al., 2016). By contrast, the C_3-C_4 *P. nudicaulis* has been reported as lacking C_4 activity (Sayre and Kennedy, 1977; Christin and Osborne, 2014), while a weak C_4 cycle was present in only some *M. verticillata* populations (Kennedy and Laetsch 1974; Christin et al., 2011). The status of *H. spergulacea* is not known, but our data suggest that it uses at most a very weak C_4 cycle. Its anatomy, which in many aspects resembles that of C_4 species and is thought to have facilitated C_4 transitions in the group (Chapter 1; Christin et al., 2011), likely mirrors a fully C_2 physiology (Sage and Stata, 2015). We conclude that the phenotypic gap between the C_3 and C_4 states was not decreased by the existence of intermediates in *Hypertelis*.

Previous phylogenetic trees based on either chloroplast or nuclear markers indicated that *H. spergulacea* is nested within otherwise C_4 species (Chapter 1; Christin et al., 2011; Thulin et al., 2016), a position that would be compatible with a reversal from C_4 to C_3-C_4 photosynthesis. However, the absence of increased expression of C_4 genes and the evidence for positive selection mostly limited to branches leading solely to C_4 species (Table 3.4; Christin et al., 2011) argue against this hypothesis. Instead, our results are compatible with two independent transitions to C_4 photosynthesis, with a third potential one in *H. walteri* that

was not sampled here. These transitions were likely facilitated by C_4-like anatomical characters in the common ancestor of the group (Chapter 1; Christin et al., 2011), but the build-up of a strong C_4 cycle seems to have occurred solely after the divergence from *H. spergulacea*. We conclude that, while the transition from C_3 to C_4-like anatomy was spread over more than 20 Ma, the C_4 cycle emerged over the last 10 Ma.

3.5.2 C_4 adaptation continued within each species.

Our analyses of genes co-opted for C_4 photosynthesis provided statistical evidence of shifts of selective pressures in C_4 species (Table 3.4). While we lack statistical power to firmly conclude to positive selection as opposed to relaxed selection, these results indicate that gene co-option was in most cases followed by adaptation of the protein sequence to fit the C_4 catalytic context. This conclusion echoes previous analyses in various groups (Christin et al., 2007; Rosnow et al., $201_{4;}$ Huang et al., 2017). Importantly, the models assuming a single episode at the base of the C_4 species were never supported, and, instead, the shift of selective pressure was always sustained in descending branches (Table 3.4). This indicates that adaptation of C_4 enzymes continued after the initial transition to C_4 photosynthesis, as previously reported in an old grass C_4 lineage (Bianconi et al., 2019). Our investigation therefore indicates that, in multiple groups of species spanning various timescales, the initial co-option of genes for C_4 photosynthesis is followed by extensive periods of adaptation. In the case of *Hypertelis*, our results indicate that this adaptation still happens within each species, mirroring the grass *Alloteropsis semialata* (Dunning et al., 2017). *Hypertelis* therefore constitutes an exciting system to study the significance of C_4 variation within a given species, its three C_4 species analysed here providing as many independent replicates.

3.5.3 Different species share some C_4 components.

The two C_4 groups analysed here were separated in previous nuclear and chloroplast phylogenies (Chapter 1; Christin et al., 2011; Thulin et al., 2016). A similar pattern is retrieved for two of the genes co-opted for C_4 photosynthesis (Fig. 3.2), confirming independent origins of some C_4 components. However, for four other genes involved in C_4 photosynthesis, the two C_4 species are sister, with *H. spergulacea* placed outside (Fig. 3.2). This pattern could result from incomplete lineage sorting, but could also stem from introgression after the split of the two species. The two processes are difficult to tease apart for specific gene trees, and genome-wide phylogenomic analyses would be needed to quantify the amount of incomplete lineage sorting in the group.

In any case, these gene trees indicate that despite representing two distinct C_4 lineages, the two species acquired genes for βCA, ASPAT and PPDK from the same source. In the case of *ppdk-1C1b* and *aspat-3C1*, there is moreover evidence of selective shift starting in the common ancestor of the two species (Table 3.4). These results indicate that C_4 adaptation started before the split of the two species. Either one of the species co-opted the genes and then transferred them to the other species, or C_4-specific genes existed as a polymorphism in the ancestor of the three *Hypertelis* (including the C_3-C_4 *H. spergulacea*) and were later sorted in a manner that differs from the species tree. The latter scenario is less likely as it would imply a photosynthetic polymorphism in the ancestral population, and we suggest that introgression boosted the early evolution of C_4 photosynthesis in *Hypertelis*, as reported in other study systems (Dunning et al., 2017). Since previous analyses of *ppc-1E1* have indicated that this gene acquired its C_4 function independently in each of the two species

(Christin et al., 2011), we propose that a weak C_4 cycle emerged twice by independent co-option of genes. Following hybridization, a better adapted gene was then introgressed and replaced the previously used genes. With the exception of the putative hybrid produced in the lab (*H. umbellata* from Arizona), phylogenetic trees do not suggest very recent hybridization and such exchanges probably happened soon after the divergence of the species.

3.6 Conclusions

In this study, we compared the transcriptomes of Molluginaceae species spanning the C_3, C_3-C_4 and C_4 phenotypes. We show that C_4-specific genes are upregulated specifically in the C_4 accessions, and C_3-C_4 intermediates do not reduce the gap to C_4 biochemistry. Signs of positive selection are found on most of the co-opted genes, but importantly, the shift of selective pressure was sustained over multiple phylogenetic branches. This result indicates that enzyme adaptation continued after the initial emergence of photosynthesis. The pattern of gene co-option via upregulation followed by adaptation of the coding sequences mirrors that reported in other C_4 systems. However, our phylogenetic analyses indicate that the history of the co-opted genes in many cases differ from the species tree. Indeed, the two distinct C_4 species acquired some of their C_4 genes independently, but others were likely passed among them during their early history. We suggest that recurrent transitions to C_4 photosynthesis in this small groups were boosted by interspecific exchanges, in addition to anatomical enablers. Together, introgression mixed with continuous adaptation led to important variations both among and within species, creating a great system to understand the ecophysiological relevance of different C_4 components.

3.7 References

Atkinson, R.R., Mockford, E.J., Bennett, C., Christin, P.A., Spriggs, E.L., Freckleton, R.P., Thompson, K., Rees, M. and Osborne, C.P. 2016. C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants.* **2**(5), 16038.

Bianconi, M.E., Hackel, J., Vorontsova, M.S., Alberti, A., Arthan, W., Burke, S.V, Duvall, M.R., Kellogg, E.A., Lavergne, S., McKain, M.R., Meunier, A., Osborne, C.P., Traiperm, P., Christin, P.A. And Besnard, G. 2020. Continued adaptation of C_4 photosynthesis after an initial burst of changes in the Andropogoneae Grasses. *Systematic Biology.*. **69**(3), pp.445–461.

Christin, P. A. and Osborne, C.P. 2014. The evolutionary ecology of C_4 plants. *New Phytologist*. **204**(4), pp.765–781.

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 C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences USA*. **110**(4), pp.1381–1386.

Christin, P. A., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. 2007. C_4 photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes. *Current Biology.* **17** (14), pp.1241–1247.

Christin, P. A., Arakaki, M., Osborne, C.P. and Edwards, E.J. 2015. Genetic enablers underlying the clustered evolutionary origins of C_4 photosynthesis in angiosperms. *Molecular Biology and Evolution.* **32** (4), pp.846–858.

Christin, P. A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R. and Sage, R.F. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in molluginaceae. *Evolution.* **65** (3), pp.643–660.

Dunn, C.W., Howison, M. and Zapata, F. 2012. Agalma: an automated de novo transcriptome assembly pipeline. *BMC Bioinformatics.* **14**, pp.330.

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 C_4 photosynthesis among close relatives. *Evolution.***71** (6), pp.1541–1555.

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., Guillaume, B., Patrik, P., Osborne, C.P., and Christin, P.A. 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences: USA.* **116** (10), pp.4416–4425.

Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* **32** (5), pp.1792–1797.

Furbank, R. and Taylor, W. 1995. Regulation of Photosynthesis in C_3 and C_4 Plants: A Molecular Approach. *The Plant Cell.* **7** (7), pp.797–807.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology.* **29** (7), pp.644–652.

GPWGII– Grass Phylogeny Working Group II. 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist*. **193**, pp.304–312.

Guindon, S. and Gascuel, O. 2003. A Simple, Fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. **52**(5), pp.696–704.

Hatch, M. 1987. C_4 Photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta*. **895**(2), pp.81–106.

Huang, C.F., Yu, C.P., Wu, Y.H., Lu, M.Y.J., Tu, S.L., Wu, S.H., Shiu, S.H., Ku, M.S.B. and Li, W.H. 2017. Elevated auxin biosynthesis and transport underlie high vein density in C_4 leaves. *Proceedings of the National Academy of Sciences: USA.* **114**(33), pp.E6884–E6891.

Kellogg, E.A. 1999. Phylogenetic aspects of the evolution of C_4 photosynthesis. In Sage, R. F., Monson. R. K. (eds) C_4 plant biology. *Academic Press: San Diego*. pp.411–444.

Kennedy, R. A., and W. M. Laetsch. 1974. Plant species intermediate for C_3, C_4 photosynthesis. *Science*. **184**, pp.1087–1089.

Langmead, B. and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* **9** (4), pp.357–359.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P. and Gowik, U. 2017. C_3 cotyledons are followed by C_4 leaves: Intra-individual transcriptome analysis of *Salsola soda* (Chenopodiaceae). *Journal of Experimental Botany*, **68** (2), pp.161–176.

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 C_3-C_4 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell and Environment*, **39** (9), pp.1874–1885.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P. and Gowik, U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife*. **2014**(3), pp.2–5.

Marazzi, B., Ané, C., Simon, M. F., Delgado-Salinas, A., Luckow, M. and Sanderson, M. J. 2012. Locating evolutionary precursors on a phylogenetic tree. *Evolution*, **66** (12), pp.3918–3930.

Mckown, A.D. and Dengler, N.G. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *American Journal of Botany*. **94**, pp.382–399.

Monson RK. 2003. Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences.* 164(S3): S43–S54.

Monson, R.K., Moore, B., Ku, M.S.B. and Edwards, G.E. 1986. Co-function of C_3- and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. *Planta*. **168**, pp.493–502.

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. and Christin, P. A. 2018. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Molecular Biology and Evolution.* **35** (1), pp.94–106.

Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G. and Sage, R.F. 2011. Characterization of C_3-C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. *Plant, Cell and Environment*. **34**(10), pp.1723–1736.

Rosnow, J.J., Edwards, G. E., and E. H. Roalson. 2014. Positive selection of Kranz and non-Kranz C_4 phosphoenolpyruvate carboxylase amino acids in Suaedoideae (Chenopodiaceae). *Journal of Experimental Botany*. **65**, pp.3595–3607.

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

Sage, R. F. 2004. The evolution of C_4 photosynthesis. *New Phytologist.* **161**, pp.341–370.

Sage, R. F., Christin, P. A. and Edwards, E. J. 2011. The C_4 plant lineages of planet earth. *Journal of Experimental Botany.* **62** (9), pp.3155–3169.

Sage, R. F. and Stata, M. 2015. Photosynthetic diversity meets biodiversity: The C_4 plant example. *Journal of Plant Physiology.* **172**, pp.104–119.

Sayre, R. T. and Kennedy, R. A. 1977. Ecotypic differences in the C_3 and C_4 photosynthetic activity in *Mollugo verticillata*, a C_3−C_4 intermediate. *Planta*. **134**, pp.257–262.

Thulin, M., Moore, A.J., El-Seedi, H., Larsson, A., Christin, P.A. and Edwards, E.J. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. **65**(August), pp.775–793.

Washburn, J. D., Schnable, J. C., Davidse, G. and Pires, J. C. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany.* **102** (9), pp.1493–1505.

Westhoff, P. and Gowik, U. 2004. Evolution of C_4 phosphoenolpyruvate carboxylase. Genes and proteins: A case study with the genus *Flaveria*. *Annals of Botany.* **93** (1), pp.13–23.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution.* **24** (8), pp.1586–1591.

Appendix for chapter 2

Table S3.1a. Transcript abundance of gene lineages encoding enzymes related to the C_4 pathway

Species	*alaat-1*	*aspat-1E1*	*aspat-2*	*aspat-3C1*	*aspat-3C2*	*nadmdh-1C1*	*nadmdh-1C2*	*nadmdh-2*
Hypertelis umbellata Mozambique	2295	638	115	1613	85	132	16	336
Hypertelis umbellata Arizona	4076	2011	122	2865	22	95	32	1394
Hypertelis cerviana Greece 1	864	303	157	153	121	45	63	252
Hypertelis cerviana Greece 2	1058	330	113	189	54	50	49	311
Hypertelis cerviana Spain 1	3300	2642	344	724	39	42	18	1157
Hypertelis cerviana Spain 2	2562	2374	352	783	35	52	24	1097
Hypertelis cerviana Spain 3	2823	2610	333	728	33	44	23	1208
Hypertelis cerviana Namibia 1	3030	1473	160	454	44	36	36	926
Hypertelis cerviana Namibia 2	2612	1068	146	625	70	61	53	794
Hypertelis cerviana Namibia 3	2664	1569	197	486	26	46	41	1007
Paramollugo nudicaulis India 1	81	79	148	148	20	85	46	248
Paramollugo nudicaulis India 2	62	66	134	149	22	72	38	281
Paramollugo nudicaulis India 3	64	55	145	167	23	66	31	231
Paramollugo nudicaulis Madagascar 1	184	131	211	112	33	61	41	378
Paramollugo nudicaulis Madagascar 2	167	143	189	93	28	67	55	481
Paramollugo nudicaulis Madagascar 3	191	129	176	118	36	53	55	555
Paramollugo nudicaulis Uganda 1	315	156	304	289	32	59	113	406
Paramollugo nudicaulis Uganda 2	209	217	313	171	7	144	67	408
Paramollugo nudicaulis Uganda 3	241	189	358	238	15	60	63	338
Trigastrotheca pentaphylla 1	81	25	98	10	20	0	28	186
Trigastrotheca pentaphylla 2	77	34	130	12	27	0	43	233
Trigastrotheca pentaphylla 3	78	28	128	13	25	0	34	190
Hypertelis spergulacea 1	213	38	168	67	63	51	32	308
Hypertelis spergulacea 2	236	73	157	84	106	36	32	331
Mollugo verticillata Brazil 1	49	230	145	61	83	35	35	468
Mollugo verticillata Brazil 2	74	236	119	53	109	28	23	356
Mollugo verticillata Brazil 3	68	262	153	68	96	45	32	422
Mollugo verticillata Montana 1	62	104	149	84	62	53	18	465
Mollugo verticillata Montana 2	63	104	152	124	63	78	16	410
Mollugo verticillata Montana 3	62	73	141	89	71	75	11	397
Mollugo verticillata Michigan 1	52	57	97	92	91	37	7	247
Mollugo verticillata Michigan 2	61	70	92	63	66	11	6	216
*Suessenguthiella scleranthoides*1	0	79	160	31	22	88	61	379
*Suessenguthiella scleranthoides*_2	0	59	157	16	23	153	38	335

Table S3.1b. Transcript abundance of gene lineages encoding enzymes related to the C_4 pathway.

Species	*nadmdh-3*	*nadme-1*	*nadme-2*	*nadpmdh-1*	*nadpmdh-2*	*nadpmdh-3*	*nadpme-1E1*	*nadpme-1E2*
Hypertelis umbellata Mozambique	229	15	1191	444	0	339	368	9
Hypertelis umbellata Arizona	404	24	2035	234	2	438	1115	38
Hypertelis cerviana Greece 1	246	122	1199	149	1	266	123	33
Hypertelis cerviana Greece 2	215	105	1069	147	1	243	139	25
Hypertelis cerviana Spain 1	524	25	3655	177	1	288	816	22
Hypertelis cerviana Spain 2	547	27	3201	215	1	265	668	44
Hypertelis cerviana Spain 3	562	24	3249	198	1	305	718	25
Hypertelis cerviana Namibia 1	255	77	2319	261	1	452	852	70
Hypertelis cerviana Namibia 2	227	70	1759	262	1	347	198	125
Hypertelis cerviana Namibia 3	277	62	1964	284	2	541	768	63
Paramollugo nudicaulis India 1	687	113	124	299	1	407	155	45
Paramollugo nudicaulis India 2	729	112	104	239	1	378	147	23
Paramollugo nudicaulis India 3	625	140	133	308	2	394	195	39
Paramollugo nudicaulis Madagascar 1	688	99	92	174	1	576	65	47
Paramollugo nudicaulis Madagascar 2	697	75	72	119	2	598	64	35
Paramollugo nudicaulis Madagascar 3	680	79	76	137	1	779	61	35
Paramollugo nudicaulis Uganda 1	349	191	167	271	0	785	115	80
Paramollugo nudicaulis Uganda 2	1008	79	116	468	0	816	111	58
Paramollugo nudicaulis Uganda 3	759	141	128	386	0	779	67	47
Trigastrotheca pentaphylla 1	641	172	92	261	1	390	105	17
Trigastrotheca pentaphylla 2	418	184	80	237	2	385	169	16
Trigastrotheca pentaphylla 3	562	181	88	232	1	363	129	14
Hypertelis spergulacea 1	643	43	86	585	2	308	123	37
Hypertelis spergulacea 2	549	57	82	487	1	276	104	55
Mollugo verticillata Brazil 1	932	46	89	481	2	825	333	76
Mollugo verticillata Brazil 2	1188	73	135	628	1	594	427	134
Mollugo verticillata Brazil 3	1253	71	140	680	2	734	403	115
Mollugo verticillata Montana 1	813	100	288	408	1	427	384	45
Mollugo verticillata Montana 2	902	124	256	455	2	352	429	62
Mollugo verticillata Montana 3	547	131	207	376	1	331	349	48
Mollugo verticillata Michigan 1	560	128	145	676	0	288	161	26
Mollugo verticillata Michigan 2	640	109	183	619	0	261	141	7
*Suessenguthiella scleranthoides*1	1025	63	98	472	0	513	141	280
*Suessenguthiella scleranthoides*_2	1206	80	71	490	0	527	109	150

Table S3.1c. Transcript abundance of gene lineages encoding enzymes related to the C_4 pathway.

Species	*nadpme-1E3*	*pck-1*	*ppc-1E1*	*ppc-1E2*	*ppc-2*	*ppdk-1C1b*
Hypertelis umbellata Mozambique	113	0	13547	8	11	11734
Hypertelis umbellata Arizona	25	12	8784	76	0	14126
Hypertelis cerviana Greece 1	130	2	1010	62	41	3524
Hypertelis cerviana Greece 2	123	3	1786	88	39	6174
Hypertelis cerviana Spain 1	18	2	10539	71	36	15328
Hypertelis cerviana Spain 2	23	3	8380	82	33	14204
Hypertelis cerviana Spain 3	19	1	9796	85	35	12948
Hypertelis cerviana Namibia 1	45	29	9882	137	17	11655
Hypertelis cerviana Namibia 2	19	36	8256	224	33	10013
Hypertelis cerviana Namibia 3	35	10	8769	134	2	7793
Paramollugo nudicaulis India 1	28	60	373	385	52	280
Paramollugo nudicaulis India 2	21	32	332	270	16	216
Paramollugo nudicaulis India 3	22	27	475	469	48	378
Paramollugo nudicaulis Madagascar 1	9	4	299	295	23	116
Paramollugo nudicaulis Madagascar 2	9	2	362	229	10	87
Paramollugo nudicaulis Madagascar 3	14	1	506	375	12	110
Paramollugo nudicaulis Uganda 1	42	6	549	310	27	216
Paramollugo nudicaulis Uganda 2	3	3	799	185	20	172
Paramollugo nudicaulis Uganda 3	6	3	730	246	28	306
Trigastrotheca pentaphylla 1	143	63	439	437	1	433
Trigastrotheca pentaphylla 2	111	62	358	306	3	597
Trigastrotheca pentaphylla 3	114	48	318	332	2	443
Hypertelis spergulacea 1	440	2	364	149	2	272
Hypertelis spergulacea 2	404	1	173	210	1	354
Mollugo verticillata Brazil 1	13	6	1071	177	0	177
Mollugo verticillata Brazil 2	41	9	1064	211	0	309
Mollugo verticillata Brazil 3	21	9	850	246	0	314
Mollugo verticillata Montana 1	47	11	105	145	0	498
Mollugo verticillata Montana 2	54	8	208	169	0	817
Mollugo verticillata Montana 3	148	11	182	144	0	593
Mollugo verticillata Michigan 1	207	6	423	164	0	623
Mollugo verticillata Michigan 2	290	8	399	108	0	321
*Suessenguthiella scleranthoides*1	45	12	614	142	1	227
*Suessenguthiella scleranthoides*_2	51	2	746	111	0	233

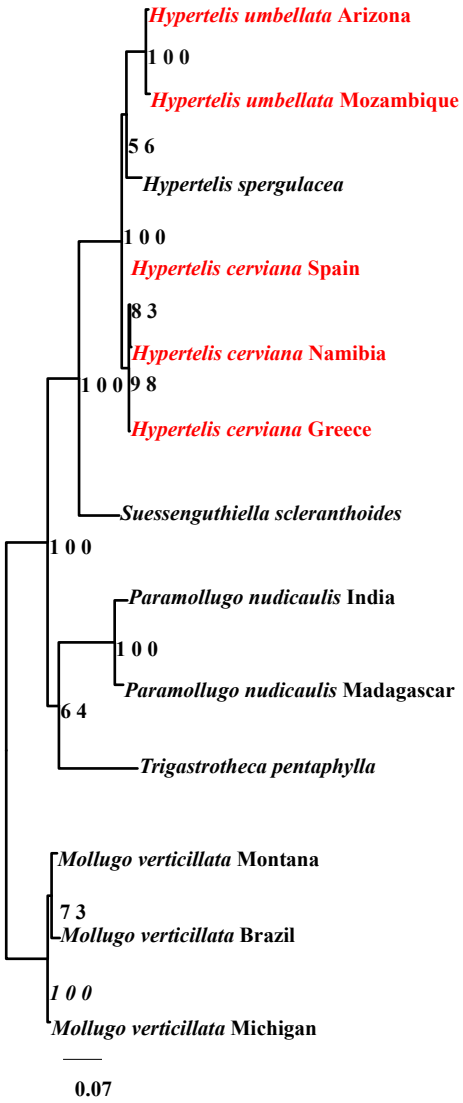

Figure S3.1.a. Phylogenetic tree of *alaat-1*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

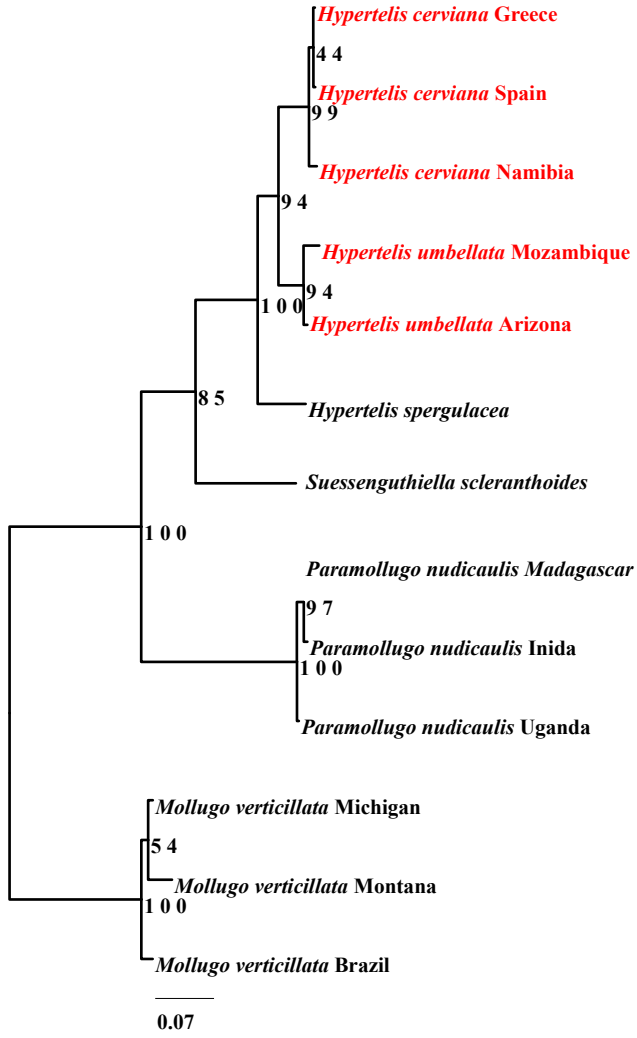

Figure S3.1.b. Phylogenetic tree of *aspat-1E1*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

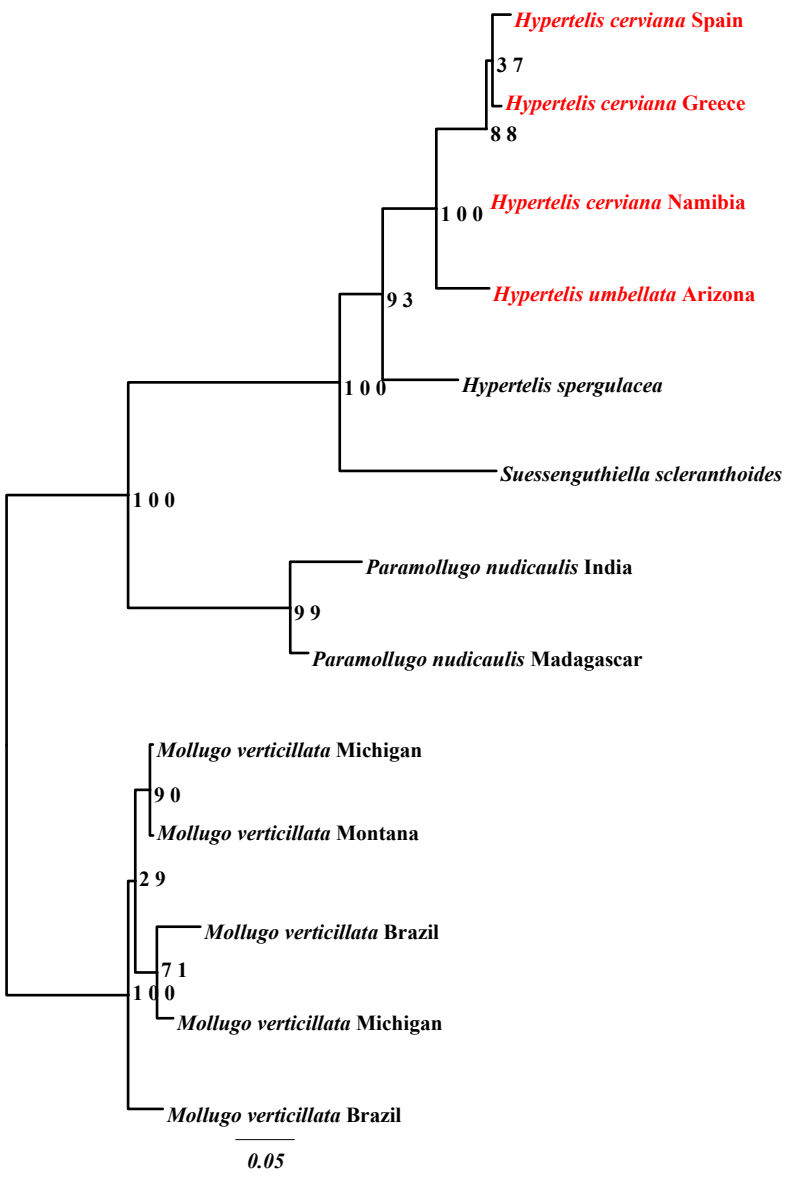

Figure S3.1.c. Phylogenetic tree of *aspat-3c1*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

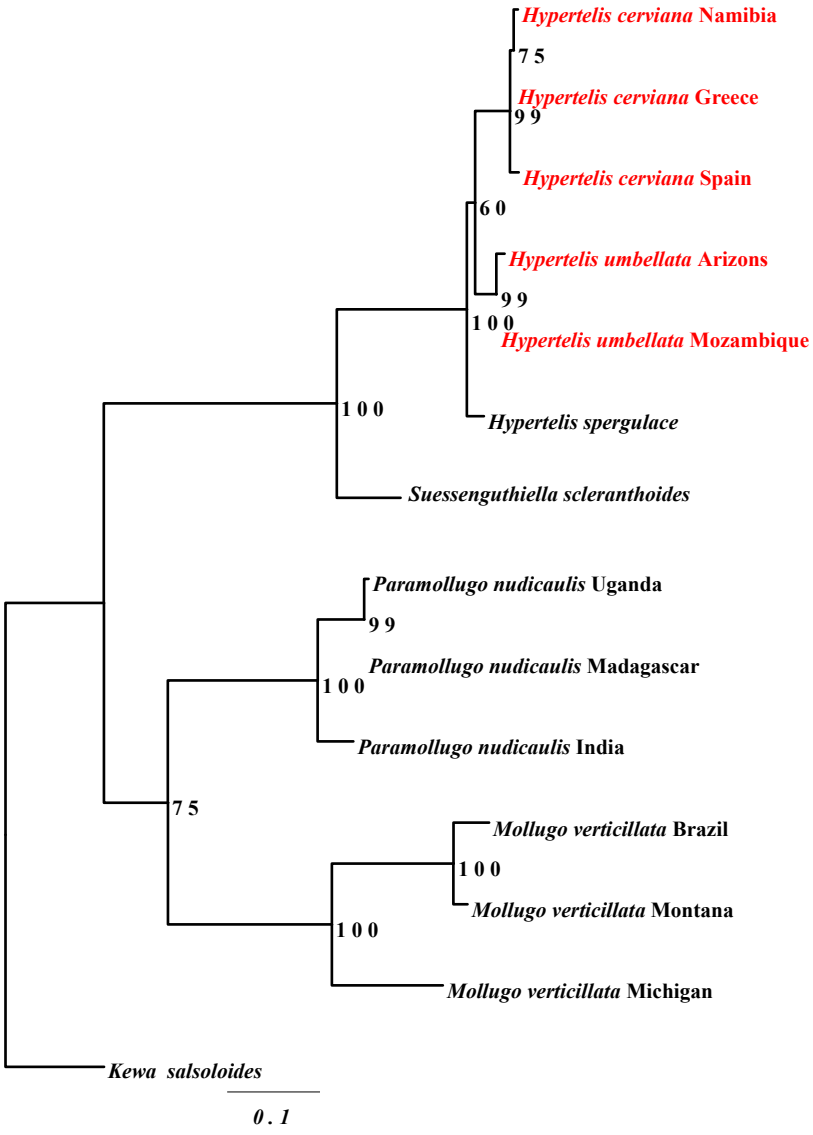

Figure S3.1.d. Phylogenetic tree of *βca-2E3*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

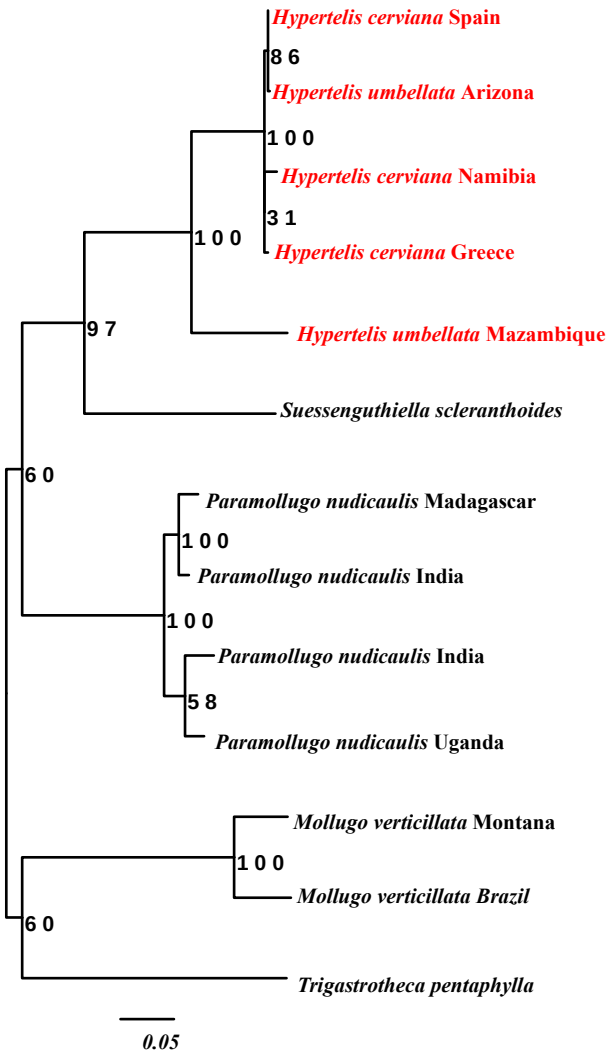

Figure S3.1.e. Phylogenetic tree of *nadme-2*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

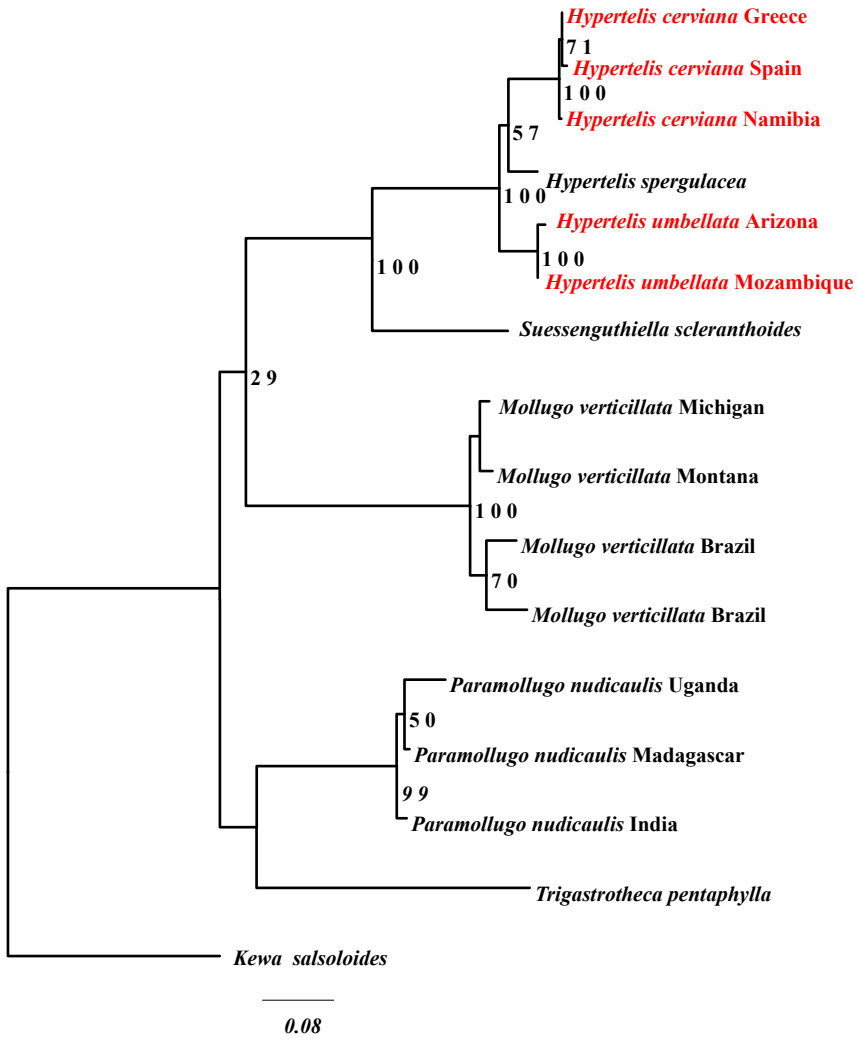

Figure S3.1.f. Phylogenetic tree of *nadmdh-2*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

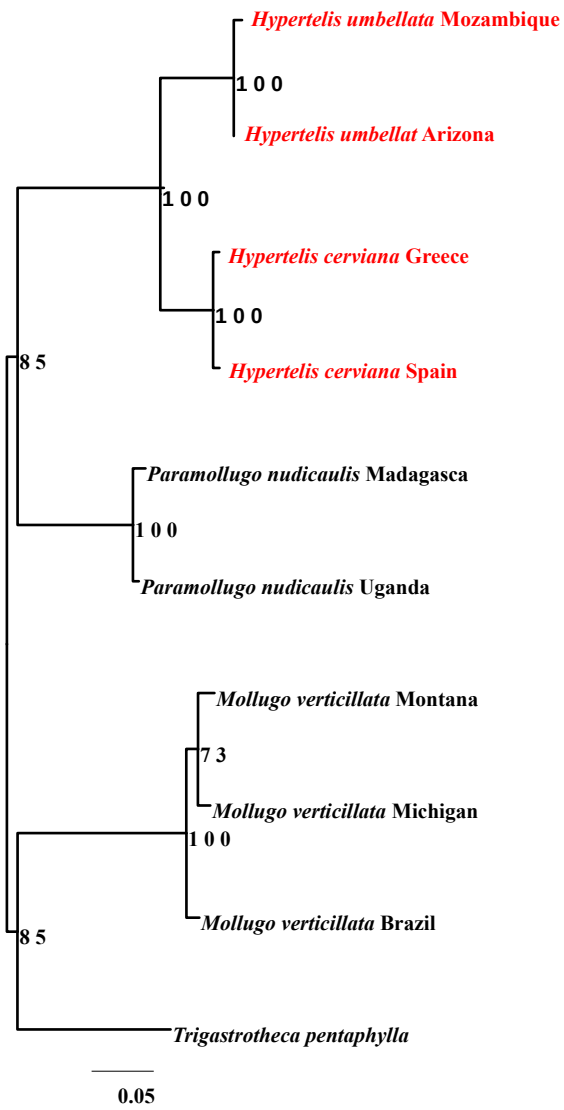

Figure S3.1.g. Phylogenetic tree of *nadpme-1E1.*

Names of C_4 accessions are in red and bootstrap values are shown near branches.

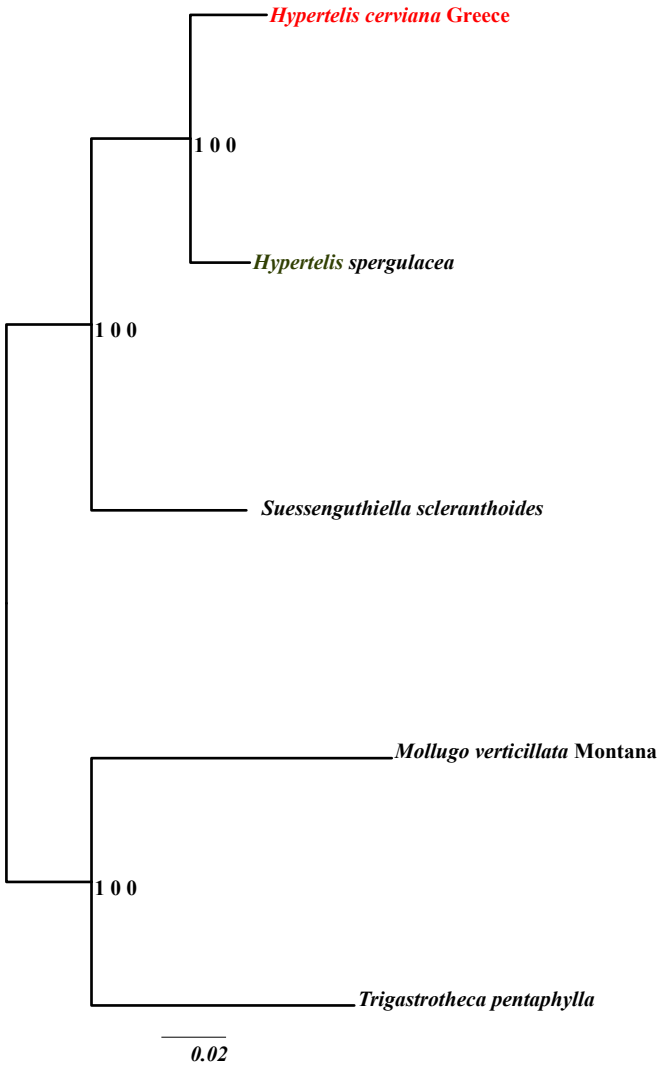

Figure S3.1.h. Phylogenetic tree of *ppc-1E1.*

Names of C_4 accessions are in red and bootstrap values are shown near branches.

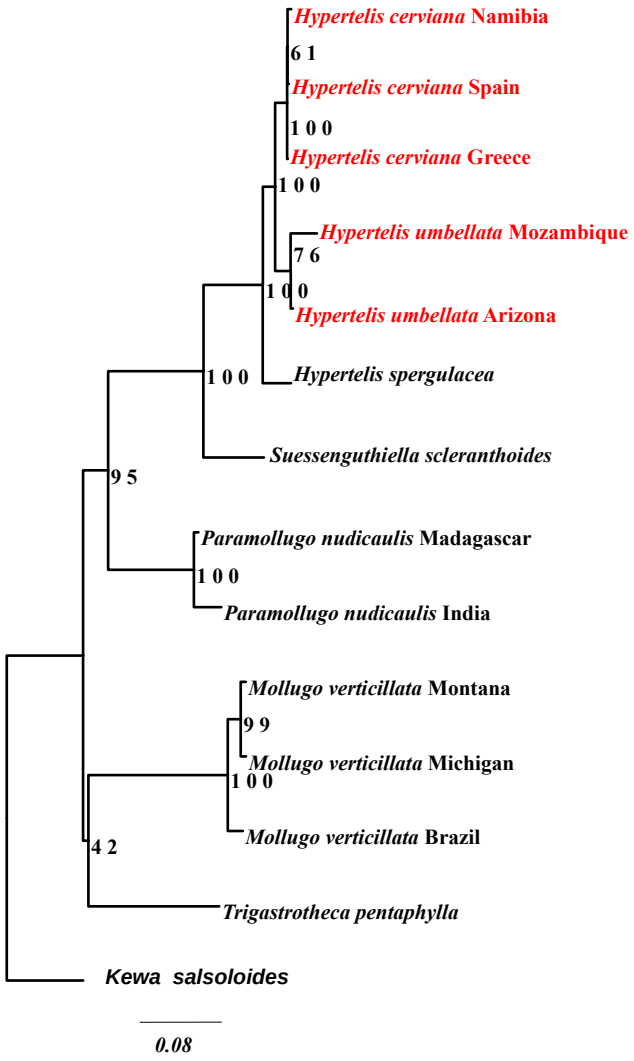

Figure S3.1.i. Phylogenetic tree of *ppdk-1C1b*.

Names of C_4 accessions are in red and bootstrap values are shown near branches.

Chapter 3:

Plastome-wide rapid evolution in a group of *Molluginaceae*

Chapter 3: Plastome-wide rapid evolution in a group of *Molluginaceae*

by Lamiaa Munshi[1]and Pascal-Antoine Christin[1]

[1] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom.

Personal contributions: [I analyzed the data and wrote the manuscript with the help of Pascal-Antoine Christin].

4.1 Abstract

Chloroplast genomes are generally well conserved among plants, accumulating substitutions at steady rates. They have consequently been the markers of choice to study relationships among plants, but cases of strong increases of evolutionary rates in some genes or lineages have been reported. In this study, we compare the rates of chloroplast evolution among members of the Molluginaceae family. While most Molluginaceae lineages have well conserved plastomes, one group of species is characterized by numerous gene losses and strongly increased rates of mutation accumulation. These high substitution rates are accompanied by numerous rearrangements and likely stem from disruption of the DNA repair machinery. The identity of genes frequently lost in Molluginaceae matches those in distant groups of plants, suggesting that these losses mainly result from relaxed selection. Conversely, the genes accumulating the most substitutions are not shared with other lineages, and we propose that they lie in repeat-rich regions undergoing numerous rearrangements. Our analyses therefore show that the fate of genes in lineages with reduced DNA repair depends on both selection pressures and the genomic localization.

Keywords: evolutionary rate, mutations, plastomes, phylogenetics, rearrangements

4.2 Introduction

Chloroplasts are responsible for photosynthesis in all autotrophic eukaryotes (Dyall et al., 2004; Wicke et al., 2011; Huang et al., 2017). They evolved via endosymbiosis between a heterotrophic eukaryote and a photosynthesising cyanobacterium (Bendich, 2004; McFadden and Van Dooren, 2004; Frailey et al., 2018). Over evolutionary time, most genes from the ancestral cyanobacterium have been lost or transferred to the nuclear or mitochondrial genome of the host (Stegemann et al., 2003; Stegemann and Bock, 2006; Olejniczak et al., 2016). The remaining chloroplast genome (plastome) is circular, haploid, and usually maternally inherited. It has around 80 genes in most plants, for a total length ranging from 120 to 160 kb in most angiosperms (Palmer, 1985; Bock, 2007; Wicke et al., 2011; Olejniczak et al., 2016; Mohanta et al., 2019). Most chloroplasts are composed of four parts; the large and small single-copy regions and the two inverted repeats (Kolodner and Tewari, 1979; Wicke et al., 2011), and the recombination between the two inverted regions helps stabilizing the structure (Martin et al., 2002; Maréchal et al., 2009).

Plant plastomes are generally highly conserved, with little variation in gene content, order, or substitution rates (Lovell and Robertson, 2010; Raubeson and Jansen 2005). Coupled with their abundance in photosynthetic cells and haploid nature, these properties made them ideal markers for genetic analyses, and they have been widely used in phylogenetics and DNA barcoding (Wolfe et al., 1987; APGIII, 2009; Moore et al., 2010; Huang et al., 2014; Nguyen et al., 2015). The accumulation of chloroplast sequences has however revealed many instances of chloroplast rearrangements, expansions, contractions or elevated rates of evolution, in a single species, large groups, and for individual genes or whole plastomes (Cai et al., 2008; Ruhlman and Jansen, 2014; Sloan et al., 2014; Park et al., 2015; Mohanta et al.,

2019). Few of these cases have been studied in details, and the mechanisms involved remain generally poorly understood.

The family Molluginaceae constitute an exciting system to understand the evolutionary dynamics of plastomes. It has species cover a variety of photosynthetic types spread across the world (Chapters 1 and 2). Previous analyses of two chloroplast markers have identified highly elevated rates of evolution in part of the family (in the *Hypertelis-Adenogramma* clade; Christin et al., 2011). Recent plastome analyses of the order Caryophyllales have confirmed that the plastomes from this clade have the highest rate of evolution of the whole order and also abundant gene losses and pseudogeneization (Yao et al., 2019). This plastome analysis however considered only two species in the *Hypertelis-Adenogramma* clade and four in the rest of the family. A denser sampling is therefore needed to understand the details of plastome dynamics in the group.

In this study, we use whole-genome sequencing to study plastome evolution in Molluginaceae. We sample 19 accessions spread across the family, with an especially dense sampling in the *Hypertelis-Adenogramma* clade. We then compare rates of evolution across species and genes to (1) delimitate the part of the family characterized by accelerated plastome evolution, (2) test whether increased evolution characterizes the whole plastomes, and (3) determine whether some genes or parts of the plastomes are more affected than others.

4.3 Methods

4.3.1 Genome sequencing.

We selected accessions from species of Molluginaceae representing different photosynthetic types and phylogenetic origins (Table 4.1). DNA was retrieved from previous studies (Christin et al., 2011) or extracted from fresh or dried leaves using the DNeasy Plant Maxi Kit (Qiagen, Hilden, Germany), following the supplier protocol. The quality and quantity of the DNA was assessed by NanoDrop (Thermo Fisher Scientific, Delaware, USA). One microgram of genomic DNA was used for library preparation using the Library Prep v. 2 Kit for Illumina (Illumina inc., California, USA), following the manufacturer's protocol. Insert sizes ranging from 350 bp to 550 bp were sequenced as 250 bp paired-end reads on the Illumina Hiseq 2500 platform at the Diagnostic Genetics Service, Sheffield Children's hospital, United Kingdom.

4.3.2 Read cleaning and assembly.

Unpaired reads and adaptor contaminations were discarded and low-quality bases ($Q > 20$) and ambiguous bases were trimmed using NxTrim v. 0.3.2 (O'Connell et al., 2015). The cleaned and filtered reads were used for reference-based plastome assembly. Reads from each accession were mapped to a previously published plastome sequence of *Mollugo verticillata* (NCBI accession MK397876_1; Yao et al., 2019) using MAFFT v. 7.402 (Katoh and Standley, 2013) with default settings. Low mapping quality reads ($Q < 5$) were removed using SAMtools v. 1.5 (Li and Durbin, 2009). For each accession, mapped reads were used to

compute a consensus sequence in Geneious v. 7.1.3 (Biomatters, Auckland, New Zealand) (Table 4.1). The assembled sequences were annotated in Geneious, and alignments corresponding to each of the 76 protein-coding genes were extracted. The gene from *Portulaca oleracea* (KY490694_1) was added to each alignment to serve as the outgroup. Genes with no reads from one accession mapped were considered as absent from this accession. Genes with less than 80% of their length covered in one accession were considered as truncated in this accession.

Table 4.1. Sampling information.

Species	Code	Voucher	Origins	Photo	Molluginaceae groups	Mapped reads
Adenogramma galoides	156	Ogburn146 (BRU)	South Africa	C_3	Adenogramma	1,700,687
Adenogramma glomerata	146	Ogburn 142 (BRU)	South Africa	C_3	Adenogramma	805,730
Hypertelis cerviana	13	-	Zambia	C_4	Hypertelis	627,525
Hypertelis cerviana	1002	Robinson 2011 (SRGH)	Zambia	C_4	Hypertelis	978,051
Hypertelis cerviana	1	-	Spain	C_4	Hypertelis	3,953,611
Hypertelis cerviana	3C	Smith 213 (CSIRO)	Australia	C_4	Hypertelis	1,146,486
Hypertelis spergulacea	5.3	JJ Moreno-Villena HYP-5.3-2 (SHD)	Namibia	C_3-C_4	Hypertelis	2,371,432
Hypertelis umbelata	3	JJ Moreno-Villena HYP-3.3-3 (SHD)	Namibia	C_4	Hypertelis	3,615,479
Hypertelis umbelata	3B	Thulin et al. 11211 (UPS)	Ethiopia	C_4	Hypertelis	5,309,321
Hypertelis walteri	7	JJ Moreno-Villena HYP-7.3-4 (SHD)	Namibia	C_4	Hypertelis	4,718,429
Mollugo verticillata	8	-	Brazil	C_3-C_4	Verticillata	3,197,079
Mollugo verticillata	9	-	Montana	C_3-C_4	Verticillata	4,204,918
Paramollugo nudicaulis	2	-	Madagascar	C_3-C_4	Paramollugo	2,110,014
Paramollugo	4	Christin 2015-18	Uganda	C_3-C_4	Paramollugo	1,100,006

nudicaulis		(SHD)				
Paramollugo nudicaulis	5	-	India	C_3-C_4	Paramollugo	1,823,355
Psammotropha quadrangularis	160	Ogburn 160 (BRU)	South Africa	C_3	Adenogramma	1,295,966
Pharnaceum confertum	153	Ogburn 163 (BRU)	South Africa	C_3	Adenogramma	1,081,341
Pharnaceum incanum	161	Ogburn 148 (BRU)	South Africa	C_3	Adenogramma	804,698
Trigastrotheca pentaphylla	-	-	India	C_3	Paramollugo	2,504,799

4.3.3 Phylogenetic analyses.

The 76 protein-coding genes were concatenated and used to infer a plastome tree (Table 4.2), using the maximum likelihood approach implemented in phyML v. 2 (Guindon and Gascuel, 2003). A substitution model GTR+G+I was used. Node support was evaluated with 100 bootstrap pseudoreplicates. Branch lengths were subsequently estimated for each gene independently, with the same settings except that the topology was fixed to that inferred from the concatenated alignment. Five genes absent from the large *Hypertelis-Adenogramma* clade were excluded from these gene-specific analyses (see Results). For each species, the sum of branch lengths from the roots was extracted from each of the 71 gene trees using the R package (R Core Team, 2020) APE (Paradis et al., 2004). To allow comparison among groups, each of these values was divided by the sum of branch lengths leading to the reference plastome of *M. verticillata*. These ratios provided gene- and accession-specific proxies of relative evolutionary rates.

Table 4.2. Functional categories of genes from Molluginaceae chloroplasts.

Functional category	Enzyme	Gene names
Photosynthesis	ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
	NADH dehydrogenase	*ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Cytochrome precursor	*petA, petB, petD, petG, petL, petN psaA,*
	Photosystem I protein	*psaA, psaB, psaC, psaI,*
	Photosystem II protein	*psbA, psbB, psbC, psbD, psbF, psbH psbJ, psbL psbN, psbT, psbZ*
Biosynthesis	Acetyl-CoA-carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Envelop membrane protein protease	*cemA*
	Protease	*clpp*
	Translational initiation	*infA*
	Maturase K	*matK*
Self-replication	Rubisco large subunit	*rbcL*
	Large subunit of ribosome	*rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rp32, rpl33, rpl36*
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
	Small subunit of ribosome	*rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19*
Unknown	Conserved open reading frames	*ycf1, ycf2, ycf3, ycf4*

4.4 Results

4.4.1 Chloroplast gene assembly and phylogenetic tree.

Sequencing reads were mapped to the assembled plastome of *Mollugo verticillata*. The coverage was excellent and continuous for species of *Mollugo*, and presented few gaps for species of *Paramollugo* and *Trigastrotheca*. By contrast, the coverage presented many gaps for species of *Hypertelis* and was very fragmented for members of the *Adenogramma* clade (*Adenogramma*, *Pharnaceum* and *Psammotropha*; Fig. 4.1). Interestingly, peaks of higher coverage, representing repeats as compared to the reference, are observed in species of the *Hypertelis-Adenogramma* clade (Fig. 4.1).

Protein-coding genes from the chloroplast genomes for 19 species of Molluginaceae representing different lineages were extracted from the mapped reads. All expected genes were detected in *Mollugo, Paramollugo* and *Trigastrotheca.* However, the five genes *accD, clpP, Infa, ycf1 and ycf2* were absent in all members of the *Hypertelis-Adenogramma clade* (Fig. 4.2). Twelve other genes were partially absent in some individuals of the *Hypertelis* and *Adenogramma* clades (Fig. 4.2). In particular, the ribosomal genes *rpl23*, *rps16* and *rps18* were truncated in most individuals within this clade, and *ndh* genes were degraded in most members of the *Adenogramma* clade (Fig. 4.2).The phylogenetic tree inferred from 76 protein coding genes was congruent with previous studies (Fig. 4.2; Christin et al., 2011; Thulin et al., 2016). However, *Mollugo verticillata* represents here the sister group to *Hypertelis-Adenogramma*, a position occupied by *Paramollugo* in previous studies. Within the *Hypertelis* clade, one sample from *H. spergulacea* was placed as sister to *H. umbelata* (Fig. 4.2), which again contradicts previous studies.

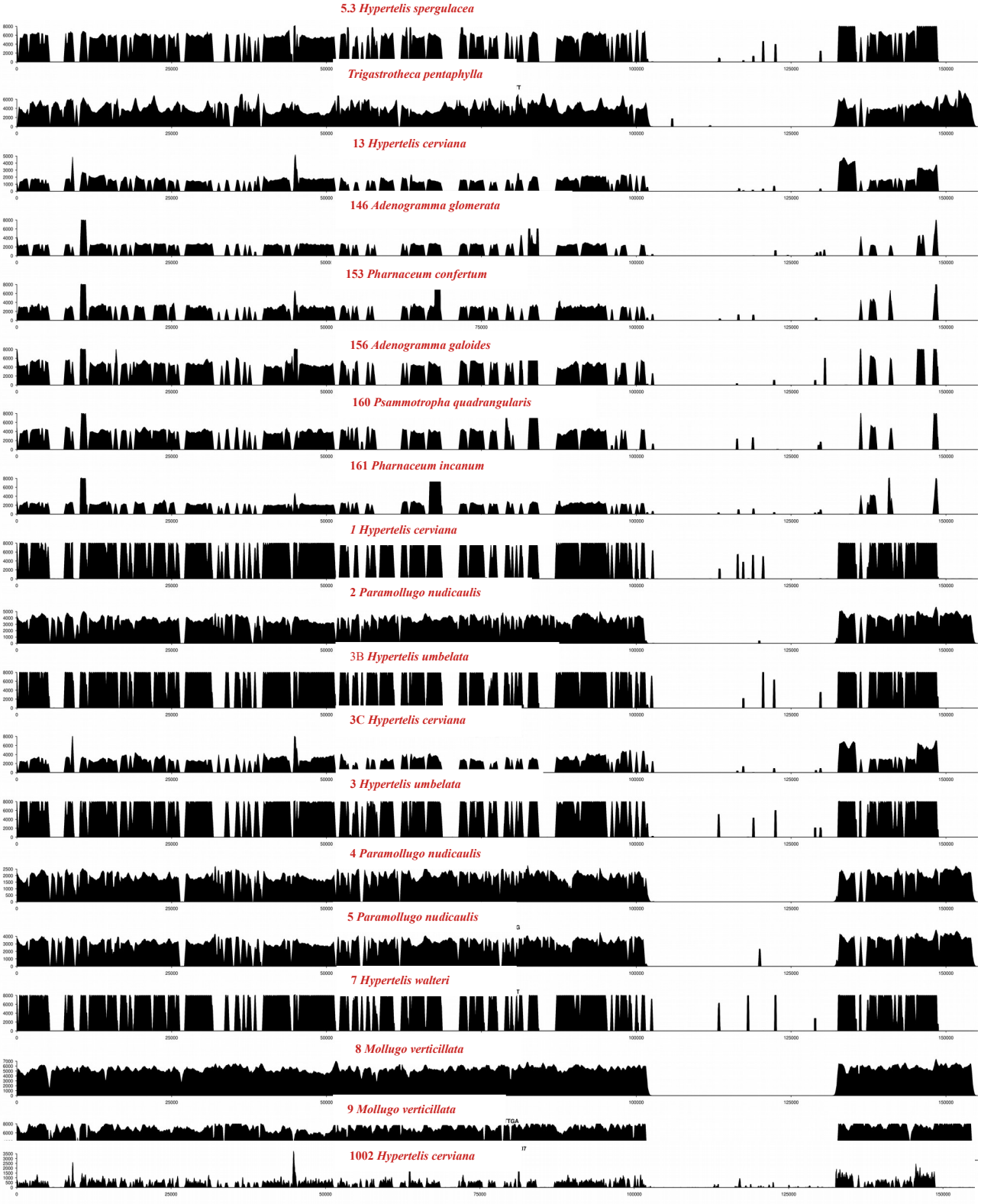

Figure 4.1. Sequencing depth of Molluginaceae plastomes.

4.4.2 Variation in chloroplast evolutionary rate among Molluginaceae.

The phylogram inferred from chloroplast protein-cording genes showed longer branches leading to the *Hypertelis-Adenogramma* group compared to the rest of the family (Fig. 4.3), as previously reported on a few markers (Christin et al., 2011) or a few individuals (Yao et al., 2019). The increased accumulation of mutations was most marked in the *Adenogramma* group (Fig. 4.3). On average, the sum of branch lengths leading to members of *Paramollugo* and *Trigastrotheca* per gene was similar to that of *Mollugo* (Fig. 4.3). In contrast, the median among genes was 4-5 times larger for members of the *Hypertelis-Adenogramma* clade, with again slightly higher values in *Adenogramma* than in *Hypertelis* (Fig. 4.3). While there are clearly strong differences among clades, there is also strong variation among genes within each individual (Fig. 4.3). Many genes have sums of branches compared to those leading to *Mollugo* that range among species from 1 to 4 (Fig. 4.4). Some genes, including in particular the ribosomal genes, were however extremely variable among species (Fig. 4.4).

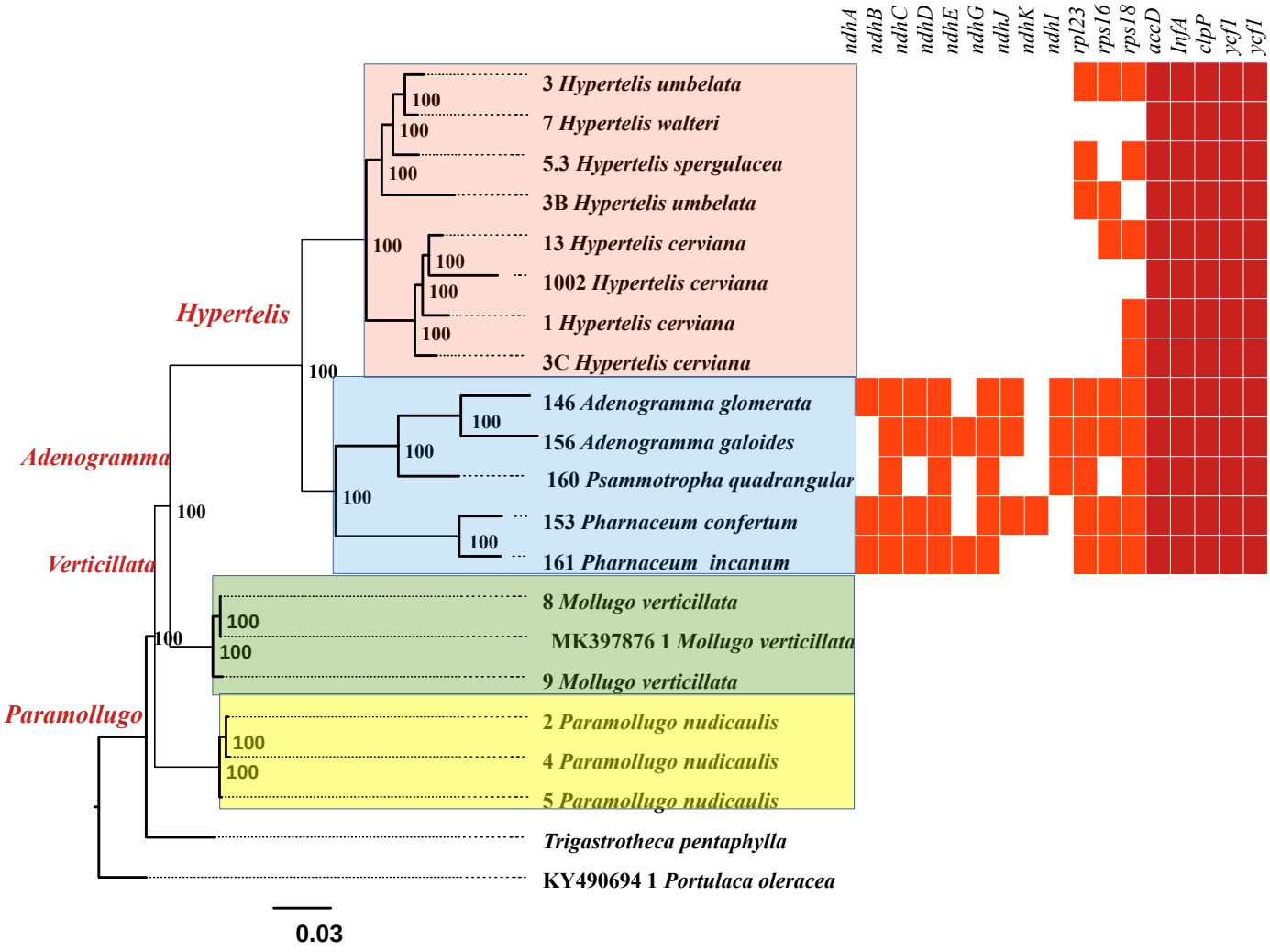

Figure 4.2. Distribution of gene losses among Molluginaceae.

The phylogeny inferred from the concatenated chloroplast alignment is indicated on the left, with bootstrap support values shown near nodes. Branches are in expected substitutions per site. Species with truncated genes are highlighted with red blocks on the right, while some with lost genes are shown with dark red blocks.

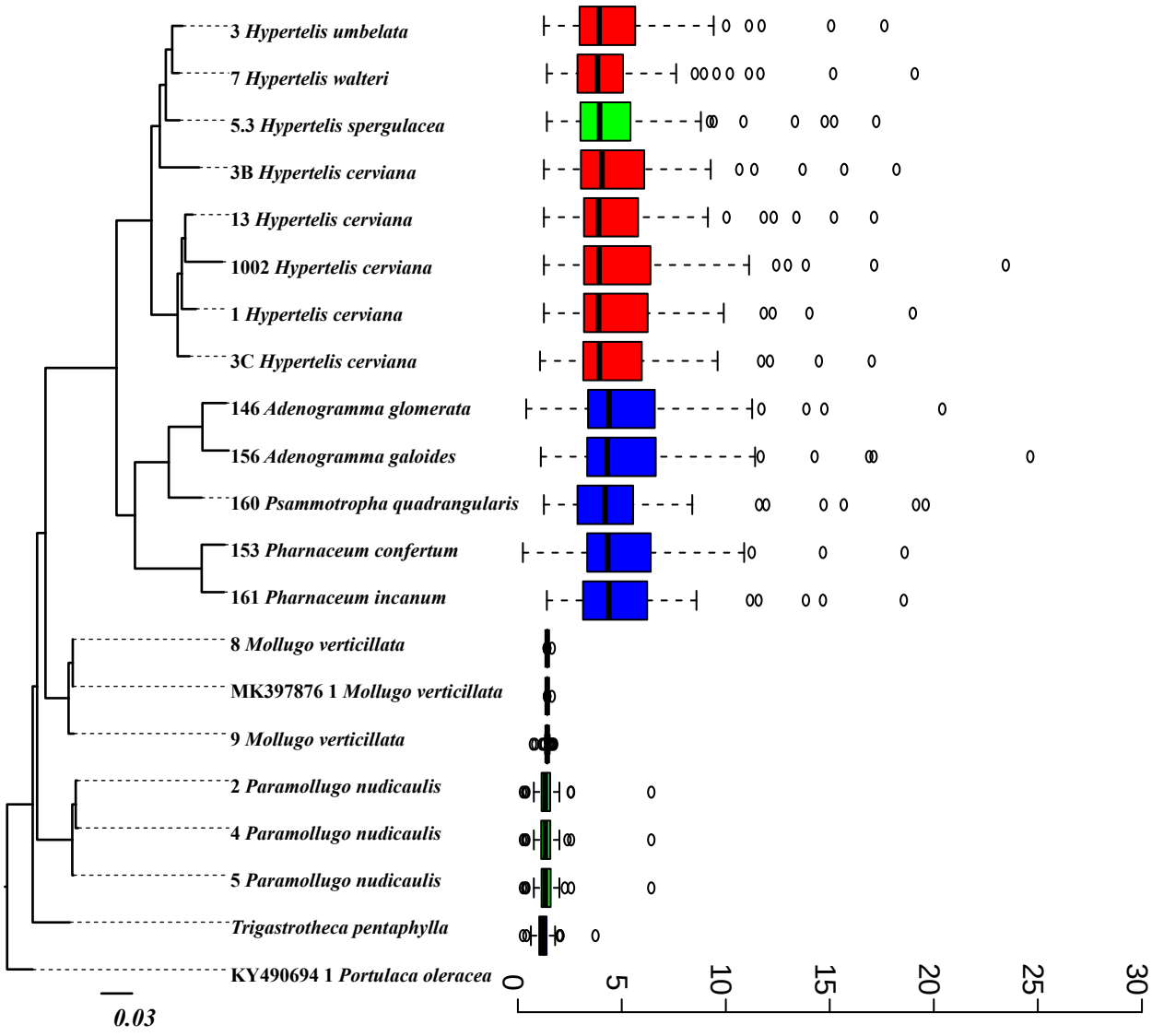

Figure 4.3. Rate of chloroplast gene evolution across Molluginaceae phylogeny.

The distribution of rates among chloroplast genes is shown for each species of Molluginaceae. Boxplots connect the interquartile range, with the median indicated. Whiskers connect the most extreme points within 1.5 times the interquartile range. Boxes are coloured according to the photosynthetic type (blue=C_3, green=C_3-C_4, red=C_4). The phylogenetic tree is shown as in Fig. 4.2.

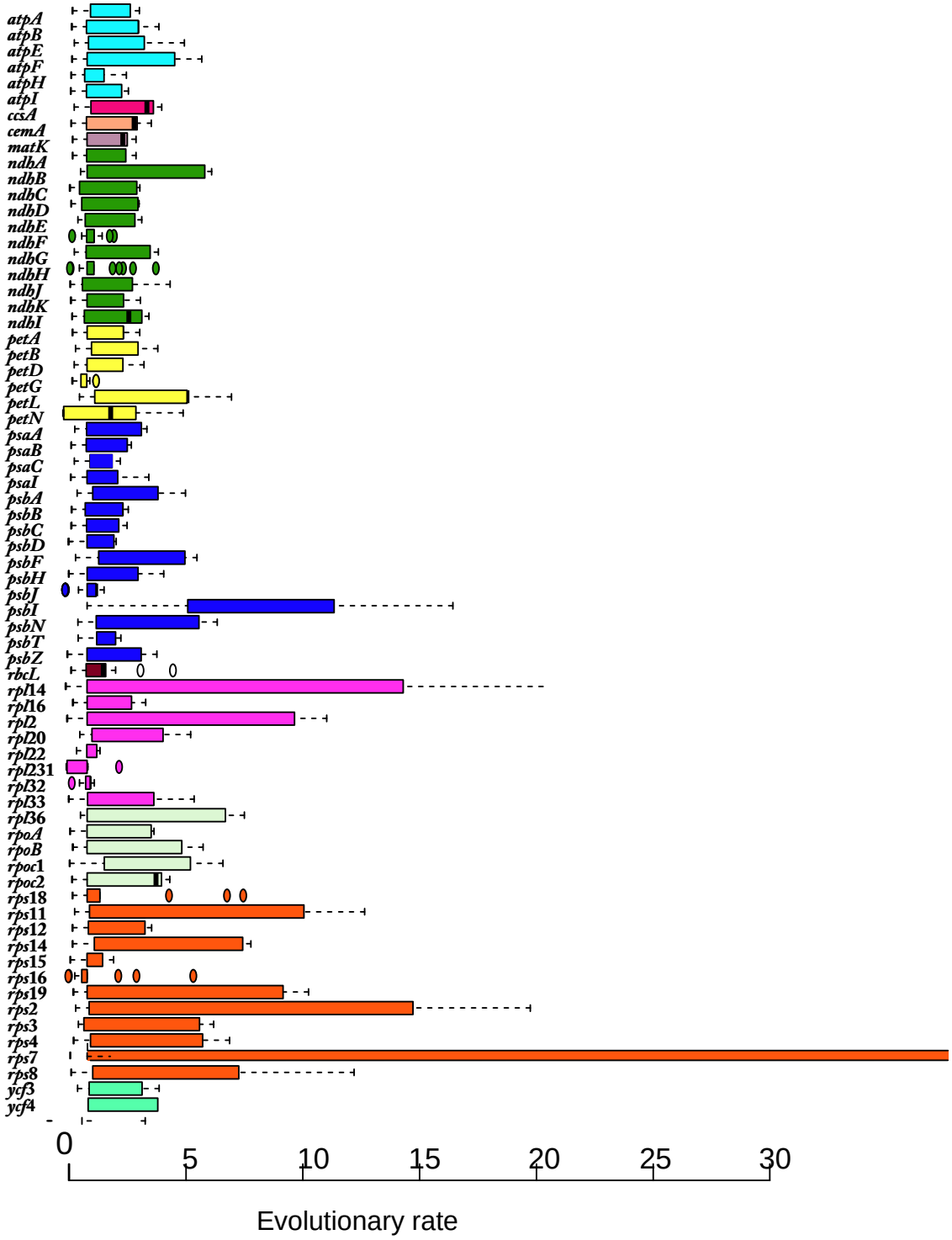

Figure 4.4. Distribution of evolutionary rates among plastome genes.

For each chloroplast gene, boxplots are used to show the distribution of rates among Molluginaceae species. Genes are coloured according to their functions (see Table 4.2).

Within the *Paramollugo* clade, most genes accumulated similar amounts of mutations as in branches leading to *M. verticillata* (Fig. 4.5). The only exception is *psbL*, which evolved faster in *Paramollugo* (Fig. 4.5). Focusing on the *Hypertelis* clade, most genes accumulated 2-4 times more mutations than in branches leading to *M. verticillata* (Fig. 4.5). The gene *psbL* and several ribosomal genes however accumulated 10-15 more mutations in *Hypertelis* than in *M. verticillata* (Fig. 4.5). All genes accumulated slightly more mutations on branches leading to members of the *Adenogramma* clade than on those leading to *Hypertelis*, and again *psbL* and some ribosomal genes represented the longest branches (Fig. 4.5). Overall, these patterns show that the increased accumulation of mutations in some Molluginaceae is driven by all chloroplast genes, but some are upper outliers even in the fast evolving species.

4.3.5 Discussion

4.5.1 Accelerated evolution characterizes the whole of the Hypertelis-Adenogramma clade.

Our analyses show that the increased rate of molecular evolution, previously reported for one *Hypertelis* and one *Pharnaceum* species (Yao et al., 2019), concerns the whole of the *Hypertelis-Adenogramma* clade (Fig. 4.3). Because all chloroplast genes are considered here, we can moreover conclude that the whole plastomes are concerned (Fig. 4.5), and not just the few genes analysed by Christin et al. (2011). An acceleration of evolution in the common ancestor of the *Hypertelis-Adenogramma* clade, followed by a return to normal rate, would lead to inflated sums of branches in all members of the *Hypertelis-Adenogramma* clade.

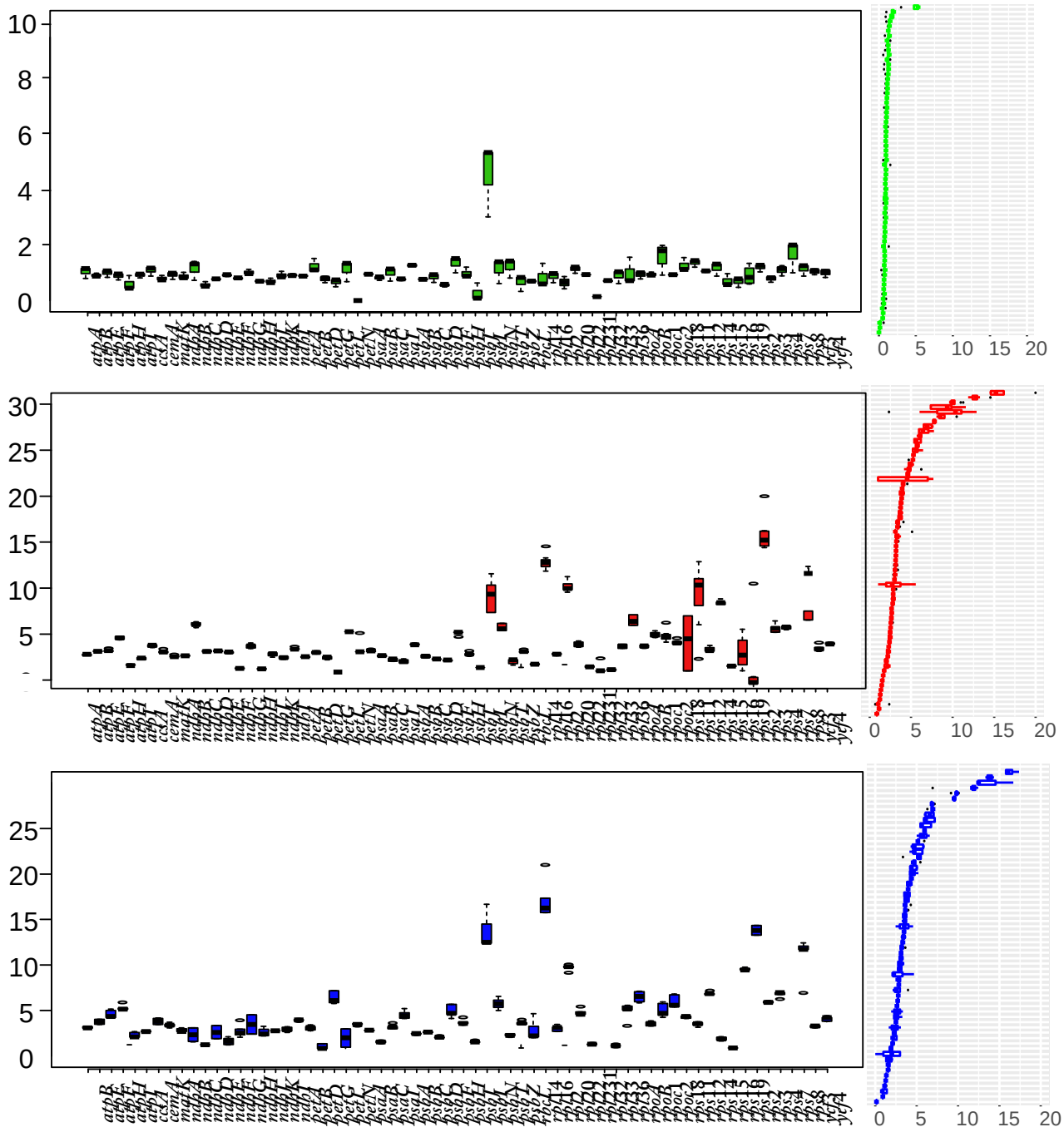

Figure 4.5. Variation of rates among genes of subgroups of Molluginaceae.

For three groups of Molluginaceae (green = *Paramollugo*, red = *Hypertelis,* blue = *Adenogramma*), the rates of chloroplast genes are shown with boxplots. The same boxplots are sorted by values in the right panels.

However, the branches within each of the *Hypertelis* and *Adenogramma* clades, or the single branch leading to the common ancestor of *Hypeterlis* and *Adenogramma*, all exceed the sum of branches leading from the root to *M. verticillata* (Fig. 4.3). These patterns unambiguously show that the increased evolutionary rate was sustained in the whole of the *Hypertelis-Adenogramma* clade. Slight variation through times could still have occurred, as illustrated by the longest branches in *Adenogramma* and *Pharnaceum* (Fig. 4.3).

The increased evolutionary rate observed on Molluginaceae chloroplast was not apparent in previous analyses of selected nuclear markers (Christin et al., 2011; Chapter 2). Such accelerated evolution specific to the organellar genomes has been reported before in Geraniaceae (Parkinson et al., 2005; Cai et al., 2008; Guisinger et al., 2011). *Plantago* (Cho et al., 2004; Weng et al., 2014) and *Silene* (Sloan et al., 2014). In Geraniaceae, the increased rates of substitution coincide with frequent recombinations linked to abundant repeats (Guisinger et al., 2011, Weng et al., 2014), while the structure of chloroplasts is usually highly conserved in land plants. The same process is apparent in Molluginaceae. We were not able to *de novo* assemble plastomes of the *Hypertelis-Adenogramma* clade, but mapping of reads showed numerous gaps (Fig. 4.1), and frequent rearrangements were suggested previously for two species (Yao et al., 2019). In Geraniaceae, the increased rate of repeat accumulation and rearrangements in chloroplasts was putatively linked to improper DNA repair, and mutations similarly disrupting the organelle repair machinery might have occurred at the base on the *Hypertelis-Adenogramma* clade.

4.5.2 Ribosomal genes are the most affected by highest evolutionary rate.

The accelerated evolution in the *Hypertelis-Adenogramma* clade concerns all chloroplast genes (Fig. 4.5), but some are disproportionately affected. Five genes were lost at the base of the *Hypertelis-Adenogramma*, and these same genes have been lost in a variety of other lineages of plants (Millen et al., 2001; McNeal et al., 2007; Guisinger et al., 2011; Weng et al., 2014; Zhitao et al., 2017; Mohanta et al., 2019). In addition, several genes are truncated in species spread across the *Hypertelis-Adenogramma* clade (Fig. 4.2), suggesting independent deletions. In some cases, our inability to detect full genes might be linked to lower sequencing depth following chloroplast-to-nuclear transfers, but this is unlikely to account for all the instances we report here.

Some of the genes partially absent from our Molluginaceae chloroplasts have been independently pseudogeneized in a larger number of Caryophyllales (Yao et al., 2019) and other groups of plants (Millen et al., 2001; Jansen et al., 2007; Jansen et al., 2008), showing that they are not essential in some conditions. Strikingly, most *ndh* genes for NADH dehydrogenase-like complex are partially absent from species of the *Adenogramma* clade (Fig. 4.2). These genes have been independently lost in a variety of land plants (Wakasugi et al., 1994; Braukmann et al., 2009; Blazier et al., 2011; Sanderson et al., 2015), and it was suggested that alternative pathways made the *ndh* genes redundant (Martín and Sabater, 2010; Sanderson et al., 2015). Overall, genes recurrently lost in Molluginaceae are physically spread across the plastome of *M. verticillata*. Coupled with their repeated losses in various groups of land plants, this suggests that they degenerate because of relaxed selection and not because they occur in mutation hotspots.

In addition to losses of some genes from the chloroplast genome, we show that other genes underwent drastically increased evolution in the *Hypertelis-Adenogramma* clade while retaining their function. This is especially the case of *psbL* in addition to multiple ribosomal genes (Fig.4.5). These genes are not consistently undergoing accelerated evolution in other groups with fast chloroplast evolution (Cai et al., 2008; Sloan et al., 2014; Zhu et al., 2016), and we conclude that the mechanisms driving the accumulation of mutations in these genes might be specific to Molluginaceae. In other groups, chloroplast regions undergoing frequent rearrangements have been associated with increased substitution rates (Magee et al., 2010; Sloan et al., 2014; Zhu et al., 2016). We consequently speculate that these genes are present in Molluginaceae near repeat sequences driving frequent rearrangements.

4.6 Conclusions

Using genome-wide sequence data, we show that the whole chloroplast genomes evolve faster in a subgroup of Molluginaceae, and propose that this reflects a disruption of the DNA repair machinery. All genes are affected, but some are more frequently degraded or lost. The genes frequently lost in Molluginaceae match those absent from other lineages of plants, and this bias likely reflects a relaxation of selection on genes that can be non essential. Other genes, which remain functional, undergo excessively high evolutionary rates in some Molluginaceae. Because these fast evolving genes are not shared with other lineages of plants, they likely correspond to those close to repeats creating frequent rearrangements in Molluginaceae. We conclude that selection and physical localization across the genome both determine the fate of individual genes following decreased DNA repair.

4.7 References

APG III (Angiosperm Phylogeny Group). 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*. **161**, pp.105–121

Bendich, A.J. 2004. Circular chloroplast chromosomes: The grand illusion. *Plant Cell*. **16**(7), pp.1661–1666.

Blazier, C.C., Guisinger, M.M. and Jansen, R.K. 2011. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Molecular Biology*. **76**(3–5), pp.263–272.

Bock, R. 2007. Structure, function, and inheritance of plastid genomes. In: Bock R (ed). Cell and Molecular Biology of Plastids. *Springer*. Berlin Heidelberg, pp 29–63.

Braukmann, T.W.A., Kuzmina, M. and Stefanović, S. 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: Extent and evolutionary significance for the seed plant phylogeny. *Current Genetics*. **55**(3), pp.323–337.

Cai, Z., Guisinger, M., Kim, H.G., Ruck, E., Blazier, J.C., McMurtry, V., Kuehl, J. V., Boore, J. and Jansen, R.K. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution*. **67**(6), pp.696–704.

Cho, Y., Mower, J.P., Qiu, Y.L. and Palmer, J.D. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences: USA*. **101**(51), pp.17741–17746.

Christin, P.A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R. and Sage, R.F. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution*. **65**(3), pp.643–660.

Dyall, S.D., Brown, M.T. and Johnson, P.J. 2004. Ancient invasions: From endosymbionts to organelles. *Science*. **304**(5668), pp.253–257.

Frailey, D.C., Chaluvadi, S.R., Vaughn, J.N., Coatney, C.G. and Bennetzen, J.L. 2018. Gene loss and genome rearrangement in the plastids of five hemiparasites in the family Orobanchaceae. *BMC Plant Biology*. **18**(1), pp.1–12.

Gao, F., Chen, C., Arab, D.A., Du, Z., He, Y. and Ho, S.Y.W. 2019. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and Evolution*. **9**(7), pp.3891–3898.

Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. **52**(5), pp.696–704.

Guisinger, M.M., Kuehl, J. V., Boore, J.L. and Jansen, R.K. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. *Molecular Biology and Evolution*. **28**(1), pp.583–600.

Huang, C.F., Yu, C.P., Wu, Y.H., Lu, M.Y.J., Tu, S.L., Wu, S.H., Shiu, S.H., Ku, M.S.B. and Li, W.H. 2017. Elevated auxin biosynthesis and transport underlie high vein density in C_4 leaves. *Proceedings of the National Academy of Sciences: USA.* **114**(33), pp.E6884–E6891.

Huang, D.Q., Yang, J.T., Zhou, C.J., Zhou, S.D. and He, X.J. 2014. Phylogenetic reappraisal of *Allium* subgenus *Cyathophora* (Amaryllidaceae) and related taxa, with a proposal of two new sections. *Journal of Plant Research*. **127**(2), pp.275–286.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J. V. and Boore, J.L. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences: USA*. **104**(49), pp.19369–19374.

Jansen, R.K., Wojciechowski, M.F., Sanniyasi, E., Lee, S.B. and Daniell, H. 2008. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Molecular Phylogenetics and Evolution*. **48**(3), pp.1204–1217.

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*. **30**(4), pp.772–780.

Kolodner R., Tewari K. 1979. Inverted repeats in chloroplast DNA from higher plants. *Proceedings of the National Academy of Sciences: USA*. **76**, pp.41–45.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**(14), pp.1754–1760.

Lovell, S.C. and Robertson, D.L. 2010. An integrated view of molecular coevolution in protein-protein interactions. *Molecular Biology and Evolution*. **27**(11), pp.2567–2575.

Magee, A.M., Aspinall, S., Rice, D.W., Cusack, B.P., Sémon, M., Perry, A.S., Stefanović, S., Milbourne, D., Barth, S., Palmer, J.D., Gray, J.C., Kavanagh, T.A. and Wolfe, K.H. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*. **20**(12), pp.1700–1710.

Maréchal, A., Parent, J.S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F. and Brisson, N. 2009. Whirly proteins maintain plastid genome stability in Arabidopsis. *Proceedings of the National Academy of Sciences: USA*. **106**(34), pp.14693–14698.

Martín, M. and Sabater, B. 2010. Plastid *ndh* genes in plant evolution. *Plant Physiology and Biochemistry*. **48**(8), pp.636–645.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. and Penny, D. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial

genes in the nucleus. *Proceedings of the National Academy of Sciences: USA.* **99**(19), pp.12246–12251.

McFadden, G.I. and Van Dooren, G.G. 2004. Evolution: Red algal genome affirms a common origin of all plastids. *Current Biology*. **14**(13), pp.514–516.

McNeal, J.R., Kuehl, J. V., Boore, J.L. and De Pamphilis, C.W. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology*. **7**, pp.1–22.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Clifford, W., Calie, P.J., Jermiin, L.S., Wolfe, K.H., Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, C.J.D., Lao, C.N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermiin, L.S. and Wolfed, K.H. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell.* **13**, pp.645–658.

Mohanta, T.K., Khan, Adil, K., Abdullatif, Abd_Allah, E.F. and Al-Harrasi, A. 2019. Gene loss and evolution of the plastome. *BioRxiv*. 676304.

Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G. and Soltis, D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences : USA.* **107**(10), pp.4623–4628.

Nguyen, N.P.D., Mirarab, S., Kumar, K. and Warnow, T. 2015. Ultra-large alignments using phylogeny-aware profiles. *Genome Biology*. **16**(1), pp.1–15.

O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M.M., Gormley, N.A. and Cox, A.J. 2015. NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics*. **31**(12), pp.2035–2037.

Olejniczak, S.A., Łojewska, E., Kowalczyk, T. and Sakowicz, T. 2016. Chloroplasts: state of research and practical applications of plastome sequencing. *Planta*. **244**(3), pp.517–527.

Paradis, E., Claude, J. and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. **20**(2), pp.289–290.

Palmer, J.D. 1985. Comparative organization of chloroplast genomes. *Annual Reviews of Genetics*. **19**, 325–354.

Park, S., Jansen, R.K. and Park, S.J. 2015. Complete plastome sequence of (Ranunculaceae) and transfer of the gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biology*. **15**(1), pp.1–12.

Parkinson, C.L., Mower, J.P., Qiu, Y.L., Shirk, A.J., Song, K., Young, N.D., DePamphilis, C.W. and Palmer, J.D. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evolutionary Biology*. **5**, pp.1–12.

Sanderson, M.J., Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J.L.M., Eguiarte, L.E., Kumar, S., Lee, H.O., Lee, J., McMahon, M., Steele, K., Wing, R., Yang, T.J., Zwickl, D. and Wojciechowski, M.F. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat 1. *American Journal of Botany*. **102**(7), pp.1115–1127.

Sloan, D.B., Triant, D.A., Forrester, N.J., Bergner, L.M., Wu, M. and Taylor, D.R. 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). *Molecular Phylogenetics and Evolution*. **72**(1), pp.82–89.

Stegemann, S. and Bock, R. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell*. **18**(11), pp.2869–2878.

Stegemann, S., Hartmann, S., Ruf, S. and Bock, R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*. **100**(15), pp.8828–8833.

Raubeson, L.A., and Jansen, R.K. 2005. Chloroplast genomes of plants. In: Henry R (ed). diversity and evolution of plants genotypic and phenotypic variation in higher plants. *CABI Publishing.* pp. 45–68.

Ruhlman, T.A., and Jansen, R.K. 2014. The plastid genomes of flowering plants. In: Maliga P, (ed). Chloroplast biotechnology. Totowa, (NJ): *Humana Press*. pp.3–38.

Thulin, M., Moore, A.J., El-Seedi, H., Larsson, A., Christin, P.A. and Edwards, E.J. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. **65**(August), pp.775–793.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T. and Sugiura, M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences of the United States of America*. **91**(21), pp.9794–9798.

Weng, M.L., Blazier, J.C., Govindu, M. and Jansen, R.K. 2014. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution*. **31**(3), pp.645–659.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F. and Quandt, D. 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*. **76**(3–5), pp.273–297.

Wolfe, K.H., Li, W.-H. and Sharp, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences: USA*. **84**, pp.9054–9058.

Yao, G., Jin, J.J., Li, H.T., Yang, J.B., Mandala, V.S., Croley, M., Mostow, R., Douglas, N.A., Chase, M.W., Christenhusz, M.J.M., Soltis, D.E., Soltis, P.S., Smith, S.A., Brockington, S.F., Moore, M.J., Yi, T.S. and Li, D.Z. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution*. **134**(February), pp.74–86.

Zhitao, N., Shuying, Z., Jiajia, P., Ludan, L., Jing, S. and Xiaoyu, D. 2017. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Scientific Reports*. **7**(1), pp.1–11.

Zhu, A., Guo, W., Gupta, S., Fan, W. and Mower, J.P. 2016. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytologist*. **209**(4), pp.1747.

General discussion

5.1 General discussion

5.1.1 Drastic anatomical changes are followed by drastic biochemical changes, but both continue when the plants are C_4.

As a complex trait, C_4 photosynthesis is thought to have evolved via the gradual accumulation of anatomical and biochemical components (Sage, 2004; Sage et al., 2012). The effects of these additions when plants are not C_4 remain debated. Indeed, some studies assume that each component leads to an increase of fitness via continuous improvements of the C_4 pathway (Brown and Hattersley, 1989; Heckmann et al., 2013), while accumulating evidence suggests that some C_4 components are initially selected for reasons unrelated to the C_4 pathway (Griffiths et al., 2013; Williams et al., 2013; Mallmann et al., 2014), and neutral evolution has even been suggested (Karki et al., 2020). Our understanding of the order of acquisition of C_4 components was strongly increased by studies of different groups with variation in photosynthetic types (Frohlich, 1978; Monson and Rawsthorne, 2000; Hilger and Diane, 2003; Sage, 2004; Martins and Scatena, 2011; Lundgren et al., 2016). As a small group with the whole spectrum of photosynthetic types, Molluginaceae represents one such system. My PhD work has tracked the evolution of both anatomical (Chapter 1) and biochemical (as estimated from gene expression; Chapter 2) components of C_4 photosynthesis across Molluginaceae.

My investigations have confirmed that some C_4-like anatomical components were acquired before the full C_4 physiology (Chapter 1), as previously suggested based on a small species sample (Christin et al., 2011). Importantly, my detailed quantification of anatomical traits showed that the C_4 anatomy emerged via the combination of properties that can all be

found in isolation in some C_3 plants (Chapter 1). These results indicate that individual anatomical components were not initially selected for the C_4 pathway, but evolved for unrelated reasons. I hypothesize that combinations of traits allowing the C_4 physiology appeared by chance or for different reasons, and in some cases were subsequently co-opted for C_4 evolution. My data do not allow establishing whether the evolution of the individual traits or their combination was driven by neutral or adaptive processes. Establishing the pressures acting on individual anatomical traits in a C_3 context would require specific investigations. For example, C_3 species with variations in these traits could be subjected to experimental evolution, to establish the environmental conditions that favour each trait. In the meantime, my anatomical work contributes to the broader picture that C_4 anatomy evolves through the combination of components that all exist in C_3 plants (Christin et al., 2013; Griffiths et al., 2013; Lundgren et al., 2014; Ermakova et al., 2020).

My investigations of Molluginaceae transcriptomes revealed that high expression of genes for core C_4 enzymes is restricted to C_4 species (Chapter 2). Indeed, the C_3-C_4 *Hypertelis spergulacea*, which is closely related to the C_4 species, does not markedly differ from other non-C_4 Molluginaceae in terms of gene expression (Chapter 2). These results suggest that the upregulation of genes required to generated the C_4 biochemistry happened once the leaf anatomy was already in place. Similar conclusions were reached before based mainly on descriptions of other study systems (Sage et al., 2004; Edwards, 2019). However, detailed quantifications of the anatomical and biochemical variations related to C_4 evolution have pointed to continuous changes of the two sets of characters (Christin et al., 2013; Williams et al., 2013; Bianconi et al., 2020). These two scenarios are not mutually exclusive. Indeed, my results indicate that the most drastic anatomical modifications happened along the branch

leading to the common ancestor of *Hypertelis*, which was not C_4 (Chapter 1). The most drastic gene upregulations then happened on branches leading to each of the C_4 species of *Hypertelis* (Chapter 2). Both anatomical and biochemical components however vary among populations of each C_4 species (Chapters 1 and 2), which shows that their modifications continued in the recent past. I conclude that drastic modifications of the anatomy predate drastic modification of the biochemistry, but adaptations of both the anatomy and the biochemistry continue when the plants are in a C_4 state. Over time, these secondary changes lead to a large diversity of species within each C_4 lineage or, in the case of the young C_4 origins of Molluginaceae, a diversity among accessions of each C_4 species.

5.1.2 On the status of C_3-C_4 intermediates in Molluginaceae.

Since their discovery, C_3-C_4 species have received the attention of numerous biologists as they are thought to represent an intermediate stage during C_4 evolution (Monson and Moore, 1989; Sage, 2004; McKown et al., 2005). Their exact status remains however debatable. It has recently been argued that they might represent hybrids between C_3 and C_4 species (Kadereit et al., 2017), while others have suggested some C_3-C_4 might represent descendants of C_4 lineages that lost some C_4 traits (Ocampo et al., 2013). The status of C_3-C_4 species of Molluginaceae is similarly ambiguous. Because they are not related to any known C_4 species and do not present signs of a hybrid origin (Chapters 1, 2 and 3), both *Mollugo verticillata* and *Paramollugo nudicaulis* likely result from transitions from C_3 to C_3-C_4 photosynthesis. Being nested in an otherwise C_4 clade, *H. spergulacea* might result from the loss of C_4 characters. My results do however not support this scenario. Indeed, the transcriptome of *H.*

spergulacea is not markedly different from that of more distantly-related non-C_4 Molluginaceae and positive selection on C_4 genes happened after the C_4 species diverged from *H. spergulaea* (Chapter 2). This species might therefore represent a true evolutionary intermediate on the way to C_4 evolution. However, more complex scenarios cannot be ruled out. First, the common ancestor of the *Hypertelis* clade might have been C_3, and then each C_4 species might have made the transition from C_3 to C_4 while *H. spergulacea* transitioned from C_3 to C_3-C_4. Second, the common ancestor of *H. spergulacea* and the C_4 species might have had a C_3-C_4 physiology that differs from that of extant *H. spergulacea*. Despite these uncertainties surrounding their origins, C_3-C_4 species can be used as proxies for possible ancestral states along the way to C_4 evolution. I will use this approach to describe how C_3-C_4 states might have helped C_4 evolution in Molluginaceae.

My results confirm that *H. spergulacea* presents a leaf anatomy similar to that of closely related C_4 Molluginaceae (Chapter 1), and C_3-C_4 in this case can consequently be said to bridge the gap between C_3 and C_4 leaf anatomy. However, the two other C_3-C_4 Molluginaceae (*M. verticillata* and *P. nudicaulis*) are more similar to C_3 than C_4 Molluginaceae (Chapter 1), showing that C_4-like anatomical characters are not inherent to the C_3-C_4 state. Physiologically, C_3-C_4 species are characterized by a CO_2-recycling mechanisms (often termed C_2 photosynthesis) based on the shuttling between mesophyll and bundle sheath cells (Hylton et al., 1988; Sage et al,. 2012). This pathway is similar to the C_4 trait in that it requires an increased concentration of chloroplasts containing Rubisco into the bundle sheath and an exchange of metabolites between the mesophyll and bundle sheath cells. It is therefore assumed to select for short distances between mesophyll and bundle sheath cells and increased investment into the bundle sheaths, two properties required for C_4 function (Sage,

2004; Sage et al., 2012; Edwards, 2019). However, my anatomical comparisons show that increased investment into bundle sheaths is found only in *H. spergulacea* and not the two other C_3-C_4 Molluginaceae (Chapter 1). I conclude that the C_3-C_4 state does not necessarily select for a more C_4-like anatomy. Instead, I suggest that it selects for a stabilization of a C_4-like anatomy once this emerges for unrelated reason. By guaranteeing that C_4-like anatomical parameters persist over million of years, the C_3-C_4 physiology would then increase opportunities to subsequently evolve C_4 photosynthesis. If the common ancestor of the *Hypertelis* was C_3-C_4, this would then have facilitated recurrent transitions to C_4 photosynthesis.

Besides anatomical traits, C_3-C_4 species have been reported to exhibit a gradual increase of expression of genes for core C_4 enzymes in *Flaveria* (Mallmann et al., 2014). While this is in some cases linked to a weak C_4 pathway acting in some C_3-C_4 (so called 'type II intermediates'), it was also reported in C_3-C_4 without any C_4 activity and was interpreted as a mechanism to rebalance nitrogen among cell types (Mallmann et al., 2014). My transcriptome analyses of Molluginaceae did not detect increased expression of genes encoding C_4 enzymes in the C_3-C_4 *H. spergulacea* (Chapter 2). I conclude that, in this case at least, the C_3-C_4 state did not bridge the gap between C_3 and C_4 biochemistries. Instead, all changes in gene expression and positive selection on coding sequences seem to have happened in each of the C_4 species (Chapter 2). The situation is however different for the other C_3-C_4 species. Indeed, some populations of *M. verticillata* and *P. nudicaulis* show increased expression of the important enzymes PEPC and/or PPDK (Chapter 2). Increased activity of some C_4 enzymes has been reported in some populations of *M. verticllata* (Sayre et al., 1979), and this was shown to lead to a weak C_4 cycle (Sayre and Kennedy, 1977). The

upregulation of some C_4 genes observed in my transcriptomes (Chapter 2) is therefore likely linked to the emergence of a weak C_4 cycle. However, *M. verticillata* is not related to any known C_4 lineage (Chapters 1, 2, and 3; Thulin et al., 2016), and its weak C_4 pathway has therefore not yet been co-opted to evolve C_4 photosynthesis.

Taken together, my anatomical and transcriptomic investigations of Molluginaceae indicate that C_3-C_4 species can exhibit C_4-like anatomical trait and gene expression independently from each other. C_4 models do indicate that both are needed for an efficient C_4 cycle, but C_4 photosynthesis in the family emerged only in the group containing C_3-C_4 species with C_4-like anatomy (i.e. *H. spergulace*a; Chapter 1). It cannot be excluded that, with more time, the C_3-C_4 with C_4-like gene expression (i.e. *M. verticillata*; Chapter 2) would have produced descendants with a C_4 physiology. For the moment, the observed patterns suggest that C_4 evolution is more likely from C_3-C_4 lineages with C_4-like anatomy rather than gene expression. It might be that C_4-like gene expression is easier to evolve so that the development of C_4-like anatomy represents a critical step for C_4 evolution, as previously suggested (Edwards, 2019). Alternatively, the selective pressure to evolve C_4 photosynthesis might be higher once a C_4-like anatomy is in place. Further studies are needed to test the hypotheses. For example, mutants of C_4 species lacking either anatomical or biochemical components could be subjected to long term experimental evolution. In the meantime, my results already show that the anatomical and biochemical components of C_4 photosynthesis do not always come together in C_3-C_4 species.

5.1.3 Increased rates of chloroplast evolution might have facilitated the diversification of the photosynthetic apparatus.

Complete genome sequencing to assemble chloroplast genomes was undertaken at the beginning of my PhD, before plastome sequences of diverse Caryophyllales (including some Molluginaceae) were published (Yao et al., 2019). I initially undertook *de novo* assemblies, using tools previously by our group to assemble chloroplast genomes of grasses (Lundgren et al., 2015; Olofsson et al., 2016). These efforts were rapidly hampered by the realization that the chloroplast genomes of some Molluginaceae were highly rearranged, as even short contigs showed distinct gene orders among closely related species. This conclusion was later confirmed by the publication of the results from Yao et al. (2019), who similarly failed to assemble complete chloroplast genomes for some Molluginaceae. I consequently settled on a mapping approach, using one chloroplast genome as the reference (Chapter 3). This approach was perfectly appropriate to estimate rates of nucleotide substitutions, and discontinuities in the mapping of reads as well as segments with very high coverage pointed to numerous rearrangements and a number of repeats (Chapter 3), confirming my observations during early attempts at *de novo* assembly. The lack of completely assembled chloroplast genomes for all analysed species however prevented me from precisely identifying the repeats and rearrangements.

All Molluginaceae samples were sequenced with Illumina short reads, an approach widely used to sequence and assemble organellar genomes. Our sequencing actually used reads that are longer than the average in the field (250 bp paired end reads). However, my investigations showed that the repeats found in the chloroplast genomes of some

Molluginaceae are longer than the insert sizes of Illumina reads (some repeats are at least 2 kb). This directly means that a complete assembly of the chloroplast genomes is impossible with the data at hand. The problem could be solve by the generation of mate pairs, with longer insert sizes or by the production of longer sequencing reads (e.g. PacBio DNA sequencing or Oxford Nanopore long reads) for the same samples. We did attempt extracting high-quality DNA for one accession of *H. cerviana* to later perform PacBio sequencing. However, the DNA extracted from large amounts of plant tissue (necessary to obtain sufficient high-quality DNA) contained an unknown compound that made it semi-solid and prevented subsequent analyses. We did try different ways without more success, and further attempts were not possible in the time of this PhD. In the future, it will be primordial to manage to obtain good quality DNA enabling long reads if we are to resolve the history of modifications of the chloroplast genomes of Molluginaceae. It is possible that other protocols that we have not tested yet would solve this issue. Alternatively, different individuals or species might not present the same difficulties. Besides enabling complete chloroplast genome assemblies, some sequencing might be able to produce complete nuclear genomes. Indeed, we estimated by flow cytometry that the nuclear genome of *H. cerviana* from Arizona has a total size of ~250 Mb. Being just twice the size of *Arabidopsis thaliana*, this genome would be an excellent resources for subsequent analyses, potentially making Molluginaceae a model system for C_4 evolution.

Despite our inability to assemble complete chloroplast genomes, my analyses have unambiguously shown that a large clade of Molluginaceae underwent increased rate of chloroplast evolution (Chapter 3). The chloroplast genome includes parts of the photosynthetic apparatus, which has in some cases been adapted for the C_4 physiology after it

evolved. As a consequence, many genes of the chloroplast genomes have been shown in other groups to be under positive selection specifically in C_4 lineages (Christin et al., 2008; Piot et al., 2018). The acceleration of chloroplast evolution characterized the whole *Adenogramma-Hypertelis* (Chapter 3), which contains C_3 species in addition to C_3-C_4 and C_4 *Hypertelis* (Chapter 1). It is consequently not directly linked to C_4 evolution. However, it is plausible that the increased mutation rate that already existed in the group that evolved C_4 photosynthesis facilitated the diversification of the chloroplast-encoded apparatus and its adaptation for the C_4 context. This hypothesis should be tested by dedicated comparative analyses of the fate of photosynthetic genes in the *Hypertelis* clade (which contains C_3-C_4 and C_4 species) versus the *Adenogramma* clade (which is completely C_3) and contrasting the results to other groups with C_3 and C_4 species that have standard rates of chloroplast evolution.

5.1.4 An effect of hybridization on C_4 evolution?

The two main C_4 species of Molluginaceae (*H. cerviana* and *H. umbellata*) were originally considered as a single species (i.e. *Mollugo cerviana*). They were distinguished based on DNA analyses (Christin et al., 2011), with morphological differences identified subsequently (Thulin et al., 2016). In addition, the closely related C_3-C_4 *H. spergulacea* was originally part of a genus that contained other species, which have been reclassified to a different family (Kewaceae) following the same DNA analyses (Christin et al., 2011; Thulin et al., 2016). This classification history highlights the morphological similarity between the two C_4 species and the paucity of morphological characters grouping them with *H. spergulacea*, although the relationships are unambiguous in the light of both chloroplast and nuclear analyses (Chapters

1, 2 and 3). My phylogenetic analyses of genes encoding C_4 enzymes has similarity shown that in some cases the two C_4 species group together (Chapter 2), questioning the history of the genus and its effect on C_4 evolution.

Incongruence among gene trees might come from different sources. A level of discordance is expected under incomplete lineage sorting, especially if successive speciation events follow each other rapidly. Molecular dating indicates that the speciation events within *Hypertelis* are relatively spread (Chapter 1), but exceptionally variable evolutionary rates (Chapter 3) could bias the results. Alternatively, the grouping of the two C_4 species in some gene trees (Chapter 2) might result from hybridization events following the evolution of C_4 photosynthesis. In this scenario, one of the lineages might have evolved the C_4 trait and then passed it to the other group. This would add to other examples of transfer of C_4 photosynthesis genes among species (e.g. Christin et al., 2012; Dunning et al., 2017). Genes responsible for morphological characters might have followed the same route, explaining earlier taxonomic treatments. Differentiating incomplete lineage sorting from hybridization can be performed via genome-wide analyses, for example using the ABBA-BABA test. I suggest that future work should focus on this, and the transcriptome (Chapter 2) and genome-wide (Chapter 3) data produced in this thesis can be used for this purpose.

5.2 Conclusions

C_4 photosynthesis provides an excellent example of complex traits that can evolve independently from a wide range of ancestral phenotypes. In this thesis, I adopted as a study system the Molluginaceae family with its different photosynthetic types to understand the enablers of complex trait evolution. Using a phylogenetic framework to reconstruct the history of modifications leading to C_4 leaf anatomy (Chapter 1) and changes in gene expression and coding sequences (Chapter 2), I showed that C_4-like anatomical characters predated the C_4 physiology, while C_4 gene upregulation was restricted to fully C_4 lineages. These two chapters together also provide new insights into the significance of so-called C_3-C_4 intermediates during C_4 evolution. Indeed, a role of C_3-C_4 states as evolutionary facilitators that bridge the anatomical gap between C_3 and C_4 leaf anatomies was seen only in one of three lineages (Chapter 1), while no C_4-like gene expression was observed in the C_3-C_4 species (Chapter 2). These results indicate that the evolutionary impacts of C_3-C_4 intermediates depends on how their physiology is realized, with only some C_3-C_4 lineages likely to give birth to C_4 descendants.

Besides their direct impacts on our understanding of C_4 evolution, my results indicate that Molluginaceae are an exciting system for detailed studies of the dynamics allowing photosynthetic adaptation. Indeed, the small size and short generation time of many Molluginaceae would make them amenable to experimental evolution. Such efforts, conducted on the long term, might be used to test the hypothesis that the C_3-C_4 species lacking C_4-like leaf traits and not directly related to any C_4 lineages are not able to produce C_4 descendants. Obviously, an experimental setting might not offer sufficient time for any C_4-like physiology to emerge in the absence of necessary mutations. However, genome editing might

be used to introduce genes for key C_4 enzymes in these C_3-C_4 species. One simple prediction would be that overexpression of PEPC in the C_3-C_4 *Hypertelis spergulacea*, which possesses C_4-like leaf characters (Chapter 1), would produce a weak C_4 cycle and a gain of fitness. Conversely, I would hypothesize that the same genes would not produce any important C_4 cycle in the C_3-C_4 *Paramollugo nudicaulis*, which lacks such C_4-like leaf traits, and only in some populations of *Mollugo verticillata* with C_4-like anatomical characters (Chapter 1). In addition, my studies have revealed important variation of the C_4 leaf anatomy and transcriptomes among populations within each of the C_4 species (Chapters 1 and 2). Coupled with their short generation times, these species have a small nuclear genome (~250 Mb for the C_4 *H. umbellata*), and they would thus constitute great systems to conduct experimental and descriptive (e.g. population genomics) studies of the impact of specific C_4 components on the physiology and ultimately fitness of the plants. Finally, I have clearly shown that the chloroplast genomes of some Molluginaceae underwent increased amounts of substitutions and rearrangements (Chapter 3).

The small family would therefore constitute an outstanding system to study the evolution of chloroplast genomes in addition to photosynthetic types. The diversity of substitution rates in some lineages may provide an opportunity to understand the relative roles of natural selection and neutral processes in shaping genome organization. Additionally, the future incorporation of nuclear genome information would provide a great system to investigate how the accelerated rates in plastids affect cytoplasmic-nuclear co-evolution. I therefore predict that my work will constitute the foundation of future work using the Molluginaceae system to fully understand the intricate interactions between anatomical and

biochemical components that dictate the evolvability of C_4 photosynthesis in different lineages and the factors that drive genome evolution.

5.3 References

Bianconi, M. E., Hackel, J., Vorontsova, M. S., Alberti, A., Arthan, W., Burke, S. V., Duvall, M. R., Kellogg, E. A., Lavergne, S., McKain, M. R., Meunier, A., Osborne, C. P., Traiperm, P., Christin, P. A. and Besnard, G. 2020. Continued Adaptation of C_4 photosynthesis after an initial burst of changes in the Andropogoneae Grasses. *Systematic Biology.* **69**(3), pp.445–461.

Brown, R.H. and Hattersley, P.W. 1989. Leaf anatomy of C_3-C_4 Species as related to evolution of C_4 photosynthesis. *Plant Physiology*. **91**(4), pp.1543–1550.

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 C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology.* **22**, pp. 445–449.

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 C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences USA*. **110**(4), pp.1381–1386

Christin, P. A., Sage, T. L., Edwards, E. J., Ogburn, R. M., Khoshravesh, R. and Sage, R. F. 2011. Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. *Evolution*. **65**(3), pp.643–660.

Christin, P. A., Salamin, N., Muasya, A. M., Roalson, E. H., Russier, F., and Besnard, G. 2008. Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. *Molecular Biology and Evolution.* **25**,pp. 2361–2368.

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 C_4 photosynthesis among close relatives. *Evolution.* **71**(6), pp.1541–1555.

Edwards, E. J. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and CAM photosynthesis. *New Phytologist*. **233**(4), pp.1742-1755.

Frohlich, M. W. 1978. Systematics of *Heliotropium* section *Orthostachys* in Mexico. PhD thesis, Harvard University, Cambridge, MA.

Griffiths, H., Weller, G., Toy, L. F. M., and Dennis, R. J. 2013. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell and Environment.* **36**(2), pp.249–261.

Ermakova, M., Danila, F. R., Furbank, R. T. and von Caemmerer, S. 2020. On the road to C_4 Rice: Advances and Perspectives. *Plant Journal.* **101**(4), pp.940–950.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., and Lercher, M. J. 2013. Predicting C_4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell.* **153**(7), pp.1579–1588.

Hilger, H.H. and Diane, N. 2003. A systematic analysis of Heliotropiaceae (Boraginales) based on *trnL* and ITS1 sequence data. *Botanische Jahrbucher.* **125**, pp.19-51.

Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A. and Woolhouse, H.W. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C_3-C_4 intermediate species. *Planta*. **175**(4), pp.452–459.

Kadereit, G., Bohley, K., Lauterbach, M., Tefarikis, D.T. and Kadereit, J.W. 2017. C_3–C_4 intermediates may be of hybrid origin a reminder. *New Phytologist.* **215**, pp.70–76.

Karki, S., Lin, H., Danila, F. R., Abu-jamous, B., Giuliani, R. and David, M. Coe1, R. A., Covshoff, S., Woodfield, H., Bagunu1, E., Thakur, V., Wanchana1, S., Slamet-Loedin, I., Cousins, A. B., Hibberd, J. M., Steven, K. and Quick, W. P. 2020. A role for neutral variation in the evolution of C_4 photosynthesis. *BioRxiv* doi:10.1101/2020.05.19.104299

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. and Christin, P. A. 2015. Photosynthetic innovation broadens the niche within a single species. *Ecology Letters.* **18**(10), pp.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 C_3-C_4 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell and Environment.* **39**(9), pp.1874–1885.

Lundgren, M.R., Osborne, C.P. and Christin, P.A. 2014. Deconstructing Kranz anatomy to understand C_4 evolution. *Journal of Experimental Botany.* **65**, pp.3357–3369.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P. and Gowik, U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife*. **2014**(3), pp.2–5.

Martins, S. and Scatena, V.L. 2011. Bundle sheath ontogeny in Kranz and non-Kranz species of Cyperaceae (Poales). *Australian Journal of Botany*. **59**, pp.554-562.

McKown, A.D., Moncalvo, J.M. and Dengler, N.G. 2005. Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *Am. J. Bot.* **92,** 1911–1928.

Monson, R.K., Moore, B.D. 1989. On the significance of C_3-C_4 intermediate photosynthesis to the evolution of C_4 photosynthesis. *Plant, Cell and Environment.* **12**, pp.689–699.

Monson, R.K. and Rawsthorne, S. 2000. CO_2 assimilation in C_3-C_4 intermediate plants. In: Leegood, R.C., Sharkey, T.D. and von Caemmerer, S. (eds). Photosynthesis: Physiology and Metabolism. Kluwer Academic Publishers. pp.533–550.

Ocampo, G., Koteyeva, N. K., Voznesenskaya, E. V., Edwards, G. E., Sage, T. L., Sage, R. F. and Travis Columbus, J. 2013. Evolution of leaf anatomy and photosynthetic pathways in Portulacaceae. *American Journal of Botany.* **100**(12), pp.2388–2402.

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**, pp.6107–6123.

Piot, A., Hackel, J., Christin, P.A., Besnard, G. 2018. One third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta*. **247**, pp.255-266.

Sage, R. F. 2004. The evolution of C_4 photosynthesis. *New Phytologist.* **161**, pp.341–370.

Sage, R. F., Sage, T. L., and Kocacinar, F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology.* **63**, pp.19–47.

Sayre, R.T. and Kennedy, R. A. 1979. Photosynthetic enzyme activities and localization in *Mollugo verticillata* populations differing in the levels of C_3 and C_4 cycle operation. *Plant Physiology*. **64**(2), pp.293–299.

Sayre, R.T. and Kennedy, R.A. 1977. Ecotypic differences in the C_3 and C_4 photosynthetic activity in *Mollugo verticillata*, a C_3-C_4 intermediate. *Planta*. **262**, pp.256–262.

Williams, B. P., Johnston, I. G., Covshoff, S. and Hibberd, J. M. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife,2*, e00961.

Yao, G., Jin, J.J., Li, H.T., Yang, J.B., Mandala, V.S., Croley, M., Mostow, R., Douglas, N.A., Chase, M.W., Christenhusz, M.J.M., Soltis, D.E., Soltis, P.S., Smith, S.A., Brockington, S.F., Moore, M.J., Yi, T.S. and Li, D.Z. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. Molecular Phylogenetics and Evolution. 134(February), pp.74–86.