White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/204166/

Version: Accepted Version

Article:
Pereira, L., Bianconi, M.E., Osborne, C.P. et al. (2023) Alloteropsis semialata as a study system for C4 evolution in grasses. Annals of Botany, 132 (3). pp. 365-382. ISSN: 0305-7364

https://doi.org/10.1093/aob/mcad078

This is a pre-copyedited, author-produced version of an article accepted for publication in Annals of Botany following peer review. The version of record Pereira, L., Bianconi, M.E., Osborne, C.P., Christin, P-A., Dunning, L.T., (2023) Alloteropsis semialata as a study system for C4 evolution in grasses, Annals of Botany, is available online at: https://doi.org/10.1093/aob/mcad078.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Alloteropsis semialata as a study system for C_4 evolution in grasses

Lara Pereira[1], Matheus E. Bianconi[1], Colin P. Osborne[1], Pascal-Antoine Christin[1], Luke T. Dunning[1*]

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

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

Abstract

Background

Numerous groups of plants have adapted to CO_2 limitations by independently evolving C_4 photosynthesis. This trait relies on concerted changes in anatomy and biochemistry to concentrate CO_2 within the leaf and thereby boost productivity in tropical conditions. The ecological and economical importance of C_4 photosynthesis has motivated intense research, often relying on comparisons between distantly related C_4 and non-C_4 plants. The photosynthetic type is fixed in most species, with the notable exception of the grass *Alloteropsis semialata*. This species includes populations exhibiting the ancestral C_3 state in southern Africa, intermediate populations in the Zambezian region and C_4 populations spread around the paleotropics.

Scope

We compile here the knowledge on the distribution and evolutionary history of the *Alloteropsis* genus as a whole and discuss how this has furthered our understanding of C_4 evolution. We then present a chromosome-level reference genome for a C_3 individual and compare the genomic architecture to that of a C_4 accession of *A. semialata*.

Conclusions

Alloteropsis semialata represents one of the best systems to investigate the evolution of C_4 photosynthesis as the genetic and phenotypic variation provides a fertile ground for comparative and population-level studies. Preliminary comparative genomic investigations show the C_3 and C_4 genomes are highly syntenic, and have undergone a modest amount of gene duplication and translocation since the different photosynthetic groups diverged. The background knowledge and publicly available genomic resources make *Alloteropsis semialata* a great model for further comparative analyses of photosynthetic diversification.

Introduction

The origin of photosynthesis more than two billion years ago was a major event in Earth history, and since then the photosynthetic apparatus has continuously been adapted in response to varying environmental conditions (Raven et al. 2003, 2008; Tcherkez et al. 2006; Cardol et al. 2008; Young et al. 2012; Dusenge et al. 2019; Kumarathunge et al. 2019). In particular, atmospheric CO_2 levels have repeatedly decreased, reaching very low levels about 30 million years ago (Pagani et al. 2005). As CO_2 represents the main substrates of photosynthesis, its depletion decreases the efficiency of carbon fixation (Ehleringer and Bjorkman 1977; Skillman 2008). Indeed, when CO_2 is limited, Rubisco, the enzyme that fixes CO_2 during photosynthesis, will fix O_2 instead, exacerbating the costly process of photorespiration (Peterhansel and Maurino 2010; Busch 2020). Plants have developed different strategies to limit photorespiration, which becomes especially important in warm and dry environments of the low-CO_2 world that prevailed since the Oligocene (Ehleringer and Monson, 1993; Ehleringer et al. 1997; Tcherkez et al. 2006). Over the last 32 million years (Christin et al. 2011), at least 60 lineages of plants have evolved the C_4 photosynthetic pathway (Sage et al., 2011), which concentrates CO_2 within the leaf before its use by Rubisco (Hatch 1987; Sage, 2004; Sage et al. 2012). C_4 plants outperform those that still use the ancestral C_3 pathway in open biomes of tropical regions, and they consequently cover vast areas of the world and account for one quarter of terrestrial primary production (Still et al. 2003; Edwards et al. 2010; Lehmann et al., 2019). C_4 plants also include major crops, such as maize, sorghum, and sugarcane, and have consequently been the focus of numerous research projects (e.g. Matsuoka et al. 2001; Langdale et al. 2011; von Caemmerer et al. 2012; Sales et al. 2021; Zhao et al. 2022). Given the complexity of the C_4 trait, its recurrent evolutionary origins are especially puzzling.

The C_4 pathway requires a specialized leaf anatomy as well as the upregulation and synchronization of numerous enzymes (Hatch 1987; Edwards et al. 2001; Sage et al. 2012). In the C_4 pathway, the initial fixation of atmospheric CO_2 is performed by the enzyme phosphoenolpyruvate carboxylase (PEPC), which unlike Rubisco has no affinity for O_2. This reaction usually happens in mesophyll cells, which are in direct contact with the atmosphere. The resulting four-carbon acids are then transformed and transported to a different cell isolated from the atmosphere, typically the bundle sheaths that surround the veins, where Rubisco is localized in C_4 plants. The largest lineages of C_4 plants (e.g. Chloridoideae grasses) evolved this trait over 30 million years ago (Christin et al. 2008), and they consequently differ broadly from C_3 plants, in many aspects besides those linked to the photosynthetic pathway (Heyduk et al. 2019). Most C_4 research has consequently utilized closely related plants with contrasted photosynthetic types, which exist in various lineages of angiosperms,

with examples from: Amaranthaceae and Chenopodiaceae (*Atriplex*) (Ehleringer and Bjorkman 1977; Kadereit et al. 2003), Asteraceae (*Flaveria*) (Vogan and Sage 2011), Cleomaceae (*Cleome*) (Koteyeva et al. 2011), Cyperaceae (*Eleocharis*) (Bruhl and Perry, 1995), Molluginaceae (*Mollugo*) (Christin et al., 2011), Poaceae (*Neurachne*) (Khoshravesh et al. 2020), and more broadly across eudicots (Muhaidat et al. 2007; Muhaidat and McKown 2013). In most cases, the photosynthetic types are however distributed among distinct species. The only known exception is the grass *Alloteropsis semialata*, which includes both C_4 and non-C_4 individuals, providing an exciting system to study the drivers of C_4 evolution.

The photosynthetic variation existing in *A. semialata* was discovered based on leaf anatomical surveys and measurements of carbon isotopes independently by Ellis (1974) and Brown (1975). The differences between photosynthetic types within this species have since then been repeatedly studied, focusing on the ecological (Ripley et al. 2007; Ibrahim et al. 2008; Osborne et al. 2008; Ripley et al. 2010b; Bateman and Johnson, 2011; Lundgren et al. 2015), cytogenetic (Frean and marks, 1988; Liebenberg et al. 2001; Lundgren et al. 2015; Bianconi et al. 2020; Olofsson et al., 2021), physiological (Frean et al. 1980, 1983a; Lundgren et al. 2016) and biochemical variation (Ueno et al. 2006; Phansopa et al. 2020), and more recently evolutionary and genomic aspects of C_4 photosynthesis (Ibrahim et al. 2009; Christin et al. 2012, 2015; Olofsson et al., 2016, 2021; Dunning et al. 2017;, 2019a; Bianconi et al. 2018, 2020; Lundgren et al. 2015, 2019; Curran et al. 2022). In this review, we compile the knowledge accumulated on this study system. First, we compile and review the distribution and evolutionary history of the *Alloteropsis* genus and its photosynthetic types. Second, we review the photosynthetic diversity discovered within the genus and discuss its importance for our understanding of C_4 evolution. Third, we present a new chromosome-level reference genome for a C_3 individual of *A. semialata* and compare it to an existing genome for a C_4 conspecific, with a special focus on synteny and gene orthology. Finally, we discuss future research directions that can build on existing knowledge and resources accumulated by different researchers over almost 50 years. We hope that this information can help motivate future research into the photosynthetic diversity of *Alloteropsis*, a unique system for studying photosynthetic diversity.

Evolutionary history of *Alloteropsis*

Evolutionary diversification of the Alloteropsis *genus*

In its latest treatment, the genus *Alloteropsis* included five recognized species (Clayton and Renvoize 1982). The species *A. cimicina*, *A. paniculata* and *A. papillosa* are all C_4 and form a monophyletic group based on chloroplast markers (Ibrahim et al. 2009). The former is an annual weed (Figure 1), native across Africa, Asia and Oceania, which is invasive in the Americas (Zuloaga and Morrone 2003; Rocha and Miranda 2012). Few samples of this group have been analyzed with molecular data, and at present it is not clear whether *A. paniculata* and *A. papillosa* represent truly distinct species from *A. cimcina*, or whether they are merely morphological variants. It has been suggested that *A. papillosa* is a possible hybrid as the linear leaf-blades resemble *A. semialata* while the inflorescence and spikelet characters are reminiscent of *A. cimicina* (Clayton and Renvoize 1982). The divergence of samples assigned to *A. cimicina* and *A. paniculata* was estimated around 2.5 Ma (Olofsson et al. 2016; Bianconi et al. 2020), and these two species will be discussed jointly as '*A. cimicina*' in this review. The *A. cimicina* group diverged from the two other species in the genus approximately 11 Ma (Figure 2; Lundgren et al. 2015; Dunning et al. 2017).

The sister species *A. angusta* and *A. semialata* have been studied in more detail, with intense species sampling increasing our confidence in the delimitation of lineages. *Alloteropsis angusta* was originally described as a small decumbent plant (Clayton and Renvoize 1982) although we discovered it also has an erect morph representing a distinct ecotype (Figure 1; Curran et al. 2022). The decumbent and erect *A. angusta* ecotypes co-occur across Central and East Africa. They are divergent on the nuclear genome, but share chloroplast haplotypes, indicating frequent hybridization (Figure 2; Curran et al. 2022). All 132 samples of *A. angusta* analysed with carbon isotopes were likely C_4 with only one anomalous non-C_4 value that could need to be verified because of limited material availability (Table S1). The karyotype of the four *A. angusta* samples analysed were likely diploid, with a haploid genome size (1C) of ~1 Gb (Curran et al. 2022). *Alloteropsis angusta* diverged from *A. semialata* about 8 Ma (Lundgren et al. 2015; Dunning et al. 2017), although nuclear genome analyses indicate the two species repeatedly hybridize where they come into contact (Figure 2; Curran et al. 2022).

The diversity of *A. semialata* was originally studied in South Africa, where both C_3 and C_4 individuals were first discovered (Ellis 1974). Comparisons of leaf anatomy from herbarium specimens suggested additional photosynthetic variation in *A. semialata* sampled from Tanzania and Zambia (Renvoize 1987), and carbon isotopes intermediate between C_3 and C_4 signatures were also reported for samples from these countries (Ellis 1981; Hattersley and Watson 1992).

Subsequent studies with a dense sampling from Central and East Africa, including live plants grown in greenhouse conditions, permitted physiological and biochemical assays, confirming that a greater photosynthetic diversity exists across Africa, and we describe here the evolutionary lineages of *A. semialata* and their photosynthetic types based on the most recent analyses.

Main lineages of Alloteropsis semialata

Chloroplast markers suggested that the species originated in Central-East Africa, and then acquired important chloroplast diversity as it migrated outside of this centre of origin (Lundgren et al. 2015; Bianconi et al. 2020). Seven main lineages of chloroplast haplotypes correspond to a reduced set of four distinct nuclear lineages, with the photosynthetic types apparently mostly consistent within each of these four groups (Figure 2; Olofsson et al. 2016; Bianconi et al. 2020). Despite evidence of recurrent hybridization among them, the four nuclear lineages are recovered in both population genetics and phylogenetic analyses (Olofsson et al. 2016, 2021; Bianconi et al. 2020), and are therefore useful to summarize the diversification of the species. Their most recent common ancestor is estimated around 2.5 Ma (Bianconi et al. 2020; Raimondeau et al. 2022), and each has since then followed a different evolutionary trajectory.

The nuclear lineage I, which encompasses chloroplast group A, results from a southern migration from its centre of origin, and is now distributed across South Africa, Mozambique, and Zimbabwe (Figure 3; Lundgren et al. 2015; Bianconi et al. 2020). All individuals so far analyzed are diploids (Lundgren et al. 2015; Bianconi et al. 2020). Based on carbon isotopes, none of lineage I individuals use the derived C_4 photosynthetic type (Figure 3; Table S2), and the live plants analysed in greenhouse conditions showed characteristics of C_3 plants, including high CO_2-compensation points, low PEPC activity in the leaves, low leaf transcript abundance of genes encoding C_4 enzymes, and low levels of C_4 enzymes themselves (Figure 4; Frean et al. 1980; Bartlett et al. 1983; Ueno and Sentoku 2006; Ripley et al. 2007; Lundgren et al. 2016, 2019; Dunning et al. 2019a). The leaf anatomy of plants from this group is moreover indicative of a C_3 photosynthetic type (Figure 4; Ellis 1974; Frean et al. 1983a, 1983b; Lundgren et al. 2019). In two individuals from this group, the inner bundle sheath (aka mestome sheath) was shown to contain glycine decarboxylase (Ueno and Sentoku 2006; Lundgren et al. 2019), an enzyme used in the recycling of the products of O_2 fixation by Rubisco. This might suggest that the plant performed a weak photorespiratory pump, as seen in other species described as “type I intermediates” (Edwards & Ku, 1987; Sage et al. 2012).

However, the abundance of glycine decarboxylase in other tissues (Ueno and Sentoku 2006; Lundgren et al. 2019) and the C_3-like physiology of the plants analyzed (Frean et al. 1980; Bartlett et al. 1983; Ripley et al. 2007; Lundgren et al. 2016) suggests such a pump is weak at best. The nuclear lineage I can thus be considered as encompassing mostly, if not only, C_3 individuals.

The nuclear lineage II is mostly associated with chloroplast groups B and C, and migrated slowly around the centre of origin of the species (Lundgren et al. 2015; Bianconi et al. 2020). It is tightly associated with the Central Zambezian miombo woodlands, where it occurs in both open and wooded grasslands (Olofsson et al. 2021), and has been identified in Burundi, Democratic Republic of Congo, Malawi, Tanzania, and Zambia (Figure 3). All individuals analyzed so far are diploid (Bianconi et al. 2020; Olofsson et al. 2021). None of the individuals from this lineage is C_4 based on carbon isotopes (Lundgren et al. 2015; Olofsson et al. 2021), although it contains isotopically intermediates that likely grew using both C_3 and C_4 pathways (Figure 3; Lundgren et al. 2016; Olofsson et al. 2021). Its individuals moreover have CO_2 compensation points, leaf anatomy, PEPC abundance, and C_4 gene transcript abundance in the leaves that are intermediate between C_3 and C_4 types (Figure 4; Lundgren et al. 2016, 2019; Dunning et al. 2019a). All the evidence suggests that individuals from lineage II assimilate some of their carbon via the C_3 cycle and some via the C_4 cycle, and they were consequently termed 'C_3+C_4' (Dunning et al. 2017). This type corresponds to 'type II intermediates' reported in a number of other lineages of plants (Monson et al. 1986; Schlüter and Weber 2016; Sage et al. 2012, 2018; Amy Lyu et al. 2022). The variation in carbon isotopes, CO_2 compensation points, and leaf anatomy within lineage II (Figure 3; Lundgren et al. 2016, 2019; Olofsson et al. 2021) moreover suggest that the strength of the C_4 cycle varies among populations and could potentially even include individuals without any C_4 cycle.

The nuclear lineage III is mostly associated with the chloroplast lineages F and G, which are sister to all other chloroplast lineages of *A. semialata* (Lundgren et al. 2015; Olofsson et al. 2016). However, its nuclear genome is consistently sister to that of lineage IV (Olofsson et al. 2016; Bianconi et al. 2020). It is hypothesised that this mismatch is a result of genomic swamping, with unidirectional pollen flow from lineage III into a distantly related maternal lineage resulting in the apparent substitution of the nuclear genome in a pattern that mirrors chloroplast capture (Bianconi et al. 2020). Lineage III is mostly restricted to the Central Zambezian miombo woodlands, where it overlaps with lineage II (Figure 3), and the two can be found in mixed populations (Olofsson et al. 2021). Its individuals include both diploids and hexaploids (Olofsson et al. 2021). All individuals

from lineage III analysed so far have carbon isotopes typical of C_4 plants (Figure 3; Lundgren et al. 2015; Olofsson et al. 2021). The populations analyzed showed high leaf transcript abundance of genes encoding C_4 enzymes (Dunning et al. 2019a). The assessed populations from Zambia and Tanzania had very low CO_2 compensation points and leaves where chloroplasts are concentrated in the inner bundle sheath, confirming they are C_4 (Lundgren et al. 2019).

The nuclear lineage IV is mostly associated with chloroplasts groups D and E (Lundgren et al. 2015). It rapidly migrated away from the centre of origin of the species to Southern Africa, Western Africa, Madagascar, Asia and Australia (Figure 3; Lundgren et al. 2015; Bianconi et al. 2020), where it is now considered a keystone species (Bateman and Johnson 2011). In South Africa, it overlaps with the C_3 lineage I, and the two form occasionally mixed populations (Frean et al. 1980). It is present in a variety of biomes, and its broader niche is likely associated with its rapid dispersal (Lundgren et al. 2015), although whether each individual possesses a broad niche or whether the group is able to rapidly adapt locally is unknown. This lineage IV contains mostly diploids in Asia, Australia and large parts of Africa, but also hexaploids in Southern Africa (Lundgren et al. 2015; Bianconi et al. 2020), and a single tetraploid reported from Australia (Olofsson et al. 2021). All individuals from this lineage have carbon isotopes indicative of C_4 photosynthesis (Figure 3; Lundgren et al. 2015; Bianconi et al. 2020). They show high PEPC abundance in the leaves, high transcript abundance of genes for C_4 photosynthesis, an abundance of inner bundle sheath cells filled with chloroplasts, and high leaf abundance of C_4 enzymes that follow the expected cellular compartmentalization (Figure 4; Frean et al. 1983a, 1983b; Ueno and Sentoku 2006; Lundgren et al. 2016, 2019; Dunning et al. 2019a). These plants however have Rubisco in the mesophyll cells, which is unexpected for C_4 plants and led to them being referred to as 'C_4-like' (Ueno and Sentoku 2006). Because the physiology and carbon isotopes are typical of C_4 plants (Figure 3), and it is not known whether their mesophyll Rubisco is active, they must be considered as performing C_4 photosynthesis.

Species status of A. semialata

Differences in photosynthetic types and ploidy levels observed in South Africa led to a suspicion that C_3 and C_4 *A. semialata* populations were different species (Frean and Marks 1988; Liebenberg and Fossey 2001), and two subspecies were defined in South Africa (Gibbs Russell 1983). The morphological characters used to recognize these two subspecies are not easily applied in other

regions of Africa (Gibbs Russell 1983), and the variation in photosynthetic type and other anatomical traits goes beyond that described in South Africa, so that many individuals would fall outside of these existing subspecies (Curran et al. 2022). More generally, the four nuclear lineages of *A. semialata* described here naturally hybridise in the wild (Figure 2; Olofsson et al. 2016, 2021; Bianconi et al. 2020) and produce healthy F1 hybrids in the greenhouse (Bianconi et al. 2020, 2022). To date, we have not been able to obtain seeds for F2 individuals despite F1 hybrids producing flowers. Some hybrid incompatibility might exist, although experimental tests are missing. Given the morphological variation existing within each lineage and the recurrence of genetic exchanges among them, they should all be considered as part of the same species complex. In the absence of morphological characterization of a large number of individuals of known genetic ancestry, we advocate for the identification of nuclear lineages based on genetic data. This is especially important as some parts of the species ranges that present genetically divergent lineages remain poorly sampled, especially in Angola and the Democratic Republic of Congo.

Recurrent polyploidization in A. semialata

While the four nuclear lineages are clearly distinct in diploids, some polyploids are phylogenetically placed outside of these lineages, with discrepancies among nuclear markers, incongruence between nuclear and organelle genomes, and even disagreement among organelles (Olofsson et al. 2016, 2019, 2021; Bianconi et al. 2020). This includes dodecaploids (12x) from Cameroon (Bianconi et al. 2020), hexaploids (6x) / octoploids (8x) from Zambia (Olofsson et al. 2021), and plants of unknown ploidy in Tanzania and the Democratic Republic of Congo (Figure 3; Olofsson et al. 2016; Bianconi et al. 2020). Among these individuals, the known polyploids present genomic contributions from multiple nuclear lineages; II+III in Zambia (Figure 2; Olofsson et al. 2021) and II+III+IV in Cameroon (Figure 2; Bianconi et al. 2020), and the hexaploids from lineages III and IV each present contributions from the other lineage (Figure 2; Olofsson et al. 2016; Bianconi et al. 2020). The affinities of other polyploids reported in the past, including tetra-, octo- and dodecaploids in addition to hexaploids (Ellis 1981; Frean and Marks 1988; Liebenberg and Fossey 2001), remains unknown in the absence of genetic analyses. The available evidence already indicates that polyploidy allows the mixing of distinct lineages, in some cases corresponding to different photosynthetic types. Based on their position in the nuclear and organelle phylogenies, polyploids emerged recurrently within *A. semialata* (Olofsson et al. 2019; Bianconi et al. 2020). Polyploidy seems to have a huge effect in some cases, with the polyploids being much more abundant than diploids in regions of Africa (Figure 3). The exact consequences of the recurrent

polyploidization on the niche and the phenotype however remain to be investigated in detail. At present the extent of gene flow among polyploids and between diploids and polyploids is not fully understood, although polyploidy has been proposed as a mechanism preventing reproductive interference among photosynthetic types in mixed populations (Olofsson et al. 2021).

Reproductive and life-history characteristics of A. semialata

Alloteropsis semialata is a perennial species, and while its lifespan has never been evaluated, we have kept some plants for ten years in climatically controlled greenhouse conditions (12 hr day, 25/20 ºC day/night temperature) without any signs of senescence. The species forms rosettes, from which long stems emerge carrying the inflorescence (Figure 1). The base of the leaf is thickened, forming bulb-like structures, and new bulbs and tillers continuously emerge from rhizomes. In some accessions, the bulbs are compressed together and form a single compact structure that cannot be separated without disrupting the bulb tissue. In others, the stolons place the bulbs further apart, and these then form separate individuals that can be repotted to produce clones. The extent of such vegetative growth has not been quantified in the wild, but we have identified wild individuals that have formed tens of bulbs, each bearing a stem and inflorescence. This suggests that clonal propagation is frequent in the wild, and it is certainly constant in the greenhouse. By comparison, sexual reproduction is uneven in the greenhouse. Some accessions seem to flower more frequently than others, and the seed set is also quite variable. Based on our own experience, polyploids are able to self-fertilize, as can some of the C_3+C_4 (lineage II), which even occasionally develop cleistogamous inflorescences among the rosette leaves. By contrast, all seedlings obtained from diploid individuals from lineages I, III and IV that were genotyped resulted from crosses between distinct accessions, suggesting a predominantly outcrossing system (Bianconi et al. 2022). The different lineages of *A. semialata* can freely interbreed to produce F1 hybrids (Bianconi et al. 2022). Inflorescences with pollen were observed in most F1 hybrids, but to date we have not been able to obtain seeds for F2 offspring. Evidence of introgression suggests that such crosses do episodically occur in the wild (Olofsson et al. 2016, 2019; Bianconi et al. 2020).

Insights into C_4 evolution

Three transitions to a full C_4 *physiology*

The *Alloteropsis* genus belongs to the Paniceae tribe of grasses, which contains multiple independent origins of C_4 photosynthesis (GPWG 2012). Its closest relatives identified so far are C_3 (*Entolasia* sp., *Amphicarpum* sp., and *Panicum pygmaeum*) (GPWG 2012), and the genus likely emerged from a C_3 lineage within the Boivinellinae (Ibrahim et al. 2009). All *Alloteropsis* accessions are C_4 except the non-C_4 *A. semialata* which are nested within the genus (Figure 2). It was therefore hypothesized that the non-C_4 *A. semialata* might represent a reversal from a C_4 ancestral state (Ibrahim et al. 2009). Further investigation of the history of individual anatomical and biochemical C_4 components, however, strongly argue against this scenario (Dunning et al. 2017). *Alloteropsis cimicina* and the pair *A. semialata*/*A. angusta* co-opted different tissues for the C_4 pathway. The former uses the outer bundle sheath to segregate Rubisco and the Calvin-Benson-Bassham cycle, with a large proportion of bundle sheath tissue achieved via a dramatic expansion of these cells (Christin et al. 2013). In stark contrast, both *A. semialata* and *A. angusta* use the inner bundle sheath to segregate Rubisco, and achieved a large proportion of bundle sheath tissue via the addition of minor veins, without drastic enlargement of bundle sheath cells (Figure 4; Ellis 1974; Brown 1977; Frean et al. 1983; Renvoize et al. 1987; Christin et al. 2013; Lundgren et al. 2019). The details of the C_4 pathway in *A. cimicina* has not been established with biochemical assays, but transcriptomes analyses indicate a high transcript abundance of genes encoding the NADP-malic enzyme (NADP-ME) and other enzymes associated to the NADP-ME type (Dunning et al. 2017, 2019a). Early analyses described some South African *A. semialata* as using the NADP-ME type, with potential changes depending on the temperature (Frean et al. 1983), although this has not been subsequently confirmed. All other samples were characterized as being of the phosphoenolpyruvate carboxykinase decarboxylase (PCK) type (Prendergast et al. 1987; Ueno and Sentoku 2006). Subsequent transcriptome analyses suggested that, at 25 ºC, all assessed C_4 accessions of either *A. semialata* or *A. angusta* use mainly the PCK C_4 type, with various amount of NADP-ME activity (Figure 4; Dunning et al. 2017, 2019a). In addition to the differences associated with the C_4 type discussed above, enzymes used for the C_4 pathway in both *A. cimicina* and *A. angusta*/*A. semialata* are not necessarily encoded for by the same gene, with different paralogs encoding the C_4 copy of aspartate aminotransferase in the two groups (Dunning et al. 2017). The numerous anatomical, biochemical, and genetic mismatches unambiguously support at least two independent C_4 origins within the *Alloteropsis* genus.

The history of photosynthetic diversification among the *A. angusta*/*A. semialata* group is more difficult to decipher as the two species use homologous anatomical and biochemical C_4

components. However, genes for the C_4 enzymes show evidence of adaptation under positive selection on the branches individually leading to each of the two species, but not in their common ancestor (Dunning et al. 2017). Biochemical adaptation therefore appears to have happened twice in the group, although it cannot be ruled out that the common ancestor of the two species possessed some C_4 components and that only the subsequent gene adaptation occurred after their split. A full C_4 pathway therefore evolved three times in this small genus, and this was helped by exchanges of C_4 genes within the genus (Dunning et al. 2017). *Alloteropsis* is not the only group in which recurrent C_4 origins appeared: C_3, C_3-C_4, C_4-like and C_4 species co-exist within *Flaveria* genus, with two independent origins of C_3-C_4 intermediacy and C_4-like (McKown et al. 2005), in Molluginaceae C_3-C_4 intermediacy evolved at least twice, and two fully developed C_4 originated twice within the same species, *Mollugo cerviana* (Christin et al. 2011), and new photosynthetic pathways have evolved repeatedly within the subtribe Neurachninae (Christin et al. 2012). Whether the most recent common ancestor of *A. semialata* was C_3+C_4 remains an open question. Such a scenario would imply that the C_3 *A. semialata* have lost their previous C_3+C_4 characters as no traces of genes previously used for a weak C_4 cycle have been detected in these populations. The exact history of components sustaining a C_3+C_4 state, such as the proliferation of chloroplasts containing active Rubisco in the bundle sheath and overexpression of some C_4 enzymes, will need to be revisited once the mutations underlying these traits are identified. Indeed, if such components existed in the ancestors of the C_3 *A. semialata*, sequences that differ from the ancestral state would be expected following mutations to revert to a C_3 physiology.

Multiple origins of C_4 *components within* A. semialata

The group consisting of lineages III and IV of *A. semialata* is wholly C_4, and it is likely that its most recent common ancestor was also C_4, leading to the inference of a single origin in this species. However, individual C_4 components unambiguously originated multiple times within the group. The best examples are given by C_4 genes acquired from distantly related species, as these are easy to identify across samples (Christin et al. 2012, Dunning et al. 2017). The main enzyme of the C_4 pathway PEPC is encoded in grasses by seven different gene lineages that emerged through repeated single gene or whole-genome duplication events (Christin et al. 2014). Of these, two are upregulated in the C_3+C_4 and some C_4 individuals of *A. semialata* (Figures 4 and 5; Dunning et al. 2017), and these co-opted genes show evidence of biochemical adaptation for the C_4 context (Phansopa et al. 2020). Other C_4 accessions of *A. semialata* (and *A. angusta*) overexpress only one of these two genes, and it is not always the same copy (Figure 5; Dunning et al. 2017), so that the

364 C_4-specific PEPC from different C_4 accessions effectively have different origins even if they
365 originated in the common ancestor from the group (Figure 5). A completely distinct gene of PEPC
366 is used by the Australian *A. semialata*, and this gene was laterally acquired from the
367 Andropogoneae *Themeda triandra* (Christin et al. 2012; Dunning et al. 2019b), with important
368 biochemical adaptations that predate the transfer (Phansopa et al. 2020). This laterally-acquired
369 copy replaced the ancestral native copies, which became pseudogenes (Figure 5). A similar process
370 happened in Africa with two other PEPC genes laterally-acquired from other groups of grasses
371 (Cenchrinae and Melinidinae; Christin et al. 2012), each present in a subset of individuals, creating
372 a mosaic of PEPC origins within the species (Olofsson et al. 2016). These PEPC examples show
373 that C_4 components can originate independently within a single C_4 lineage. The C_4 *A. semialata* also
374 present important C_4 anatomical variation (Lundgren et al. 2019) and in the expression of C_4-related
375 genes (Dunning et al. 2019a), offering a great system to understand the secondary adaptations that
376 happened among sub-lineages after the initial transition to C_4 photosynthesis.

377

378 *Studies of* A. semialata *suggest an initial C_4 trait can emerge via few changes*

379 The C_4 pathway observed in older C_4 lineages is often complex, relying on numerous anatomical
380 and biochemical specialities (Hatch 1987; Sage 2004; Sage et al. 2012). Anatomically, a higher
381 proportion of vascular bundle sheath tissue, as a result of greater number of veins and larger size of
382 the bundle sheath cells, strongly increases C_4 evolvability in grasses (Christin et al. 2013).
383 Biochemical changes are numerous and very diverse depending on the photosynthetic subtype
384 (Burgess and Hibberd 2015). Such features likely encompass those that were needed to evolve the
385 C_4 trait (primary adaptations), but also those that evolved later (secondary adaptations) and those
386 that evolved for unrelated reasons (Heyduk et al. 2019). In fact, most comparative transcriptomic
387 studies among C_4 and C_3 relatives identify several hundreds of differentially expressed genes, many
388 of which would likely represent secondary adaptations (Bräutigam et al. 2011, 2014; Lyu et al.
389 2021). Because they share a recent common ancestor, comparisons of the different groups of *A.*
390 *semialata* offer an opportunity to distinguish the primary adaptations that were acquired during the
391 initial transition to C_4 photosynthesis from those that evolve later. In such comparative endeavours,
392 it is especially important to sample the diversity within each group to identify those features
393 restricted to a subsample of C_4 plants, which represent secondary adaptations. Leaf anatomy
394 comparisons revealed that the only characters that consistently differ between C_4 and non-C_4 *A.*
395 *semialata* is the proliferation of minor veins (Figure 4), which are therefore sufficient to support the
396 transition between C_3+C_4 and C_4 in this species (Lundgren et al. 2019). As in *A. semialata*, vein

density seems to be an important factor for C_4 evolvability in *Neurachne* (Khoshravesh et al. 2020), in contrast with C_3 and C_4 relatives within the Chenopodiaceae, which present similar vein density (Voznesenskaya et al. 2013; Freitag and Kadereit 2014). Certain *A. semialata* C_4 individuals differ in other leaf anatomy characters (e.g. bundle sheath cell enlargement) from the non-C_4 *A. semialata* (Lundgren et al. 2019), but these likely represent secondary adaptations. In terms of leaf transcriptomes, only three genes encoding known C_4 enzymes (aspartate aminotransferase, phosphoenolpyruvate carboxykinase, and phosphoenolpyruvate carboxylase) were consistently upregulated in the C_3+C_4 compared to the C_4 *A. semialata*, and only pyruvate orthophosphate dikinase was upregulated in the C_4 compared to the C_3+C_4 *A. semialata* (Figure 4; Dunning et al. 2019a). Other C_4-related genes were upregulated solely in a subset of C_4 populations (e.g. malate dehydrogenase), suggesting they represent secondary adaptations that were not involved in the initial transition to C_4 photosynthesis.

Not all aspects of the C_4 trait have been studied in *A. semialata*. For example, the intracellular anatomical variation has been investigated in very few individuals, and the enzyme kinetics and regulation are largely unknown (but see Phansopa et al. 2020 for PEPC). There are, therefore, likely to be more differences between the groups of *A. semialata* than those reported here. Nevertheless, the evidence available to date suggests that the transition to complete reliance on the C_4 pathway involves relatively few primary adaptations and can emerge via the upregulation of only four genes (listed above) accompanied with a single key leaf development alteration involving the proliferation of minor veins. Many other features traditionally associated with C_4 photosynthesis instead represent secondary adaptations, which improved existing C_4 traits and might allow this species to perform extremely well in a diversity of ecological niches. These secondary adaptations can moreover evolve in distinct populations, and be later combined following genetic exchanges among distinct lineages and adaptive introgression (Olofsson et al. 2016). The presence of distinct populations occupying different environments and accumulating novel mutations might thus accelerate the evolutionary diversification of photosynthesis in some groups (Olofsson et al. 2016; Dunning et al. 2017).

Hybridization and the origin of C_3+C_4 A. semialata

C_3+C_4 plants are found in a number of angiosperm lineages (Sage et al. 2012, 2018; Lundgren and Christin 2017), and it has been hypothesized that they might result from natural crosses between C_3

and C_4 plants (Kadereit et al. 2017). For example, within Salsoleae, the "type I intermediate" species *Salsola divaricata* was proposed to be of hybrid origin (Tefarikis et al. 2022). In *Flaveria*, F_1 hybrids obtained by artificially crossing C_3, C_3-C_4 and C_4 species showed a varying degree of inheritance of C_4 phenotypes, suggesting that a hybrid origin of natural C_3-C_4 intermediates was unlikely (Kadereit et al. 2017). Furthermore, coordinated changes in gene expression, protein sequences and physiological and anatomical traits along the C_4 evolutionary pathway defended the hypothesis of C_3-C_4 species being evolutionary intermediate steps (Lyu et al. 2021). In contrast, a recent study using transcriptomes from 17 *Flaveria* species showed recurrent hybridization among them, that might have a certain impact into the development of C_4 photosynthesis in the lineage (Morales-Briones and Kadereit 2022).

In *A. semialata*, hybrids between C_3 and C_4 indeed resemble naturally occurring C_3+C_4 *A. semialata* in their physiology, although their C_3+C_4 state relies on different components (Bianconi et al. 2022). In particular, hybrids between C_3 and C_4 parents in semi-controlled greenhouse conditions possess some of the minor veins that characterize the C_4 *A. semialata* (Bianconi et al. 2022), as do a few naturally occurring C_3+C_4 x C_4 hybrids identified in the wild (Figure S1). Such minor veins are generally absent from the natural C_3+C_4 populations (Lundgren et al. 2019). While these differences indicate that the origin of C_3+C_4 is not likely to simply stem from an ancient hybridization between C_3 and C_4 parents, gene flow between the different *A. semialata* lineages may have played a role in the diversification of the photosynthetic trait. A multigene coalescent species tree analysis highlights that the placement of the C_3+C_4 clade is not fully resolved, with approximately equal proportions of gene trees placing this clade as sister to the C_3 or C_4 lineages (Bianconi et al. 2020). This pattern points to an episode of hybridization, which is likely relatively ancient as the genes encoding the C_4 enzymes in the C_3+C_4 intermediates that are sister to the C_4 clade (e.g. aspartate aminotransferase) generally lack the traces of adaptive amino acid replacement seen in the C_4 populations (Dunning et al. 2017). These gene tree analyses indicate that enzyme adaptation happened after the hybridization event. Gene flow continues to periodically connect the C_4 and C_3+C_4 lineages, as evidenced by the more recent transfer of a gene encoding PCK between lineages (Dunning et al. 2017).

Shared history, photosynthetic diversification and secondary adaptations all influence the ecophysiological behaviour of A. semialata

C_4 photosynthesis brings a suite of benefits for CO_2-fixation under hot, dry and high-light environments, including improved light-, water- and nitrogen-use efficiencies (Pearcy & Ehleringer,

1984; Sage & Pearcy, 1987; Long, 1999). However, it has long been recognised that the photosynthetic pathway is only one of a suite of ecophysiological characters needed to succeed in any particular environment. These may be inherited from non-C_4 ancestors or evolve as environmental adaptations during C_4 diversification (Osmond et al., 1982; Pearcy & Ehleringer, 1984). *Alloteropsis semialata* has therefore been used as a study system to investigate how the photosynthetic pathway interacts with ancestral and novel functional traits to determine ecological behaviour in relation to temperature, water-deficits, fire and atmospheric CO_2 concentration (Table S3). Two populations of *A. semialata*, a C_3 diploid belonging to nuclear lineage I and a C_4 hexaploid from nuclear lineage IV, have been used for detailed ecophysiological studies. These plants grow in mixed populations close to their range limits in South Africa, but their ecological behaviour differs significantly.

In *A. semialata*, the rate of leaf photosynthesis under light-saturated conditions, the quantum efficiency under light-limited conditions, and the rate of leaf growth, were all greater in the C_4 than C_3 type at high temperatures (Osborne et al., 2008), as expected from theory and past work in other species (Monson and Jaeger 1991; Kephart et al. 1992). However, the temperature threshold for a C_4 advantage occured at a lower temperature (15-17 °C) than is typically observed in comparisons of older C_4 species with C_3 counterparts (e.g. Ehleringer & Bjorkman, 1977). Furthermore, the benefit of C_3 photosynthesis expected at lower temperatures was compromised by exposure to chilling in the range 10-15 °C, which causes photodamage to PSII in both C_3 and C_4 types (Osborne et al., 2008). This shared failure in C_3 and C_4 lineages to tolerate chilling likely reflected their shared recent tropical history (Osborne et al., 2008). In contrast, tolerance of freezing differs markedly between these C_3 and C_4 populations because the C_3 lineage has evolved a physiological response that protects leaves from freezing damage after a period of cold acclimation (Osborne et al., 2008). A common garden experiment in South Africa showed that differential freezing tolerance leads to complete canopy senescence under field conditions during winter in the C_4 type, but allowed the C_3 to retain functioning leaves (Ibrahim et al., 2008). Differential leaf survival was significant because daytime leaf temperatures on cloudless winter days exceed 25 °C, at which C_4 photosynthesis would provide a significant performance advantage (Ibrahim et al., 2008). In this species, adaptations (or a lack of adaptations) to avert chilling and freezing damage were therefore more important determinants of ecophysiological behaviour in cool conditions than the differential limitation of C_3 and C_4 photosynthesis by low temperatures.

C_3 and C_4 South African populations of *A. semialata* were also differentially impacted by water availability. C_4 leaves have a higher water-use efficiency of photosynthesis than the C_3 type under well-watered conditions (Ibrahim et al., 2008). However, drought conditions offset this benefit by causing a greater non-stomatal (i.e. metabolic) limitation of photosynthesis in the C_4 than C_3 type (Ripley et al., 2007). In a common garden, this difference in ecophysiological behaviour negated the C_4 photosynthetic advantage over the C_3 during drought events (Ibrahim et al., 2008). The general susceptibility of C_4 photosynthesis to drought limitation observed in multiple species is thought to be non-stomatal and metabolic in origin as photosynthetic inhibition is independent of ambient CO_2 concentration (Ghannoum et al., 2003; Ghannoum, 2009). In other wild grass species, the metabolic limitation to CO_2 assimilation doubled the time taken in C_4 compared to C_3 types to recover photosynthetic potential after rewatering (Ripley et al., 2010a). Evidence from *A. semialata* showed that the depression in CO_2 assimilation is not caused by the restriction of alternative electron sinks when photochemistry is limited by water deficits (Ripley et al., 2007), but the mechanism remains unknown. In the case of drought responses, the differences in ecophysiological behaviour between C_3 and C_4 populations of *A. semialata* therefore relate directly to photosynthetic diversification.

Life history and biomass allocation also differ between South African C_3 and C_4 *A. semialata*. C_4 plants allocate a smaller proportion of biomass to leaves and roots than the C_3 type, and a greater proportion to bulbs and flowers (Ripley et al., 2008). This difference in allocation strategy may reflect the higher nitrogen-use efficiency of the C_4 than C_3 type, which means that biomass productivity is greater and may be more flexibly allocated to storage and reproduction for a given amount of nitrogen. Alternatively, it may represent a secondary adaptation to disturbance by fire in the C_4 type, which tends to occupy more fire-prone habitats (Ripley et al., 2008, 2010b). A controlled burning experiment supported the hypothesis that the studied C_4 population was better fire adapted than the C_3. After an experimental fire during the winter dry season, spring growth was little impacted in the C_4 due to the remobilization of belowground resources and faster aboveground productivity, but was significantly impaired in the C_3 (Ripley et al., 2010b). These growth patterns have important ecological implications that must only be indirectly associated with C_4 photosynthesis.

Finally, experiments under controlled environmental conditions have evaluated the photosynthetic and growth responses of C_3 and C_4 *A. semialata* to historical CO_2 concentrations corresponding to glacial and interglacial levels (Ripley et al., 2013). C_4 photosynthesis provides the greatest benefits

for CO_2-fixation at the lowest, glacial CO_2 level, with increases in photosynthetic capacity in the C_3 type compensating for CO_2-limitation at the interglacial level (Ripley et al., 2013). This photosynthetic acclimation in the C_3 leaves was associated with significant increases in nitrogen concentration. However, the total pool of nitrogen in C_3 plants was unchanged by CO_2 treatments, such that the acclimation response was associated with lower nitrogen-use efficiency and biomass. In contrast, tissue nitrogen concentrations within C_4 *A. semialata* were little impacted by CO_2 (Ripley et al., 2013). The work indicated that leaf acclimation to CO_2 is mediated by whole plant resource-use.

Work using *A. semialata*, therefore, showed that ecophysiological behaviour arose from the interactions of shared history with photosynthetic diversification and novel secondary adaptations. Several aspects of physiological function are clearly related to the diversification of photosynthetic pathway within this species, especially drought and CO_2 responses. However, the work emphasised the importance of adaptations (or a lack of adaptations) in non-photosynthetic characters for ecophysiological behaviour, even for environmental factors closely associated with C_4 performance such as temperature. Comparison of closely related C_3 and C_4 populations of *A. semialata* showed that such secondary adaptations may arise rapidly during photosynthetic pathway diversification.

Additional genomic research using C_3 and C_4 *A. semialata*

Resources for comparative genomics will become essential for further dissection of the evolution of C_4 photosynthesis, as the precise genetic changes responsible for the evolution of this complex phenotype are characterised. As part of this review, we present additional genomic resources for *A. semialata*, which now includes *de novo* reference genomes for each of the four nuclear lineages, two of which (one C_3 and one C_4) are assembled at the chromosome scale. We conduct some preliminary analyses that highlight the type of questions which can be asked with these data (i.e. the role of structural rearrangements and gene duplication in the emergence of C_4 photosynthesis). We hope that the background information gathered here, and the publication of additional resources, will encourage other researchers to use *A. semialata* as a model for C_4 evolution.

Reference genomes for each of the four nuclear lineages in A. semialata

Similarly to the rest of the Paniceae tribe, the genome of *A. semialata* is composed of nine distinct chromosomes (Frean and Marks 1988). The haploid genome size (1C) is around 1 Gb in diploid individuals, and slightly lower in polyploid individuals (Olofsson et al. 2016, 2021), suggesting genome downsizing occurs post polyploidization. We previously assembled a chromosome-level reference genome for one diploid accession from lineage IV originating from Australia (AUS1; Dunning et al. 2019b). We further generated assemblies of varying contiguity for one representative from each of the other three lineages, all diploid accessions (Raimondeau et al. 2022). Here, we present Omni-C sequence data for one of the C_3 individuals from lineage I for which a draft genome assembly was already available (individual RSA5-3; Raimondeau et al. 2022). We use these data for HiRise scaffolding, obtaining a chromosome-level genome assembly for this C_3 RSA5-3 individual that we then compare to the C_4 AUS1 (See supplementary information for full details of the methods and supplementary results). In addition to the reference genomes, there is publicly available whole-genome re-sequencing data for almost 100 individuals, RAD data for several hundred others representing the entirety of the *A. semialata* diversity sampled to date, and transcriptome data (Available from GenBank Sequence Read Archive under BioProject PRJNA401220, PRJNA481434 PRJNA560360, PRJNA649872, PRJNA666779, PRJNA752516, PRJNA715711 and PRJNA824797).

The role of structural variants in the emergence of C_4 photosynthesis

Structural rearrangements (e.g. inversions) reduce recombination in that region of the genome and can therefore lock together beneficially loci (Rieseberg 2001; Hipp et al. 2010; Huang and Rieseberg 2020). These structural rearrangements can drive adaptive shifts (Hoffman and Rieseberg, 2008), and they could potentially play a role in the emergence of C_4 photosynthesis. To identify structural rearrangements, the first step is to establish the syntenic regions between genomes. Both assemblies consist of nine chromosomes and differ in their overall size (RSA5-3 =621 Mb and AUS1 = 747 Mb). This difference largely reflects the estimated 1C genome sizes (Table 1) and is partially explained by the expansion of "copy and paste" transposable elements, especially Ty3-retrotransposons (Table S4). The physical locations of orthologous genes on the nine chromosomes shows that synteny is well conserved between the C_3 and C_4 genomes, with most matches placed along the diagonal (Figure 6). Indeed, only 1,480 one-to-one orthologs were not in syntenic positions (Figure S2 and S3; Table S5 and S6) indicating that only about 10% of genes changed physical location since the divergence of the two genomes . The synteny plots include several duplicated segments, for instance between chromosome 1 and both chromosomes 4 and 7

(Figure 6) which likely represent remnants of an ancient Poales whole-duplication event which is also visible in other grasses (Paterson et al. 2004, 2009). The plots also show several more recent rearrangements, including inversions (e.g. on chromosomes 2, 3, and 6; Figure 6) and translocations (e.g. on chromosomes 5 and 9; Figure 6). Further data is needed to confirm these represent true chromosomal changes and not assembly artefacts, but comparisons with the *S. italica* genome indicate that three of them are specific to RSA5-3 (i.e. they differ between both RSA5-3 and AUS1 and RSA5-3 and *S. italica*), while two are specific to AUS1 (Figure 6). A further five inversions exist between both *A. semialata* and *S. italica* (Figure 6).

To evaluate the role of structural rearrangements in C_4 evolution, we selected the most relevant genes involved in C_4 photosynthesis in *A. semialata*, based on gene expression levels in leaves, and located them on their physical position in both assemblies (Dunning et al. 2019a, Moreno-Villena et al. 2018). All C_4 genes are present in both genomes, except *tpt-1P1*, that is absent in the C_3 accession, and most of them show a conserved chromosome location and copy number (Figure 7). Focusing on the four genes that were consistently upregulated in C_3+C_4 and C_4 versus C_3 (Figure 4), we observed that they are located in genomic regions that rearranged between AUS1 and RSA5-3. The gene *aspat-3P4* is located on chromosome 5 in both individuals, but seems to be affected by a large inversion or translocation (Figure 7). The gene *ppdk-1P2* is surprisingly duplicated in RSA5-3, the C_3 individual (Figure 7). The genes *pck-1P1* and *ppc-1P3* were laterally transferred into the C_4 accession AUS1 from other C_4 grasses (Christin et al. 2012) and are, therefore, located on chromosomes 7 and 8 in AUS1 genome but absent in RSA5-3 (Figure 7). However, native copies still persist in both genomes. For *ppc-1P3*, the native copy is located on chromosome 4 in both accessions, and RSA5-3 has an additional duplication on chromosome 6 (Figure 7). For *pck-1P1*, both native copies are located on chromosome 9 but translocated to different chromosome arms (Figure 7). These large variants usually have a significant effect in gene regulation and gene function (Alonge et al. 2020, Hämälä et al. 2021, Yuan et al. 2021), but further work is required to determine whether these structural rearrangements have played a role in the evolution of C_4 photosynthesis in *A. semialata*.

Gene duplication plays no lasting role in C_4 evolution

The evolution of C_4 involves the increased expression of numerous enzymes, and gene duplication may play an important role in gene expression through a dosage effect. Indeed, it has previously

been shown that there is a dosage effect on the expression of two C_4 genes (*pck* and *pepc*) in *A. semialata* (Bianconi et al. 2018). However, differences in gene copy number appear to be transient, being rendered obsolete by the fixation of regulatory mutations increasing expression levels (Bianconi et al. 2018). When comparing the C_3 and C_4 reference genomes, we see that gene duplication is common, with the number of genes increasing by gene duplication being 24% in RSA5-3 and 21% in AUS1 since the split of the two accessions. However, only a few C_4 genes are duplicated, and these are in fact in the C_3 individual. This confirms previous work that while it may be important for the initial emergence of C_4 photosynthesis, these copy number variants are not maintained over longer evolutionary periods (Bianconi et al. 2018). However, it is important to note that some recent duplicates creating gene copy number variation might be collapsed in reference genomes and therefore hidden in the absence of read depth analyses (Bianconi et al. 2018).

Furthermore, neofunctionalization of retained duplicated genes may contribute to the evolution of novel traits. Most known C_4 genes do not seem to have followed this pathway (Gowik and Westhoff 2011; Williams et al. 2012; van den Bergh et al. 2014), however, comparisons of seven monocot genomes led to 21 orthologous genes that were duplicated and retained in parallel in two distinct C_4 origins (Emms et al. 2016), suggesting that these might be newly identified genes that contributed to the evolution or optimization of C_4 photosynthesis through duplication and neofunctionalization.

Conclusions

The genus *Alloteropsis* in general, and the species *A. semialata* in particular, constitute outstanding systems for retracing the events that led to C_4 evolution, and the ecological and physiological consequences of this gradual transition. Compared to what is known in other species, *A. semialata* has maintained an unparalleled diversity of photosynthetic traits, and across the wider genus there is extensive variation in C_4 anatomy, biochemistry, and physiology. This phenotypic variation is accompanied by important genetic diversity, both among and within groups. This diversity provides the perfect ground for comparative analyses, enabling the identification of the traits that consistently differ among groups. Such an approach has been applied to C_4 ecology, physiology, anatomy, and transcriptomics, and this system would also be highly suited for quantitative genetics (Simpson et al. 2022). In addition, this diversity provides ample information to reconstruct the genomic history of the group and the genes within them using phylogenetic approaches. The genome provided here

for a C_3, together with the previous one we published for a C_4 individual (Dunning et al. 2019b), will enable future research into the genetic architecture of C_4 evolution. Properties that are beneficial for comparative analyses, such as slow growth allowing the maintenance of accessions and the retention of diversity among accessions, however, represent limitations for experimental work. To our knowledge, no transformation has been attempted with any *Alloteropsis* species. In addition, *A. semialata* infrequently sets flowers and seeds in the greenhouse, hampering controlled crosses. If these limitations were to be overcome, the group would represent one of the best systems to study the genomic factors that control the expression of the C_4 trait in grasses.

Acknowledgements

LP is supported by a NERC Standard Grant (NE/V000012/1), PAC was funded by the Royal Society (grants RGF\EA\181050 and URF\R\180022), and LTD is supported by a Natural Environment Research Council Independent Research Fellowship (NE/T011025/1). We thank Vanja Milenkovic for her help in generating carbon isotope ratios for samples of *Alloteropsis angusta*.

Data availability

Raw sequencing data for AUS1 and RSA5-3 are available from the NCBI Sequence Read Archive under BioProject PRJNA481434 and PRJNA824797. Genome assemblies, annotations, extracted coding sequence and proteins for AUS1 and RSA5-3 generated as part of this review are available from dryad (doi:10.5061/dryad.c866t1gb1).

References

Alonge, M, Wang, X, Benoit, et al. 2020. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* 182(1):145-161.e23.

Amy Lyu MJ, Tang Q, Wang Y, Essemine J, et al. 2022. Evolution of gene regulatory network of C4 photosynthesis in the genus *Flaveria* reveals the evolutionary status of C_3-C_4 intermediate species. *Plant Communications*. 4(1): 100426.

Barrett, DR, Frean, ML, Cresswell, CF. 1983. C_3 and C_4 Photosynthetic and Anatomical forms of Alloteropsis semialata (R. Br.) Hitchcock: I. Variability in Photosynthetic Characteristics, Water Utilization Efficiency and Leaf Anatomy. *Annals of Botany* 51(6):801-809.

Bateman, BL, Johnson, CN. 2011. The influences of climate, habitat and fire on the distribution of cockatoo grass (*Alloteropsis semialata*) (Poaceae) in the Wet Tropics of northern Australia. *Australian Journal of Botany* 59(4):315-323.

Bianconi, ME, Dunning, LT, Moreno-Villena, JJ, Osborne, CP, Christin, PA. 2018. Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus *Alloteropsis*. *Journal of Experimental Botany* 69(8):1967-1980.

Bianconi, ME, Dunning, LT, Curran, EV, et al. 2020. Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B* 287(1938):20201960.

Bianconi, ME, Sotelo, G, Curran, EV et al. 2022. Upregulation of C_4 characteristics does not consistently improve photosynthetic performance in intraspecific hybrids of a grass. *Plant, Cell & Environment* 45(5):1398-1411.\

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Maß J, Lercher MJ, Westhoff P, Hibberd JM, Weber APM. 2011. An mRNA Blueprint for C_4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C_3 and C_4 Species. *Plant Physiology* 155(1):142–156

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM. 2014. Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. *Journal of Experimental Botany* 65(13):3579–3593.

Brown, WV. 1975. Variations in anatomy, associations, and origins of Kranz tissue. *American Journal of Botany* 62(4):395-402.

Brown, WV. 1977. The Kranz syndrome and its subtypes in grass systematics. *Memoirs of the Torrey Botanical Club* 1-97.

Bruhl, JJ, Perry, S. 1995. Photosynthetic pathway-related ultrastructure of C_3, C_4 and C_3-like C_3-C_4 intermediate sedges (Cyperaceae), with special reference to *Eleocharis*. *Functional Plant Biology* 22(4):521-530.

Burgess SJ, Hibberd JM. 2015. Insights into C_4 metabolism from comparative deep sequencing. *Current Opinion in Plant Biology* 25:138–144.

Busch, FA. 2020. Photorespiration in the context of Rubisco biochemistry, CO_2 diffusion and metabolism. *The Plant Journal* 101(4):919-939.

Cardol, P, Bailleul, B, Rappaport, F, et al. 2008. An original adaptation of photosynthesis in the marine green alga *Ostreococcus*. *Proceedings of the National Academy of Sciences* 105(22):7881-7886.

Christin, PA, Edwards, EJ, Besnard, G, et al. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology* 22(5):445-449.

Christin, PA, Sage, TL, Edwards, EJ, Ogburn, RM, Khoshravesh, R, Sage, RF. 2011. Complex evolutionary transitions and the significance of C_3–C_4 intermediate forms of photosynthesis in *Molluginaceae*. *Evolution: International Journal of Organic Evolution* 65(3):643-660.

Christin, PA, Osborne, CP, Chatelet, DS, et al. 2013. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences* 110(4):1381-1386.

Christin, PA, Arakaki, M, Osborne, CP et al. 2014. Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. *Journal of Experimental Botany* 65(13):3609-3621.

Clayton WD, Renvoize SA. 1982. Flora of Tropical East Africa. Gramineae (Part 3). Rotterdam: AA Balkema

Curran, EV, Scott, MS, Olofsson, JK et al. 2022. Hybridization boosts dispersal of two contrasted ecotypes in a grass species. *Proceedings of the Royal Society B* 289(1967):20212491.

Dunning, LT, Lundgren, MR, Moreno-Villena, JJ et al. 2017. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* 71(6):1541-1555.

Dunning, LT, Moreno-Villena, JJ, Lundgren, MR et al. 2019a. Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *Journal of Experimental Botany* 70(12):3255-3268.

739 Dunning, LT, Olofsson, JK, Parisod, C et al. 2019b. Lateral transfers of large DNA fragments
740 spread functional genes among grasses. *Proceedings of the National Academy of Sciences*
741 116(10):4416-4425.

742 Dusenge, M.E., Duarte, A.G. and Way, D.A. 2019. Plant carbon metabolism and climate change:
743 elevated CO2 and temperature impacts on photosynthesis, photorespiration and respiration. *New*
744 *Phytol,* 221: 32-49.

745 Edwards, GE, Furbank, RT, Hatch, MD, Osmond, CB. 2001. What does it take to be C_4? Lessons
746 from the evolution of C_4 photosynthesis. *Plant Physiology* 125(1):46-49.

747 Edwards EJ, Osborne CP, Strömberg CA, Smith SA, C_4 Grasses Consortium. 2010. The origins of
748 C_4 grasslands: integrating evolutionary and ecosystem science. *Science* 328(5978):587-91.

749 Ehleringer, J, Björkman, O. 1977. Quantum yields for CO_2 uptake in C_3 and C_4 plants: dependence
750 on temperature, CO_2, and O_2 concentration. *Plant Physiology* 59(1):86-90.

751 Ehleringer, JR, Monson, RK. 1993. Evolutionary and ecological aspects of photosynthetic pathway
752 variation. *Annual Review of Ecology and Systematics* 411-439.

753 Ehleringer, JR, Cerling, TE, Helliker, BR. 1997. C_4 photosynthesis, atmospheric CO_2, and climate.
754 *Oecologia* 112(3):285-299.

755 Ellis, RP. 1974. Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades.
756 *Bothalia* 11(3):273-275.

757 Ellis RP. 1981. Relevance of Comparative Leaf Anatomy in Taxonomic and Functional Research
758 on the South African Poaceae. DSc Thesis, University of Pretoria, South Africa.

759 Emms, DM, Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome
760 comparisons dramatically improves orthogroup inference accuracy. *Genome biology* 16(1):1-14.

761 Hämälä, T, Wafula, EK, Guiltinan, MJ, Ralph, PE, dePamphilis, CW and Tiffin P. 2021. Genomic
762 structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*,
763 the chocolate tree. *Proceedings of the National Academy of Sciences* 118(35):e2102914118.

764 Huang K and Rieseberg LH. 2020. Frequency, Origins, and Evolutionary Role of Chromosomal
765 Inversions in Plants. *Front Plant Sci*. 11:296.

766 Frean, M, Barrett, D, Cresswell, CF. 1980. Variability in leaf surface features and water efficiency
767 utilisation in C_3 and C_4 forms of *Alloteropsis semialata* (R. Br.) hitchc. *Proceedings of the Annual*
768 *Congresses of the Grassland Society of Southern Africa* 15:99–103.

769 Frean, ML, Ariovich, D and Cresswell, CF. 1983a. C_3 and C_4 Photosynthetic and anatomical forms
770 of *Alloteropsis semialata* (R. Br.) Hitchcock: 2. A comparative investigation of leaf ultrastructure
771 and distribution of chlorenchyma in the two forms. *Annals of Botany* 51(6):811-821.

772 Frean, ML, Barrett, DR, Ariovich, D, Wolfson, M, Cresswell, CF. 1983b. Intraspecifíc variability in
773 *Alloteropsis semialata* (R. Br.) Hitchc. *Bothalia*, 14(3/4):901-913.

774 Frean, ML, Marks, E. 1988. Chromosome numbers of C_3 and C_4 variants within the species
775 *Alloteropsis semialata* (R. Br.) Hitchc. (Poaceae). *Botanical journal of the Linnean Society*,
776 97(3):255-259.

777 Freitag, H., Kadereit, G. C_3 and C_4 leaf anatomy types in Camphorosmeae (Camphorosmoideae,
778 Chenopodiaceae). 2014. *Plant Syst Evol* 300:665–687.

779 Ghannoum, O, Conroy, JP, Driscoll, SP, Paul, MJ, Foyer, CH and Lawlor, DW. 2003. Nonstomatal
780 limitations are responsible for drought-induced photosynthetic inhibition in four C_4 grasses. *New*
781 *Phytologist* 159:599-608.

782 Ghannoum, O. 2009. C_4 photosynthesis and water stress, *Annals of Botany* 103(4):635–644.

783 Gibbs Russell, G. 1983. The taxonomic position of C_3 and C_4 *Alloteropsis semialata* (Poaceae) in
784 southern Africa. *Bothalia* 14(2):205-213.

785 Giussani, LM, Cota-Sánchez, JH, Zuloaga, FO, Kellogg, EA. 2001. A molecular phylogeny of the
786 grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. *American*
787 *Journal of Botany*, 88(11):1993-2012.

788 Grass Phylogeny Working Group II. 2012., New grass phylogeny resolves deep evolutionary
789 relationships and discovers C_4 origins. *New Phytologist*, 193: 304-312.

790 Gowik U, Westhoff P. 2011. The path from C-3 to C-4 photosynthesis. *Plant Physiol*. 155:56–63.

791 Hatch, MD. 1987. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and
792 ultrastructure. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics* 895(2):81-106.

793 Hattersley PW, Watson L. 1992. Diversification of photosynthesis. In: Chapman, GP, eds. *Grass*
794 *Evolution and Domestication*. Cambridge: Cambridge University Press: 38–116.

795 Heyduk, K, Moreno-Villena, JJ, Gilman, IS, Christin, PA, Edwards, EJ. 2019. The genetics of
796 convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics* 20(8):485-493.

797 Hipp AL, Rothrock PE, Whitkus R, Weber JA. (2010), Chromosomes tell half of the story: the
798 correlation between karyotype rearrangements and genetic diversity in sedges, a group with
799 holocentric chromosomes. *Molecular Ecology*, 19: 3124-3138.

800 Hoffmann, AA, Rieseberg, LH. 2008. Revisiting the impact of inversions in evolution: from
801 population genetic markers to drivers of adaptive shifts and speciation? *Annual review of ecology,*
802 *evolution, and systematics*. 39:21-42.

803 Ibrahim, DG, Gilbert, ME, Ripley, BS and Osborne, CP. 2008. Seasonal differences in
804 photosynthesis between the C_3 and C_4 subspecies of *Alloteropsis semialata* are offset by frost and
805 drought. *Plant, Cell & Environment* 31:1038-1050.

806 Ibrahim, DG, Burke, T, Ripley, BS, Osborne, CP. 2009. A molecular phylogeny of the genus
807 *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3
808 photosynthesis. *Annals of Botany* 103(1):127-136.

809 Kadereit, G, Borsch, T, Weising, K, Freitag, H. 2003. Phylogeny of Amaranthaceae and
810 Chenopodiaceae and the evolution of C_4 photosynthesis. *International journal of plant sciences*
811 164(6):959-986.

812 Kadereit, G, Bohley, K, Lauterbach, M, Tefarikis, DT and Kadereit, JW. 2017. C_3–C_4 intermediates
813 may be of hybrid origin–a reminder. *New Phytologist* 215(1):70-76.

814 Kephart, KD, Buxton, DR, Taylor, SE. 1992. Growth of C_3 and C_4 Perennial Grasses under
815 Reduced Irradiance. *Crop Science*, 32: 1033-1038.

816 Khoshravesh, R, Stata, M, Busch, FA, Saladié, M, Castelli, JM, Dakin, N, Hattersley, PW,
817 Macfarlane, TD, Sage, RF, Ludwig, M, Sage, TL. 2020. The evolutionary origin of C_4
818 photosynthesis in the grass subtribe Neurachninae. *Plant Physiology* 182(1):566-83.

819 Koteyeva, NK, Voznesenskaya, EV, Roalson, EH, Edwards, GE. 2011. Diversity in forms of C_4 in
820 the genus *Cleome* (Cleomaceae). *Annals of Botany* 107(2):269-283.

821 Kumarathunge, D.P., Medlyn, B.E., Drake, J.E., Tjoelker, M.G., Aspinwall, M.J., Battaglia, M.,
822 Cano, F.J., Carter, K.R., Cavaleri, M.A., Cernusak, L.A., Chambers, J.Q., Crous, K.Y., De Kauwe,
823 M.G., Dillaway, D.N., Dreyer, E., Ellsworth, D.S., Ghannoum, O., Han, Q., Hikosaka, K., Jensen,
824 A.M., Kelly, J.W.G., Kruger, E.L., Mercado, L.M., Onoda, Y., Reich, P.B., Rogers, A., Slot, M.,
825 Smith, N.G., Tarvainen, L., Tissue, D.T., Togashi, H.F., Tribuzy, E.S., Uddling, J., Vårhammar, A.,
826 Wallin, G., Warren, J.M. and Way, D.A. 2019. Acclimation and adaptation components of the
827 temperature dependence of plant photosynthesis at the global scale. *New Phytol*, 222: 768-784.

828 Langdale, JA. 2011. C_4 cycles: past, present, and future research on C_4 photosynthesis. *The Plant*
829 *Cell* 23(11):3879-3892.

830 Lehmann, CE, Griffith, DM, Simpson, KJ, Anderson, TM, Archibald, S, Beerling, DJ, Bond, WJ,
831 Denton, E, Edwards, EJ, Forrestel, EJ, Fox, et al DL. 2019. Functional diversification enabled
832 grassy biomes to fill global climate space. *BioRxiv*

833 Liebenberg, EJL, Fossey, A. 2001. Comparative cytogenetic investigation of the two subspecies of
834 the grass *Alloteropsis semialata* (Poaceae). *Botanical Journal of the Linnean Society* 137(3):243-
835 248.

836 Lundgren, MR, Besnard, G, Ripley, BS et al. 2015. Photosynthetic innovation broadens the niche
837 within a single species. *Ecology Letters* 18(10):1021-1029.

838 Lundgren, MR, Christin, PA, Escobar, EG et al. 2016. Evolutionary implications of C_3–C_4
839 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell & Environment* 39(9):1874-1885.

840 Lundgren, MR, Christin, PA. 2017. Despite phylogenetic effects, C_3–C_4 lineages bridge the
841 ecological gap to C_4 photosynthesis. *Journal of Experimental Botany* 68(2):241-254.

842 Lundgren, MR, Dunning, LT, Olofsson, JK, et al. 2019. C_4 anatomy can evolve via a single
843 developmental change. *Ecology letters* 22(2):302-312.

844 Lyu, MJ.A., Gowik, U., Kelly, S. et al. 2021. The coordination of major events in C_4 photosynthesis
845 evolution in the genus *Flaveria*. *Sci Rep* 11, 15618.

846 Matsuoka, M, Furbank, RT, Fukayama, H, Miyao, M. 2001. Molecular engineering of C_4
847 photosynthesis. *Annual review of plant biology* 52(1):297-314.

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

850 Monson, RK, Moore, BD, Ku, MSB, Edwards, GE. 1986. Co-function of C_3-and C_4-photosynthetic
851 pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. *Planta* 168(4):493-502.

852 Monson, RK and Jaeger, CH. 1991. Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria*
853 *floridana* (Asteraceae) in natural habitats: evidence of advantages to C_3-C_4 photosynthesis at high
854 leaf temperatures. *American Journal of Botany*, 78: 795-800.

855 Morales-Briones DF, Kadereit G. 2022. Exploring the possible role of hybridization in the 1
856 evolution of photosynthetic pathways in *Flaveria* (Asteraceae), the prime model of C_4
857 photosynthesis evolution. *bioRxiv*

858 Moreno-Villena, JJ, Dunning, LT, Osborne, CP, Christin, PA. 2018. Highly Expressed Genes Are
859 Preferentially Co-Opted for C_4 Photosynthesis. *Molecular Biology and Evolution* 35(1):94–106.

860 Muhaidat, R, Sage, RF, Dengler, NG. 2007. Diversity of Kranz anatomy and biochemistry in C_4
861 eudicots. *American journal of botany* 94(3):362-381.

862 Riyadh Muhaidat, Athena D. McKown. 2013. Significant involvement of PEP-CK in carbon
863 assimilation of C_4 eudicots. *Annals of Botany* 111(4):577–589.

864 Olofsson, JK, Bianconi, M, Besnard, G et al. 2016. Genome biogeography reveals the intraspecific
865 spread of adaptive mutations for a complex trait. *Molecular ecology* 25(24):6107-6123.

866 Olofsson, JK, Dunning, LT, Lundgren, MR et al. 2019. Population-specific selection on standing
867 variation generated by lateral gene transfers in a grass. *Current Biology* 29(22):3921-3927.

868 Olofsson, JK, Curran, EV, Nyirenda, F et al. 2021. Low dispersal and ploidy differences in a grass
869 maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology*
870 30(9):2116-2130.

871 Osborne CP, Wythe EJ, Ibrahim DG, Gilbert ME, Ripley BS. 2008. Low temperature effects on leaf
872 physiology and survivorship in the C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of*
873 *Experimental Botany* 59(7):1743-54.

874 Osmond, CB, Winter, K, Ziegler, H. 1982. Functional Significance of Different Pathways of CO_2
875 Fixation in Photosynthesis. In: Lange, OL, Nobel, PS, Osmond, CB, Ziegler, H, eds. *Physiological*
876 *Plant Ecology II. Encyclopedia of Plant Physiology*, vol 12 B. Berlin: Springer.

877 Pagani, M, Zachos, JC, Freeman, KH, Tipple, B, Bohaty, S. 2005. Marked decline in atmospheric
878 carbon dioxide concentrations during the Paleogene. *Science* 309(5734):600-603.

879 Paterson, AH, Bowers, JE, Chapman, B. 2004. Ancient polyploidization predating divergence of the
880 cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of*
881 *Sciences*, 101(26):9903-9908.

882 Paterson, AH, Bowers, JE, Bruggmann, R et al. 2009. The *Sorghum bicolor* genome and the
883 diversification of grasses. *Nature* 457(7229):551-556.

884 Pearcy, RW, Ehleringer, J. 1984. Comparative ecophysiology of C_3 and C_4 plants. *Plant, Cell &*
885 *Environment* 7:1-13.

886 Peterhansel, C, Maurino, VG. 2011. Photorespiration redesigned. *Plant physiology* 155(1):49-55.

887 Phansopa, C, Dunning, LT, Reid, JD, Christin, PA. 2020. Lateral gene transfer acts as an
888 evolutionary shortcut to efficient C_4 biochemistry. *Molecular biology and evolution* 37(11):3094-
889 3104.

890 Prendergast, HDV, Hattersley, PW, Stone, NE. 1987. New structural/biochemical associations in
891 leaf blades of C_4 grasses (Poaceae). *Functional Plant Biology* 14(4):403-420.

892 Raimondeau, P, Bianconi, ME, Pereira, L, Parisod, C, Christin, PA, Dunning, LT. (In Prep) High
893 turnover of lateral gene transfers diversifies a grass pangenome.

894 Raven JA and Geider RJ. 2003. Adaptation, acclimation and regulation in algal photosynthesis. In:
895 Larkum AWD, Douglas SE and Raven JA, eds. *Photosynthesis in Algae. Advances in*
896 *Photosynthesis and Respiration, vol 14*. Dordrecht: Springer, 385-412.

897 Raven, J.A. and Geider, R.J., 2003. Adaptation, acclimation and regulation in algal photosynthesis.
898 In *Photosynthesis in algae* (pp. 385-412). Springer, Dordrecht.

899 Raven, JA, Cockell, CS, De La Rocha, CL. 2008. The evolution of inorganic carbon concentrating
900 mechanisms in photosynthesis. *Philosophical Transactions of the Royal Society B: Biological*
901 *Sciences* 363(1504):2641-2650.

902 Renvoize SA. 1987. A survey of leaf-blade anatomy in grasses XI. Paniceae. *Kew Bulletin* 42:739–
903 768.

904 Rieseberg LH. 2001. Chromosomal rearrangements and speciation. Trends in Ecology & Evolution
905 16(7):351-358.

906 Ripley, BS, Gilbert, ME, Ibrahim, DG, Osborne, CP. 2007. Drought constraints on C_4
907 photosynthesis: stomatal and metabolic limitations in C_3 and C_4 subspecies of *Alloteropsis*
908 *semialata*. *Journal of experimental botany* 58(6):1351-1363.

909 Ripley BS, Abraham TI, Osborne CP. 2008. Consequences of C_4 photosynthesis for the partitioning
910 of growth: a test using C_3 and C_4 subspecies of *Alloteropsis semialata* under nitrogen-limitation.
911 *Journal of Experimental Botany* 59(7):1705-14.

912 Ripley B, Frole K, Gilbert M. 2010a. Differences in drought sensitivities and photosynthetic
913 limitations between co-occurring C_3 and C_4 (NADP-ME) Panicoid grasses. *Annals of Botany*
914 105(3):493-503.

915 Ripley, B, Donald, G, Osborne, CP, Abraham, T, Martin, T. 2010b. Experimental investigation of
916 fire ecology in the C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of Ecology* 98: 1196-
917 1203.

Ripley, BS, Cunniff, J, Osborne, CP. 2013. Photosynthetic acclimation and resource use by the C_3 and C_4 subspecies of *Alloteropsis semialata* in low CO_2 atmospheres. *Global Change Biology* 19:900-910.

Rocha AES, Miranda IS. 2012. Nova ocorrência de Poaceae para a América do Sul: *Alloteropsis* (Panicoideae/Poaceae). *Acta Amazonica* 42:457–460.

Sage, RF, Pearcy, RW. 1987. The Nitrogen Use Efficiency of C_3 and C_4 Plants: I. Leaf Nitrogen, Growth, and Biomass Partitioning in *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiology* 84(3):954–958.

Sage, RF. 2004. The evolution of C_4 photosynthesis. *New phytologist* 161(2):341-370.

Sage, RF, Sage, TL, Kocacinar, F. 2012. Photorespiration and the evolution of C_4 photosynthesis. *Annual review of plant biology* 63:19-47.

Sage, RF, Monson, RK, Ehleringer, JR, Adachi, S, Pearcy, RW. 2018. Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia* 187(4):941-966.

Sales, CR, Wang, Y, Evers, JB, Kromdijk, J. 2021. Improving C_4 photosynthesis to increase productivity under optimal and suboptimal conditions. *Journal of Experimental Botany* 72(17):5942-5960.

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

Simpson, CJ, Reeves, G, Tripathi, A, Singh, P, Hibberd, JM. 2022. Using breeding and quantitative genetics to understand the C_4 pathway. *Journal of experimental botany* 73(10):3072-3084.

Skillman, JB. 2008. Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark. *Journal of experimental botany* 59(7):1647-1661.

Still, CJ, Berry, JA, Collatz, GJ, DeFries, RS. 2003. Global distribution of C_3 and C_4 vegetation: carbon cycle implications. *Global biogeochemical cycles* 17(1):1006.

Tcherkez, GG, Farquhar, GD, Andrews, TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* 103(19):7246-7251.

Tefarikis DT, Morales-Briones DF, Yang Y, Edwards G. and Kadereit, G. 2022. On the hybrid origin of the C_2 *Salsola divaricata* agg. (Amaranthaceae) from C_3 and C_4 parental lineages. *New Phytol*, 234: 1876-1890.

948 Ueno, O, Sentoku, N. 2006. Comparison of leaf structure and photosynthetic characteristics of C_3
949 and C_4 *Alloteropsis semialata* subspecies. *Plant, Cell & Environment* 29(2):257-268.

950 van den Bergh E, Külahoglu C, BräutigamA, Hibberd JM, Weber AP, Zhu XG, Schranz ME. 2014.
951 Gene and genome duplications and the origin of C 4 photosynthesis: Birth of a trait in the
952 Cleomaceae. *Curr Plant Biol*. 1:2–9.

953 Vogan, PJ, Sage, RF. 2011. Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate
954 species of Flaveria Juss. (Asteraceae). *Plant, cell & environment* 34(9):1415-1430.

955 von Caemmerer, S, Quick, WP, Furbank, RT. 2012. The development of C_4 rice: current progress
956 and future challenges. *Science* 336(6089):1671-1672.

957 Voznesenskaya EV, Koteyeva NK, Akhani H, Roalson EH, Edwards GE. 2013. Structural and
958 physiological analyses in Salsoleae (Chenopodiaceae) indicate multiple transitions among C_3,
959 intermediate, and C_4 photosynthesis. *J Exp Bot.* 64(12):3583-604.

960 Young, JN, Rickaby, REM, Kapralov, MV, Filatov, DA. 2012. Adaptive signals in algal Rubisco
961 reveal a history of ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal*
962 *Society B: Biological Sciences* 367(1588):483-492.

963 Yuan Y, Bayer PE, Batley J, Edwards D. 2021. Current status of structural variation studies in
964 plants. *Plant Biotechnology Journal* 19(11):2153-2163.

965 Zhao H, Wang Y, Lyu MJA, Zhu XG. 2022. Two major metabolic factors for an efficient NADP-
966 malic enzyme type C_4 photosynthesis. *Plant Physiology*, 189(1):84–98

967 Zuloaga, FO, Morrone, O, Davidse, O et al. 2003. *Alloteropsis*. In: R.J. Soreng & S.J. Pennington,
968 eds. *Catalogue of New World Grasses (Poaceae): III. Subfamilies Panicoideae, Aristidoideae,*
969 *Arundinoideae and Danthonioideae v.46*. Washington: Smithsonian Institution, 16.

970

971 **Table 1: Comparison of chromosome-level reference genomes of *A. semialata***

	ASEM_RSA5	**ASEM_AUS1**
Lineage	I - C_3	IV - C_4
Origin	South Africa (-32.70 27.53)	Australia (-19.62 146.96)
1C genome size	870 Mb	1,100 Mb
Assembled genome size	621 Mb	747 Mb
Number of scaffolds	1323	687
Size across nine chromosomes	582 Mb	782 Mb
Annotated genes	71,593	75,709
Annotated genes with Panicoideae homologs	30,060	29,126

972

Figure captions

Figure 1: Diversity of *Alloteropsis*

Press herbarium specimens are shown for the main groups of *Alloteropsis*. From top to bottom and then left to right: *A. cimicina* (Nyirenda, Curran, Christin ZAM2065-05; SHD), *A. angusta* erect (Lundgren 2015-3-3; SHD) and decumbent (Nyirenda, Curran, Bianconi, Christin ZAM1933a; SHD), *A. semialata* lineage I (Mapaura, Lundgren, Olofsson 4; SHD), lineage II (Nyirenda, Curran, Bianconi, Christin ZAM1936-H1; SHD), lineage III (Nyirenda, Curran, Bianconi, Christin ZAM1934; SHD), and lineage IV (Dunning, Yakandawala, Ariyarathne-06; SHD).

Figure 2: Relationships among *Alloteropsis* lineages

a) The phylogenetic relationships among the main lineages of *Alloteropsis* are shown, based on the nuclear genome analyses of Raimondeau et al. (2022). The approximate age of the root is indicated and the names of the group and their photosynthetic types are given on the right; dec = decumbent, ere = erect. Arrows connecting branches indicate reported episodes of genetic exchanges (Olofsson et al. 2021; Curran et al. 2022). **b)** The genetic contributions of genetic groups are represented for polyploid individuals analysed in detail (Olofsson et al. 2016, 2021; Bianconi et al. 2020). For each polyploid, the approximate proportion of their genome originating from each diploid lineage (shown at the top) is represented with a corresponding colour. Ploidy levels are indicated on the left, and the name of a representative population is given for each polyploid.

Figure 3: Geographic distribution and photosynthetic diversity of *Alloteropsis semialata*

(A) For each of the four main lineages of *A. semialata* (lineage I in blue, lineage II in green, lineage III in orange, lineage IV in red), the distribution of known samples is shown, with shapes indicating their ploidy level (see Table S2 for data). Individuals of mixed ancestry unassigned to these lineages are shown separately. The grey area on each map shows the putative range of the species, based on known individuals plus reported samples on GBIF (https://www.gbif.org/). (B) The distribution of carbon isotope ratios ($\delta^{13}C$) is shown with histograms for each group. In the case of individuals with mixed ancestries, the relative parental contributions of the four lineages are shown with pie charts.

Figure 4: Photosynthetic types within *A. semialata*

The figure shows the (A) C_4 gene expression patterns, (B) putative carbon acquisition pathways, and (C) leaf anatomy for a C_3 individual of *A. semialata* (RSA5, left column), a C_3+C_4 individual (TAN1, middle column), and a C_4 individual (AUS1, right column). (A) The expression levels of genes encoding enzymes involved in the C_4 pathway in reads per milliom of mappable reads per

kilobase (rpkm) were retrieved from Dunning et al. (2017); 1 = carbonic anhydrase (gene *βca-2P3*), 2 = phosphoenolpyruvate carboxylase [sum of genes *ppc-1P3* and *ppc-1P6* (shown in orange), including the laterally-acquired *ppc-1P3_LGT-A* present in AUS1 shown in black], 3 = aspartate aminotransferase (gene *aspat-3P4*), 4 = phosphoenolpyruvate carboxykinase (laterally-acquired gene *pck-1P1_LGT-C*), 5 = NADP-malate dehydrogenase (gene *nadpmdh-3P4*), 6 = NADP-malic enzyme (gene *nadpme-1P4*), 7 = alanine aminotransferase (*alaat-1P5*), and 8 = pryuvate, phosphate dikinase (*ppdk-1P2*). Asterisks below gene numbers indicate those that were consistently upregulated in the C_3+C_4 versus C_3 or C_4 versus C_3+C_4 comparisons (see Dunning et al. 2019a). (B) The putative C_4 cycles show the metabolites across the mesophyll (in pink) and bundle sheath (in yellow) cells, with enzymes represented by circles and numbered as in (A) and the PCK pathway in blue and the NADP-ME pathway in orange; OAA = oxaloacetate, ASP = aspartate, PEP = phosphoenolpyruvate, mal = malate, pyr = pyruvate, ALA = alanine, CBB = Calvin-Benson-Bassham cycle. (C) The leaf cross sections were retrieved from Bianconi et al. 2022. The black arrows point to the inner bundle sheath of each vein.

Figure 5: History of PEPC co-option in *Alloteropsis*

Each bar approximates the expression level of one of the seven genes encoding PEPC existing in grasses, in black for functional non-C_4 genes, in orange for genes co-opted in C_4 photosynthesis, in red for laterally-acquired C_4 genes, and in grey for non-functional pseudogenes. The name of an individual illustrating each pattern is indicated on the right (Dunning et al.; 2017; Phansopa et al. 2020).

Figure 6: Synteny comparisons based on coding DNA

Pairs of synteny blocks (i.e. genes in the same position in both comparisons) are shown along the nine chromosomes of *A. semialata* (genomes RSA5-3 and AUS1) and *Setaria italica* v2.2. High-quality matches (at least eight reciprocal best matches in a nine match window) are shown in black or red, with the others in grey. Rectangles highlight major rearrangement, in blue for those specific to RSA5-3, in magenta for those specific to AUS1 and in green for those that differ between *S. italica* and both *A. semialata* assemblies.

Figure 7: Genomic distribution of C_4 genes

The two *A. semialata* assemblies are represented by chromosomes, and expressed C_4 genes (Dunning et al. 2019a) are located according to their genomic positions. The genes differentially expressed among C_3, C_3+C_4 and C_4 groups are highlighted in pink, and other C_4-related genes are

represented in grey. The two laterally-transferred copies are marked with an asterisk. Gene names are abbreviated according to Moreno-Villena et al. (2018): ppc = Phosphoenolpyruvate carboxylase; pck = Phosphoenolpyravute carboxykinase; ppdk = Pyruvate, phosphate dikinase; aspat = Aspartate aminotransferase; alaat = Alanine aminotransferase; nadpme = NADP-malic enzyme; nadpmdh = NADP-malate dehydrogenase; ®ca = β-carbonic anhydrase; sbas = Sodium bile acid symporter family; pgk = Phosphoglycerate Kinase; tpt = Triosephosphate-phosphate translocator; dic = Dicarboxylate carrier; ak = Adenylate kinase protein; gapdh = Glyceraldehyde-3-phosphate Dehydrogenase.

This manuscript also includes supplementary information.

Supplementary Materials and Methods

Supplementary Figures (all available in the main SI document):

Figure S1: Leaf cross section of a naturally occurring C_3+C_4 x C_4 hybrid identified in the wild showing the intermittent presence of minor veins.

Figure S2: Genomic distribution of one-to-one orthologs.

Figure S3: Origins of genes in the two genomes of *Alloteropsis semialata.*

Supplementary Tables (All available as individual tabs in the supplementary tables file)**:**

Table S1: Carbon isotope ratios and coordinates of Alloteropsis angusta samples

Table S2: Sampled populations of *Alloteropsis semialata*

Table S3: Mean environmental parameters for C_3, C_3+C_4 and C_4 *Alloteropsis semialata* populations extracted from Lundgren and Christin (2017).

Table S4: Statistics of TE annotation for AUS1 and RSA5-3 genome assemblies.

Table S5: One-to-one orthologs between the two genomes of *Alloteropsis semialata*

Table S6: Position and syntenic relationship among pairs of genes belonging to the same orthogroup. At least one of the pair has to be located on one of the nine chromosomes