Annals of Botany **132**: 365–381, 2023
https://doi.org/10.1093/aob/mcad078, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

Alloteropsis semialata as a study system for C_4 evolution in grasses

Lara Pereira[1], Matheus E. Bianconi[1], Colin P. Osborne[2], Pascal-Antoine Christin[1,] and Luke T. Dunning[*,1]

[1]*Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN,UK*
[2]*Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK*
[*]*For correspondence. E-mail l.dunning@sheffield.ac.uk*

Received: 17 February 2023 Returned for revision: 2 June 2023 Editorial decision: 20 June 2023 Accepted: 7 July 2023

• **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 economic 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 palaeotropics.
• **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 with that of a C_4 accession of *A. semialata*.
• **Conclusions** *Alloteropsis semialata* is one of the best systems in which to investigate the evolution of C_4 photosynthesis because the genetic and phenotypic variation provides a fertile ground for comparative and population-level studies. Preliminary comparative genomic investigations show that 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 *A. semialata* a great model for further comparative analyses of photosynthetic diversification.

Key words: Poaceae, photosynthesis, adaptation.

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 been adapted continuously in response to varying environmental conditions (Raven and Geider, 2003; Tcherkez *et al.*, 2006; Cardol *et al.*, 2008; Raven *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 ~30 million years ago (Ma) (Pagani *et al.*, 2005). Given that CO_2 represents the main substrate 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, 2011; 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 after the Oligocene (Ehleringer and Monson, 1993; Ehleringer *et al.*, 1997; Tcherkez *et al.*, 2006). Over the last 32 million years (Christin *et al.*, 2011), >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 consequently, they 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, 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 in addition to 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 that is isolated from the atmosphere, typically in 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 >30 Ma (Christin *et al.*, 2008), and consequently, they 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 therefore used 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, however, the photosynthetic types are 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 in which 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 been repeatedly studied since then, focusing on the ecological (Ripley *et al.*, 2007, 2010*b*; Ibrahim *et al.*, 2008; Osborne *et al.*, 2008; Bateman and Johnson, 2011; Lundgren *et al.*, 2015), cytogenetic (Frean and Marks, 1988; Liebenberg and Fossey, 2001; Lundgren *et al.*, 2015; Bianconi *et al.*, 2020; Olofsson *et al.*, 2021), physiological (Frean *et al.*, 1980, 1983*a*; Lundgren *et al.*, 2016) and biochemical variation (Ueno and Sentoku, 2006; Phansopa *et al.*, 2020), and more recently, on evolutionary and genomic aspects of C_4 photosynthesis (Ibrahim *et al.*, 2009; Christin *et al.*, 2012; Lundgren *et al.*, 2015, 2019; Olofsson *et al.*, 2016, 2021; Dunning *et al.*, 2017, 2019*a*; Bianconi *et al.*, 2018, 2020; Curran *et al.*, 2022). In this review, we consolidate 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 with 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 to 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 *Alloteropsis cimicina*, *A. paniculata* and *A. papillosa* are all C_4 and form a monophyletic group based on chloroplast markers (Ibrahim *et al.*, 2009). *Alloteropsis cimicina* is an annual weed (Fig. 1), native across Africa, Asia and Oceania, which is invasive in the Americas (Zuloaga *et al.*, 2003; Rocha and Miranda, 2012). Few samples of this group have been analysed 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, because 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 as ~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 other two species in the genus ~11 Ma (Fig. 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 that it also has an erect morph representing a distinct ecotype (Fig. 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 (Fig. 2; Curran *et al.*, 2022). All 132 samples of *A. angusta* analysed with carbon isotopes were probably C_4, with only one anomalous non-C_4 value that could need to be verified because of limited availability of material (Supplementary Data Table S1). The karyotype of the four *A. angusta* samples analysed was probably diploid, with a haploid genome size (1C) of ~1 Gb (Curran *et al.*, 2022). *Alloteropsis angusta* diverged from *A. semialata* ~8 Ma (Lundgren *et al.*, 2015; Dunning *et al.*, 2017), although nuclear genome analyses indicate that the two species hybridize repeatedly where they come into contact (Fig. 2; Curran *et al.*, 2022).

The diversity of *A. semialata* was studied originally 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 *A. semialata*

Chloroplast markers suggested that the species originated in Central-East Africa, 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 appearing to be mostly consistent within each of these four groups (Fig. 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

FIG. 1. Diversity of *Alloteropsis*. Pressed herbarium specimens are shown for the main groups of *Alloteropsis*. From top to bottom, 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).

of the species. Their most recent common ancestor is estimated at ~2.5 Ma (Bianconi *et al.* 2020; Raimondeau *et al.*, 2022), and each has followed a different evolutionary trajectory since then.

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 (Fig. 3; Lundgren *et al.*, 2015; Bianconi *et al.* 2020). All individuals analysed so far are diploids (Lundgren *et al.*, 2015; Bianconi *et al.*, 2020). Based on carbon isotopes, no lineage I individuals use the derived C_4 photosynthetic type (Fig. 3; Supplementary Data 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 (Fig. 4; Frean *et al.*, 1980; Barrett *et al.*, 1983; Ueno and Sentoku, 2006; Ripley *et al.*, 2007; Lundgren *et al.*, 2016, 2019; Dunning *et al.*, 2019*a*). The leaf anatomy of plants from this group is, moreover, indicative of a C_3 photosynthetic type (Fig. 4; Ellis, 1974; Frean *et al.*, 1983*a*, *b*; Lundgren *et al.*, 2019). In two individuals from this group, the inner bundle sheath (also known as the mestome sheath) was shown to contain glycine decarboxylase (Ueno and Sentoku, 2006; Lundgren *et al.*, 2019), an enzyme used in 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 and 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 analysed (Frean *et al.*, 1980; Barrett *et al.*, 1983; Ripley *et al.*, 2007; Lundgren *et al.*, 2016) suggest that such a pump is weak, at best. Nuclear lineage I can thus be considered as encompassing mostly, if not only, C_3 individuals.

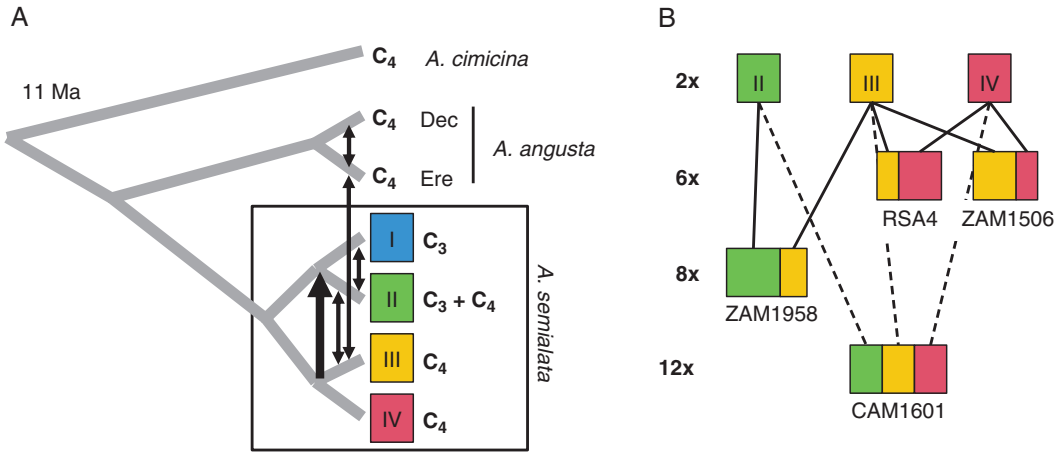

FIG. 2. Relationships among *Alloteropsis* lineages. (A) The phylogenetic relationships among the main lineages of *Alloteropsis* are shown, based on the nuclear genome analyses by Raimondeau *et al.* (2022). The approximate age of the root is indicated, and the names of the group and their photosynthetic types are given on the right. Abbreviations: 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 by a corresponding colour. Ploidy levels are indicated on the left, and the name of a representative population is given for each polyploid.

The nuclear lineage II is associated mostly 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 (Fig. 3). All individuals analysed 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 the lineage contains isotopic intermediates that probably grew using both C_3 and C_4 pathways (Fig. 3; Lundgren *et al.*, 2016; Olofsson *et al.*, 2021). Moreover, its individuals 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 (Fig. 4; Lundgren *et al.*, 2016, 2019; Dunning *et al.*, 2019*a*). 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 consequently, they were 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; Sage *et al.*, 2012, 2018; Schlüter and Weber, 2016; Lyu *et al.*, 2022). Moreover, the variation in carbon isotopes, CO_2 compensation points and leaf anatomy within lineage II (Fig. 3; Lundgren *et al.*, 2016, 2019; Olofsson *et al.*, 2021) 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 associated mostly 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 hypothesized 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 restricted mostly to the Central Zambezian miombo woodlands, where it overlaps with lineage II (Fig. 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 (Fig. 3; Lundgren *et al.*, 2015; Olofsson *et al.*, 2021). The populations analysed showed high leaf transcript abundance of genes encoding C_4 enzymes (Dunning *et al.*, 2019*a*). The assessed populations from Zambia and Tanzania have very low CO_2 compensation points and leaves where chloroplasts are concentrated in the inner bundle sheath, confirming that they are C_4 (Lundgren *et al.*, 2019).

The nuclear lineage IV is associated mostly with chloroplast groups D and E (Lundgren *et al.*, 2015). It migrated rapidly away from the centre of origin of the species to Southern Africa, Western Africa, Madagascar, Asia and Australia (Fig. 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 occasionally form mixed populations (Frean *et al.*, 1980). It is present in a variety of biomes, and its broader niche is likely to be associated with its rapid dispersal (Lundgren *et al.*, 2015), although whether each individual possesses a broad niche or whether the group is able to adapt rapidly to local conditions 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 (Fig. 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 (Fig. 4; Frean *et al.*, 1983*a*, *b*; Ueno and Sentoku, 2006; Lundgren *et al.*, 2016,

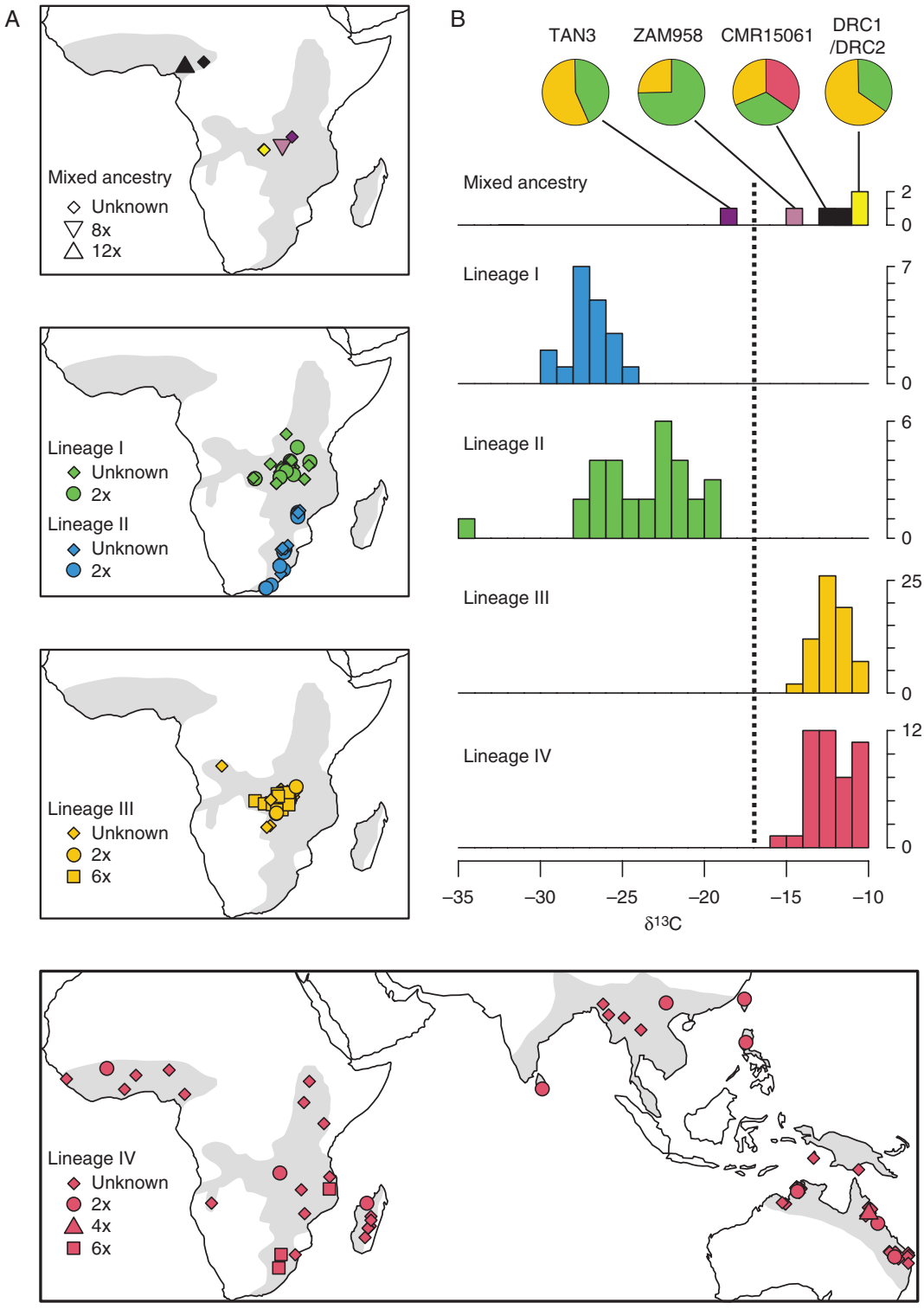

FIG. 3. Geographical distribution and photosynthetic diversity of *Alloteropsis semialata*. (A) For each of the four main lineages of *A. semialata* (lineage I in blue, lineage II in green, lineage III in orange and lineage IV in red), the distribution of known samples is shown, with shapes indicating their ploidy level (for data, see Supplementary Data Table S2). 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.

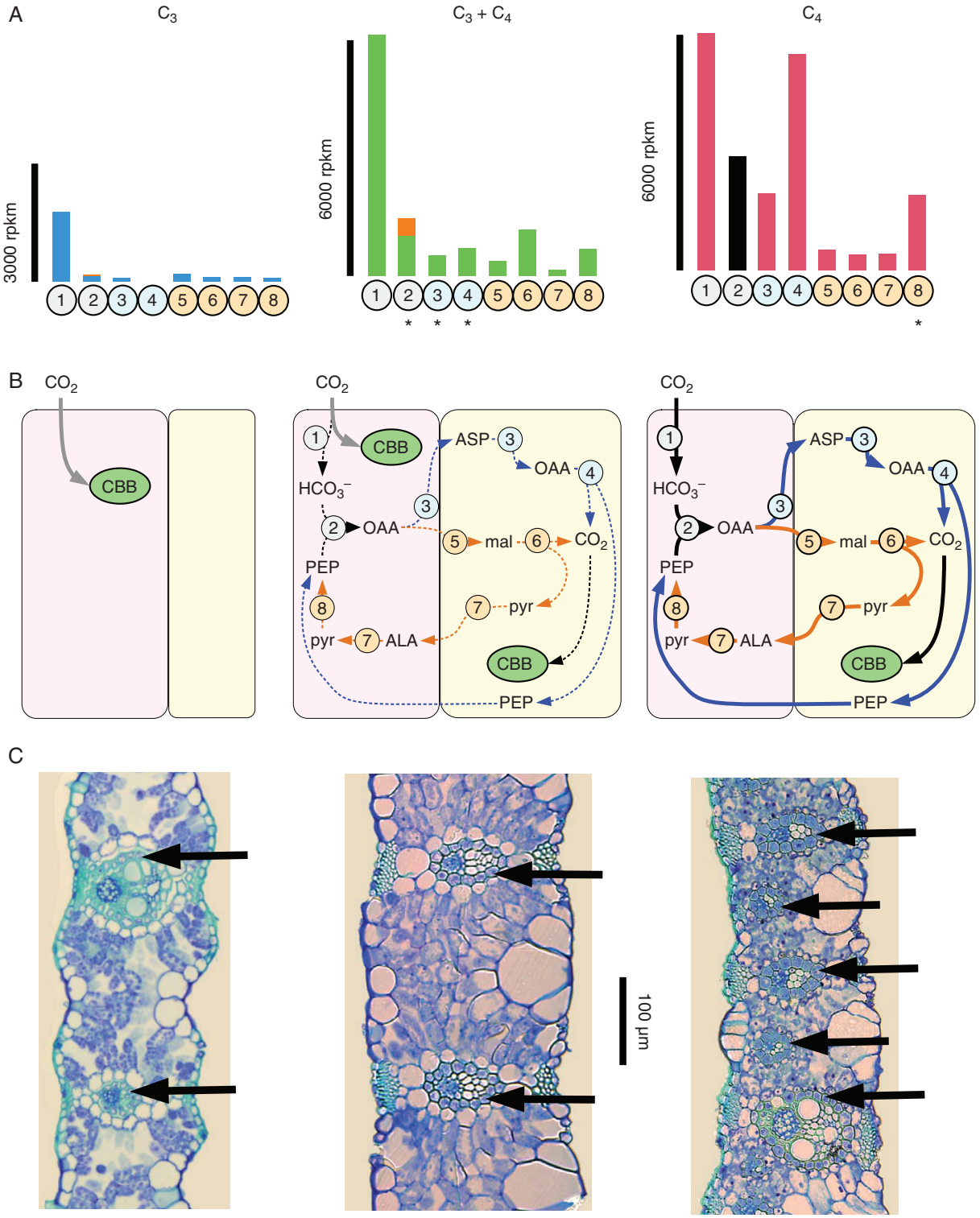

FIG. 4. Photosynthetic types within *Alloteropsis semialata*. The figure shows: (A) the C_4 gene expression patterns; (B) putative carbon acquisition pathways; and (C) leaf anatomy for a C_3 individual of *A. semialata* (RSA5, left column), a $C_3 + C_4$ individual (TAN1, middle column) and a C_4 individual (AUS1, right column). (A) The expression levels of genes encoding enzymes involved in the C_4 pathway [in reads per million 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.*, 2019*a*). (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), the PCK pathway in blue and the NADP-ME pathway in orange. Abbreviations: ALA, alanine; ASP, aspartate; CBB, Calvin–Benson–Bassham cycle; mal, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; pyr, pyruvate. (C) The leaf cross-sections were retrieved from the paper by Bianconi *et al.* (2022). The black arrows point to the inner bundle sheath of each vein.

2019; Dunning *et al.*, 2019*a*). 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). Given that the physiology and carbon isotopes are typical of C_4 plants (Fig. 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 applied easily 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, meaning that many individuals would fall outside these existing subspecies (Curran *et al.*, 2022). More generally, the four nuclear lineages of *A. semialata* described here hybridize naturally in the wild (Fig. 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 lacking. 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 because 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*

Although the four nuclear lineages are clearly distinct in diploids, some polyploids are placed phylogenetically outside 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 (12×) from Cameroon (Bianconi *et al.*, 2020), hexaploids (6×)/octoploids (8×) from Zambia (Olofsson *et al.*, 2021) and plants of unknown ploidy in Tanzania and the Democratic Republic of Congo (Fig. 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 (Fig. 2; Olofsson *et al.*, 2021) and II + III + IV in Cameroon (Fig. 2; Bianconi *et al.*, 2020); and the hexaploids from lineages III and IV each present contributions from the other lineage (Fig. 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), remain 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 (Fig. 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 although its lifespan has never been evaluated, we have kept some plants for 10 years in climatically controlled greenhouse conditions (12 h daylight, 25/20 °C day/night temperature) without any signs of senescence. The species forms rosettes, from which long stems emerge carrying the inflorescence (Fig. 1). The base of the leaf is thickened, forming bulb-like structures, and new bulbs and tillers emerge continuously 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.

In comparison, sexual reproduction is uneven in the greenhouse. Some accessions seem to flower more frequently than others, and the seed set is also fairly variable. Based on our own experience, polyploids are able to self-fertilize, as can some of the $C_3 + C_4$ (lineage II), which occasionally even develop cleistogamous inflorescences among the rosette leaves. In 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 interbreed freely 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 occur episodically 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 is likely to have 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 (Fig. 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, argues strongly 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 (Fig. 4; Ellis, 1974; Brown, 1977; Frean *et al.*, 1983a; Renvoize, 1987; Christin *et al.*, 2013; Lundgren *et al.*, 2019). The details of the C_4 pathway in *A. cimicina* have not been established with biochemical assays, but transcriptomic analyses indicate a high transcript abundance of genes encoding the NADP-malic enzyme (NADP-ME) and other enzymes associated with the NADP-ME type (Dunning *et al.*, 2017, 2019*a*). Early analyses described some South African *A. semialata* as using the NADP-ME type, with potential changes depending on the temperature (Frean *et al.*, 1983b), although this has not been confirmed subsequently. All other samples were characterized as being of the phosphoenolpyruvate carboxykinase decarboxylase (PCK) type (Prendergast *et al.*, 1987; Ueno and Sentoku, 2006). Subsequent transcriptomic analyses suggested that, at 25 °C, all assessed C_4 accessions of either *A. semialata* or *A. angusta* use mainly the PCK C_4 type, with various amounts of NADP-ME activity (Fig. 4; Dunning *et al.*, 2017, 2019*a*). 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 paralogues 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 because 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 the genus *Flaveria*, 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, because 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, because 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* (Figs 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 (Fig. 5; Dunning *et al.*, 2017), meaning that the C_4-specific PEPC from different C_4 accessions effectively have different origins even if they originated in the common ancestor from the group (Fig. 5). A completely distinct gene of PEPC is used by the Australian *A. semialata*, and this gene was acquired laterally, from the Andropogoneae *Themeda triandra* (Christin *et al.*, 2012; Dunning *et al.*, 2019*b*), with important biochemical adaptations that pre-date the transfer (Phansopa *et al.*, 2020). This laterally acquired copy replaced the ancestral native copies, which became pseudogenes (Fig. 5). A similar process happened in Africa with two other PEPC genes acquired laterally from other groups of grasses (Cenchrinae and Melinidinae; Christin *et al.*, 2012), each present in a subset of individuals, creating a mosaic of PEPC origins within the species (Olofsson *et al.*, 2016). These PEPC examples show that C_4 components can originate independently within a single C_4 lineage. The C_4 *A. semialata* also present important C_4 anatomical variation (Lundgren *et al.*, 2019) and variation in the expression of C_4-related genes (Dunning *et al.*, 2019*a*), offering an excellent system in which to unravel the secondary

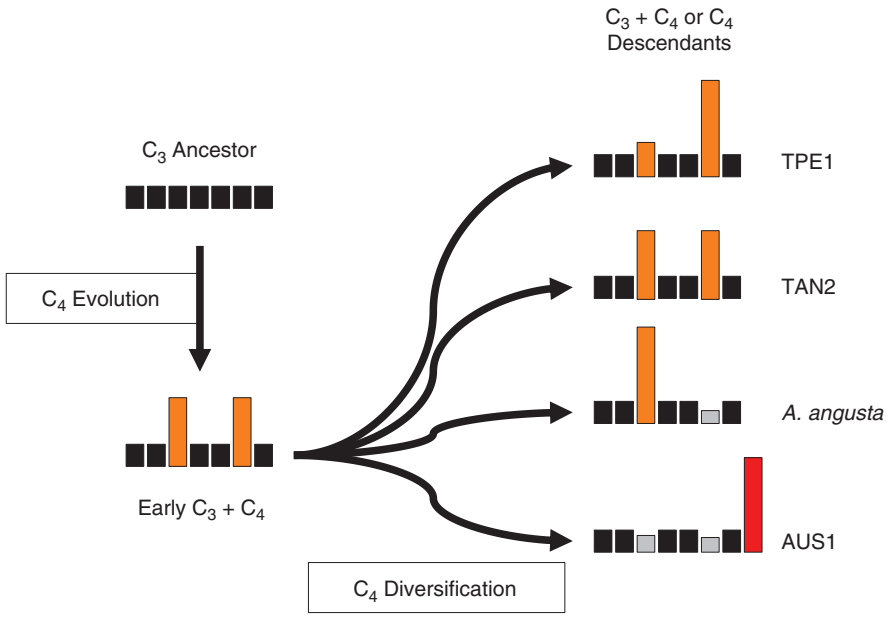

FIG. 5. History of phosphoenolpyruvate carboxylase (PEPC) co-option in *Alloteropsis*. Each bar approximates the expression level of one of the seven genes encoding PEPC existing in grasses, in black for functional non-C_4 genes, in orange for genes co-opted in C_4 photosynthesis, in red for laterally 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).

adaptations that happened among sub-lineages after the initial transition to C_4 photosynthesis.

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

The C_4 pathway observed in older C_4 lineages is often complex, relying on numerous anatomical and biochemical specialities (Hatch, 1987; Sage, 2004; Sage *et al.*, 2012). Anatomically, a higher proportion of vascular bundle sheath tissue, as a result of a greater number of veins and larger size of the bundle sheath cells, strongly increases C_4 evolvability in grasses (Christin *et al.*, 2013). Biochemical changes are numerous and very diverse depending on the photosynthetic subtype (Burgess and Hibberd, 2015). Such features are likely to encompass those that were needed to evolve the C_4 trait (primary adaptations), but also those that evolved later (secondary adaptations) and those that evolved for unrelated reasons (Heyduk *et al.*, 2019). In fact, most comparative transcriptomic studies among C_4 and C_3 relatives identify several hundreds of differentially expressed genes, many of which would probably represent secondary adaptations (Bräutigam *et al.*, 2011, 2014; Lyu *et al.*, 2021).

Given that they share a recent common ancestor, comparisons of the different groups of *A. semialata* offer an opportunity to distinguish the primary adaptations that were acquired during the initial transition to C_4 photosynthesis from those that evolved later. In such comparative endeavours, it is especially important to sample the diversity within each group to identify those features restricted to a subsample of C_4 plants, which represent secondary adaptations. Leaf anatomy comparisons revealed that the only character that differs consistently between C_4 and non-C_4 *A. semialata* is the proliferation of minor veins (Fig. 4), which is therefore sufficient to support the 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 to 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 are likely to 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 with the C_4 *A. semialata*, and only pyruvate orthophosphate dikinase was upregulated in the C_4 compared with the $C_3 + C_4$ *A. semialata* (Fig. 4; Dunning *et al.*, 2019*a*). Other C_4-related genes were upregulated only in a subset of C_4 populations (e.g. malate dehydrogenase), suggesting that 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 for PEPC, see Phansopa *et al.*, 2020). 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 (*aspat*, *pck*, *ppc* and *ppdk*) accompanied by a single key alteration in leaf development 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 combined later, following genetic exchanges among distinct lineages and adaptive introgression (Olofsson *et al.*, 2016). The presence of distinct populations occupying different environments and accumulating new 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*, F1 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 evolutionarily intermediate steps (Lyu *et al.*, 2021). In contrast, a recent study using transcriptomes from 17 *Flaveria* species showed recurrent hybridization among them, which 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 $C_3 + C_4 \times C_4$ hybrids identified in the wild (Supplementary Data Fig. S1). Such minor veins are generally absent from the natural $C_3 + C_4$ populations (Lundgren *et al.*, 2019). Although these differences indicate that the origin of $C_3 + C_4$ is not likely to stem simply from an ancient hybridization between C_3 and C_4 parents, gene flow between the different *A. semialata* lineages might have played a role in 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 to be relatively ancient because 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 connect the C_4 and $C_3 + C_4$ lineages periodically, 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 in hot, dry and high-light environments, including improved light-, water- and nitrogen-use efficiencies (Pearcy and Ehleringer, 1984; Sage and Pearcy, 1987; Long, 1999). However, it has long been recognized that the photosynthetic pathway is only one of a suite of ecophysiological characters needed to succeed in any particular environment. These might be inherited from non-C_4 ancestors or evolve as environmental adaptations during C_4 diversification (Osmond *et al.*, 1982; Pearcy and 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 relationship to temperature, water deficits, fire and atmospheric CO_2 concentration (Supplementary Data 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 in light-saturated conditions, the quantum efficiency in light-limited conditions and the rate of leaf growth were all greater in the C_4 than the 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 occurred 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 and 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 photosystem II 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 is likely to reflect 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 acclimatization (Osborne *et al.*, 2008). A common garden experiment in South Africa showed that differential freezing tolerance led to complete canopy senescence in field conditions during winter in the C_4 type, but allowed the C_3 type to retain functioning leaves (Ibrahim *et al.*, 2008). Differential leaf survival was significant because daytime leaf temperatures on cloudless winter days exceeded 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 are 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* are also differentially impacted by water availability. C_4 leaves have a higher water-use efficiency of photosynthesis than the C_3 type in 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 the 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 because 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 with C_3 types to recover photosynthetic potential after rewatering (Ripley *et al.*, 2010*a*). 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 might reflect the higher nitrogen-use efficiency of the C_4 than the C_3 type, which means that biomass productivity is greater and might be allocated more flexibly to storage and reproduction for a given amount of nitrogen. Alternatively, it might 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, 2010*b*). A controlled burning experiment supported the hypothesis that the studied C_4 population was better fire adapted than the C_3 type. After an experimental fire during the winter dry season, spring growth was little impacted in the C_4 type, owing to the remobilization of belowground resources and faster aboveground productivity, but was significantly impaired in the C_3 type (Ripley *et al.*, 2010*b*). These growth patterns have important ecological implications that must only be associated with C_4 photosynthesis indirectly.

Finally, experiments in 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 acclimatization 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 acclimatization 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 acclimatization 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 new secondary adaptations. Several aspects of physiological function are clearly related to diversification of the photosynthetic pathway within this species, especially the drought and CO_2 responses. However, the work has emphasized the importance of adaptations (or 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* have shown that such secondary adaptations might 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 characterized. 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 types of questions that 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*

Like 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 ~1 Gb in diploid individuals and slightly lower in polyploid individuals (Olofsson *et al.*, 2016, 2021), suggesting that genome downsizing occurs after polyploidization. We previously assembled a chromosome-level reference genome for one diploid accession from lineage IV originating from Australia (AUS1; Dunning *et al.*, 2019*b*). We also 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, which we then compare with the C_4 AUS1 (for full details of the methods and results, see Supplementary Data Methods S1 and Results). In addition to the reference genomes, there are publicly available whole-genome resequencing 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 loci beneficial loci together (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 explained, in part, by the expansion of 'copy and paste' transposable elements, especially Ty3-retrotransposons (Supplementary Data 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 (Fig. 6). Indeed, only 1480 one-to-one orthologues were not in syntenic positions (Supplementary Data Figs S2 and S3; Tables S5 and S6), indicating that only ~10% of genes have 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 (Fig. 6), which probably represent remnants of an ancient Poales whole-duplication event that 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; Fig. 6) and translocations (e.g. on chromosomes 5 and 9; Fig. 6). Further data are needed to confirm that these represent true chromosomal changes and not assembly artefacts, but comparisons with the *Setaria italica* genome indicate that three of them are specific to RSA5-3 (i.e. they differ both between RSA5-3 and AUS1 and between RSA5-3 and *S. italica*), while two are specific to AUS1 (Fig. 6). A further five inversions exist between both *A. semialata* and *S. italica* (Fig. 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 (Moreno-Villena *et al.*, 2018; Dunning *et al.*, 2019*a*). All C_4 genes are present in both genomes, except *tpt-1P1*, which is absent in the C_3 accession, and most of them show a conserved chromosome location and copy number (Fig. 7). Focusing on the four genes that were consistently upregulated in $C_3 + C_4$ and C_4 versus C_3 (Fig. 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 (Fig. 7). The gene *ppdk-1P2* is surprisingly duplicated in RSA5-3, the C_3 individual (Fig. 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 the AUS1 genome but absent in RSA5-3 (Fig. 7). However, native copies 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 (Fig. 7). For *pck-1P1*, both native copies are located on chromosome 9 but translocated to different chromosome arms (Fig. 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 can 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

Table 1. *Comparison of chromosome-level reference genomes of* Alloteropsis semialata

Parameter	**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 (Mb)	870	1100
Assembled genome size (Mb)	621	747
Number of scaffolds	1323	687
Size across nine chromosomes (Mb)	582	782
Annotated genes	71 593	75 709
Annotated genes with Panicoideae homologues	30 060	29 126

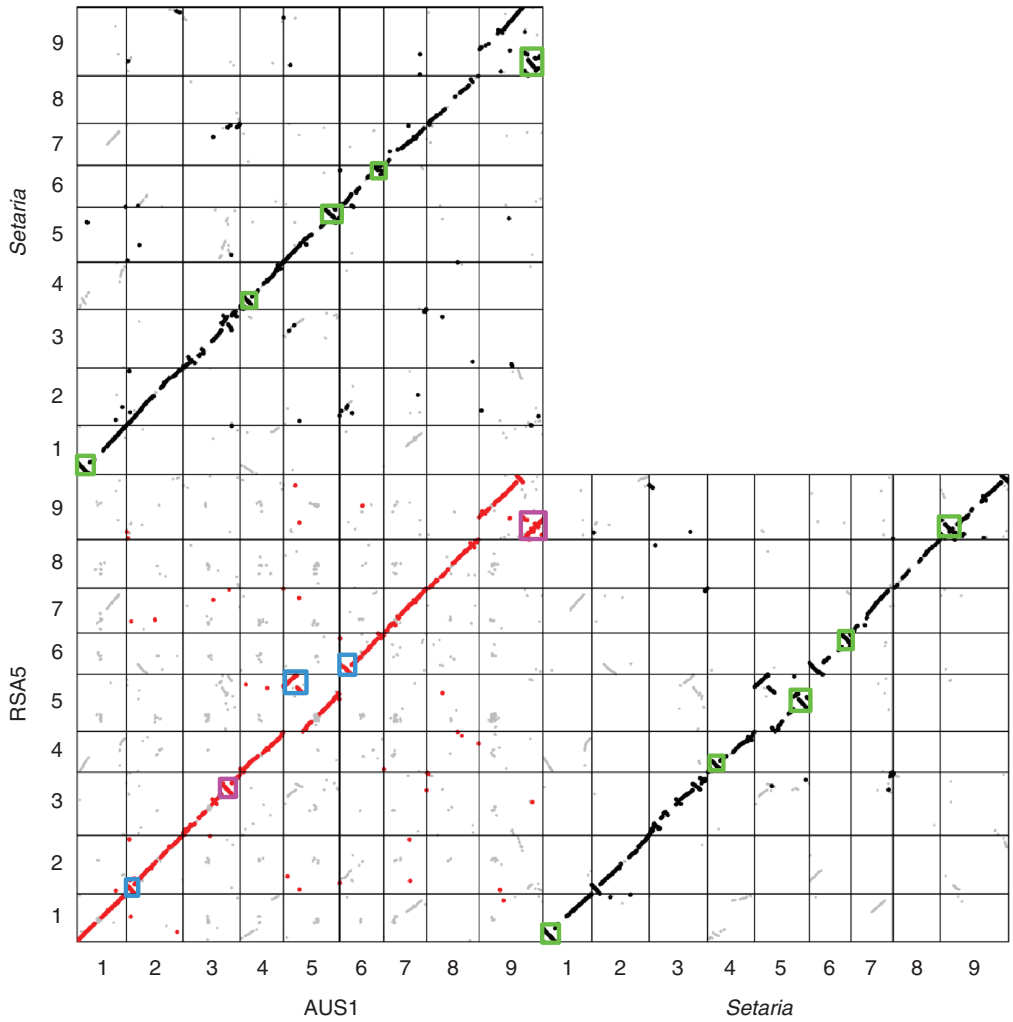

FIG. 6. Synteny comparisons based on coding DNA. Pairs of synteny blocks (i.e. genes in the same position in both comparisons) are shown along the nine chromosomes of *Alloteropsis semialata* (genomes RSA5-3 and AUS1) and *Setaria italica* v2.2. High-quality matches (at least eight reciprocal best matches 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.

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 the C_3 individual. This confirms previous work that although it might 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 can 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 and Kelly, 2015), 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. In comparison 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 differ consistently among groups. Such an approach has been applied to C_4 ecology, physiology, anatomy and transcriptomics, and this system would also be very suitable for quantitative

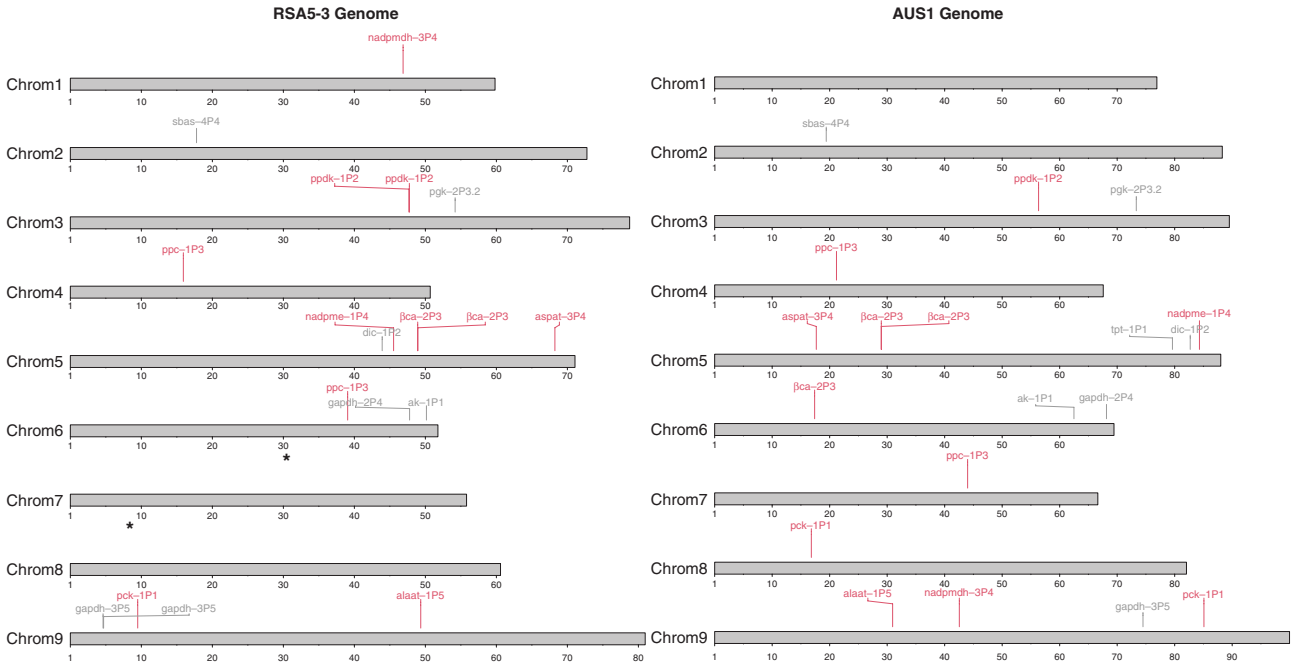

FIG. 7. Genomic distribution of C_4 genes. The two *Alloteropsis semialata* assemblies are represented by chromosomes, and expressed C_4 genes (Dunning *et al.*, 2019*a*) are located according to their genomic positions. The genes differentially expressed among C_3, $C_3 + C_4$ and C_4 groups are shown in pink, with other C_4-related genes in grey. The two laterally transferred copies are marked with an asterisk. Gene names are abbreviated according to Moreno-Villena *et al.* (2018). Abbreviations: ak, adenylate kinase protein; alaat, alanine aminotransferase; aspat, aspartate aminotransferase; βca, β-carbonic anhydrase; dic, dicarboxylate carrier; gapdh, glyceraldehyde-3-phosphate dehydrogenase; nadpmdh, NADP-malate dehydrogenase; nadpme, NADP-malic enzyme; pck, phosphoenolpyravute carboxykinase; pgk, phosphoglycerate kinase; ppc, phosphoenolpyruvate carboxylase; ppdk, pyruvate, phosphate dikinase; sbas, sodium bile acid symporter family; tpt, triosephosphate-phosphate translocator.

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 individual, together with the previous one we published for a C_4 individual (Dunning *et al.*, 2019*b*), will enable future research into the genetic architecture of C_4 evolution. Properties that are beneficial for comparative analyses, such as slow growth that allows 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 in which to study the genomic factors that control expression of the C_4 trait in grasses.

SUPPLEMENTARY DATA

Supplementary data are available at *Annals of Botany* online and consist of the following. Methods S1. Figure S1: leaf cross-section of a naturally occurring $C_3 + C_4 \times C_4$ hybrid identified in the wild, showing the intermittent presence of minor veins. Figure S2: genomic distribution of one-to-one orthologues. Figure S3: origins of genes in the two genomes of *Alloteropsis semialata*. 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 transposable element annotation for AUS1 and RSA5-3 genome assemblies. Table S5: one-to-one orthologues between the two genomes of *Alloteropsis semialata*. Table S6: position and syntenic relationship among pairs of genes belonging to the same orthogroup.

FUNDING

L.P. is supported by a NERC Standard Grant (NE/V000012/1), P.-A.C. was funded by the Royal Society (grants RGF\EA\181050 and URF\R\180022), and L.T.D. is supported by a Natural Environment Research Council Independent Research Fellowship (NE/T011025/1).

ACKNOWLEDGEMENTS

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).

LITERATURE CITED

Alonge M, **Wang X**, **Benoit**, ***et al***. **2020**. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**: 145–161.e23.

Lyu M-JA, **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**: 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**: 801–809. doi:10.1093/oxfordjournals.aob.a086531.

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**: 315–323. doi:10.1071/bt10266.

van den Bergh E, **Külahoglu C**, **Bräutigam A**, ***et al***. **2014**. Gene and genome duplications and the origin of C_4 photosynthesis: birth of a trait in the Cleomaceae. *Current Plant Biology* **1**: 2–9.

Bianconi ME, **Dunning LT**, **Moreno-Villena JJ**, **Osborne CP**, **Christin P-A. 2018**. Gene duplication and dosage effects during the early emergence of C_4 photosynthesis in the grass genus *Alloteropsis*. *Journal of Experimental Botany* **69**: 1967–1980. doi:10.1093/jxb/ery029.

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: Biological Sciences* **287**: 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**: 1398–1411. doi:10.1111/pce.14301.

Bräutigam A, **Kajala K**, **Wullenweber J**, ***et al***. **2011**. An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiology* **155**: 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**: 3579–3593. doi:10.1093/jxb/eru100.

Brown WV. 1975. Variations in anatomy, associations, and origins of Kranz tissue. *American Journal of Botany* **62**: 395–402. doi:10.1002/j.1537-2197.1975.tb14062.x.

Brown WV. 1977. The Kranz syndrome and its subtypes in grass systematics. *Memoirs of the Torrey Botanical Club* **10**: 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**: 521–530. doi:10.1071/pp9950521.

Burgess SJ, **Hibberd JM. 2015**. Insights into C_4 metabolism from comparative deep sequencing. *Current Opinion in Plant Biology* **25**: 138–144. doi:10.1016/j.pbi.2015.05.017.

Busch FA. 2020. Photorespiration in the context of Rubisco biochemistry, CO_2 diffusion and metabolism. *The Plant Journal* **101**: 919–939. doi:10.1111/tpj.14674.

von Caemmerer S, **Quick WP**, **Furbank RT. 2012**. The development of C_4 rice: current progress and future challenges. *Science* **336**: 1671–1672.

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 of the United States of America* **105**: 7881–7886.

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**: 643–660. doi:10.1111/j.1558-5646.2010.01168.x.

Christin P-A, **Edwards EJ**, **Besnard G**, ***et al***. **2012**. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology* **22**: 445–449. doi:10.1016/j.cub.2012.01.054.

Christin P-A, **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 of the United States of America* **110**: 1381–1386.

Christin P-A, **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**: 3609–3621. doi:10.1093/jxb/eru087.

Christin PA, **Besnard G**, **Samaritani E**, ***et al***. **2008**. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology* **18**: 37–43.

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: Biological Sciences* **289**: 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; International Journal of Organic Evolution* **71**: 1541–1555.

Dunning LT, **Moreno-Villena JJ**, **Lundgren MR**, ***et al***. **2019***a*. Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *Journal of Experimental Botany* **70**: 3255–3268. doi:10.1093/jxb/erz149.

Dunning LT, **Olofsson JK**, **Parisod C**, ***et al***. **2019***b*. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 4416–4425. doi:10.1073/pnas.1810031116.

Dusenge ME, **Duarte AG**, **Way DA. 2019**. Plant carbon metabolism and climate change: elevated CO_2 and temperature impacts on photosynthesis, photorespiration and respiration. *New Phytologist* **221**: 32–49.

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

Edwards GE, **Ku MSB**. **1987**. Biochemistry of C_3–C_4 intermediates. In: **Hatch MD**, **Boardman NK**, eds. *The biochemistry of plants*, Vol. **10**. New York: Academic Press, 275–325.

Edwards EJ, **Osborne CP**, **Strömberg CA**, ***et al***.; **C_4 Grasses Consortium**. **2010**. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science* **328**: 587–591. doi:10.1126/science.1177216.

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

Ehleringer JR, **Monson RK. 1993**. Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics* **24**: 411–439.

Ehleringer JR, **Cerling TE**, **Helliker BR. 1997**. C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia* **112**: 285–299. doi:10.1007/s004420050311.

Ellis RP. 1974. Anomalous vascular bundle sheath structure in *Alloteropsis semialata* leaf blades. *Bothalia: African Biodiversity and Conservation* **11**: 273–275. doi:10.4102/abc.v11i3.1460.

Ellis RP. 1981. *Relevance of comparative leaf anatomy in taxonomic and functional research on the South African Poaceae*. DSc Thesis, University of Pretoria, South Africa.

Emms DM, **Kelly S. 2015**. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**: 157.

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

Frean M, **Barrett D**, **Cresswell CF. 1980**. Variability in leaf surface features and water efficiency utilisation in C_3 and C_4 forms of *Alloteropsis semialata* (R. Br.) hitchc. *Proceedings of the Annual Congresses of the Grassland Society of Southern Africa* **15**: 99–103. doi:10.1080/00725560.1980.9648893.

Frean ML, **Ariovich D**, **Cresswell CF. 1983***a*. C_3 and C_4 photosynthetic and anatomical forms of *Alloteropsis semialata* (R. Br.) Hitchcock: 2. A comparative investigation of leaf ultrastructure and distribution of chlorenchyma in the two forms. *Annals of Botany* **51**: 811–821. doi:10.1093/oxfordjournals.aob.a086532.

Frean ML, **Barrett DR**, **Ariovich D**, **Wolfson M**, **Cresswell CF. 1983***b*. Intraspecific variability in *Alloteropsis semialata* (R. Br.) Hitchc. *Bothalia: African Biodiversity and Conservation* **14**: 901–913. doi:10.4102/abc.v14i3/4.1260.

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

Ghannoum O. 2009. C_4 photosynthesis and water stress. *Annals of Botany* **103**: 635–644. doi:10.1093/aob/mcn093.

Ghannoum O, **Conroy JP**, **Driscoll SP**, **Paul MJ**, **Foyer CH**, **Lawlor DW. 2003**. Nonstomatal limitations are responsible for drought-induced photosynthetic inhibition in four C_4 grasses. *New Phytologist* **159**: 599–608. doi:10.1046/j.1469-8137.2003.00835.x.

Gibbs Russell GE. 1983. The taxonomic position of C_3 and C_4 *Alloteropsis semialata* (Poaceae) in southern Africa. *Bothalia: African Biodiversity & Conservation* **14**: 205–213.

Gowik U, **Westhoff P. 2011**. The path from C-3 to C-4 photosynthesis. *Plant Physiology* **155**: 56–63. doi:10.1104/pp.110.165308.

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

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

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

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

Heyduk K, **Moreno-Villena JJ**, **Gilman IS**, **Christin PA**, **Edwards EJ. 2019**. The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics* **20**: 485–493. doi:10.1038/s41576-019-0107-5.

Hipp AL, **Rothrock PE**, **Whitkus R**, **Weber JA. 2010**. Chromosomes tell half of the story: the correlation between karyotype rearrangements and genetic diversity in sedges, a group with holocentric chromosomes. *Molecular Ecology* **19**: 3124–3138. doi:10.1111/j.1365-294X.2010.04741.x.

Hoffmann AA, **Rieseberg LH. 2008**. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* **39**: 21–42. doi:10.1146/annurev.ecolsys.39.110707.173532.

Huang K, **Rieseberg LH. 2020**. Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science* **11**: 296. doi:10.3389/fpls.2020.00296.

Ibrahim DG, **Gilbert ME**, **Ripley BS**, **Osborne CP. 2008**. Seasonal differences in photosynthesis between the C_3 and C_4 subspecies of *Alloteropsis semialata* are offset by frost and drought. *Plant, Cell & Environment* **31**: 1038–1050. doi:10.1111/j.1365-3040.2008.01815.x.

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

Kadereit G, **Borsch T**, **Weising K**, **Freitag H. 2003**. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences* **164**: 959–986. doi:10.1086/378649.

Kadereit G, **Bohley K**, **Lauterbach M**, **Tefarikis DT**, **Kadereit JW. 2017**. C_3–C_4 intermediates may be of hybrid origin – a reminder. *New Phytologist* **215**: 70–76. doi:10.1111/nph.14567.

Kephart KD, **Buxton DR**, **Taylor SE. 1992**. Growth of C_3 and C_4 perennial grasses under reduced irradiance. *Crop Science* **32**: 1033–1038. doi:10.2135/cropsci1992.0011183x003200040040x.

Khoshravesh R, **Stata M**, **Busch FA**, ***et al***. **2020**. The evolutionary origin of C_4 photosynthesis in the grass subtribe Neurachninae. *Plant Physiology* **182**: 566–583.

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

Kumarathunge DP, **Medlyn BE**, **Drake, JE**, ***et al***. **2019**. Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. *New Phytologist* **222**: 768–784.

Langdale JA. 2011. C_4 cycles: past, present, and future research on C_4 photosynthesis. *The Plant Cell* **23**: 3879–3892. doi:10.1105/tpc.111.092098.

Lehmann CE, **Griffith DM**, **Simpson KJ**, ***et al***. **2019**. Functional diversification enabled grassy biomes to fill global climate space. *BioRxiv*. doi:10.1101/583625.

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

Long SP. **1999**. Environmental responses. In: **Sage RF**, **Monson RK**, eds. *C_4 Plant Biology*. San Diego: Academic Press, 215–249.

Lundgren MR, **Christin PA. 2017**. Despite phylogenetic effects, C_3–C_4 lineages bridge the ecological gap to C_4 photosynthesis. *Journal of Experimental Botany* **68**: 241–254. doi:10.1093/jxb/erw451.

Lundgren MR, **Besnard G**, **Ripley BS**, ***et al***. **2015**. Photosynthetic innovation broadens the niche within a single species. *Ecology Letters* **18**: 1021–1029. doi:10.1111/ele.12484.

Lundgren MR, **Christin PA**, **Escobar EG**, ***et al***. **2016**. Evolutionary implications of C_3–C_4 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell & Environment* **39**: 1874–1885. doi:10.1111/pce.12665.

Lundgren MR, **Dunning LT**, **Olofsson JK**, ***et al***. **2019**. C_4 anatomy can evolve via a single developmental change. *Ecology Letters* **22**: 302–312. doi:10.1111/ele.13191.

Lyu M-JA, **Gowik U**, **Kelly S**, ***et al***. **2021**. The coordination of major events in C_4 photosynthesis evolution in the genus *Flaveria*. *Scientific Reports* **11**: 15618.

Matsuoka M, **Furbank RT**, **Fukayama H**, **Miyao M. 2001**. Molecular engineering of C_4 photosynthesis. *Annual Review of Plant Biology* **52**: 297–314.

McKown AD, **Moncalvo J-M**, **Dengler NG. 2005**. Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *American Journal of Botany* **92**: 1911–1928.

Monson RK, **Jaeger CH. 1991**. Photosynthetic characteristics of C_3-C_4 intermediate *Flaveria floridana* (Asteraceae) in natural habitats: evidence of advantages to C_3-C_4 photosynthesis at high leaf temperatures. *American Journal of Botany* **78**: 795–800. doi:10.1002/j.1537-2197.1991.tb14481.x.

Monson RK, **Moore BD**, **Ku MSB**, **Edwards GE. 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**: 493–502. doi:10.1007/BF00392268.

Morales-Briones DF, **Kadereit G**. **2022**. Exploring the possible role of hybridization in the 1 evolution of photosynthetic pathways in *Flaveria* (Asteraceae), the prime model of C_4 photosynthesis evolution. *bioRxiv*. doi:10.1101/2022.01.31.478436.

Moreno-Villena JJ, **Dunning LT**, **Osborne CP**, **Christin PA. 2018**. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. *Molecular Biology and Evolution* **35**: 94–106. doi:10.1093/molbev/msx269.

Muhaidat R, **McKown AD. 2013**. Significant involvement of PEP-CK in carbon assimilation of C_4 eudicots. *Annals of Botany* **111**: 577–589. doi:10.1093/aob/mct017.

Muhaidat R, **Sage RF**, **Dengler NG. 2007**. Diversity of Kranz anatomy and biochemistry in C_4 eudicots. *American Journal of Botany* **94**: 362–381. doi:10.3732/ajb.94.3.362.

Olofsson JK, **Bianconi M**, **Besnard G**, ***et al***. **2016**. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology* **25**: 6107–6123. doi:10.1111/mec.13914.

Olofsson JK, **Dunning LT**, **Lundgren MR**, ***et al***. **2019**. Population-specific selection on standing variation generated by lateral gene transfers in a grass. *Current Biology* **29**: 3921–3927.e5. doi:10.1016/j.cub.2019.09.023.

Olofsson JK, **Curran EV**, **Nyirenda F**, ***et al***. **2021**. Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology* **30**: 2116–2130. doi:10.1111/mec.15871.

Osborne CP, **Wythe EJ**, **Ibrahim DG**, **Gilbert ME**, **Ripley BS. 2008**. Low temperature effects on leaf physiology and survivorship in the C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of Experimental Botany* **59**: 1743–1754. doi:10.1093/jxb/ern062.

Osmond, CB, **Winter, K**, **Ziegler, H. 1982**. Functional significance of different pathways of CO_2 fixation in photosynthesis. In: **Lange OL**, **Nobel PS**, **Osmond CB**, **Ziegler H**, eds. *Physiological Plant Ecology II. Encyclopedia of Plant Physiology,* **Vol. 12 B**. Berlin: Springer, 479–547.

Pagani M, **Zachos JC**, **Freeman KH**, **Tipple B**, **Bohaty S. 2005**. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**: 600–603. doi:10.1126/science.1110063.

Paterson AH, **Bowers JE**, **Chapman B. 2004**. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9903–9908.

Paterson AH, **Bowers JE**, **Bruggmann R**, ***et al***. **2009**. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551–556. doi:10.1038/nature07723.

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

Peterhansel C, **Maurino VG. 2011**. Photorespiration redesigned. *Plant Physiology* **155**: 49–55. doi:10.1104/pp.110.165019.

Phansopa C, **Dunning LT**, **Reid JD**, **Christin PA. 2020**. Lateral gene transfer acts as an evolutionary shortcut to efficient C_4 biochemistry. *Molecular Biology and Evolution* **37**: 3094–3104. doi:10.1093/molbev/msaa143.

Prendergast HDV, **Hattersley PW**, **Stone NE. 1987**. New structural/biochemical associations in leaf blades of C_4 grasses (Poaceae). *Functional Plant Biology* **14**: 403–420. doi:10.1071/pp9870403.

Raimondeau P, **Bianconi ME**, **Pereira L**, **Parisod C**, **Christin PA**, **Dunning LT**. **2022**. Variation in the rate lateral gene transfers accumulate in a grass lineage. *bioRxiv*, 2022-10.

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

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

Renvoize SA. 1987. A survey of leaf-blade anatomy in grasses XI. Paniceae. *Kew Bulletin* **42**: 739–768. doi:10.2307/4110087.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16**: 351–358. doi:10.1016/s0169-5347(01)02187-5.

Ripley BS, **Gilbert ME**, **Ibrahim DG**, **Osborne CP. 2007**. Drought constraints on C_4 photosynthesis: stomatal and metabolic limitations in C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of Experimental Botany* **58**: 1351–1363. doi:10.1093/jxb/erl302.

Ripley BS, **Abraham TI**, **Osborne CP. 2008**. Consequences of C_4 photosynthesis for the partitioning of growth: a test using C_3 and C_4 subspecies of *Alloteropsis semialata* under nitrogen-limitation. *Journal of Experimental Botany* **59**: 1705–1714. doi:10.1093/jxb/erm210.

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

Ripley B, **Donald G**, **Osborne CP**, **Abraham T**, **Martin T. 2010***b*. Experimental investigation of fire ecology in the C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of Ecology* **98**: 1196–1203. doi:10.1111/j.1365-2745.2010.01700.x.

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. doi:10.1111/gcb.12091.

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

Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytologist* **161**: 341–370.

Sage RF, **Christin PA**, **Edwards EJ**. **2011**. The C_4 plant lineages of planet Earth. *Journal of Experimental Botany* **62**: 3155–3169.

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**: 954–958. doi:10.1104/pp.84.3.954.

Sage RF, **Sage TL**, **Kocacinar F. 2012**. Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology* **63**: 19–47. doi:10.1146/annurev-arplant-042811-105511.

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**: 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**: 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**: 881–889. doi:10.1093/pcp/pcw009.

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**: 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**: 1647–1661. doi:10.1093/jxb/ern029.

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**: 6-1–6-14. doi:10.1029/2001gb001807.

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 of the United States of America* **103**: 7246–7251.

Tefarikis DT, **Morales-Briones DF**, **Yang Y**, **Edwards G**, **Kadereit G. 2022**. On the hybrid origin of the C_2 *Salsola divaricata* agg. (Amaranthaceae) from C_3 and C_4 parental lineages. *New Phytologist* **234**: 1876–1890. doi:10.1111/nph.18098.

Ueno O, **Sentoku N. 2006**. Comparison of leaf structure and photosynthetic characteristics of C_3 and C_4 *Alloteropsis semialata* subspecies. *Plant, Cell & Environment* **29**: 257–268. doi:10.1111/j.1365-3040.2005.01418.x.

Vogan PJ, **Sage RF. 2011**. Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of *Flaveria* Juss. (Asteraceae). *Plant, Cell & Environment* **34**: 1415–1430. doi:10.1111/j.1365-3040.2011.02340.x.

Voznesenskaya EV, **Koteyeva NK**, **Akhani H**, **Roalson EH**, **Edwards GE. 2013**. Structural and physiological analyses in Salsoleae (Chenopodiaceae) indicate multiple transitions among C_3, intermediate, and C_4 photosynthesis. *Journal of Experimental Botany* **64**: 3583–3604. doi:10.1093/jxb/ert191.

Williams BP, **Aubry S**, **Hibberd JM**. **2012**. Molecular evolution of genes recruited into C_4 photosynthesis. *Trends in Plant Science* **17**: 213–220.

Young JN, **Rickaby REM**, **Kapralov MV**, **Filatov DA. 2012**. Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 483–492. doi:10.1098/rstb.2011.0145.

Yuan Y, **Bayer PE**, **Batley J**, **Edwards D. 2021**. Current status of structural variation studies in plants. *Plant Biotechnology Journal* **19**: 2153–2163. doi:10.1111/pbi.13646.

Zhao H, **Wang Y**, **Lyu MJA**, **Zhu XG. 2022**. Two major metabolic factors for an efficient NADP-malic enzyme type C_4 photosynthesis. *Plant Physiology* **189**: 84–98. doi:10.1093/plphys/kiac051.

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