



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/239859/>

Version: Published Version

Article:

Pereira, L., Alenazi, A.S., Mian, S. et al. (2026) Gene turnover in the common ancestor of all C4 grasses. *Plants People Planet*. ISSN: 2572-2611

<https://doi.org/10.1002/ppp3.70206>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

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

RESEARCH ARTICLE

Gene turnover in the common ancestor of all C₄ grassesLara Pereira¹  | Ahmed S. Alenazi²  | Sahr Mian³  | Ilia J. Leitch³  |
Pascal-Antoine Christin¹  | Colin P. Osborne⁴  | Luke T. Dunning¹ ¹Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield, UK²Department of Biological Sciences, Northern Border University, Arar, Saudi Arabia³Trait Diversity and Function, Royal Botanic Gardens, Kew, Richmond, UK⁴Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, Western Bank, Sheffield, UK

Correspondence

Luke T. Dunning, Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

Email: l.dunning@sheffield.ac.uk

Funding information

Natural Environment Research Council, Grant/Award Numbers: NE/T011025/1, NE/V000012/1; Royal Society, Grant/Award Number: URF\R\180022; Deanship of Scientific Research at Northern Border University, Grant/Award Number: NBU-FFR-2026-630-01

Societal Impact Statement

Understanding how plants evolve more efficient photosynthesis is important in a warming world where improving crop productivity and resilience is a global priority. By generating the first reference genomes for an early-diverging group of grasses called the Aristidoideae, we were able to reconstruct the genetic makeup of the last common ancestor of all grasses that use C₄ photosynthesis that is more efficient in hot climates. We show that early genetic changes in this lineage may have laid the groundwork for the later evolution of C₄ photosynthesis. These findings help explain how complex biological innovations can repeatedly arise yet often remain restricted to particular evolutionary lineages.

Summary

- In grasses, almost all species belong to two evenly sized clades: BOP (Bambusoideae, Oryzoideae and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae). Yet the >20 independent origins of C₄ photosynthesis are confined to the PACMAD lineage. This pattern suggests that PACMAD grasses may harbour genetic features that primed them for the repeated evolution of this trait. To investigate this possibility, we examine gene turnover to identify genetic components that potentially facilitated the repeated evolution of C₄ photosynthesis and were present at the base of the PACMAD clade, representing the last common ancestor of all C₄ grasses.
- We generated the first reference genomes for Aristidoideae species (*Aristida adscensionis* and *Stipagrostis hirtigluma*), the sister lineage of all other PACMAD grasses. In combination with 34 other Poales genomes, we identify genes gained at the base of the PACMAD clade and genes lost in the sister BOP lineage.
- Gene expansion at the base of the PACMAD lineage includes one with a known C₄ function (β-carbonic anhydrase), as well as genes involved in amino acid and nitrate transport, carbon metabolism, oxidative stress management and transcription regulation.

Lara Pereira and Ahmed S. Alenazi contributed equally to the work.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

- Gene turnover in the PACMAD stem lineage generated novel variation in the last common ancestor of all C_4 grasses. Although this predates the emergence of C_4 photosynthesis itself, these changes may have generated the genetic substrate and requirements for the repeated emergence of C_4 photosynthesis in this clade.

KEYWORDS

C_4 photosynthesis, evolution, gene duplication, grasses, neofunctionalisation, phylogenomics

1 | INTRODUCTION

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

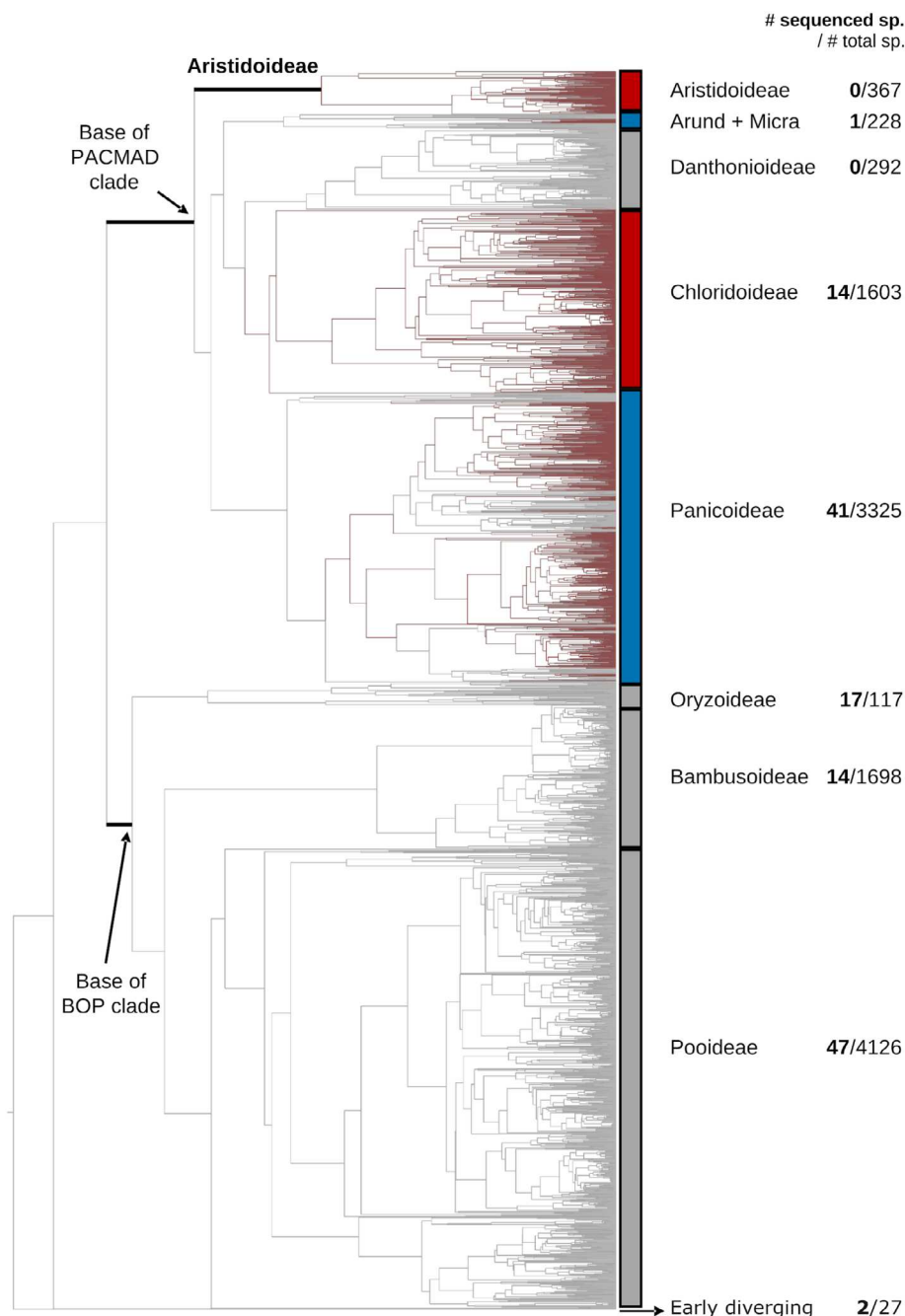
The C_4 photosynthetic pathway is a complex trait that increases the efficiency of carbon fixation in hot and high light environments when compared to the ancestral C_3 pathway (Atkinson et al., 2016; Ehleringer, 1978; Ehleringer et al., 1991, 1997; Ehleringer & Björkman, 1977; Sage et al., 2012). C_4 photosynthesis requires the coordinated action of numerous anatomical and biochemical components to concentrate CO_2 around the carbon-fixing enzyme Rubisco (Hatch, 1987). Gene duplication and subsequent neofunctionalization have previously been proposed as a mechanism for creating C_4 expression patterns (Monson, 2003). However, there is a lack of evidence for the immediate role of gene duplication in C_4 evolution (Gowik & Westhoff, 2011; van den Bergh et al., 2014; Williams et al., 2012), with the recruitment of core metabolic genes based on their ancestral expression patterns and catalytic properties (Aubry et al., 2011; Christin et al., 2015; Christin, Boxall, et al., 2013; Dunning et al., 2019; Emms et al., 2016; Hibberd & Covshoff, 2010; Moreno-Villena et al., 2018; Wang et al., 2009). Beyond the recruitment of core C_4 enzymes, whole genome duplication is thought to have enabled the evolution of C_4 anatomy and biochemistry in the eudicot *Gynandropsis gynandra* (Hoang et al., 2021), and a single photorespiratory gene duplication underpins the development of the photorespiratory CO_2 pump in the C_3 - C_4 intermediate eudicot *Flaveria* species (Schulze et al., 2013). In the grass *Alloteropsis semialata*, post co-option duplication and increased dosage seem to have played a role in reinforcing an initially weak C_4 cycle (Bianconi et al., 2018). However, distinguishing whether genetic changes enabled the emergence of the C_4 pathway, or

occurred as a consequence of its establishment, remains a major challenge (Dunning et al., 2017).

Despite its apparent complexity, C_4 photosynthesis has evolved over 60 times in angiosperms, representing a remarkable example of convergent evolution (Heyduk et al., 2019). The origins of C_4 photosynthesis are phylogenetically clustered (Sage et al., 2011), with 22–24 C_4 origins in grasses alone (Grass Phylogeny Working Group II, 2012), and modelling studies indicating a stepwise evolution from C_3 to a C_3 - C_4 intermediate (photorespiratory CO_2 pump, C_2 photosynthesis) and subsequently to full C_4 being the most likely path (Blätke & Bräutigam, 2019; Heckmann et al., 2013; Mallmann et al., 2014). Even within the grass family (Poaceae), the multiple origins of C_4 photosynthesis are clustered, only occurring in the PACMAD clade (six subfamilies: Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae and Danthonioideae), which is sister to the BOP clade (three subfamilies: Bambusoideae, Oryzoideae and Pooideae), where C_4 species are conspicuously absent (Grass Phylogeny Working Group II, 2012; Figure 1). These two clades contain over 99% of described grass species and are equally sized (BOP = 5941 species; PACMAD = 5815 species) (Soreng et al., 2022).

The absence of C_4 photosynthesis in BOP grasses suggests that species from the PACMAD clade carry precursors that make them prone to evolve C_4 (Figure 1), driven by either gene gains specific to the PACMAD clade or gene losses in the BOP lineage. A previous phylogenomic study using transcriptome data found no evidence for a whole genome duplication event at the base of the PACMAD clade (Zhang et al., 2024), indicating that small-scale duplications are more likely to be the source of these precursors. Relevant precursors could be associated with the biochemical pathway itself, related with leaf anatomy differences (e.g., proportion of the vascular bundle sheath layer) known to exist between the BOP and PACMAD clades (Christin et al., 2015; Christin, Osborne, et al., 2013), or even novel changes beyond previously hypothesised expectations. The genetic control of leaf development has only been partially elucidated (Lewis & Hake, 2016; Sedelnikova et al., 2018), preventing a comprehensive inspection of specific genes that might be responsible for these preadaptations. A global analysis using comparative genomics to identify gene families that have been duplicated and retained in two independent C_4 origins in a single grass subfamily (Panicoideae) unveiled 41 orthologous genes with heterogeneous functions, including transcription factors, transporters and genes involved in hormone metabolism (Emms et al., 2016), with these genes potentially acting as genetic precursors to C_4 evolution. Although there are high-quality reference genomes for more than

FIGURE 1 Phylogeny, C_4 origins and sequenced genomes. The dated phylogeny is from Spriggs et al. (2014), with C_3 species represented in grey and C_4 species in red. Bars for each clade indicate if they are exclusively C_3 (grey), C_4 (red) or mixed (blue). The number of species per subfamily was extracted from Soreng et al. (2022), and the number of sequenced genomes per subfamily was searched in NCBI (15/11/2024). Arund = Arundinoideae; BOP = Bambusoideae, Oryzoideae, and Pooideae; Micra = Micraioideae; PACMAD = Panicoideae, Arundinoideae, Chloridoideae, Micraioideae, Aristidoideae and Danthonioideae.



100 grass species (Figure 1), they are only available for one of the two groups resulting from the earliest split within the PACMAD lineage. Genomes for representatives of the Aristidoideae subfamily are lacking, preventing a global analysis of the specific precursors present in the last common ancestor of all C_4 origins in grasses.

In this study, we aimed to identify gene gains or losses that may represent preadaptations to C_4 photosynthesis present in the last common ancestor of the PACMAD clade. To achieve this, we generated reference genomes for two members of the Aristidoideae family, *Aristida adscensionis* and *Stipagrostis hirtigluma*. We then used these genomes, together with publicly available assemblies for other grass species, to identify duplication and de novo gene births at the base of the

PACMAD clade and gene losses at the base of the BOP lineage. We found that the duplicated genes were enriched for transmembrane transporters and genes associated with resistance to oxidative damage.

2 | MATERIALS AND METHODS

2.1 | Tissue collection and DNA extractions

A. adscensionis seeds were obtained from the USDA Germplasm Resources Information Network (GRIN) (Accession Number = PI 269867; collection location = Pakistan) and *S. hirtigluma* var. *patula*

seeds from the Royal Botanic Gardens, Kew Millennium Seed Bank (Accession Number = 74902; collection location = Botswana). Seeds were germinated and grown in a glasshouse at the Arthur Willis Environment Centre, University of Sheffield (United Kingdom) under semi-controlled conditions (12 h daylight, 25/20°C day/night temperature). The 1C genome size for each species was estimated by flow cytometry using the one-step protocol (Doležel et al., 2007) with minor modifications (see Clark et al., 2016). *Petroselinum crispum* 'Champion Moss Curled' (2C = 4.50 pg; Obermayer et al., 2002) was used as the calibration standard and nuclei were isolated in the General Purpose Buffer (GPB; Loureiro et al., 2007), supplemented with 3% PVP and 0.08% (v/v) beta-mercaptoethanol.

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

2.2 | Organelle assembly

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

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

contigs with GeSeq. Only contigs containing at least one gene model were included in the draft mitochondrial contigs.

2.3 | Nuclear assembly

The estimated coverage and ploidy for each species were calculated using the k-mer spectrum generated with GenomeScope 2.0 (Ranallo-Benavidez et al., 2020). Prior to nuclear genome assembly, we removed reads belonging to the organelles by mapping raw HiFi reads to the plastid and mitochondrial genomes with minimap2 v2-2.24 (Li, 2018), discarding those where ≥80% of the read aligned. The organelle-filtered HiFi reads were then de novo assembled using hifiasm, using default parameters except for the number of haplotypes in *S. hirtigluma* being set to four as it was inferred to be tetraploid (see Section 3.1). The assembly completeness was checked using BUSCO v5.4.2 (Simão et al., 2015) with the poales_odb10 database, and Inspector v1.2 (Chen et al., 2021) was used to estimate assembly quality values (QV).

2.4 | Genome annotation

Extensive de novo TE Annotator v1.9.4 (EDTA) (Ou et al., 2019) was used to annotate transposable elements (TEs) and RepeatMasker v4.1.2-p1 to annotate simple tandem repeats and low complexity (polypurine, AT-rich) regions, with default parameters. The identified TEs and simple repeats were then soft-masked in each draft genome using BEDtools maskfasta (Quinlan & Hall, 2010). To annotate the protein coding genes in the soft-masked reference genome, we used Helixer v0.3.2, a tool that uses a deep learning model to annotate large eukaryotic genomes (Holst et al., 2023; Stiehler et al., 2020). Helixer outperforms most other de novo annotation tools despite only using the reference genome (i.e., no RNA or protein data required) (Holst et al., 2023), a conclusion further supported by our re-annotation of publicly available genomes used in this study (see below) having higher BUSCO scores with Helixer (Table S1).

2.5 | Identifying orthologous genes and duplication events

We used OrthoFinder v2.5.4 (Emms & Kelly, 2019) to identify groups of orthologs and duplication events in grasses. We included the 2 reference genomes produced here in addition to 32 published genomes from across the Poales (15 × *C₄* PACMAD, 4 × *C₃* PACMAD, 9 × BOP, 2 × early diverging grasses and 4 × non-grass Poales) with the aim of representing as many subfamilies and *C₄* origins as possible (Figure 2 and Table S2). The 15 *C₄* genomes represent 8 independent *C₄* origins, out of the 22–24 proposed by the Grass Phylogeny Working Group II (2012), with the remaining *C₄* lineages missing due to a lack of high-quality reference genomes. The unsampled *C₄* origins are distributed across three subfamilies: Micrairoideae (one origin), Chloridoideae (Centropodia [one origin]) and Panicoideae (Paspaleae [nine

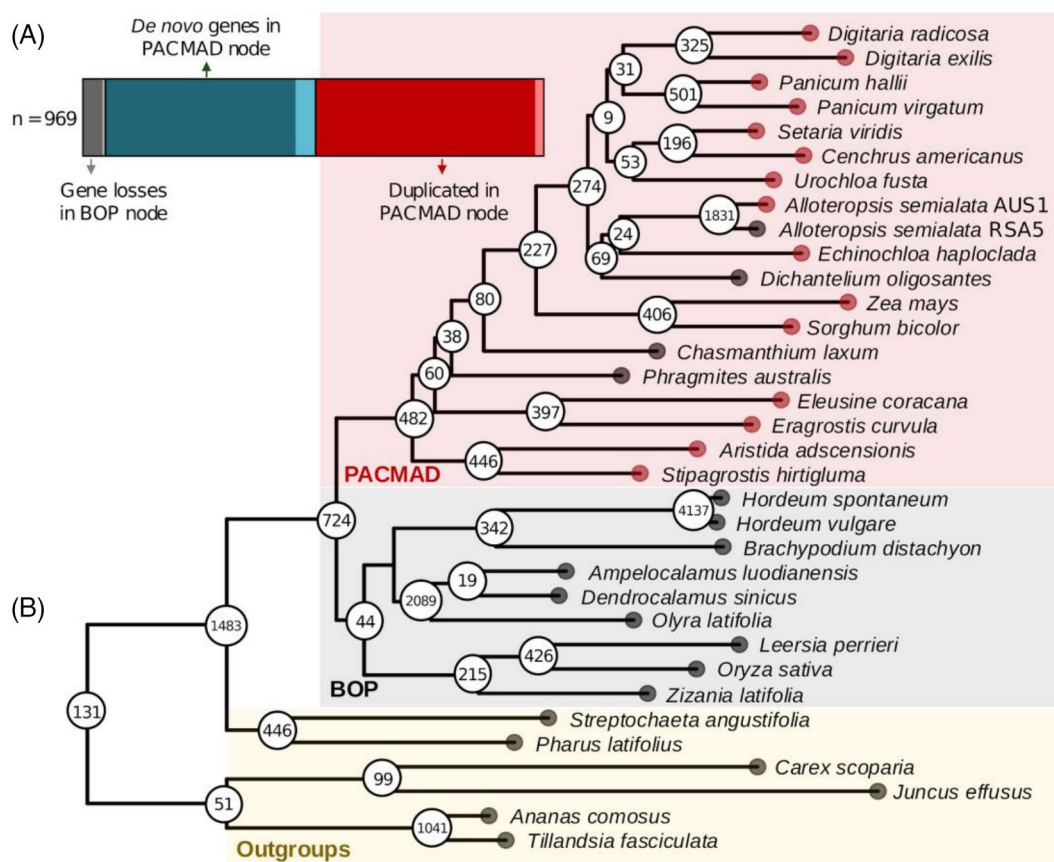


FIGURE 2 Candidate genetic precursors of C₄ photosynthesis in grasses. (a) The proportion of candidate genes arising through gene duplication, gene loss and de novo origin. The lighter portion of each group represents the core set for each group; that is, present in all 15 C₄ PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) species used in this study. (b) Species tree and number of duplications per node generated by OrthoFinder given in the circles. C₄ species represented by a red circle at the tip and C₃ by a black circle. BOP = Bambusoideae, Oryzoideae and Pooideae.

origins], Tristachyideae [one origin], Neurachninae [one origin] and *Alloteropsis* [two origins]) (GPWG II, 2012). To avoid overestimating the importance of any single C₄ origin, we included a maximum of two representative species per C₄ origin, except for the large MPC clade (Melinidinae + Panicinae + Cenchrinae). Although this clade was initially proposed to represent a single C₄ origin (GPWG II, 2012), it is now thought to comprise several closely related but independent C₄ origins (Washburn et al., 2015). Accordingly, we sampled two genomes from each of the three subtribes, taking the number of independent C₄ origins sampled in this study to 10.

To reduce the effect of annotation quality on our interpretations, we re-annotated the published genomes with Helixer when the extracted coding sequences (CDS) were <90% complete according to BUSCO analysis (Table S2). Poaceae genomes had to have ≥90% complete BUSCO annotation scores to be included in the final analysis (Table S2); an exception was made for one of the *Alloteropsis semialata* genomes (>89%), as there are relatively few C₃ PACMAD species with genomes available, and this species is a model for C₄ evolution in grasses (Pereira et al., 2023). For Poales outgroups, the completeness threshold was lowered (*Carex scoparia* [Cyperaceae] 79.8% complete BUSCO; *Juncus effusus* [Juncaceae] 78.1% complete BUSCO) as it has been

previously suggested that their relatively low BUSCO scores are a result of only Poaceae and Bromeliaceae species being used in the construction of the Poales BUSCO gene database, meaning these genes are not necessarily conserved across the whole order (Planta et al., 2022).

Orthologs across all 36 genomes were identified with OrthoFinder using the CDS, aligned using MUSCLE v3.2 (Edgar, 2004), and maximum likelihood trees inferred using FastTree v2 (Price et al., 2010). OrthoFinder generates a species tree inferred from all orthogroups using the 'Species Tree Inference from All Genes' (STAG) method (Emms & Kelly, 2018), rooted using 'Species Tree Root Inference from Gene Duplication Events' (STRIDE) (Emms & Kelly, 2017). OrthoFinder then reconciles gene duplications against the inferred STAG species tree to determine the branch on which the duplication occurred, with a minimum support value of ≥50% (i.e., ≥50% of species that share a common ancestor on that branch have retained both copies of the duplicated gene). Of particular interest for this study are the orthogroups inferred to have been duplicated at the base of the PACMAD clade, the branch representing the last common ancestor of all grass C₄ origins. To determine if the number of duplications along the PACMAD branch was higher than expected, we calculated the rate of duplication across the STAG tree using divergence times for

homologous branches extracted from a previously dated phylogeny (Christin et al., 2014).

The PACMAD specific duplicates were further filtered to a core set; that is, they had to be [1] present in all 15 C_4 species included in the analysis and [2] verified in phylogenetic trees constructed independently of OrthoFinder. To infer individual gene trees, we first identified homologous sequences in the 36 species with BLASTn v2.16.0 using all coding sequences from the orthogroup of interest as a query. Blast matches had to be ≥ 300 bp long, and we only retained the longest match for each sequence in the database. The BLASTn matching regions were then aligned with mafft v7.525 (Katoh & Standley, 2013), set to automatically choose the best alignment algorithm (–auto). A phylogenetic tree was then inferred with PhyML v2.1031022 (Guindon et al., 2010), with the best substitution model identified using Smart Model Selection (SMS) v.1.8.1 (Lefort et al., 2017).

Finally, for the verified set of PACMAD duplicates, we tested if this resulted in a significantly increased gene copy number at the orthogroup level. For each orthogroup, the gene copy number was normalised by ploidy level, and it was determined if the data was normally distributed using the Shapiro–Wilk test ($p < .05$). Differences in gene copy number between BOP and PACMAD species were assessed using the Wilcoxon signed-rank test (non-normal data), with Benjamini–Hochberg correction of multiple testing (significance at adjusted $p < .05$). All analyses were done in R.

2.6 | Gene losses in the BOP clade and de novo genes in the PACMAD clade

Orthogroups lost in the BOP clade were identified as those present in at least one of the six outgroup species and in $\geq 50\%$ of PACMAD species (including at least one representative from each of the Panicoideae, Aristidoideae and Chloridoideae subfamilies) but absent in all nine BOP species analysed. PACMAD-specific de novo orthogroups were identified as those absent in the outgroup and BOP species and present in $\geq 50\%$ of PACMAD species (including at least one representative from each of the Panicoideae, Aristidoideae and Chloridoideae subfamilies). Both orthogroups lost in the BOP clade and PACMAD-specific orthogroups were further filtered to a core set (i.e., present in all 15 C_4 species included in the analysis).

2.7 | Gene ontology

To determine if certain Gene Ontology (GO) terms (Ashburner et al., 2000) are overrepresented in the genes that are either duplicated at the base of the PACMAD clade, lost in the BOP clade, or are PACMAD-specific de novo genes, we used the statistical over-representation test in the Panther online tool v.19 (<https://pantherdb.org/>) (Mi et al., 2012; Thomas et al., 2021). Maize was used as the background reference genome for the analysis, as it is the best characterised PACMAD species, and the target lists used either [1] all maize loci in the orthogroups of interest (BOP losses and de novo PACMAD) or [2] one randomly selected duplicate from each of the PACMAD specific

duplication events. We only used one of the duplicates for the second analysis to test for enrichment compared to the ‘ancestral’ genome prior to this duplication, as each duplication only represents a net increase of one gene in the genome. A Fisher's exact test was used to determine which PANTHER GO-Slim terms were significantly enriched ($FDR < 0.05$) in the target lists. The results were further restricted to only include those GO terms that applied to 10–500 genes in the maize genome, to exclude overly general categories and those with low statistical support. Finally, QuickGO (<https://www.ebi.ac.uk/QuickGO/>) was used to generate a hierarchical representation of the significant GO terms.

The function of the PACMAD-specific gene complement (i.e., PACMAD-specific duplicates, PACMAD de novo genes, and BOP losses) was further investigated using BLASTp v2.16.0 against the curated SwissProt database, a high-quality database of protein sequences with extensive, expert-reviewed annotations. A randomly selected coding sequence from each orthogroup (or duplicate) was translated with EMBOSS Transeq v.6.6.0 (Rice et al., 2000) and used as a protein query, with `-max_target_seqs` set to 3 and an E-value cut-off of 1×10^{-5} . We further supported the function of the core sets of PACMAD duplicates, BOP losses and de novo PACMAD genes using Ensembl Plants (Howe et al., 2020), the maize database MaizeGBD (Woodhouse et al., 2021), the *Arabidopsis thaliana* database TAIR (Berardini et al., 2015) and literature searches.

2.8 | Tissue-specific C_4 expression patterns

We used a chi-squared test to determine if genes resulting from PACMAD-specific duplications are more likely to be differentially expressed between mesophyll (M) and bundle sheath (BS) cells than the genome-wide proportion. To do this, we cross-referenced our candidate gene list with the results from previous studies in maize (Li et al., 2010) and sorghum (Döring et al., 2016) that identified genes significantly differentially expressed between M and BS cells. Because these previous studies used legacy versions of maize annotation, if the gene models for our candidate loci had changed in the current maize annotation (only happened in seven maize genes; for details, see Table S3), we applied the same expression pattern to the fused or split gene models used in our analysis. We also included any genes arising from further duplications of these loci that may have occurred after PACMAD lineages began diverging. To infer whether some of these candidates changed their pattern of expression after duplication, we also compared the expression (M vs. BS) of the orthologous genes in rice using an available transcriptomics dataset (Hua et al., 2021).

3 | RESULTS

3.1 | Reference genomes for two Aristidoideae grasses

We assembled reference genomes for two divergent Aristidoideae species, *A. adscensionis* and *S. hirtigluma*, representing independent realisations of the C_4 phenotype. For *A. adscensionis*, we generated

TABLE 1 Assembly statistics for the two Aristidoideae genomes.

		<i>Aristida adscensionis</i>	<i>Stipagrostis hirtigluma</i>
Sequencing	Number of reads	1,889,280	4,507,740
	Average read length	16.39 kb	15.01 kb
	Sequence data volume	30.97 Gb	67.68 Gb
Assembly	Ploidy	Diploid	Tetraploid
	Estimated 1C genome size	0.58 Gb	0.48 Gb ^a
	Total assembly length	0.45 Gb	0.61 Gb
	Number of contigs	186	458
	Contig N50	5.6 Mb	12.14 Mb
	Longest contig	17.85 Mb	31 Mb
	Assembly quality value	63.55	25.06
	BUSCO complete (duplicated)	98.1% (2.4%)	98.7% (35.1%)
	BUSCO fragmented	0.9%	0.8%
Annotation	Transposable elements	53.0%	48.7%
	Simple tandem repeats	0.81%	0.72%
	Low complexity regions	0.08%	0.07%
	Number of gene models	32,006	50,517

^aMonoploid genome size (1Cx).

31.0 Gb of HiFi data, which is approximately 53× coverage of the nuclear genome based on the flow cytometry estimate for 1C of 0.58 Gb (Table 1). The analysis of the k-mer spectrum suggested that the *A. adscensionis* individual sequenced was diploid (Figure S1), and the final assembly was 0.45 Gb (77.8% of the genome size estimated using flow cytometry), relatively contiguous (186 contigs and N50 = 5.6 Mb), and 98.1% of BUSCOs were complete (Table 1). Genome annotation showed that 53.0% of the genome is made up of TEs, 0.81% simple tandem repeats and 0.08% low complexity regions, and we were able to identify 32,006 protein coding genes (Table 1).

We generated 67.7 Gb of HiFi data for *S. hirtigluma* (Table 1). The analysis of the k-mer spectrum suggested it was a tetraploid (unknown if allo- or auto-polyploid) (Figure S2); therefore, the monoploid genome size (i.e., the 1Cx value corresponding to the amount of DNA in one chromosome set calculated from the genome size measured by flow cytometry) was estimated to be 0.48 Gb (corresponding to 141× coverage). The final assembly was 0.61 Gb, which is higher than the monoploid flow cytometry estimate, possibly because divergent sequences from different subgenomes are assembled separately and concatenated in the final assembly. The assembly is more fragmented (458 contigs, N50 = 12.1 Mb) than the *A. adscensionis* assembly, despite the higher coverage and larger N50 obtained. This is most likely due to the added complexity of assembling a polyploid species. In terms of completeness, 98.7% of BUSCOs were complete (Table 1), and 35.1% were duplicated, further supporting the partial redundancy of the assembly. Genome annotation showed that 48.72% of the genome is made up of TEs, 0.72% simple tandem repeats, 0.07% low complexity regions, and we were able to identify 50,517 protein coding genes (Table 1).

3.2 | Gene turnover in the PACMAD stem lineage

The PACMAD clade has evolved C_4 photosynthesis 22–24 times (Figure 1), whereas the sister BOP clade has not. This evolutionary pattern is potentially driven by genetic enablers that are present in the last common ancestor of the PACMAD grasses, facilitating the recurrent emergence of the photorespiratory pump and/or C_4 in this clade. We identified groups of orthologs and duplication events across the grass phylogeny with OrthoFinder v.2.5.4 (Emms & Kelly, 2019), which reconciles gene duplications against the species tree to determine the branch on which the duplication occurred. Recent duplications within a genome (including those due to polyploidy or mis-assembly) would be resolved to have occurred at the tip of the phylogeny and would not affect our results. We then selected the genes absent in the BOP clade that evolved prior to PACMAD divergence (Figure S3). In total, we identified 969 gene turnover events (Figure 2a), including 482 genes duplicated in the last common ancestor of PACMAD (Table S4), 440 orthogroups that formed de novo in the last common ancestor of PACMAD (Table S5) and 47 orthogroups that were lost in the last common ancestor of BOP (Table S6). The number of predicted duplications varied across the phylogeny (Figure 2b and Table S7), with a mean rate of duplication for each branch being 24.7 duplications per Ma (SD = 18.4). The branch leading to the PACMAD clade does not have a significantly higher duplication rate than the rest of the phylogeny (38.8 duplications per Ma; 79th percentile), ruling out a burst of gene duplications at the base of the PACMAD clade.

Significant enrichment of GO terms was detected for 23 categories in the genes duplicated at the base of the PACMAD clade, a single GO term in the de novo PACMAD genes (GO:0008171—O-methyltransferase activity) and none for the BOP losses. The

TABLE 2 Gene Ontology (GO) terms significantly overrepresented in the PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) specific gene duplicates.

GO term	Description	Class	#Ref	#Dup	FE	Padj
GO terms associated with transporter activity						
GO:0008028	Monocarboxylic acid transmembrane transporter activity	MF	18	3	16.46	1.47E−02
GO:0015179	L-amino acid transmembrane transporter activity	MF	22	3	13.47	2.49E−02
GO:0015171	Amino acid transmembrane transporter activity	MF	89	7	7.77	1.42E−03
GO:0019829	ATPase-coupled monoatomic cation transmembrane transporter activity	MF	54	4	7.32	3.70E−02
GO:0046943	Carboxylic acid transmembrane transporter activity	MF	109	7	6.34	3.27E−03
GO:0005342	Organic acid transmembrane transporter activity	MF	109	7	6.34	3.11E−03
GO:0042626	ATPase-coupled transmembrane transporter activity	MF	133	7	5.20	9.58E−03
GO:0008514	Organic anion transmembrane transporter activity	MF	142	7	4.87	1.30E−02
GO:0015399	Primary active transmembrane transporter activity	MF	151	7	4.58	1.72E−02
GO:0006865	Amino acid transport	BP	86	7	8.04	3.84E−02
GO:0046942	Carboxylic acid transport	BP	107	7	6.46	3.89E−02
GO:0015849	Organic acid transport	BP	107	7	6.46	3.11E−02
GO terms associated with catalytic activity (including response to oxidative stress)						
GO:0004364	Glutathione transferase activity	MF	72	6	8.23	2.50E−03
GO:0035251	UDP-glucosyltransferase activity	MF	121	9	7.35	2.08E−04
GO:0016616	Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	MF	192	12	6.17	6.91E−05
GO:0008194	UDP-glycosyltransferase activity	MF	149	9	5.97	1.02E−03
GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	MF	199	12	5.96	7.21E−05
GO:0046527	Glucosyltransferase activity	MF	158	9	5.63	1.05E−03
GO:0016765	Transferase activity, transferring alkyl or aryl (other than methyl) groups	MF	151	8	5.23	3.78E−03
GO:0016758	Hexosyltransferase activity	MF	274	10	3.61	1.13E−02
GO:0016757	Glycosyltransferase activity	MF	381	11	2.85	3.34E−02
GO:0008171	O-methyltransferase activity	MF	35	5	13.47	4.00E−03
GO:0006749	Glutathione metabolic process	BP	66	6	8.98	3.88E−02
GO:0006790	Sulphur compound metabolic process	BP	180	9	4.94	4.56E−02

Abbreviations: #Dup = number of duplicated genes; #Ref = number of genes in reference list; FE = fold enrichment compared to the expected value; Padj = Benjamini–Hochberg adjusted *p* value.

23 significantly enriched GO terms in the PACMAD duplicates were divided into five groups with functions clustering around amino acid transport (GO:0015171—amino acid transport), transport against gradients (GO:0015399—primary active transmembrane transporter activity), oxidoreductases (GO:0016616—oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor), glycosyltransferases (GO:0016757—glycosyltransferase activity) and glutathione transferases (GO:0004364—glutathione transferase activity) (Table 2 and Figure S4), with the specific genes comprising each enriched orthogroup listed in Table S4.

3.3 | PACMAD gene turnover events retained in all *C*₄ species

Among the gene turnover events in the PACMAD stem lineage (PACMAD duplicates, PACMAD de novo genes and BOP losses),

those consistently retained across all *C*₄ species are more likely to have contributed to the repeated evolution of *C*₄ photosynthesis in grasses. In total, both copies resulting from 3.7% (*n* = 18) of PACMAD gene duplications, 9.5% (*n* = 42) of PACMAD de novo orthogroups and 12.7% (*n* = 6) of BOP orthogroup losses were retained in all 15 *C*₄ species used in this study (Table S8).

The core set of 18 PACMAD duplicated and *C*₄ retained candidates was further validated using additional gene trees (Figure S5 and Dataset S1), with 15 duplication events validated and 3 unconfirmed: (i) OG0003655 (Non-specific lipid transfer protein GPI-anchored 16) contained too few sequences (*n* = 26), all from PACMAD species; (ii) OG0001524 (Small ribosomal subunit protein uS15) topology was inconsistent with the species tree, with very low support; and (iii) OG0002361 (Rho GTPase-activating protein REN1) duplication possibly occurred prior to the BOP and PACMAD clades diverging (Dataset S1). Finally, we looked at the overall gene copy number in the orthogroups that contained the 15 core PACMAD-specific duplication

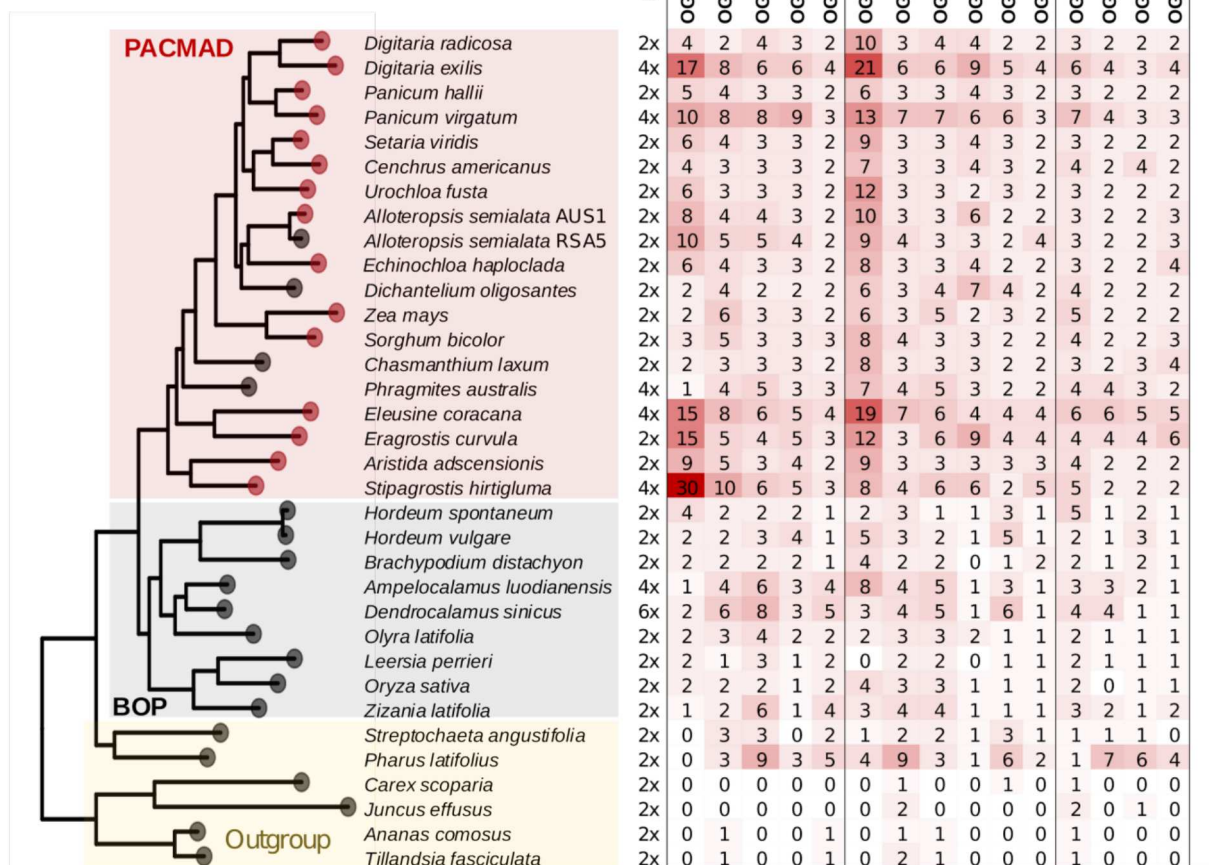


FIGURE 3 Gene copy number for each of the orthogroups that contain one of the 15 validated duplication events that occurred at the base of the PACMAD (Panicoidae, Arundinoideae, Chloridoideae, Micraioideae, Aristidoideae and Danthonioideae) lineage and subsequently retained in all C_4 species (these orthogroups may contain additional duplications and/or losses). Ploidy is also indicated, and the asterisks denote orthogroups, which have a significantly higher gene copy number in the PACMAD than BOP (Bambusoideae, Oryzoideae and Pooideae) clade (adjusted p value < .05). Additional information about the functional annotation of these orthogroups is shown in Table S8.

events (Figure 3). There was a lot of variation between species which presumably results from other independent duplications and/or losses (Figure S6). However, there was a significantly higher (adjusted p value < .05) gene copy number in PACMAD compared to BOP in 11 of these orthogroups (Table S9 and Figure S6). In C_3 PACMAD species, copy numbers seem to follow patterns of retention and expansion more similar to C_4 PACMAD than to BOP species (Figure 3).

A majority of the 15 core PACMAD duplicated genes (retained in all C_4 lineages and validated) are involved with transport, carbon metabolism or reducing the effects of oxidative damage (Figure 3 and Table S8). The five transporters move amino acids (OG0002111 [cationic amino acid transporter 6, chloroplastic] and OG0000861 [organic cation/carnitine transporter 7]), nitrates (OG0000391 [protein NRT1/ PTR FAMILY 4.4] and OG0000379 [protein NRT1/PTR FAMILY 3.1]), and toxic compounds (OG0000180 [Protein DETOXIFICATION 16]). The three carbon metabolism enzymes are phosphoglycerate mutase-like (PGML) protein (OG0001190)—glycolysis/ gluconeogenesis and Calvin-Benson-Bassham cycle; (DL)-glycerol-

3-phosphatase (OG0003194)—glycolysis and potentially photorespiration; and a chloroplastic aspartyl proteinase NANA protein (OG0000505) that potentially plays a role in photoassimilate partitioning and communication between chloroplast and nucleus (Paparelli et al., 2012). The three enzymes with functions relating to reducing oxidative damage are a peroxidase 2 (OG0000809) linked to lignin biosynthesis (Fernández-Pérez et al., 2015); a salutaridine reductase (SALR) (OG0000101) that is orthologous to an Arabidopsis NAD(P)-binding Rossmann-fold superfamily protein involved in detoxification and responding to reactive carbonyl species (Ksas et al., 2024); and a phosphoethanolamine N-methyltransferase 1 (PEAM1) involved in phosphatidylcholine biosynthesis (OG0000461) (Liu et al., 2019). The remaining four genes have diverse functions relating to DNA repair (OG0002653 [crossover junction endonuclease MUS81]), telomere protection (OG0002585 [single myb histone 6]), signalling (OG0002789 [influenza vs. NS1A-binding protein]) and a mildew locus O (MLO)-like protein 1 (OG0000548) linked to abiotic stress tolerance (Acevedo-Garcia et al., 2014).

Given their reduced taxonomic distribution, it is not surprising that many of the core de novo PACMAD genes are uncharacterised proteins (31%) or lack meaningful descriptions (Table S8). The annotated genes encode a number of transcription factors including: NAC domain-containing protein (OG0004173), homeodomain-like superfamily protein (OG0005708), ethylene-responsive transcription factors (OG0009413), two-component response regulator (OG0016229), LNK1 protein (OG0018191), FRIGIDA-like protein (OG0018794), a WRKY domain-containing protein (OG0020072) and an orphan transcription factor (OG0019167). There are also a few core de novo transporters that move lipids (OG0019414 [non-specific lipid transfer protein-like 1]) and small solutes across cell membranes in response to chemiosmotic gradients (OG0020033 [major facilitator superfamily (MFS) profile domain-containing protein]) (Table S8). There were only six core BOP orthogroup losses retained in all 15 C_4 species, five with annotations (Table S8). These genes encoded a protein-serine/threonine phosphatase (OG0014342), lipid transporter (OG0019164 [protein FLOURY 1]), microtubule regulatory protein (OG0018324 [protein WVD2-like 3]), cupredoxin superfamily protein that catalyses inter-molecular electron transfer reactions (OG0018546) and a membrane attack complex/perforin (MACPF) pathogen resistance protein (OG0017688).

3.4 | Little turnover of known C_4 genes

Among the key C_4 metabolic enzymes, only a gene encoding β -carbonic anhydrase (OG0001853) was duplicated in the PACMAD grasses. β -carbonic anhydrase is the entry step of the C_4 cycle where it is responsible for the reversible conversion of CO_2 into bicarbonate. The duplication corresponds to the specific gene lineage co-opted for C_4 photosynthesis in grasses (β -carbonic anhydrase 2P3; Moreno-Villena et al., 2018), and although it was duplicated at the base of the PACMAD clade as previously hypothesised (Besnard et al., 2018), both duplicates were only retained in 13 of the 15 C_4 species analysed (Figure 4).

Another potential link to the evolution of C_4 photosynthesis involves enzymes that are part of the photorespiration cycle, which is thought to be a precursor to evolving a C_4 physiology (Bräutigam & Gowik, 2016). A gene encoding (DL)-glycerol-3-phosphatase (OG0003194) was duplicated at the base of PACMAD, with both copies retained in all C_4 species used in this study. This enzyme is classically associated with glycerol metabolism and lipid biosynthesis, including the provision of carbon skeletons for triacylglycerol synthesis. However, it has been considered a putative 2-phosphoglycolate phosphatase (PGLP) based on protein sequence homology to prokaryotic PGLPs, and therefore, it may play a role in catalysing the first step of photorespiration, although this has not been functionally verified (Schwarte & Bauwe, 2007).

Finally, two de novo PACMAD genes are annotated as Scarecrow-like proteins (OG0002850 and OG0017141; Table S5), although these only remain present in 10 of the 15 C_4 species analysed. Scarecrow-like proteins belong to the GRAS gene family, a

group of transcription factors shown to play a key role in C_4 leaf anatomy and photosynthetic development in maize (Hughes & Langdale, 2020; Slewinski et al., 2012).

3.5 | Differential gene expression in mesophyll and bundle sheath cells

The C_4 cycle involves the recruitment and compartmentalisation of multiple enzymes across mesophyll (M) and bundle sheath (BS) cells. Gene duplication may facilitate the substrate for cell-specific subfunctionalisation, enabling loci to separate their role in C_4 photosynthesis from other functions that they may perform. We used published transcriptomic data in grasses to test whether subfunctionalisation happened more commonly than expected in our PACMAD-specific duplicates. In maize, 10.6% of the 32,540 gene models were significantly differentially expressed between M and BS cells (Li et al., 2010), whereas in sorghum, it was 4.9% of the 34,211 genes (Döring et al., 2016). There was no significant difference in the proportion of genes differentially expressed between M and BS in our list of PACMAD-specific duplicates compared to the genome-wide levels in sorghum (1035 duplicates, 54 [5.2%] differentially expressed, p value = .789) (Table S10) and a significant decrease in our maize candidates (917 duplicates, 66 [7.2%] differentially expressed, p value = .003) (Table S3). There was also no significant difference in the proportion of genes differentially expressed between M and BS when only considering the core PACMAD duplicates retained in all C_4 species compared to the genome-wide proportion in both sorghum (45 duplicates, 4 [8.9%] differentially expressed, p value = .261) and maize (40 duplicates, 5 [12.5%] differentially expressed, p value = .890).

There were 14 PACMAD-specific duplicates that contained at least one ortholog that was significantly differentially expressed between M and BS cells in both maize and sorghum, 4 of which were core candidates retained in all C_4 lineages: PEAM1 (OG0000461), MLO1 (OG0000548), PGML (OG0001190) and SALR (OG0000101) (Table 3). Over half ($n = 8$) of the differentially expressed gene duplicates shared between maize and sorghum were not differentially expressed in rice (Tables 3 and S11). The remaining six were differentially expressed in all three species, two of which were more highly expressed in the same tissue (including β -carbonic anhydrase [OG0001853]) and two with opposite tissue-specific expression patterns between rice and maize + sorghum (including PEAM1 [OG0000461] and PGML [OG0001190]).

4 | DISCUSSION

C_4 photosynthesis is a remarkable example of convergent evolution, yet the numerous C_4 origins are clustered in a relatively small number of major families and infrafamily groups (Christin et al., 2015). In grasses, 99% of species are found in 2 equally sized clades (BOP = 5941 species; PACMAD = 5815 species; Soreng et al., 2022),

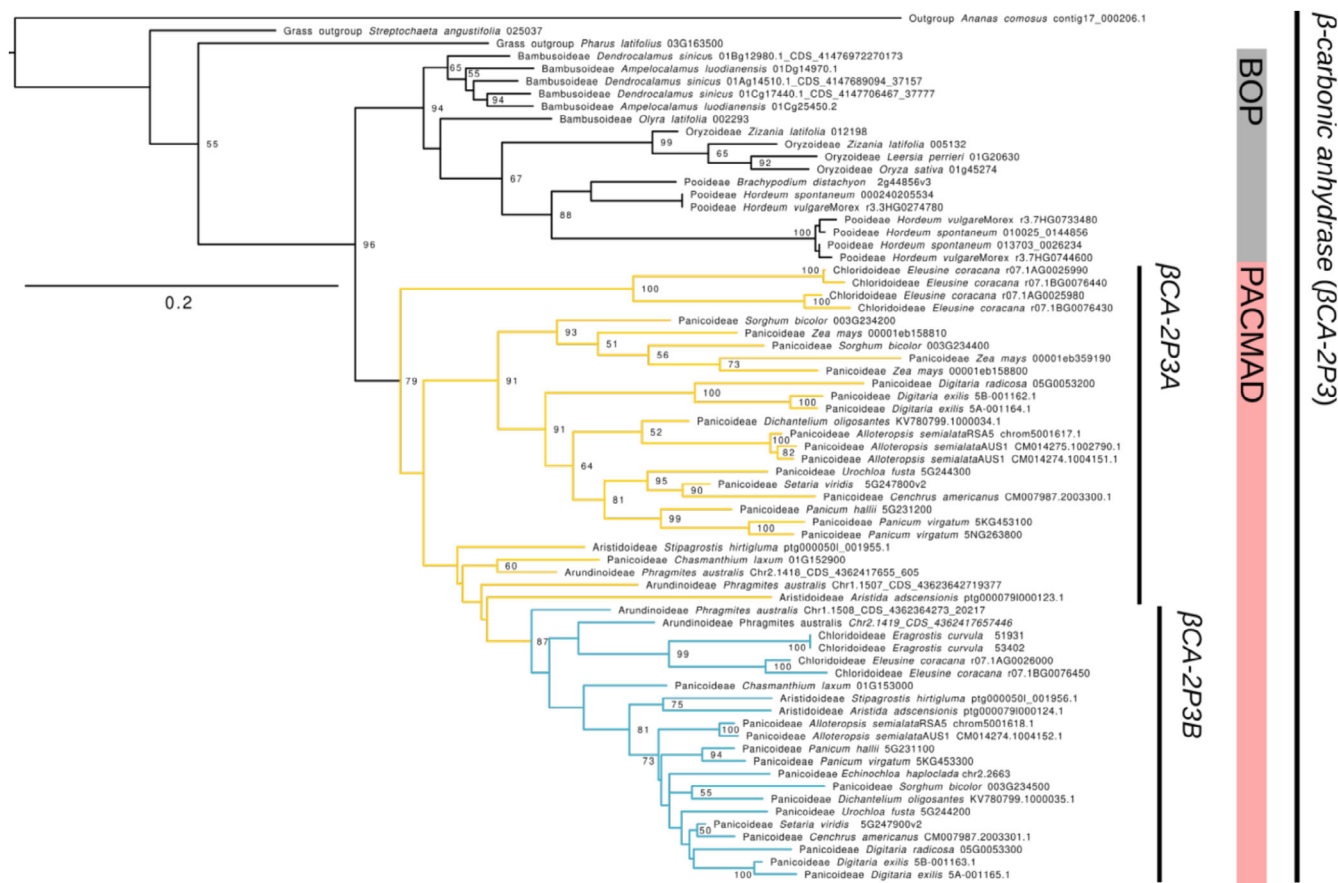


FIGURE 4 Maximum likelihood phylogenetic trees showing the PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micraroideae, Aristidoideae and Danthonioideae) duplication of β -carbonic anhydrase 2P3 (β CA-2P3), showing Copy A in yellow and Copy B in blue. Bootstrap support values $\geq 50\%$ shown, and a OrthoFinder-resolved gene tree (OG0001853) was used to classify gene copies into paralogous groups. BOP = Bambusoideae, Oryzoideae and Pooideae.

and the 22–24 origins of C_4 in this family are restricted to the PACMAD clade (Grass Phylogeny Working Group II, 2012). Here, we assembled genomes for the Aristidoideae (*A. adscensionis* and *S. hirtigluma*), a subfamily of grasses that includes over 350 species and is the sister subfamily to all other PACMAD grasses. We then used comparative genomics to identify genes that were duplicated or originated at the base of PACMAD or lost in the BOP clade, because these potentially represent important changes that increase the chances of C_4 evolution.

4.1 | No increase in gene duplication or subfunctionalization at the base of PACMAD

A significant increase in the rate of gene duplication is one explanation as to why genetic precursors for C_4 photosynthesis arose specifically in the PACMAD lineage. However, there is no evidence for a whole genome duplication event at the base of this clade (Zhang et al., 2024) nor do we detect a significantly increased duplication rate along this branch. Furthermore, genes duplicated at the base of PACMAD show no increase in subfunctionalization, with a similar

proportion of differential expression between C_4 photosynthetic cells (mesophyll and bundle sheath) to genome-wide levels. However, duplication and subfunctionalization may still play a role in the evolution of C_4 photosynthesis in grasses, with specific examples identified in this (Table 3) and previous (Emms et al., 2016) studies.

The PACMAD-specific gene duplications supplement existing variation to potentially create a pool of paralogues that can later be independently recruited. This means the initial duplication events that are foundational for the later emergence of the trait were not directly driven by selection for a C_4 function. A similar scenario is observed in *Solanum* species where paralog diversification generates the genetic substrate and contingency (the dependence of evolutionary outcomes on pre-existing genetic variation) for convergent crop trait evolution (Benoit et al., 2025).

4.2 | Duplication of β -carbonic anhydrase and carbon metabolism enzymes

The evolution of C_4 photosynthesis involves co-opting the functions of numerous enzymes to increase the efficiency and rate of carbon

TABLE 3 PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) duplicates that contain at least one maize and sorghum ortholog that is significantly differentially expressed between mesophyll (M) and bundle sheath (BS) cells.

Orthogroup	Sorghum	D.	Maize	D.	Differential expression			SwissProt annotation
					Sorghum	Maize	Rice	
OG0000101 ^a	SORBI_3006G158600	2	Zm00001eb427310	2	M 1424.5x	M 2.9x	M 3.7x	Salutaridine reductase (SALR)
OG0000348	SORBI_3005G004800	2	Zm00001eb404880	2	BS 19.1x	M 3.2x	N.E.	Xylan O-acetyltransferase 12 (XOATC)
OG0000461 ^a	SORBI_3003G266500	1	Zm00001eb155940	1	BS 18.4x	BS 5.8x	M 1.6x	Phosphoethanolamine
			Zm00001eb360800	2		BS 4.5x		N-methyltransferase 1 (PEAM1)
OG0000528	SORBI_3003G228100	1	Zm00001eb159310 ^b	1	M 9.7x	M 2.6x	N.S.	Cytochrome P450 CYP72A616 (72A61)
	SORBI_3003G228000	2	Zm00001eb358770	2	M 17.2x	M 3.5x		
OG0000548 ^a	SORBI_3009G140300	2	Zm00001eb288190	1	BS 5.7x	M 10.5x	N.S.	MLO-like protein 1 (MLO1)
OG0000565	SORBI_3006G195800	2	Zm00001eb072510	2	BS 15.3x	M 2.5x	M 2.7x	ABC transporter C family member 9 (AB9C)
OG0000658	SORBI_3009G003700	1	Zm00001eb073810	2	BS 4.5x	M 2.3x	N.S.	Protein argonaute 1B (AGO1B)
OG0000951	SORBI_3005G036600	1	Zm00001eb196930	2	BS 98.2x	BS 3.4x	N.S.	IQ domain-containing protein IQM3 (IQM3)
OG0001027	SORBI_3002G355300	2	Zm00001eb107930	2	BS 4.3x	M 1.7x	N.S.	Heterogeneous nuclear ribonucleoprotein 1 (RNP1)
OG0001105	SORBI_3002G176200	1	Zm00001eb310450	1	BS 5.3x	M 1.7x	BS 7.0x	Sulphite exporter TauE/SafE family protein 3 (TAUE3)
OG0001190 ^a	SORBI_3006G002000	2	Zm00001eb418980	2	BS 5.9x	BS 3.2x	M 4.2x ^c	Phosphoglycerate mutase-like protein 1 (PGML)
OG0001509	SORBI_3005G033100	2	Zm00001eb404990	1	M 102.9x	BS 2.2x	N.S.	Long chain acyl-CoA synthetase 6, peroxisomal (LACS6)
OG0001853	SORBI_3003G234500	1	Zm00001eb158810	2	M 13.6x	M 4.7x	M 4.3x	Carbonic anhydrase, chloroplastic (β CA)
	SORBI_3003G234200	2	Zm00001eb158800	2	M 30.1x	M 3.9x		
	SORBI_3003G234400	2	Zm00001eb359190	2	M 60.7x	M 3.6x		
OG0002656	SORBI_3001G464900	1	Zm00001eb007650	1	M 34.6x	M 5.4x	N.S.	Calcium-transporting ATPase 10, plasma membrane-type (ACA10)

Note: The corresponding expression pattern for rice ortholog is also shown.

Abbreviations: D., PACMAD duplicate group (ortholog = same number; paralogs = different number); N.E., not expressed; N.S., not significant.

^aCore set (i.e., duplicated and retained in all *C₄* PACMAD species used in this study).

^bSplit annotation, also including Zm00001eb159320.

^cOnly significant in one of the two analyses conducted.

fixation in high temperature and light environments. Duplication and neofunctionalization are not thought to give rise directly to *C₄* enzymes (Gowik & Westhoff, 2011; van den Bergh et al., 2014; Williams et al., 2012). However, older duplications in a gene family can result in paralogs that diverge in expression levels and patterning through drift or selection. The likelihood of recruiting the *C₄* specific copy then depends on the ancestral expression patterns and catalytic properties that are closest to the *C₄* optima (Aubry et al., 2011; Christin et al., 2015; Christin, Boxall, et al., 2013; Dunning et al., 2019; Emms et al., 2016; Hibberd & Covshoff, 2010; Moreno-Villena et al., 2018; Wang et al., 2009). In the last common ancestor of the PACMAD grasses, only a single gene encoding a core *C₄* enzyme was duplicated: β -carbonic anhydrase (β CA).

β CA accounts for up to 1–2% of soluble leaf protein in seed plants (Tiwari et al., 2005), and it catalyses the first step of the *C₄* cycle where CO₂ is reversibly converted into bicarbonate in the

mesophyll cells. β CA activity is essential for the establishment of the *C₄* cycle (Chatterjee et al., 2021; DiMario et al., 2022). Although β CA was originally thought to play a role in the ancestral *C₃* photosynthetic pathway, CRISPR mutants in tobacco have shown that β CA is not required for *C₃* photosynthesis but instead plays a crucial role in leaf development (Hines et al., 2021). β CA is encoded by three main paralogs in grasses (β CA-1P1, β CA-2P2 and β CA-2P3), with only β CA-2P3 recruited for *C₄* photosynthesis (Moreno-Villena et al., 2018). The PACMAD-specific duplicate is of the *C₄* β CA-2P3 copy, with the availability of reference genomes from Aristidoideae confirming that the duplication event happened in the last common ancestor of the PACMAD clade and not during the early diversification of this clade (Besnard et al., 2018). The duplication of this gene may have later facilitated independent neofunctionalization of a cytosolic *C₄* copy via the loss of chloroplast transit peptide, leaving the other copy to maintain the non-photosynthetic function (Ludwig, 2012) (Table 3).

However, whether a duplication of β CA is required for the evolution of C_4 photosynthesis in grasses is unclear. Furthermore, not all C_4 species within PACMAD have retained both duplicates. One of the β CA-2P3 daughter lineages has been lost independently in *Echinochloa haploclada* and maize (Figure 4), although in the latter case, the remaining lineage has undergone multiple independent duplications, potentially again highlighting the role of genetic contingency in convergent trait evolution.

The duplication of other carbon metabolism enzymes may also indirectly influence the emergence of C_4 photosynthesis. The evolution of C_4 photosynthesis is thought to be preceded by the establishment of a photorespiratory CO_2 pump that is partitioned between cell types (Bräutigam & Gowik, 2016). In *Flaveria* (Asteraceae), a gene duplication facilitated the establishment of this CO_2 pump because multiple copies enabled subfunctionalization and bundle sheath specificity (Schulze et al., 2013). Potentially, a similar evolutionary scenario could occur in the PACMAD grasses, which all possess a duplicated gene encoding (DL)-glycerol-3-phosphatase (OG0003194). This enzyme shows sequence similarity to known PGLPs that catalyse the first step of photorespiration (Schwarte & Bauwe, 2007). However, no photorespiratory activity has been functionally demonstrated for this protein in plants (Schwarte & Bauwe, 2007). Furthermore, although dual glycerol-3-phosphate and phosphoglycerate phosphatase activities have been reported for homologous enzymes in mammalian systems (Possik et al., 2017), it remains unknown whether such functional versatility occurs in plant PGLPs, which are adapted for high photorespiratory flux. However, the duplication of this gene in the PACMAD lineage may have facilitated subfunctionalisation of its roles.

The C_4 and C_3 photosynthetic pathways are also able to share carbon flux by the interchange of phosphoenolpyruvate (PEP) and 3-phosphoglycerate using the phosphoglycerate mutase and enolase pathway (Furbank & Kelly, 2021; Furbank & Leegood, 1984; von Caemmerer & Furbank, 2016). A phosphoglycerate mutase-like (PGML) protein was duplicated at the base of the PACMAD clade and is retained in all C_4 species (Table S4), and one of the two daughter copies has increased expression in the bundle sheath (Table 3). However, the increased tissue specificity of one of these copies may not be relevant to C_4 photosynthesis as the mutase-enolase shunt needs to be active in the mesophyll chloroplasts (Furbank & Kelly, 2021; Furbank & Leegood, 1984; von Caemmerer & Furbank, 2016). Phosphoglycerate mutase may also influence carbon partitioning by contributing to the control of 3-phosphoglycerate for non-photorespiratory serine synthesis, thereby linking photosynthetic carbon flux to biosynthetic pathways. The duplication of PGML and other carbon metabolism enzymes (Table S4) may generate the genetic substrate facilitating the spatial partitioning of carbohydrate metabolism in C_4 species, with sucrose predominantly synthesised in the mesophyll and starch in the bundle sheaths (Furbank & Kelly, 2021), although future studies are needed to confirm the functional consequences of these duplications.

4.3 | Repeated duplications of transmembrane transporters

Genes duplicated at the base of the PACMAD lineage are significantly enriched for transmembrane amino acid and organic acid transporters (Table 2) and a quarter of the 15 core PACMAD validated duplicates (i.e., retained in all C_4 species) are transporters, two of which are low-affinity transporters in the NRT1/PTR family (PTR15 [OG0000391] and PTR18 [OG0000379]). They are likely involved in nitrate transport (Plett et al., 2010), although in *Arabidopsis* can also transport other substrates such as plant hormones (Corratgé-Faillie & Lacombe, 2017). Another two move amino acids (cationic amino acid transporter 6 CAT6—OG0002111) and small amines such as carnitine (organic cation/carnitine transporter 7—OG0000861). In C_3 – C_4 intermediate plants that operate a glycine shuttle, there is a nitrogen imbalance caused by the release of ammonia (NH_3) in the bundle sheath. Indeed, this potential problem has been cited as one of the reasons why intermediate species might rapidly transition to a full C_4 cycle (Bräutigam & Gowik, 2016; Mallmann et al., 2014). Although these genes do not play a known role in C_4 photosynthesis itself, the duplicated nitrate, amine, and amino acid transporters may play a role in mitigating the nitrogen and amino-group imbalance during the incipient stages of C_4 evolution. These transporters could also represent previously unrecognised exaptations, facilitating the redistribution of metabolites such as serine and glycine between tissues once the photorespiratory pump is established, thereby subtly influencing metabolic gradients during C_4 evolution (Walsh et al., 2023).

4.4 | Enhancing defence against oxidative stress in bundle sheath cells

Oxygenic photosynthesis generates abundant reactive oxygen species (ROS). Although ROS can play key signalling roles, they can also damage cells and impair photosynthesis. To maintain redox homeostasis, plants have evolved several enzymatic and non-enzymatic strategies for ROS (Turkan et al., 2018). Glutathione is one of the key cellular pools of antioxidants (Mittler et al., 2004), and genes duplicated at the base of PACMAD are enriched for glutathione transferase activity (GO:0004364) and glutathione metabolic process (GO:0006749). The evolutionary trajectory to C_4 should overall reduce photorespiration and ROS production, but the dynamics of ROS production in the different leaf compartments require a tailored solution (Turkan et al., 2018). Comparisons between different C_4 and C_3 species show that regulatory and redox-based systems appear to be quite variable and are likely to be specific to the decarboxylation enzyme used by the plant for C_4 photosynthesis (Bräutigam et al., 2014; Turkan et al., 2018).

Out of the 15 core PACMAD validated duplicates, three have functions directly relating to limiting oxidative damage: a peroxidase (OG0000809), SALR (OG0000101) and PEAM1 (OG0000461). The duplication of phosphoethanolamine N-methyltransferase 1 (PEAM1)

is correlated with an expression shift from being mesophyll (C_3 rice) to bundle sheath (C_4 maize and sorghum) specific (Table 3). The same gene was also identified in a genome-wide association study (GWAS) in *Alloteropsis semialata* for the inner bundle sheath fraction (Alenazi et al., 2024), which is the proportion of the leaf responsible for refixing carbon through the C_4 cycle (Alenazi et al., 2023). PEAM1 catalyses the synthesis of phosphatidylcholine, a major membrane phospholipid (Chen et al., 2018), and orthologs in *Arabidopsis* are involved in regulating salt tolerance (He et al., 2022) and root cell differentiation (Zou et al., 2019) by balancing ROS production and abscisic acid or auxin, respectively. SALR (OG0000101) is mesophyll-specific (Table 3) and orthologous to an *Arabidopsis* NAD(P)-binding Rossmann-fold superfamily protein involved in detoxification and responding to reactive carbonyl species (Ksas et al., 2024). The repeated duplication of genes associated with oxidative stress at the base of PACMAD created redundancy in the genetic basis of oxidative stress management. Although these enzymes are not directly part of the core C_4 photosynthetic pathway, the contingency generated by the expansion of oxidoreductase enzymes may have helped mitigate the altered oxidative stress management required for C_4 photosynthesis to evolve.

4.5 | Multiple PACMAD-specific transcription regulators

Genes co-opted for C_4 photosynthesis become highly expressed in leaf tissue, but it still remains unclear how many of these are regulated, particularly for trans-acting factors (Hibberd & Covshoff, 2010; Lyu et al., 2025; Schlüter & Weber, 2020). A recent study on the evolution of C_4 photosynthesis in *Flaveria* showed that the emergence of this trait was correlated with the expansion of ethylene response factor (ERF) transcription factors (Lyu et al., 2025). In PACMAD, we identify a number of de novo transcription factor orthogroups, including a group of ERFs retained in all C_4 lineages (OG0009413). One possible explanation why C_4 photosynthesis is restricted to the PACMAD clade is the de novo origin of transcription factors that could be readily recruited for this process.

In this study, we examined patterns of gene birth, duplication and loss to identify differences in gene complements between PACMAD and BOP grasses. The most recent common ancestor of the PACMAD clade also represents the last common ancestor of all C_4 origins in grasses, suggesting that the restricted emergence of this trait may be associated with some of the gene complement differences identified here. Several of the candidate genes have independent lines of evidence supporting their relevance to C_4 evolution, including established biological functions (e.g., β CA), overlap with GWAS hits for C_4 -related anatomical traits (e.g. PEAM1), and cell specific expression (Table 3). However, the PACMAD stem lineage persisted for approximately 7.51 million years (2.63–12.57 95% CI; Christin et al., 2014) before diverging, providing ample time for other differences to accumulate either through selection or drift. It is therefore likely that only a subset of changes highlighted here played

a role in facilitating the recurrent emergence of the photorespiratory pump and/or C_4 photosynthesis within this clade. These changes also did not trigger the emergence of C_4 photosynthesis itself, which occurred much later, but may have generated the required redundancy and contingency that enabled selection to assemble the components of the C_4 pathway. The genetic turnover of transporters, oxidoreductases, transferases and transcription regulators may have increased metabolic permissiveness by reducing the fitness cost of early C_4 evolution and facilitating enhanced developmental plasticity. Overall, these changes potentially increased the probability of successful C_4 evolution without being directly involved in the C_4 biochemical pathway itself. Future comparative analyses, single-cell transcriptomics and functional genomics studies within the PACMAD clade will be critical to refine the list of candidate gene families and identify the foundational steps in C_4 evolution. In particular, it will be important to test if any of these duplications were lost in PACMAD lineages that never evolved C_4 photosynthesis, for example, the Danthonioideae.

5 | CONCLUSION

The assembly of the first genomes for species belonging to the subfamily Aristidoideae permits the inference of genetic changes in the last common ancestor of the PACMAD clade that may have enabled the repeated evolution of C_4 photosynthesis. Gene duplication, de novo formation and loss generates standing variation in gene content that could foreshadow the repeated evolution of C_4 photosynthesis in the PACMAD clade. This includes an additional copy of a core C_4 enzyme (β -carbonic anhydrase), as well as genes with functions including carbon metabolism, amino acid and nitrate transport, oxidative stress management and transcription regulation. This gene turnover altered the ancestral genomic landscape, potentially predisposing this lineage for the evolution of C_4 photosynthesis. This supports the idea that the repeated evolution of C_4 photosynthesis in PACMAD grasses was not driven by a single genetic event but rather by the opportunistic utilisation of pre-existing genetic diversity.

AUTHOR CONTRIBUTIONS

Lara Pereira, Pascal-Antoine Christin, Colin P. Osborne and Luke T. Dunning conceptualised and designed the study. Lara Pereira generated the genome assemblies. Ahmed S. Alenazi and Lara Pereira performed the duplication and gene loss analyses. Sahr Mian and Ilia J. Leitch obtained genome size estimates. All authors helped interpret the data. Lara Pereira, Ahmed S. Alenazi and Luke T. Dunning wrote the manuscript with the help of all authors.

ACKNOWLEDGEMENTS

We thank the staff of the Arthur Willis Environment Centre and all members of the Dunning and Osborne labs for their help with plant growth and maintenance. We thank Lucy Knowles for her support in DNA quality assessment. Sequencing data generation was carried out by the Centre for Genomic Research, which is based at the University

of Liverpool. This work was funded by the Natural Environment Research Council (Grant No. NE/V000012/1). PAC was funded by a Royal Society University Research Fellowship (Grant No. URF\R\180022). LTD is funded by a NERC Fellowship (Grant No. NE/T011025/1). The authors also extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA, for funding this research work (Project No. NBU-FFR-2026-630-01).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The Aristidoideae raw data, genomes and annotation files are available in NCBI (Bioproject PRJNA1254624). Tree files and Helixer re-annotations of published genomes are available on dryad (doi: [10.5061/dryad.xgxd254t2](https://doi.org/10.5061/dryad.xgxd254t2)). The genome assembly and annotation code are available in GitHub (<https://github.com/Sheffield-Plant-Evolutionary-Genomics>).

ORCID

Lara Pereira  <https://orcid.org/0000-0001-5184-8587>

Ahmed S. Alenazi  <https://orcid.org/0000-0003-4105-2539>

Sahr Mian  <https://orcid.org/0000-0001-7828-0629>

Ilia J. Leitch  <https://orcid.org/0000-0002-3837-8186>

Pascal-Antoine Christin  <https://orcid.org/0000-0001-6292-8734>

Colin P. Osborne  <https://orcid.org/0000-0002-7423-3718>

Luke T. Dunning  <https://orcid.org/0000-0002-4776-9568>

REFERENCES

- Acevedo-Garcia, J., Kusch, S., & Panstruga, R. (2014). Magical mystery tour: MLO proteins in plant immunity and beyond. *New Phytologist*, 204, 273–281. <https://doi.org/10.1111/nph.12889>
- Alenazi, A. S., Bianconi, M. E., Middlemiss, E., Milenkovic, V., Curran, E. V., Sotelo, G., Lundgren, M. R., Nyirenda, F., Pereira, L., Christin, P. A., Dunning, L. T., & Osborne, C. P. (2023). Leaf anatomy explains the strength of C₄ activity within the grass species *Alloteropsis semialata*. *Plant, Cell & Environment*, 46, 2310–2322. <https://doi.org/10.1111/pce.14607>
- Alenazi, A. S., Pereira, L., Christin, P.-A., Osborne, C. P., & Dunning, L. T. (2024). Identifying genomic regions associated with C₄ photosynthetic activity and leaf anatomy in *Alloteropsis semialata*. *New Phytologist*, 243, 1698–1710. <https://doi.org/10.1111/nph.19933>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25, 25–29. <https://doi.org/10.1038/75556>
- Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M., & Osborne, C. P. (2016). C₄ photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants*, 2, 16038. <https://doi.org/10.1038/nplants.2016.38>
- Aubry, S., Brown, N. J., & Hibberd, J. M. (2011). The role of proteins in C₃ plants prior to their recruitment into the C₄ pathway. *Journal of Experimental Botany*, 62, 3049–3059. <https://doi.org/10.1093/jxb/err012>
- Benoit, M., Jenike, K. M., Satterlee, J. W., Ramakrishnan, S., Gentile, I., Hendelman, A., Passalacqua, M. J., Suresh, H., Shohat, H., Robitaille, G. M., Fitzgerald, B., Alonge, M., Wang, X., Santos, R., He, J., Ou, S., Golan, H., Green, Y., Swartwood, K., ... Lippman, Z. B. (2025). *Solanum* pan-genetics reveals paralogues as contingencies in crop engineering. *Nature*, 640, 135–145. <https://doi.org/10.1038/s41586-025-08619-6>
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis*, 53, 474–485. <https://doi.org/10.1002/dvg.22877>
- Besnard, G., Bianconi, M. E., Hackel, J., Manzi, S., Vorontsova, M. S., & Christin, P. A. (2018). Herbarium genomics retraces the origins of C₄-specific carbonic anhydrase in Andropogoneae (Poaceae). *Botany Letters*, 165, 419–433. <https://doi.org/10.1080/23818107.2018.1469429>
- Bianconi, M. E., Dunning, L. T., Moreno-Villena, J. J., Osborne, C. P., & Christin, P. A. (2018). Gene duplication and dosage effects during the early emergence of C₄ photosynthesis in the grass genus *Alloteropsis*. *Journal of Experimental Botany*, 69, 1967–1980. <https://doi.org/10.1093/jxb/ery029>
- Blätke, M. A., & Bräutigam, A. (2019). Evolution of C₄ photosynthesis predicted by constraint-based modelling. *eLife*, 8, e49305. <https://doi.org/10.7554/eLife.49305>
- Bräutigam, A., & Gowik, U. (2016). Photorespiration connects C₃ and C₄ photosynthesis. *Journal of Experimental Botany*, 67, 2953–2962. <https://doi.org/10.1093/jxb/erw056>
- Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C. P., & Weber, A. P. (2014). Towards an integrative model of C₄ photosynthetic subtypes: Insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C₄ species. *Journal of Experimental Botany*, 65, 3579–3593.
- Bussotti, G., Raineri, E., Erb, I., Zytnecki, M., Wilm, A., Beaudoin, E., Bucher, P., & Notredame, C. (2011). BlastR—Fast and accurate database searches for non-coding RNAs. *Nucleic Acids Research*, 39, 6886–6895. <https://doi.org/10.1093/nar/gkr335>
- Chatterjee, J., Coe, R. A., Acebron, K., Thakur, V., Yennamalli, R. M., Danila, F., Lin, H.-C., Balahadia, C. P., Bagunu, E., Padhma, P. P. O. S., Bala, S., Yin, X., Rizal, G., Dionora, J., Furbank, R. T., von Caemmerer, S., & Quick, W. P. (2021). A low CO₂-responsive mutant of *Setaria viridis* reveals that reduced carbonic anhydrase limits C₄ photosynthesis. *Journal of Experimental Botany*, 72(8), 3122–3136. <https://doi.org/10.1093/jxb/erab039>
- Chen, W., Salari, H., Taylor, M. C., Jost, R., Berkowitz, O., Barrow, R., Qiu, D., Branco, R., & Masle, J. (2018). NMT1 and NMT3 N-methyltransferase activity is critical to lipid homeostasis, morphogenesis, and reproduction. *Plant Physiology*, 177, 1605–1628. <https://doi.org/10.1104/pp.18.00457>
- Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., & Chong, Z. (2021). Accurate long-read de novo assembly evaluation with inspector. *Genome Biology*, 22, 312. <https://doi.org/10.1186/s13059-021-02527-4>
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18, 170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Christin, P.-A., Arakaki, M., Osborne, C. P., & Edwards, E. J. (2015). Genetic enablers underlying the clustered evolutionary origins of C₄ photosynthesis in angiosperms. *Molecular Biology and Evolution*, 32, 846–858. <https://doi.org/10.1093/molbev/msu410>
- Christin, P. A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., & Osborne, C. P. (2013). Parallel recruitment of multiple genes into C₄ photosynthesis. *Genome Biology and Evolution*, 5, 2174–2187. <https://doi.org/10.1093/gbe/evt168>
- Christin, P. A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., Garrison, L. M., Vorontsova, M. S., & Edwards, E. J. (2013). Anatomical enablers and the evolution of C₄ photosynthesis in grasses. *Proceedings of the National Academy of Sciences*, 110, 1381–1386. <https://doi.org/10.1073/pnas.1216777110>

- Christin, P.-A., Spriggs, E., Osborne, C. P., Strömberg, C. A., Salamin, N., & Edwards, E. J. (2014). Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63, 153–165. <https://doi.org/10.1093/sysbio/syt072>
- Clark, J., Hidalgo, O., Pellicer, J., Liu, H., Marquardt, J., Robert, Y., Christenhusz, M., Zhang, S., Gibby, M., Leitch, I. J., & Schneider, H. (2016). Genome evolution of ferns: Evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist*, 210, 1072–1082. <https://doi.org/10.1111/nph.13833>
- Corratgé-Faillie, C., & Lacombe, B. (2017). Substrate (un)specificity of Arabidopsis NRT1/PTR FAMILY (NPF) proteins. *Journal of Experimental Botany*, 68(12), 3107–3113. <https://doi.org/10.1093/jxb/erw499>
- Darwin, C. (1859). *On the origin of species by means of natural selection*. John Murray.
- DiMario, R. J., Giuliani, R., Ubierna, N., Slack, A. D., Cousins, A. B., & Studer, A. J. (2022). Lack of leaf carbonic anhydrase activity eliminates the C₄ carbon-concentrating mechanism requiring direct diffusion of CO₂ into bundle sheath cells. *Plant, Cell & Environment*, 45, 1382–1397. <https://doi.org/10.1111/pce.14291>
- Doležel, J., Greilhuber, J., & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, 2, 2233–2244. <https://doi.org/10.1038/nprot.2007.310>
- Döring, F., Streubel, M., Bräutigam, A., & Gowik, U. (2016). Most photosynthetic genes are preferentially expressed in the bundle sheath cells of the C₄ grass *Sorghum bicolor*. *Journal of Experimental Botany*, 67, 3053–3064. <https://doi.org/10.1093/jxb/erw041>
- Dunning, L. T., Lundgren, M. R., Moreno-Villena, J. J., Namaganda, M., Edwards, E. J., Nosil, P., Osborne, C. P., & Christin, P. A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C₄ photosynthesis among close relatives. *Evolution*, 71, 1541–1555. <https://doi.org/10.1111/evo.13250>
- Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J. K., Mapaura, A., Grundy, I. M., Kayombo, C. J., Dunning, L. A., Kentatchime, F., Ariyaratne, M., Yakandawala, D., Besnard, G., Quick, W. P., Bräutigam, A., Osborne, C. P., & Christin, P. A. (2019). Key changes in gene expression identified for different stages of C₄ evolution in *Allotheropsis semialata*. *Journal of Experimental Botany*, 70, 3255–3268. <https://doi.org/10.1093/jxb/erz149>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.
- Ehleringer, J., & Björkman, O. (1977). Quantum yields for CO₂ uptake in C₃ and C₄ plants: Dependence on temperature, CO₂, and O₂ concentration. *Plant Physiology*, 59, 86–90. <https://doi.org/10.1104/pp.59.1.86>
- Ehleringer, J. R. (1978). Implications of quantum yield differences on the distributions of C₃ and C₄ grasses. *Oecologia*, 31, 255–267. <https://doi.org/10.1007/BF00346246>
- Ehleringer, J. R., Cerling, T. E., & Helliker, B. R. (1997). C₄ photosynthesis, atmospheric CO₂, and climate. *Oecologia*, 112, 285–299. <https://doi.org/10.1007/s004420050311>
- Ehleringer, J. R., Sage, R. F., Flanagan, L. B., & Pearcy, R. W. (1991). Climate change and the evolution of C₄ photosynthesis. *Trends in Ecology & Evolution*, 6, 95–99. [https://doi.org/10.1016/0169-5347\(91\)90183-X](https://doi.org/10.1016/0169-5347(91)90183-X)
- Emms, D. M., Covshoff, S., Hibberd, J. M., & Kelly, S. (2016). Independent and parallel evolution of new genes by gene duplication in two origins of C₄ photosynthesis provides new insight into the mechanism of phloem loading in C₄ species. *Molecular Biology and Evolution*, 33, 1796–1806. <https://doi.org/10.1093/molbev/msw057>
- Emms, D. M., & Kelly, S. (2017). STRIDE: Species tree root inference from gene duplication events. *Molecular Biology and Evolution*, 34, 3267–3278. <https://doi.org/10.1093/molbev/msx259>
- Emms, D. M., & Kelly, S. (2018). STAG: Species tree inference from all genes. *BioRxiv* 267914. <https://doi.org/10.1101/267914>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 1–14.
- Fernández-Pérez, F., Pomar, F., Pedreño, M. A., & Novo-Uzal, E. (2015). The suppression of AtPrx52 affects fibers but not xylem lignification in *Arabidopsis* by altering the proportion of syringyl units. *Physiologia Plantarum*, 154, 395–406. <https://doi.org/10.1111/ppl.12310>
- Furbank, R. T., & Kelly, S. (2021). Finding the C₄ sweet spot: Cellular compartmentation of carbohydrate metabolism in C₄ photosynthesis. *Journal of Experimental Botany*, 72, 6018–6026. <https://doi.org/10.1093/jxb/erab290>
- Furbank, R. T., & Leegood, R. C. (1984). Carbon metabolism and gas exchange in leaves of *Zea mays* L. interaction between the C₃ and C₄ pathways during photosynthetic induction. *Planta*, 162, 457–462. <https://doi.org/10.1007/BF00393459>
- Gowik, U., & Westhoff, P. (2011). The path from C₃ to C₄ photosynthesis. *Plant Physiology*, 155, 56–63. <https://doi.org/10.1104/pp.110.165308>
- Grass Phylogeny Working Group II. (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C₄ origins. *New Phytologist*, 193, 304–312. <https://doi.org/10.1111/j.1469-8137.2011.03972.x>
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hatch, M. D. (1987). C₄ photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta*, 895, 81–106. [https://doi.org/10.1016/s0304-4173\(87\)80009-5](https://doi.org/10.1016/s0304-4173(87)80009-5)
- He, Q. Y., Jin, J. F., Lou, H. Q., Dang, F. F., Xu, J. M., Zheng, S. J., & Yang, J. L. (2022). Absciscic acid-dependent PMT1 expression regulates salt tolerance by alleviating absciscic acid-mediated reactive oxygen species production in *Arabidopsis*. *Journal of Integrative Plant Biology*, 64, 1803–1820. <https://doi.org/10.1111/jipb.13326>
- Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P., & Lercher, M. J. (2013). Predicting C₄ photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell*, 153, 1579–1588. <https://doi.org/10.1016/j.cell.2013.04.058>
- Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P. A., & Edwards, E. J. (2019). The genetics of convergent evolution: Insights from plant photosynthesis. *Nature Reviews Genetics*, 20, 485–493. <https://doi.org/10.1038/s41576-019-0107-5>
- Hibberd, J. M., & Covshoff, S. (2010). The regulation of gene expression required for C₄ photosynthesis. *Annual Review of Plant Biology*, 61, 181–207. <https://doi.org/10.1146/annurev-arplant-042809-112238>
- Hines, K. M., Chaudhari, V., Edgeworth, K. N., Owens, T. G., & Hanson, M. R. (2021). Absence of carbonic anhydrase in chloroplasts affects C₃ plant development but not photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2107425118. <https://doi.org/10.1073/pnas.2107425118>
- Hoang, C. F., Liu, W. Y., Lu, M. Y. J., Chen, Y. H., Ku, M. S., & Li, W. H. (2021). Whole-genome duplication facilitated the evolution of C₄ photosynthesis in *Gynandropsis gynandra*. *Molecular Biology and Evolution*, 38, 4715–4731.
- Holst, F., Bolger, A., Günther, C., Maß, J., Triesch, S., Kindel, F., Kiel, N., Saadat, N., Ebenhö, O., Usadel, B., Schwacke, R., Bolger, M., Weber, A. P. M., & Denton, A. K. (2023). Helixer-de novo prediction of primary eukaryotic gene models combining deep learning and a hidden markov model. *BioRxiv* 2023–2002.
- Howe, K. L., Contreras-Moreira, B., de Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolger, D. M., Cambell, L., & Carbajo, M. (2020). Ensembl genomes 2020—Enabling non-vertebrate genomic research. *Nucleic Acids Research*, 48, D689–D695. <https://doi.org/10.1093/nar/gkz890>
- Hua, L., Stevenson, S. R., Reyna-Llorens, I., Xiong, H., Kopriva, S., & Hibberd, J. M. (2021). The bundle sheath of rice is conditioned to play

- an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis. *The Plant Journal*, 107, 268–286. <https://doi.org/10.1111/tpj.15292>
- Hughes, T. E., & Langdale, J. A. (2020). SCARECROW gene function is required for photosynthetic development in maize. *Plant Direct*, 4, e00264. <https://doi.org/10.1002/pld3.264>
- Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 16, 294. <https://doi.org/10.1186/s13059-015-0849-0>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Ksas, B., Chiarenza, S., Dubourg, N., Ménard, V., Gilbin, R., & Havaux, M. (2024). Plant acclimation to ionising radiation requires activation of a detoxification pathway against carbonyl-containing lipid oxidation products. *Plant, Cell & Environment*, 47, 3882–3898. <https://doi.org/10.1111/pce.14994>
- Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart model selection in PhyML. *Molecular Biology and Evolution*, 34, 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Lenski, R. E., Ofria, C., Pennock, R. T., & Adami, C. (2003). The evolutionary origin of complex features. *Nature*, 423, 139–144. <https://doi.org/10.1038/nature01568>
- Lewis, M. W., & Hake, S. (2016). Keep on growing: Building and patterning leaves in the grasses. *Current Opinion in Plant Biology*, 29, 80–86. <https://doi.org/10.1016/j.pbi.2015.11.008>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., Kebrom, T. H., Provart, N., Patel, R., Myers, C. R., Reidel, E. J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., & Brutnell, T. P. (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, 42, 1060–1067. <https://doi.org/10.1038/ng.703>
- Liu, Y. C., Lin, Y. C., Kanehara, K., & Nakamura, Y. (2019). A methyltransferase trio essential for phosphatidylcholine biosynthesis and growth. *Plant Physiology*, 179, 433–445. <https://doi.org/10.1104/pp.18.01408>
- Loureiro, J., Rodriguez, E., Doležel, J., & Santos, C. (2007). Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. *Annals of Botany*, 100, 875–888. <https://doi.org/10.1093/aob/mcm152>
- Ludwig, M. (2012). Carbonic anhydrase and the molecular evolution of C₄ photosynthesis. *Plant, Cell & Environment*, 35, 22–37. <https://doi.org/10.1111/j.1365-3040.2011.02364.x>
- Lyu, M. J. A., Du, H., Yao, H., Zhang, Z., Chen, G., Huang, Y., Ni, X., Chen, F., Zhao, Y. Y., Tang, Q., & Miao, F. (2025). A dominant role of transcriptional regulation during the evolution of C₄ photosynthesis in *Flaveria* species. *Nature Communications*, 16, 1643. <https://doi.org/10.1038/s41467-025-56901-y>
- Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P., & Gowik, U. (2014). The role of photorespiration during the evolution of C₄ photosynthesis in the genus *Flaveria*. *eLife*, 3, e02478. <https://doi.org/10.7554/eLife.02478>
- Meléndez-Hevia, E., Waddell, T. G., & Cascante, M. (1996). The puzzle of the Krebs citric acid cycle: Assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *Journal of Molecular Evolution*, 43, 293–303. <https://doi.org/10.1007/BF02338838>
- Mi, H., Muruganujan, A., & Thomas, P. D. (2012). PANTHER in 2013: Modelling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1), D377–D386. <https://doi.org/10.1093/nar/gks118>
- Mittler, R., Vandersaunders, S., Gollery, M., & van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends in Plant Science*, 9, 490–498. <https://doi.org/10.1016/j.tplants.2004.08.009>
- Monson, R. K. (2003). Gene duplication, neofunctionalization, and the evolution of C₄ photosynthesis. *International Journal of Plant Sciences*, 164, S43–S54.
- Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P., & Christin, P. A. (2018). Highly expressed genes are preferentially co-opted for C₄ photosynthesis. *Molecular Biology and Evolution*, 35, 94–106. <https://doi.org/10.1093/molbev/msx269>
- Obermayer, R., Leitch, I. J., Hanson, L., & Bennett, M. D. (2002). Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Annals of Botany*, 90, 209–217. <https://doi.org/10.1093/aob/mcf167>
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, 20, 275. <https://doi.org/10.1186/s13059-019-1905-y>
- Ourisson, G., & Nakatani, Y. (1994). The terpenoid theory of the origin of cellular life: The evolution of terpenoids to cholesterol. *Chemistry & Biology*, 1, 11–23. [https://doi.org/10.1016/1074-5521\(94\)90036-1](https://doi.org/10.1016/1074-5521(94)90036-1)
- Paparelli, E., Gonzali, S., Parlanti, S., Novi, G., Giorgi, F. M., Licausi, F., Kosmacz, M., Feil, R., Lunn, J. E., Brust, H., van Dongen, J. T., Steup, M., & Perata, P. (2012). Misexpression of a chloroplast aspartyl protease leads to severe growth defects and alters carbohydrate metabolism in *Arabidopsis*. *Plant Physiology*, 160, 1237–1250. <https://doi.org/10.1104/pp.112.204016>
- Pereira, L., Bianconi, M. E., Osborne, C. P., Christin, P. A., & Dunning, L. T. (2023). *Alloteropsis semialata* as a study system for C₄ evolution in grasses. *Annals of Botany*, 132, 365–382. <https://doi.org/10.1093/aob/mcad078>
- Planta, J., Liang, Y. Y., Xin, H., Chansler, M. T., Prather, L. A., Jiang, N., Jiang, J., & Childs, K. L. (2022). Chromosome-scale genome assemblies and annotations for Poales species *Carex cristatella*, *Carex scoparia*, *Juncus effusus*, and *Juncus inflexus*. G3: Genes, Genomes, Genetics, 12, jkac211. <https://doi.org/10.1093/g3journal/jkac211>
- Plett, D., Toubia, J., Garnett, T., Tester, M., Kaiser, B. N., & Baumann, U. (2010). Dichotomy in the NRT gene families of dicots and grass species. *PLoS ONE*, 5(12), e15289. <https://doi.org/10.1371/journal.pone.0015289>
- Possik, E., Madiraju, S. M., & Prentki, M. (2017). Glycerol-3-phosphate phosphatase/PGP: Role in intermediary metabolism and target for cardiometabolic diseases. *Biochimie*, 143, 18–28.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5, e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). Genome-Scope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, 16, 276–277. [https://doi.org/10.1016/S0168-9525\(00\)00204-2](https://doi.org/10.1016/S0168-9525(00)00204-2)
- Sage, R. F., Christin, P. A., & Edwards, E. J. (2011). The C₄ plant lineages of planet Earth. *Journal of Experimental Botany*, 62, 3155–3169. <https://doi.org/10.1093/jxb/err048>
- Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C₄ photosynthesis. *Annual Review of Plant Biology*, 63, 19–47. <https://doi.org/10.1146/annurev-arplant-042811-105511>

- Schlüter, U., & Weber, A. P. (2020). Regulation and evolution of C₄ photosynthesis. *Annual Review of Plant Biology*, 71, 183–215. <https://doi.org/10.1146/annurev-arplant-042916-040915>
- Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., Gowik, U., & Westhoff, P. (2013). Evolution of C₄ photosynthesis in the genus *Flaveria*: Establishment of a photorespiratory CO₂ pump. *The Plant Cell*, 25, 2522–2535. <https://doi.org/10.1105/tpc.113.114520>
- Schwarte, S., & Bauwe, H. (2007). Identification of the photorespiratory 2-phosphoglycolate phosphatase, PGLP1, in *Arabidopsis*. *Plant Physiology*, 144, 1580–1586. <https://doi.org/10.1104/pp.107.099192>
- Sedelnikova, O. V., Hughes, T. E., & Langdale, J. A. (2018). Understanding the genetic basis of C₄ Kranz anatomy with a view to engineering C₃ crops. *Annual Review of Genetics*, 52, 249–270. <https://doi.org/10.1146/annurev-genet-120417-031217>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Slewinski, T. L., Anderson, A. A., Zhang, C., & Turgeon, R. (2012). Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant and Cell Physiology*, 53, 2030–2037. <https://doi.org/10.1093/pcp/pcs147>
- Sloan, D. B., Alverson, A. J., Chuckalovcak, J. P., Wu, M., McCauley, D. E., Palmer, J. D., & Taylor, D. R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology*, 10, e1001241. <https://doi.org/10.1371/journal.pbio.1001241>
- Soreng, R. J., Peterson, P. M., Zuloaga, F. O., Romaschenko, K., Clark, L., Clark, L. G., Teisher, J. K., Gillespie, L. J., Barberá, P., Welker, C. A. D., Kellogg, E. A., Li, D.-Z., & Davidse, G. (2022). A worldwide phylogenetic classification of the Poaceae (Gramineae) III: An update. *Journal of Systematics and Evolution*, 60, 476–521. <https://doi.org/10.1111/jse.12847>
- Spriggs, E. L., Christin, P.-A., & Edwards, E. J. (2014). C₄ photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE*, 9, e97722. <https://doi.org/10.1371/journal.pone.0097722>
- Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A. P., & Denton, A. K. (2020). Helixer: Cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics*, 36, 5291–5298. <https://doi.org/10.1093/bioinformatics/btaa1044>
- Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L., & Mi, H. (2021). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science*, 31, 8–22. <https://doi.org/10.1002/pro.4218>
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45, W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Tiwari, A., Kumar, P., Singh, S., & Ansari, S. A. (2005). Carbonic anhydrase in relation to higher plants. *Photosynthetica*, 43, 1–11. <https://doi.org/10.1007/s11099-005-1011-0>
- Turkan, I., Uzilday, B., Dietz, K. J., Bräutigam, A., & Ozgur, R. (2018). Reactive oxygen species and redox regulation in mesophyll and bundle sheath cells of C₄ plants. *Journal of Experimental Botany*, 69, 3321–3331. <https://doi.org/10.1093/jxb/ery064>
- Uliano-Silva, M., Ferreira, J. G., Krashennnikova, K., Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy, S. A. (2023). Mitohifi: A python pipeline for mitochondrial genome assembly from Pacbio high fidelity reads. *BMC Bioinformatics*, 24, 288.
- van den Bergh, E., Külahoglu, C., Bräutigam, A., Hibberd, J. M., Weber, A. P. M., Zhu, X. G., & Eric Schranz, M. (2014). Gene and genome duplications and the origin of C₄ photosynthesis: Birth of a trait in the Cleomaceae. *Current Plant Biology*, 1, 2–9. <https://doi.org/10.1016/j.cpb.2014.08.001>
- von Caemmerer, S., & Furbank, R. T. (2016). Strategies for improving C₄ photosynthesis. *Current Opinion in Plant Biology*, 31, 125–134. <https://doi.org/10.1016/j.cpb.2016.04.003>
- Walsh, C. A., Bräutigam, A., Roberts, M. R., & Lundgren, M. R. (2023). Evolutionary implications of C₂ photosynthesis: How complex biochemical trade-offs may limit C₄ evolution. *Journal of Experimental Botany*, 74, 707–722. <https://doi.org/10.1093/jxb/erac465>
- Wang, X., Gowik, U., Tang, H., Bowers, J. E., Westhoff, P., & Paterson, A. H. (2009). Comparative genomic analysis of C₄ photosynthetic pathway evolution in grasses. *Genome Biology*, 10, R68. <https://doi.org/10.1186/gb-2009-10-6-r68>
- Washburn, J. D., Schnable, J. C., Davidse, G., & Pires, J. C. (2015). Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany*, 102, 1493–1505. <https://doi.org/10.3732/ajb.1500222>
- Williams, B. P., Aubry, S., & Hibberd, J. M. (2012). Molecular evolution of genes recruited into C₄ photosynthesis. *Trends in Plant Science*, 17, 213–220. <https://doi.org/10.1016/j.tplants.2012.01.008>
- Woodhouse, M. R., Cannon, E. K., Portwood, J. L., Harper, L. C., Gardiner, J. M., Schaeffer, M., & Andorf, C. (2021). A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biology*, 21, 1–10.
- Zhang, T., Huang, W., Zhang, L., Li, D. Z., Qi, J., & Ma, H. (2024). Phylogenomic profiles of whole-genome duplications in Poaceae and landscape of differential duplicate retention and losses among major Poaceae lineages. *Nature Communications*, 15, 3305. <https://doi.org/10.1038/s41467-024-47428-9>
- Zou, Y., Zhang, X., Tan, Y., Huang, J. B., Zheng, Z., & Tao, L. Z. (2019). Phosphoethanolamine N-methyltransferase 1 contributes to maintenance of root apical meristem by affecting ROS and auxin-regulated cell differentiation in *Arabidopsis*. *New Phytologist*, 224, 258–273. <https://doi.org/10.1111/nph.16028>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pereira, L., Alenazi, A. S., Mian, S., Leitch, I. J., Christin, P.-A., Osborne, C. P., & Dunning, L. T. (2026). Gene turnover in the common ancestor of all C₄ grasses. *Plants, People, Planet*, 1–18. <https://doi.org/10.1002/ppp3.70206>