This is a repository copy of *Continued adaptation of C4 photosynthesis after an initial burst of changes in the andropogoneae grasses*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/152332/

Version: Published Version

Article:

Continued Adaptation of C$_4$ Photosynthesis After an Initial Burst of Changes in the Andropogoneae Grasses

MATHEUS E. BIANCONI[1]*, JAN HACKEL[2,3]*, MARIA S. VORONTSOVA[3], ADRIANA ALBERTI[4], WATCHARA ARTHAN[5], SEAN V. BURKE[6], MELVIN R. DUVALL[6], ELIZABETH A. KELLOGG[7], SÉBASTIEN LAVERGNE[8], MICHAEL R. MCKAIN[9], ALEXANDRE MEUNIER[2], COLIN P. OSBORNE[1], PAWEENA TRAIPERM[5], PASCAL-ANTOINE CHRISTIN[1], GUILLAUME BESNARD[2]**

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

[2] Laboratoire Evolution & Diversité Biologique (EDB, UMR 5174), CNRS/IRD/Université Toulouse III, 118 route de Narbonne, 31062 Toulouse, France

[3] Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW93AB, UK

[4] CEA - Institut de Biologie Francois-Jacob, Genoscope, 2 Rue Gaston Cremieux 91057 Evry Cedex, France

[5] Department of Plant Science, Faculty of Science, Mahidol University, King Rama VI Road, Bangkok 10400, Thailand

[6] Plant Molecular and Bioinformatics Center and Department of Biological Sciences, Northern Illinois University, 1425 W. Lincoln Hwy, DeKalb, IL 60115-2861, USA

[7] Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, Missouri 63132, USA

[8] Laboratoire d'Ecologie Alpine, CNRS – Université Grenoble Alpes, UMR 5553, Grenoble, France

[9] Department of Biological Sciences, The University of Alabama, 500 Hackberry Lane, Tuscaloosa, AL 35487

* These authors contributed equally to this work

** Corresponding author: guillaume.besnard@univ-tlse3.fr

Abstract

C_4 photosynthesis is a complex trait that sustains fast growth and high productivity in tropical and subtropical conditions and evolved repeatedly in flowering plants. One of the major C_4 lineages is Andropogoneae, a group of ~ 1,200 grass species that includes some of the world's most important crops and species dominating tropical and some temperate grasslands. Previous efforts to understand C_4 evolution in the group have compared a few model C_4 plants to distantly related C_3 species, so that changes directly responsible for the transition to C_4 could not be distinguished from those that preceded or followed it. In this study, we analyse the genomes of 66 grass species, capturing the earliest diversification within Andropogoneae as well as their C_3 relatives. Phylogenomics combined with molecular dating and analyses of protein evolution show that many changes linked to the evolution of C_4 photosynthesis in Andropogoneae happened in the Early Miocene, between 21 and 18 Ma, after the split from its C_3 sister lineage, and before the diversification of the group. This initial burst of changes was followed by an extended period of modifications to leaf anatomy and biochemistry during the diversification of Andropogoneae, so that a single C_4 origin gave birth to a diversity of C_4 phenotypes during 18 million years of speciation events and migration across geographic and ecological spaces. Our comprehensive approach and broad sampling of the diversity in the group reveals that one key transition can lead to a plethora of phenotypes following sustained adaptation of the ancestral state.

One of the major goals of evolutionary biology is to understand the origins of key innovations underlying the ecological success of particular groups. This requires the study of adaptive traits in a phylogenetic context, using comparisons of species differing in character states (e.g. Bond et al. 2014; Rainford et al. 2014; McGee et al. 2015; Sánchez-García and Matheny 2017). Because species differ in numerous ways, such comparisons must capture the diversity that emerged after the transition in addition to the diversity that preceded it, to precisely identify the properties directly involved in the origin of the trait of interest (e.g. Sprent 2007; Yukawa 2009; Endress 2011; Puttick 2014; Marek and Moore 2015; Clark et al. 2018). Among flowering plants, C_4 photosynthesis represents an adaptive novelty with significant ecological consequences (Sage 2004; Edwards et al. 2010; Christin and Osborne 2014).

The C_4 physiology results from multiple anatomical and biochemical modifications of the ancestral C_3 photosynthetic metabolism, which include (1) the confinement of the primary enzyme of the photosynthetic carbon reduction pathway, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), to a compartment isolated from the atmosphere, and (2) increased and cell-specific activity of several enzymes that concentrate CO_2 at the site of Rubisco (Hatch 1987; von Caemmerer and Furbank 2003; Sage 2004). The concentration of CO_2 around Rubisco boosts photosynthetic efficiency, and therefore growth, particularly in high-light, warm and dry conditions (Long 1999; Atkinson et al. 2016).

Although the C_4 trait requires the modification of multiple components, it has evolved at least 62 times independently during the diversification of flowering plants (Sage et al. 2011). The grass family (Poaceae) encompasses almost half of the C_4 origins, including some with particular ecological and economic relevance, such as the Andropogoneae (Sage 2017). The roughly 1,200 species of this tribe are all C_4, making it the second-most speciose (Sage et al. 2011) and the most dominant C_4 lineage (Lehmann et al. 2019). Andropogoneae include some of the world's most important cereal and fuel crop plants, such as maize, sorghum, sugarcane and *Miscanthus* spp. (e.g. silver grass), as well as numerous species that dominate tropical savannas and some temperate

grasslands, including the tallgrass prairies of the Great Plains (Hartley 1958; Bond et al. 2008; Edwards et al. 2010; Kellogg 2015). Besides generating some of the most productive plants in the world, their C_4 trait also increased the diversification of Andropogoneae, which in turn has shaped ecosystems around the world (Osborne 2008; Edwards et al. 2010; Forrestelet al. 2014; Spriggs et al. 2014; Sage and Stata 2015). Because (1) they are separated from other C_4 grass lineages in the phylogeny by several C_3 branches (GPWG II 2012), and (2) the different C_4 lineages differ in the underlying genetic changes, Andropogoneae are accepted as a C_4 origin independent from those in other groups of grasses (Sinha and Kellogg 1996; Christin et al. 2008, 2010; Vicentini et al. 2008; Edwards and Smith 2010; Sage et al. 2011; GPWG II 2012; Emms et al. 2016; Gallaher et al. 2019; Niklaus and Kelly 2019).

Due to their economic and ecological importance, Andropogoneae have been included in most studies addressing the evolutionary origins of C_4 photosynthesis in grasses. In particular, efforts to determine the genomic changes involved in the transition to C_4 photosynthesis have focused on comparisons between the two C_4 model Andropogoneae species – maize and sorghum – and distantly related C_3 model grasses (e.g. rice and *Dichanthelium*; Paterson et al. 2009; Wang et al. 2009; Emms et al. 2016; Studer et al. 2016; Huang et al. 2017). Such a narrow taxon sampling neither covers the diversity of anatomical and biochemical components observed among C_3 grasses (e.g. Hattersley 1984; Christin et al. 2013; Lundgren et al. 2014) nor among C_4 grasses within the Andropogoneae themselves (Renvoize 1982a; Ueno 1995; Sinha and Kellogg 1996). Yet, sampling this diversity is crucial for distinguishing those modifications involved in the early origin of the C_4 pathway in the group as opposed to its subsequent diversification (Christin et al. 2010; Christin and Osborne 2014; Dunning et al. 2017a; Heyduk et al. 2019). Differentiating these scenarios is necessary to determine whether changes for C_4 evolution were concentrated during the initial history of large C_4 groups, or whether they were sustained throughout the diversification of large C_4 clades, as suggested for young C_4 lineages (Dunning et al. 2017a) and hypothesized based on previously available evidence (Christin and Osborne 2014; Heyduk et al. 2019).

The first divergence within Andropogoneae *sensu* Kellogg (2015) separates the subtribe Arundinellinae from Andropogoneae *s.s.* (tribes Arundinelleae and Andropogoneae, respectively, in Soreng et al. 2017), the latter of which includes the model species maize and sorghum. Until recently, the closest known C_3 relative of Andropogoneae belonged to a different tribe that also included several C_4 groups, and the branch separating them was consequently long (Christin et al. 2008; Vicentini et al. 2008; GPWG II 2012; Kellogg 2015; Soreng et al. 2017). However, the C_3 genera *Jansenella* and *Chandrasekharania* have been recently suggested as the sister group of Andropogoneae based on individual chloroplast or nuclear markers (Besnard et al. 2018; Hackel et al. 2018). High-quality genomes are not available for species from these genera or from Arundinellinae, but low-coverage genome data have recently provided insights into the evolution of the nuclear genome in other non-model grasses (Besnard et al. 2014, 2018; Olofsson et al. 2016; Bianconi et al. 2018; Dunning et al. 2019). Capitalizing on the availability of such genomic datasets as a side-product of plastome sequencing (e.g. Washburn et al. 2015; Burke et al. 2016; Arthan et al. 2017; Piot et al. 2018), we are now able to phylogenetically track the modifications underlying one of the major innovations of flowering plants.

In this study, we analyse genome-skimming data for 66 grass species covering the diversity of C_3 relatives of Andropogoneae, as well as the earliest diversification within the group, to test the hypothesis that C_4 evolution was sustained throughout the history of old, large C_4 lineages. First, we generate plastome and genome-wide nuclear phylogenetic trees of grasses to confirm the relationships between Andropogoneae and their C_3 relatives, and estimate the age of C_4 photosynthesis in the group from a time-calibrated plastome phylogeny. Anatomical traits are then mapped onto the time-calibrated phylogeny to infer the timing of leaf structural transitions in the group. Finally, we look for signatures of adaptive evolution in key C_4 enzymes, testing whether shifts in selective pressures on protein sequences occurred (i) in a C_3 context and therefore predated the origin of Andropogoneae, (ii) at the base of the clade, during a short period of time representing the initial transition from C_3 to C_4 photosynthesis, or (iii) were sustained throughout the history of

the group, representing a prolonged period of gradual innovation within the monophyletic C_4 Andropogoneae. Overall, our study presents a comprehensive approach to dissecting a complex adaptive trait and inferring the tempo of key phenotypic transitions in a large group of ecological importance.

MATERIAL AND METHODS

Species Sampling and Sequencing

A dataset of whole-genome sequences of varied sequencing depth was assembled that covers: the main lineages of Andropogoneae including the subtribe Arundinellinae and the Andropogoneae *s.s.* (*sensu* Kellogg 2015), which represents the earliest known split within this C_4 group (GPWG II 2012); their putative closest C_3 relatives; a variety of other C_3 and C_4 Panicoideae; and representatives of the other grass subfamilies (Table S1). In total, genomic data for 59 grass species were retrieved from previous studies (Besnard et al. 2013, 2018; Lundgren et al. 2015; Burke et al. 2016; Arthan et al. 2017; Dunning et al. 2017b, 2019; Olofsson et al. 2016; Piot et al. 2017; Silva et al. 2017), and similar data for seven species were generated here (Table S1). For the latter, low-coverage sequencing was performed using Illumina technology. Genomic DNA (gDNA) was isolated from ca. 5–10 mg of leaf material using the BioSprint 15 DNA Plant Kit (Qiagen). Four herbarium samples were sequenced at the Genopole platform (Toulouse, France) while the three remaining samples were extracted from silica-preserved material and sequenced at the Genoscope platform (Evry, France). For all samples analysed at the Genopole, between 100 and 500 ng of double stranded DNA were used to construct sequencing libraries with the Illumina TruSeq Nano DNA LT Sample Prep kit (Illumina, San Diego, CA, USA), following the manufacturer's instructions (for more details, see Besnard et al. 2018). Each sample was multiplexed with samples from the same or different projects and paired-end sequenced on 1/24th of an Illumina HiSeq3000 lane (Table S1). At the Genoscope, libraries were constructed using 250 ng of sonicated gDNA. Fragments were end-repaired and 3'-adenylated. NextFlex DNA barcodes

(Bioo Scientific Corporation, Austin, TX, USA) were then added using the NEBNext DNA Modules Products (New England Biolabs, MA, USA) followed by clean up with 1x AgencourtAMPure XP (Beckman Coulter, Brea, CA, USA). The ligated product was amplified with 12 PCR cycles using Kapa HiFi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA) followed by a 0.6x Agencourt AMPure XP purification. Each sample was multiplexed with samples from a different project and paired-end sequenced on 1/48th of an Illumina HiSeq2000 lane (Illumina, USA; Table S1).

Plastome Analyses

A total of 51 plastome sequences were retrieved from NCBI and another 15 were assembled in this study using the genomic datasets (Table S1). For those assembled here, published plastomes of closely related species (same genus) were retrieved from NCBI and used as references for read mapping using Geneious v.9.1.8 (Kearse et al. 2012; Biomatters Ltd., Auckland, NZ) with the Geneious Read Mapper and default sensitivity. A consensus sequence was then called using the highest-quality base criterion, with indels being manually extended/reduced by the assembly of iteratively mapped reads. In cases where no plastome models were available for congeners, a *de novo* strategy was applied using the software Org.Asm v.1.0 (https://git.metabarcoding.org/org-asm/org-asm) with default parameters. Potential errors in the *de novo* assembly were corrected by mapping the genomic reads to the assembled sequence using Geneious following the strategy described above.

The 66 plastome sequences were aligned with MAFFT v.7.13 (Katoh and Standley 2013), after excluding the second inverted repeat region to avoid representing the same sequence twice. Plastome phylogenetic trees were inferred independently for coding and non-coding regions of the alignment using MrBayes v.3.2.6 (Ronquist et al. 2012) with the GTR+G model. Two analyses were run in parallel and were stopped after reaching a standard deviation of splits < 0.01. A consensus tree was obtained after a burn-in period of 25% and trees were rooted on the BOP clade

(Bambusoideae, Oryzoideae and Pooideae), which is sister to the large PACMAD clade that contains the Andropogoneae (GPWG II 2012). All plastome alignments are available on the Dryad data repository (doi: 10.5061/dryad.q114rt1). Plastome sequences assembled here were deposited in NCBI (see Table S1 for accession numbers).

Genome-Wide Nuclear Analyses

Because many of the genomic datasets used here have sequencing depths below the minimum required by existing software (Table S1 and S2; Bertels et al. 2014; Allen et al. 2018), we have adapted pipelines previously used to obtain nuclear phylogenetic trees from genome skimming data (Olofsson et al. 2016, 2019; Dunning et al. 2017b, 2019). The general approach consists of extracting nucleotide sequences from each genomic dataset by mapping reads onto a reference, which is analogous to existing approaches (Allen et al. 2018). However, all positions receiving mapped reads were considered here regardless of the coverage, to allow genome skimming samples to be incorporated. Sets of coding sequences (CDSs) were used as references, as they represent the portion of the genome that is sufficiently conserved to allow mapping among distant relatives (Olofsson et al. 2016, 2019), and are present at low copy numbers.

A genome-wide reference dataset of putative orthologous sequences of grasses was prepared using the complete CDS datasets of three model grasses representing different degrees of divergence from the Andropogoneae: *Sorghum bicolor* (part of the focus group)*, Setaria italica* (different tribe in the same subfamily as the focus group) and *Brachypodium distachyon* (different subfamily). These CDSs were retrieved from Phytozome v.12 (Goodstein et al. 2012). Putative one-to-one orthologs were identified using the BLAST reciprocal best hits (RBH) tool as implemented in Galaxy (Cock et al. 2015). Only CDSs that corresponded to the intersection of the RBH among the three species and which were > 500 bp were retained. Genes potentially transferred from organelles to the nuclear genome were identified via BLAST searches (e-value = 10^{-6}) using *S. bicolor* organellar genomes as reference, and subsequently removed from this dataset. The nuclear genome-wide reference dataset

consisted of 9,161 putative orthologs. Each of these genes is expected to descend from a single gene in the common ancestor of the two main groups of grasses, the BOP and PACMAD clades, but might have been lost or duplicated in some derived groups. Collapsing such duplicates allows the extraction of phylogenetically useful markers. Downstream analyses were conducted using the *S. italica* sequence for each group of orthologs, which is closely related, yet outside of the focus group and therefore maximizes the evenness of the coverage in the group (Fig. S1).

To minimize the amount of missing data, the nuclear analyses focused on the species of Andropogoneae, Jansenelleae, Paspaleae for which the estimated sequencing depth was above 1x. *Arundinella nepalensis* was added despite an estimated depth of 0.8x since it is one of only three Arundinellinae for which sequence data were available. Four species outside of Panicoideae for which high-coverage data were available were added to root the tree. Gene models corresponding to each of the 9,161 putative orthologs were assembled independently for each of the 37 grass species included in this reduced dataset. First, raw genomic datasets were filtered using the NGSQC Toolkit v.2.3.3 (Patel and Jain 2012) to retain only high-quality reads (i.e. > 80% of the bases with Phred quality score > 20), and to remove adaptor contamination and reads with ambiguous bases. The retained reads were subsequently trimmed from the 3' end to remove bases with Phred score < 20. The cleaned genomic datasets were then mapped as unpaired reads to the genome-wide CDS reference using Bowtie2 v.2.3.2 (Langmead 2012) with default parameters, which map reads identical on 90% or more bases, independently of the read length. Consensus sequences were called based on variant call format (VCF) files from read alignments with mapping quality score > 20 using the *mpileup* function of Samtools v.1.5 (Li et al. 2009) implemented in a bash-scripted pipeline, modified from Olofsson et al. (2016, 2019; available on the Dryad data repository). Sites with nucleotide variation among mapped reads were coded as ambiguous bases following IUPAC codes. Consensus sequences shorter than 200 bp were removed from the dataset. Sites within each alignment with more than 50% missing data were then trimmed using trimAl v1.4 (Capella-

226 Gutiérrez et al. 2009). Only gene alignments ≥ 300 bp (with ≥ 200 bp per individual sequence) and

227 containing ≥ 50% of the total number of species after trimming were retained for subsequent

228 analyses. Maximum-likelihood (ML) trees were then inferred for each gene alignment using

229 RAxML v.8.2.4 (Stamatakis 2014), with a GTR+CAT substitution model and 100 bootstrap

230 pseudoreplicates. To remove poorly informative markers, gene trees with < 50% of branches with

231 bootstrap support ≥ 50% were discarded. A multigene coalescent tree using Astral v.5.6.2 (Mirarab

232 et al. 2015) was then inferred using the filtered set of gene trees after collapsing branches with

233 bootstrap support values < 50. Because phylogenomic analyses can be biased by the reference and

234 the amount of missing data (Bertels et al. 2015; Xi et al. 2016; Olofsson et al. 2019), we repeated

235 the mapping and filtering with different filtering stringencies and an alternative reference species (*S.*

bicolor; Fig. S1).

In addition to the genome-wide dataset, eight individual nuclear markers previously used to

infer grass phylogenies (GPWG 2001; Bomblies and Doebley 2005; Doust et al. 2007; Christin et

al. 2012a; Estep et al. 2012, 2014) were investigated, namely *aberrant panicle organization 1*

(*apo1*), the gene encoding arogenate dehydrogenase (*arodeh*), the DELLA protein-encoding gene

dwarf 8 (*dwarf8*), *floricaula/leafy-like* (*floricaula*), *knotted 1* (*kn1*), the gene encoding phytochrome

B (*phyB*), *retarded palea 1* (*rep1*) and the gene encoding granule-bound starch synthase 1 (GBSSI

or *waxy*). Sequences of these genes were manually assembled for *Garnotia stricta* var. *longiseta*

(Arundinellinae) and the putative C_3 sister group of Andropogoneae (i.e. *Jansenella* and

Chandrasekharania) using the reference-guided approach from Besnard et al. (2018). In brief,

CDSs from *S. bicolor* were used as seeds to map reads, which were assembled into contigs by

recursively incorporating pairs of reads that overlapped on at least 30 bp. All sequences of nuclear

markers assembled here were deposited in NCBI (see Table S3 for accession numbers). Preliminary

visualization of read alignments for *Jansenella neglecta* suggested two divergent copies for all

genes, but the low sequencing depth for this accession prevented phasing the reads into distinct

copies. We therefore did not include *J. neglecta* in the phylogenetic analyses of nuclear markers. It

is worth mentioning, however, that the short segments that were recovered indicated that one of the

copies was very similar to the sequence of *J. griffithiana,* suggesting a hybrid (e.g. allopolyploid) origin of *J. neglecta*. The assembled genomic sequences were aligned with additional data retrieved from NCBI nucleotide databases using MAFFT. Phylogenetic trees were inferred for each of the eight markers using MrBayes, running two parallel chains for 40,000,000 generations. Run convergence and appropriateness of the burn-in period were verified using Tracer v.1.6 (Rambaut and Drummond 2007). The burn-in period was then set to 10,000,000, and a majority-rule consensus was inferred from the posterior trees.

Molecular Dating

Divergence times were estimated for the plastome dataset using a relaxed molecular clock as implemented in BEAST v.1.8.4 (Drummond and Rambaut 2007). The plastome alignment was reduced to coding sequences (57,239 bp), to remove intergenic spacers that undergo a large number of insertions and deletions, and are more difficult to align. The phylogenetic tree was time-calibrated by fixing the age of the split between the PACMAD and BOP clades to 51.2 Ma (based on Christin et al. 2014), using a normal distribution with standard deviation of 0.0001. This age represents the scenario based on macrofossils only, but we also report the equivalent ages from a dating scenario including phytoliths (82.4 Ma for the same node; Christin et al. 2014). These microfossils are abundant in the fossil record, but assigning them to modern lineages of grasses is complicated by their restricted number of characters (Prasad et al. 2005, 2011; Christin et al. 2014; Stromberg et al. 2018; for a discussion on the fossil record of grasses, see Kellogg 2015). The GTR+G substitution model was used, with the Yule model as speciation prior and a lognormal uncorrelated relaxed clock (Drummond et al. 2006). Three MCMC chains were run in parallel for at least 250 million generations , sampling every 10,000 generations. The runs were monitored using Tracer v.1.6 (Rambaut et al. 2013), checking for convergence and effective sample sizes > 100 for all parameters. The burn-in period was set to the point of convergence of the runs (25%) and all trees sampled after that were combined. Median ages were summarized on the maximum clade

credibility tree.

Carbon Isotopes and Leaf Anatomy

Photosynthetic types for most species were retrieved from the literature (Osborne et al. 2014; Table S4). The photosynthetic type of *J. griffithiana* was verified through analysis of carbon isotopes. Leaf fragments from the sequenced herbarium specimen were analysed using an ANCA GSL preparation module coupled to a Sercon 20-20 stable isotope ratio mass spectrometer (PDZ Europa, Cheshire, UK). The carbon isotopic ratio ($\delta^{13}C$, in ‰) was reported relative to the standard Pee Dee Belemnite (PDB). Values of $\delta^{13}C$ ranging from -33 to -24‰ are typical of C_3 plants, and values higher than -17‰ indicate that the plants grew using a C_4 pathway (O'Leary 1988).

Leaf anatomical phenotypes were recorded for members of Andropogoneae and their C_3 relatives, using data from the literature (Table S4; Renvoize 1982a, 1982b, 1982c, 1985; Watson et al. 1992; Ueno 1995; Zuloaga et al. 2000; Christin et al. 2013). In addition, new leaf cross sections were prepared for the herbarium samples of *J. griffithiana* and *G. stricta* used for genome sequencing. A leaf fragment (ca. 2 cm) was rehydrated by warming the sample in dH_2O up to 60°C followed by immersion in 1% KOH overnight. The rehydrated fragment was then dehydrated through an ethanol series from 10% to 100% EtOH, with steps of 30 min, and resin-infiltrated with Technovit 7100 (Heraeus Kulzer GmbH, Wehrheim, Germany). Cross sections of 9 μm were obtained using a microtome (Leica RM 2245, Leica Biosystems Nussloch GmbH, Nussloch, Germany) and stained with Toluidine Blue O (Sigma-Aldrich, St. Louis, MO). Micrographs were obtained using an Olympus BX51 microscope coupled to an Olympus DP71 camera (Olympus Corporation, Tokyo, Japan). A number of qualitative and quantitative leaf characters related to the C_4 function were measured on the cross sections following Christin et al. (2013): number of bundle sheath layers, distance between the centres of consecutive veins (interveinal distance), minimal distance between the bundle sheaths of consecutive veins (bundle sheath distance), fraction of the mesophyll plus bundle sheath area represented by the inner bundle sheath (% inner sheath area),

presence/absence of distinctive cells (*sensu* Tateoka 1958; Renvoize 1982b), and localization of starch production.

Analyses of Protein Sequence Evolution

To test for episodes of adaptive evolution of C_4 enzymes during different periods of the history of Andropogoneae, branch model tests using the ratio of nonsynonymous to synonymous substitutions rates (dN/dS; Yang 1998; Yang and Nielsen 1998) were conducted on alignments of five genes encoding proteins known to play important roles in the C_4 pathway (Hatch 1987; Huang et al. 2017): NADP-malate dehydrogenase (NADP-MDH; gene *nadpmdh-1P1*), NADP-malic enzyme (NADP-ME; gene *nadpme-1P4*), phosphoenolpyruvate carboxykinase (PCK; gene *pck-1P1*), phosphoenolpyruvate carboxylase (PEPC; gene *ppc-1P3*) and pyruvate, phosphate dikinase (PPDK; gene *ppdk-1P2*). To test whether shifts in selective pressures could be related to processes other than C_4 evolution, an alternative set of 12 genes not known to be involved in C_4 photosynthesis were used as negative controls. These included some paralogs of the same core C_4 genes for which sequences were available in NCBI database and the individual nuclear markers used for phylogenetic analyses (see above; except *apo1*, *rep1* and *floricaula*, for which no C_3 PACMAD species besides *Jansenella* and *Chandrasekharania* was available). For each of these 17 genes, complete or partial coding sequences for the putative C_3 sister group of Andropogoneae and *G. stricta* were manually assembled using the approach described above. Additional sequences were extracted using BLAST (e-value $= 10^{-9}$) from (1) the CDS dataset of seven published genomes (*S. bicolor, Zea mays, S. italica, Panicum hallii, Panicum virgatum*, *B. distachyon* and *Oryza sativa*) retrieved from Phytozome v.12, (2) the NCBI nucleotide database, and (3) the transcriptomes of 34 PACMAD species retrieved from Washburn et al. (2017). All sequences from each gene were aligned using MAFFT and the alignment was visually inspected. Low-confidence alignment regions containing indels were removed to avoid erroneously inflating estimates of nonsynonymous substitutions. Sequences from multiple accessions of the same species, paralogs and sequences

containing stop codons or frameshift mutations were also removed before the analysis. The 3rd

positions of codons were used for phylogenetic inference to decrease biases due to adaptive

evolution (Christin et al. 2012b). Phylogenetic trees were obtained using Bayesian inference with

MrBayes as described above for the individual nuclear markers. Branch model tests were conducted

using the consensus gene trees without collapsing unsupported nodes. These tests were repeated on

the species tree obtained from the multigene coalescent analysis (see above), after pruning species

for which sequences of the gene were not available. The duplication events inferred from the gene

tree were incorporated into this phylogeny by duplicating the corresponding branches. C_4 species

outside Andropogoneae were pruned from all trees before analyses to avoid either (a) inflating the

dN/dS estimate for the background branches as a result of independent selection signals in other C_4

groups, or (b) underestimating dN/dS in the foreground branches by misidentifying the paralog used

for C_4 photosynthesis in these other taxa.

A number of branch models were optimized using *codeml* as part of PAML v. 4.9 (Yang

2008). The null model, which assumes a single dN/dS ratio for all branches, was compared to

several branch models that hypothesized a different dN/dS ratio (i.e. shift in the selective pressure)

in a set of foreground branches defined *a priori*: (1) the branch leading to Andropogoneae and its C_3

sister group (shifts in selective pressures before the transition to C_4); (2) the branch leading to

Andropogoneae (shifts in selective pressures during the transition to C_4); and (3) the branches

leading to each of the two main Andropogoneae groups Arundinellinae and Andropogoneae *s.s.*

(two independent shifts in selective pressures just after the transition to C_4). Each model was

repeated with a sustained shift in selective pressures from the selected branches to all descendants.

The best model was selected using the Akaike Information Criterion (AIC), after verifying that it

was significantly better than the null model (at a significance level of 5%) as assessed via a

likelihood ratio test, with a *p*-value adjusted for multiple testing using the Bonferroni correction.

The number of amino acid substitutions through time was assessed by estimating via ML

the branch lengths on the amino acid alignment while constraining the topology to that obtained on

3rd positions of codons. This was performed for all genes analysed here, using IQ-tree v.1.6.1 (Nguyen et al. 2015) with an automated selection of the model of protein sequence evolution.

RESULTS

Plastome and Nuclear Datasets

The plastome alignment of 66 species was 140,427 bp long, of which 56,991 bp corresponded to CDS. The mean estimated sequencing depth for the plastomes ranged from 90 to 4,602 reads per site across species. The nuclear dataset consisted of 37 species and 365 genes. The alignments were on average 701 bp long (95% range = 370-1532 bp, total = 255,870 bp) and 70% complete (95% range = 63-80%), with an average of 64 parsimony informative sites (95% range = 27-163, total = 23,367; Fig. S1). As expected, less stringent filtering parameters allowed more genes to be retained, and more parsimony informative sites per gene, but resulted in higher amounts of missing data (Fig. S1). The number of genes producing resolved phylogenetic trees was drastically reduced when using the full set of 66 species (Fig. S1f-j). The number of genes retained after filtering when using an alternative reference species (*Sorghum bicolor*) was on average 62% higher, but similar patterns of missing data across datasets were observed.

Plastome and Nuclear Phylogenetic Trees

The phylogenetic trees inferred from plastomes and nuclear genomes were largely congruent with previous studies, with discrepancies between the two types of markers as previously reported (GPWG II 2012; Washburn et al. 2017; Moreno-Villena et al. 2018; Dunning et al. 2019). The multigene coalescent tree was generally congruent with the plastid phylogeny. This nuclear analysis revealed gene discordance for many nodes, which indicates incomplete lineage sorting and possibly hybridization in some parts of the family (Dunning et al. 2019). The relationship between Andropogoneae and its C_3 relatives was however consistent among the plastome and nuclear datasets, with the C_3 genera *Jansenella* and *Chandrasekharania* forming a strongly supported group

sister to Andropogoneae (Figs 1, S2, S3 and S4). This relationship was also highly supported by all nuclear datasets obtained with different filtering thresholds (Table S5; Fig. S3), as well as in all trees inferred from individual nuclear markers (Fig. S4), except for two genes, in which *Jansenella* and *Chandrasekharania* formed a paraphyletic group (*apo1*; Fig. S4a) or were nested within Andropogoneae (*floricaula*; Fig. S4d). Our data and analyses therefore provide strong evidence that the clade formed by *Jansenella* and *Chandrasekharania* (hereafter Jansenelleae) is the extant C_3 lineage most closely related to the Andropogoneae grasses.

Within Andropogoneae, the genera *Garnotia* and *Arundinella* (subtribe Arundinellinae) either formed a group that was sister to Andropogoneae *s.s.* (Figs S2, S3, S4f,h and S5b,c), or were paraphyletic (Figs 1 and S4c,d,g and S5a). Short internal branches, incongruence in the multigene coalescent trees (Figs 1 and S3), and low posterior probability support values (Fig. S4) within Andropogoneae *s.s.* are associated with high incongruence between nuclear and plastome trees, suggesting a complex history for the group, which might be related to a rapid radiation and frequent hybridization (Estep et al. 2014). In particular, nuclear and plastome trees identify different taxa as sister to the rest of Andropogoneae *s.s.* (*Zea mays* in the multigene coalescent trees, *Lasiurus scindicus* in the plastome trees; Figs 1, S2 and S3).

Divergence Time Estimates

The confirmation of the sister relationship between Jansenelleae and Andropogoneae allows for refined divergence time estimates, as the most recent divergence from a C_3 relative (stem group node) represents the upper bound for the origin of a trait that could have evolved at any point along the branch leading to the most recent common ancestor of all species with the trait of interest (crown group node). Divergence times were estimated using the plastome dataset (coding sequences). Based on a secondary calibration considering only macrofossils, the divergence between Andropogoneae and its C_3 sister lineage was estimated at 21.1 (95% HPD = 14.6 – 27.6) Ma (Table 1; Fig. S5). These dates would be pushed back to 34 (95% HPD = 23.5 – 44.4) Ma if a

microfossil dating scenario was followed. The first split within Andropogoneae was estimated at 17.9 (95% HPD = 12.2 – 23.7) Ma (28.8 Ma under a microfossil dating scenario).

Anatomical Changes During the Early Diversification of Andropogoneae

Anatomical and biochemical characters linked to C_4 photosynthesis were recorded based on the literature and on new measurements for *G. stricta* and *J. griffithiana* (Table 2; Table S4). Our carbon isotope analysis confirmed that *J. griffithiana* is a C_3 plant ($\delta^{13}C$ = -27.28‰). Its leaf anatomy is typical of C_3 grasses, with two layers of bundle sheath cells (Fig. 2a), which contradicts previous reports of a single sheath (Metcalfe 1960; Türpe 1970). As with other C_3 species, it has a large distance between consecutive bundle sheaths, and no minor veins or distinctive cells (Fig. 2b). In addition, the proportion of the leaf occupied by the inner bundle sheath falls within the range observed for other C_3 grasses (Christin et al. 2013). The leaf anatomy of *G. stricta* was similar to that previously reported for other Arundinellinae (Renvoize 1982c, 1986; Watson et al. 1992). Its veins are surrounded by a single bundle sheath and are separated by a large number of mesophyll cells (Fig. 2b). Multiple distinctive cells separate the veins, and staining suggests starch production in both bundle sheaths and distinctive cells (Fig. 2b). Similar anatomical structures are observed in other Arundinellinae (Table S4; Renvoize 1982c, 1986; Watson et al. 1992), but also in the genus *Arthraxon* (Ueno 1995), which is an Andropogoneae *s.s.* representative that diverged early from the rest of the group (e.g. GPWG II 2012; Estep et al. 2014). By contrast, most Andropogoneae *s.s.* lack distinctive cells and decrease the distance between consecutive veins via the proliferation of minor veins (Fig. 2b; Table 2; Table S4).

Positive Selection in C_4 Enzymes

Phylogenetic trees for genes encoding C_4 enzymes inferred from 3rd positions of codons were compatible with plastome and nuclear genome trees, with most of the variation being observed within Andropogoneae *s.s.* and Paniceae (Fig. S6). In all cases, Jansenelleae was sister to

Andropogoneae, except in *pck-1P1*, where Jansenelleae formed a group with Arundinellinae that was sister to Andropogoneae *s.s.* (Fig. S6c). Arundinellinae and Andropogoneae *s.s.* represented the first split within Andropogoneae in *nadpmdh-1P1*, *nadpme-1P4* and *ppc-1P3*, but not in *ppdk-1P2*, in which Arundinellinae is paraphyletic (Fig. S6e) and *pck-1P1*, as mentioned above. Lineage-specific duplications are observed within Andropogoneae *s.s.* and Arundinellinae species for *nadpme-1P4*, and only in Andropogoneae *s.s.* for *nadpmdh-1P1* (Fig. S6a,b), as previously reported (Rondeau et al. 2005; Christin et al. 2009a; Wang et al. 2009).

The inferred trees were used to track shifts in selective pressures in Andropogoneae, independently for each gene. In three core C_4 genes analysed (*nadpmdh-1P1*, *nadpme-1P4* and *ppc-1P3*), the best model inferred an increase of the dN/dS ratio after the split between Andropogoneae and Jansenelleae, which was sustained in the descendant branches (Table 3). As compared to the background branches, the estimated dN/dS ratio was 2 to 3.7 times higher in Andropogoneae for these genes. In *pck-1P1* and *ppdk-1P2*, the best model indicates two independent three- and four-fold increases of the dN/dS ratio at the base of each of Arundinellinae and Andropogoneae *s.s.*, which was sustained in the descendant branches in *pck-1P1*, but not in *ppdk-1P2*.

A shift of the dN/dS ratio was identified in three of the 12 other genes used as negative controls (Table 3). In the case of *nadpme-1P3*, a gene encoding a NADP-ME isoform not involved in the C_4 cycle of any of the previously screened species (Moreno-Villena et al. 2018), the best model indicates an increase in dN/dS in the branch leading to the most recent common ancestor of Andropogoneae and Jansenelleae (Table 3). In *dwarf8*, an increase of dN/dS occurred on branches leading to each of the Arundinellinae and Andropogoneae *s.s.* and was sustained in the descendant branches. Finally, the best model for *ppc-1P7*, a gene encoding a PEPC isoform also not co-opted for the C_4 cycle in any species previously analysed (Moreno-Villena et al. 2018), assumed a decrease of dN/dS in the two branches leading to each of the Arundinellinae and Andropogoneae *s.s.* and their descendants. Therefore, out of the 12 negative controls, only *dwarf8* presents an increase of dN/dS that coincides with C_4 evolution in the group. This gene, which affects flowering

time (Thornsberry et al. 2001), has been linked to the adaptation of some Andropogoneae to varying climates (Camus-Kulandaivelu et al. 2006).

All positive selection tests were repeated using the multigene coalescent species tree topology (Fig. 1) instead of the trees inferred from 3rd positions of codons (Table S6). The results mostly confirmed those reported above, except that no shift of selective pressure was observed in the core C$_4$ genes *nadpmdh-1P1* and *pck-1P1*, nor in *nadpme-1P3* and *dwarf8*. Because sequences of the genes were available for species not included in the species tree, the sampling was reduced in these analyses compared to those based on the gene trees. The reduced evidence for positive shifts might therefore reflect a smaller statistical power of the tests based on the species tree.

To visualize the amount of amino acid substitutions during different periods of the Andropogoneae history, we estimated branch lengths from amino acid sequences after excluding C$_4$ species outside of Andropogoneae. Overall, numerous substitutions occurred in *nadpmdh-1P1* and *ppc-1P3* at the base of Andropogoneae, and increased rates compared to non-C$_4$ species on these genes were sustained throughout Andropogoneae (Figs 3 and 4). By contrast, bursts of amino acid substitutions in *nadpme-1P4* and *ppdk-1P2* occurred at the base of both Arundinellinae and Andropogoneae *s.s.* lineages, with the first coinciding with events of gene duplication. An increased number of substitutions is also observed in *pck-1P1*, but it was restricted to a few branches within these groups (Figs 3 and 4). The same patterns were observed when C$_4$ species outside Andropogoneae were included in the analyses (Fig. S7), as increased rates of amino acid substitution in all five genes characterize most C$_4$ grasses, which highlights the highly convergent nature of C$_4$ evolution in grasses. Similar bursts of amino acid substitutions were observed in the negative controls *nadpme-1P3* (on the branch leading to Jansenelleae and Andropogoneae) and *dwarf8* (in a few derived groups within Andropogoneae *s.s.*; Fig. S7).

DISCUSSION

A Single Origin of the New C$_4$ Physiology Followed by Continued Anatomical Changes

In previous grass phylogenetic trees, Andropogoneae formed a large clade entirely composed of C_4 species, and its closest known C_3 relatives belonged to a different group containing multiple independent C_4 lineages (GPWG II 2012; Gallaher et al. 2019). The branch leading to Andropogoneae was therefore long, preventing the precise inference of changes leading to C_4 evolution in this group. We confirm here that *Jansenella* and *Chandrasekharania* form the sister group of Andropogoneae, both based on plastomes and on markers spread across the nuclear genomes (Figs 1, S2-S5). This, combined with a distinctive morphology, supports their recognition as a separate tribe, Jansenelleae (Appendix 1). We further confirm that the group is C_3, as previously suggested (Türpe 1970; Renvoize 1985, 1986), providing a shorter branch connecting the last known C_3 ancestor of Andropogoneae (most recent ancestor shared with Jansenelleae) and the first split within the group. The anatomy of *Jansenella* is typical of C_3 grasses, with a large distance between consecutive veins, a double bundle sheath and no minor veins or distinctive cells (Fig. 2). In addition, the genes encoding C_4-related enzymes from *Jansenella* and *Chandrasekharania* are similar to those of other C_3 grasses, with no trace of positive selection or increased rates of amino acid replacement (Figs 3, 4, S7; Table 3). We therefore conclude that the last common ancestor of Jansenelleae and Andropogoneae was a typical C_3 plant, with the anatomical and genetic characteristics common to all PACMAD grasses (Christin et al. 2013; Emms et al. 2016; Moreno-Villena et al. 2018). The changes responsible for the emergence of a C_4 pathway therefore happened after the divergence between Andropogoneae and Jansenelleae. Previous studies comparing C_3 and C_4 anatomical traits or genomes typically sampled only a few Andropogoneae species, preventing them from assigning changes to different phases of C_4 evolution (Christin et al. 2013; Emms et al. 2016; Huang et al. 2017), as enabled here thanks to our denser species sampling.

The comparison of anatomical types suggests multiple modifications during the early diversification of Andropogoneae. All species in this group have a single bundle sheath (Renvoize 1982), which is ontogenetically equivalent to the inner sheath of C_3 grasses (i.e. the mestome

sheath; Dengler et al. 1985). The large distance between consecutive veins, as observed in *Jansenella* (Table 2), is reduced in Arundinellinae by the insertion of one or multiple distinctive cells, where Rubisco can be segregated (Fig. 2; Dengler and Dengler 1990; Sinha and Kellogg 1996). While these distinctive cells are shared by some Andropogoneae *s.s.* (Ueno 1995), most use a different strategy to reduce the distance between consecutive veins, which consists of the proliferation of minor veins (Table 2; Lundgren et al. 2014, 2019). Distinctive cells and minor veins have similar developmental patterns (Dengler et al. 1996), and the former could be precursors of the latter, in which case minor veins could represent the specialization of ancestral distinctive cells after the split of Andropogoneae *s.s.* from Arundinellinae. Alternatively, the ancestral state of the group could be minor veins that later degenerated in Arundinellinae and some Andropogoneae *s.s.*, or else these specializations evolved multiple times during the early diversification of the group. In all cases, the phylogenetic distribution of distinctive cells and minor veins shows that changes following the initial transition to C_4 led to diverse anatomical solutions for the effective segregation of biochemical reactions.

Modifications of C₄ Enzymes Occurred Throughout the Diversification of Andropogoneae

The emergence of a C_4 pathway generally requires the co-option of multiple enzymes already existing in the C_3 ancestor via their massive upregulation (Hibberd and Covshoff 2010; Moreno-Villena et al. 2018). This is followed by adaptation of their kinetics for the new catalytic context through numerous amino acid replacements (Blasing et al. 2002; Tausta et al. 2002; Christin et al. 2007; Huang et al. 2017). Tests of shifts in selective pressures conducted here for multiple C_4-encoding genes from Andropogoneae and other grasses confirm that the evolution of C_4 genes in this group involved an increased fixation of nonsynonymous mutations (Table 3; Christin et al. 2007, 2009; Wang et al. 2009; Huang et al. 2017). Genes for the key enzyme of the C_4 pathway, PEPC, underwent convergent changes in numerous groups of grasses, and most were shared between *Arundinella* and Andropogoneae *s.s.* (Christin et al. 2007). However, only a fraction of the

changes are also observed in *Garnotia stricta* (data not shown), indicating that the enzyme underwent adaptive changes both before and after the diversification of major Andropogoneae lineages. Consistent with this, the branch models did not favour increased amino acid replacements solely at the base of the whole clade, but a sustained increase throughout the diversification of the group (Table 3). A sustained shift in selective pressure on the branch leading to Andropogoneae after the split from Jansenelleae is also supported for genes encoding NADP-MDH and NADP-ME, while independent shifts are observed at the base of Arundinellinae and Andropogoneae *s.s.* for PCK (Table 3). A comparison of branch lengths indeed shows increased rates of amino acid replacements at the respective branches for all five core C_4 genes (Fig. 3). Our analyses therefore confirm that massive changes happened at the base of Andropogoneae, and models assuming that increased fixation of nonsynonymous mutations persisted after early shifts in selective pressures are strongly favoured for four out of the five core C_4 genes analysed (Table 3). In addition, increased rates of sustained amino acid replacements are observed on many branches within the group (Figs 3 and 4). Increased rates of amino acid replacements were also detected in genes not known to be directly involved in C_4 biochemistry, such as the NADP-ME paralog *nadpme-1P3*, and *dwarf8*. While the selective drivers for changes in *nadpme-1P3* are not known, the branches with elevated rates of nonsynonymous mutations do not strictly coincide with the C_4 phenotype (Fig. S7; Table 3). The gene *dwarf8* is linked to flowering time in maize, and selective sweeps in the genomic region including *dwarf8* have been associated with climatic adaptations in maize (Camus-Kulandaivelu et al. 2006, 2008). We conclude that, while other genes undergo elevated rates of amino acid substitutions for different reasons, important alterations of enzymes for the initial build-up of a C_4 cycle at the base of Andropogoneae were followed by continued adaptation throughout the diversification of the group.

While some enzymes participate in all biochemical variants of the C_4 cycle (Kanai and Edwards 1999), the identity of the enzyme(s) responsible for the decarboxylation of CO_2 in the bundle sheath varies among C_4 lineages (Prendergast et al. 1987; Sage et al. 2011). Our analyses

concordantly indicate that the decarboxylating enzyme PCK underwent rounds of amino acid replacements only in some derived groups within Andropogoneae (Figs 3 and 4), without evidence of positive selection at the base of the whole group (Table 3). This conclusion was reached previously (Christin et al. 2009b) and supports later additions of a PCK-catalyzed decarboxylation reaction in some of the Andropogoneae (Gutierrez et al. 1974; Walker et al. 1997; Wingler et al. 1999). However, our data also indicate that NADP-ME, which is the main decarboxylating enzyme in all Andropogoneae, similarly acquired its C_4 properties relatively late in the history of the group. Again, the best model assumed adaptive evolution throughout Andropogoneae (Table 3). The gene *nadpme-1P4* for NADP-ME was duplicated independently in Andropogoneae *s.s.*, *Garnotia* and *Arundinella*, and amino acid replacements are especially prevalent in one of the copies in each group (Figs. 3 and S7; Christin et al. 2009a). These observations point to independent adaptation of the enzyme kinetics, but the expression patterns also likely evolved independently in Andropogoneae *s.s.* and Arundinellinae. Indeed, modifications of the promoter regions allowing the C_4-specific binding of a transcription factor are restricted to one of the Andropogoneae *s.s.* duplicates that fulfils the C_4 function (Borba et al. 2018), which evolved after the split from Arundinellinae. We therefore hypothesize that the common ancestor of the Andropogoneae performed a C_4 cycle based on several decarboxylating enzymes relatively abundant in many C_3 grasses (Moreno-Villena et al. 2018), with some amino acid changes in the other C_4 enzymes. Further modifications, which canalized the use of the NADP-ME encoded by *nadpme-1P4*, added a PCK shuttle and/or improved the action of PEPC, PPDK and NADP-MDH happened later during the diversification of the group, so that its numerous C_4 species represent a diversity of realizations of the C_4 pathway. Similar conclusions were reached for small groups that evolved the C_4 trait more recently (Dunning et al. 2017a), but we show here for the first time that the continuous adaptation of the C_4 trait can be sustained over long evolutionary periods, leaving traces even within one of the largest C_4 groups.

C_4 Physiology Evolved During the Early Miocene in Andropogoneae

Besides inferring the changes underlying C_4 evolution in Andropogoneae, our plastome phylogeny encompassing a diversity of Andropogoneae and their closest C_3 relatives shed new light on the age of C_4 photosynthesis in the group. Our molecular dating estimated the split between Jansenelleae and Andropogoneae at roughly 22 Ma, with the first split within Andropogoneae at 18-19 Ma. While older ages would be inferred if disputed microfossils dates are considered (see Results), these dates represent the interval in which C_4 most likely evolved in this group, and are consistent with those obtained from previous studies (Christin et al. 2008, 2014; Vicentini et al. 2008; Estep et al. 2014; Spriggs et al. 2014; Dunning et al. 2017b).

Reconstructing the ancient biogeography of Andropogoneae is complicated by their diversity and presumably numerous long-distance dispersals, but India represents the centre of diversity of both Andropogoneae and Jansenelleae (Bor 1955; Hartley 1958; Nair et al. 1982; Yadav et al. 2010), suggesting an origin on the subcontinent. The three species of Jansenelleae occur in open habitats (Bor 1955; Nair et al. 1982; Yadav et al. 2010), including some that regularly burn (Shilla and Tiwari 2015), calling for more research to establish which ecological traits now typical for Andropogoneae had already emerged before the C_3 to C_4 transition and which only appeared afterwards.

The contrast between the sister groups Jansenelleae and Andropogoneae is striking. While the former has only three known species, two of them restricted to small regions of India, the latter encompasses roughly 1,200 species spread around the world, many of which are dominant in savanna ecosystems (Hulbert 1988; Solbrig 1996; Bond et al. 2003; Kellogg 2015). This difference is partially explained by the divergence of photosynthetic types, but the expansion of C_4 grasslands happened 7-15 Ma after C_4 originated in Andropogoneae (Edwards et al. 2010), and increased diversification occurred only in some of its subclades (Spriggs et al. 2014). While the initial C_4 trait might have played the role of a key innovation broadening the niche of early Andropogoneae (Lundgren et al. 2015; Aagesen et al. 2016), the later diversification and dominance of some

subgroups, their rapid dispersal across large distances (Dunning et al. 2017b) and into different ecosystems (Watcharamongkol et al. 2018) were likely enabled by the acquisition of additional attributes. Traits only partially related or entirely unrelated to C_4 photosynthesis, such as frequent allopolyploidy, herbivore resistance and fire tolerance have previously been used to explain the success of some Andropogoneae (Stebbins 1975; Bond et al. 2003; Edwards et al. 2010; Visser et al. 2012; Estep et al. 2014;Forrestel et al. 2014; Ripley et al. 2015; Linder et al. 2018). We suggest that the diversity of C_4 phenotypes revealed here might also contribute to variation among Andropogoneae. For instance, the addition of a PCK shuttle, which happened recurrently in some derived Andropogoneae, is predicted to increase tolerance to fluctuating light conditions (Bellasio and Griffiths 2014; Wang et al. 2014). Other anatomical and biochemical variations observed here might alter the hydraulic efficiency and growth rates of the different Andropogoneae (Osborne and Sack 2012). Overall, we conclude that, because of continuous adaptive reinforcement following a key physiological transition, descendants of a lineage sharing the derived trait should not all be considered as functionally equivalent.

CONCLUSION

Using plastome and nuclear phylogenomics, we confirmed a rare Asian C_3 lineage, Jansenelleae, as sister to the C_4 Andropogoneae grasses. This opens new avenues for comparative analyses of C_4 evolution, which were explored here. The C_4 pathway in Andropogoneae most likely evolved in the Early Miocene between roughly 22 and 18 Ma, and many adaptive changes in C_4 enzymes happened during this 4-My period, while many more occurred during the next 18 million years of lineage diversification. The group including Andropogoneae apparently originated on the Indian subcontinent, and the evolutionary diversification of the C_4 phenotype after its origin might have facilitated the spread of Andropogoneae into novel niches and to different regions of the globe, contributing to the success of this emblematic group of savanna grasses.

C.P.O., and P.T. contributed data. M.E.B., J.H., A.M. and G.B. analyzed the data. M.S.V. did the

taxonomic treatment. M.E.B., J.H., P.A.C. and G.B. wrote the paper with the help of all co-authors.

REFERENCES

Aagesen L., Biganzoli F., Bena J., Godoy-Bürki A.C., Reinheimer R., Zuloaga F.O. 2016. Macro-

climatic distribution limits show both niche expansion and niche specialization among C_4

Panicoids. PLoS One 11:e0151075.

Allen J.M., LaFrance R., Folk R.A., Johnson K.P., Guralnick R.P. 2018. aTRAM 2.0: An improved,

flexible locus assembler for NGS data. Evol. Bioinform. 14:1–4.

Arthan W., McKain M.R., Traiperm P., Welker C.A.D., Teisher J.K., Kellogg E.A. 2017.

Phylogenomics of Andropogoneae (Panicoideae: Poaceae) of mainland Southeast Asia. Syst.

Bot. 42:418–431.

Bellasio C., Griffiths H. 2014. The operation of two decarboxylases, transamination, and

partitioning of C_4 metabolic processes between mesophyll and bundle sheath cells allows light

capture to be balanced for the maize C_4 pathway. Plant Physiol. 164:466–480.

Bertels F., Silander O.K., Pachkov M., Rainey P.B., van Nimwegen E. 2014. Automated

reconstruction of whole-genome phylogenies from short-sequence reads. Mol. Biol. Evol.

31:1077–1088.

Besnard G., Christin P.-A., Malé P.-J.G., Coissac E., Ralimanana H., Vorontsova M.S. 2013.

Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from

Madagascar. Ann. Bot. 112:1057–1066.

Besnard G., Christin P.-A., Malé P.-J.G., Coissac E., Lhuillier E., Lauzeral C., Vorontsova M.S.

2014. From museums to genomics: old herbarium specimens shed light on a C_3 to

C_4transition. J. Exp. Bot. 65:6711–6721.

Besnard G., Bianconi M.E., Hackel J., Manzi S., Vorontsova M.S., Christin P.-A. 2018. Herbarium

genomics retraces the origins of C_4-specific carbonic anhydrase in Andropogoneae (Poaceae).

Bot. Lett. 165:419–433.

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 C_4 photosynthesis in the grass genus *Alloteropsis*. J. Exp. Bot. 69:1967–1980.

Bläsing O.E., Ernst K., Streubel M., Westhoff P., Svensson P. 2002. The non-photosynthetic phosphoenolpyruvate carboxylases of the C_4 dicot *Flaveria trinervia* – implications for the evolution of C_4 photosynthesis. Planta 215:448–456.

Bomblies K., Doebley J.F. 2005. Molecular evolution of FLORICAULA/LEAFY orthologs in the Andropogoneae (Poaceae). Mol. Biol. Evol. 22:1082–1094.

Bond W.J., Midgley G.F., Woodward F.I. 2003. What controls South African vegetation - climate or fire? S. Afr. J. Bot. 69:79–91.

Bond J.E., Garrison N.L., Hamilton C.A., Godwin R.L., Hedin M., Agnarsson I. 2014. Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. Curr. Biol. 24:1765–1771.

Bond W.J., Silander J.A., Ranaivonasy J., Ratsirarson J. 2008. The antiquity of Madagascar's grasslands and the rise of C_4 grassy biomes. J. Biogeogr. 35:1743–1758.

Bor N.L. 1955. Notes on Asiatic grasses: XXIII. *Jansenella* Bor, a new genus of Indian grasses. Kew Bull. 10:93.

Borba A.R., Serra T.S., Górska A., Gouveia P., Cordeiro A. M., Reyna-Llorens I.,Kneřová J., Barros P.M., Abreu I.A., Oliveira M.M., Hibberd J.M., Saibo N.J.M. 2018. Synergistic binding of bHLH transcription factors to the promoter of the maize NADP-ME gene used in C_4 photosynthesis is based on an ancient code found in the ancestral C_3 state. Mol. Biol. Evol. 35:1690–1705.

Burke S.V., Wysocki W.P., Zuloaga F.O., Craine J.M., Pires J.C., Edger P.P., Mayfield-Jones D., Clark L.G., Kelchner S.A., Duvall M.R. 2016. Evolutionary relationships in panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 16:140.

Camus-Kulandaivelu L., Veyrieras J.B., Madur D., Combes V., Fourmann M., Barraud S., Dubreuil P., Gouesnard B., Manicacci D., Charcosset A. 2006. Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *dwarf8* gene. Genetics172:2449–2463.

Camus-Kulandaivelu L., Chevin L.-M., Tollon-Cordet C., Charcosset A., Manicacci D., Tenaillon M.I. 2008. Patterns of molecular evolution associated with two selective sweeps in the *tb1-dwarf8* region in maize. Genetics 180:1107–1121.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Christin P.-A., Salamin N., Savolainen V., Duvall M.R., Besnard G. 2007. C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr. Biol. 17:1241–1247.

Christin P.-A., Besnard G., Samaritani E, Duvall M.R., Hodkinson T.R., Savolainen V., Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr. Biol. 18:37–43.

Christin P.-A., Samaritani E., Petitpierre B., Salamin N., Besnard G. 2009a. Evolutionary insights on C_4photosynthetic subtypes in grasses from genomics and phylogenetics. Genome Biol. Evol. 1:221–230.

Christin P.-A., Petitpierre B., Salamin N., Büchi L., Besnard G. 2009b. Evolution of C_4 phosphoenolpyruvate carboxykinase in grasses, from genotype to phenotype. Mol. Biol. Evol. 26:357–365.

Christin P.-A., Freckleton R.P., Osborne C.P. 2010. Can phylogenetics identify C_4 origins and reversals? Trends Ecol. Evol. 25:403–409.

Christin P.-A., Edwards E.J., Besnard G., Boxall S.F., Gregory R., Kellogg E.A., Hartwell J., Osborne C.P. 2012a. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. Curr. Biol. 22:445–449.

Christin P.-A., Besnard G., Edwards E.J., Salamin N. 2012b. Effect of genetic convergence on phylogenetic inference. Mol. Phylogenet. Evol. 62:921–927.

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 C_4 photosynthesis in grasses. Proc. Natl. Acad. Sci. USA 110:1381–1386.

Christin P.-A., Osborne C.P. 2014. The evolutionary ecology of C_4 plants. New Phytol. 204:765–781.

Christin P.-A., Spriggs E., Osborne C.P., Strömberg C.A.E., Salamin N., Edwards E.J. 2014. Molecular dating, evolutionary rates, and the age of the grasses. Syst. Biol. 63:153–165.

Clark C.J., McGuire J.A., Bonaccorso E., Berv J.S., Prum R.O. 2018. Complex coevolution of wing, tail, and vocal sounds of courting male bee hummingbirds. Evolution 72:630–646.

Cock P.J.A., Chilton J.M., Grüning B., Johnson J.E., Soranzo N. 2015. NCBI BLAST+ integrated into Galaxy. GigaScience 4:39.

Cotton J.L., Wysocki W.P., Clark L.G., Kelchner S.A., Pires J.C., Edger P.P., Mayfield-Jones D., Duvall M.R. 2015. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. BMC Plant Biol. 15:178.

Dengler N.G., Dengler R.E., Hattersley P.W. 1985. Differing ontogenetic origins of PCR ("Kranz") sheaths in leaf blades of C_4 grasses (Poaceae). Am. J. Bot. 72:284–302.

Dengler N.G., Donnelly P.M., Dengler R.E. 1996. Differentiation of bundle sheath, mesophyll, and distinctive cells in the C_4 grass *Arundinella hirta* (Poaceae). Am. J. Bot. 83:1391–1405.

Dengler R.E., Dengler N.G. 1990. Leaf vascular architecture in the atypical C_4 NADP − malic enzyme grass *Arundinella hirta*. Can. J. Bot. 68:1208–1221.

Doust A.N., Penly A.M., Jacobs S.W.L., Kellogg E.A. 2007. Congruence, conflict, and polyploidization shown by nuclear and chloroplast markers in the monophyletic "bristle clade" (Paniceae, Panicoideae, Poaceae). Syst. Bot. 32:531–544.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:699–710.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.

BMC Evol. Biol. 7:214.

Dunning L.T., Lundgren M.R., Moreno-Villena J.J., Namaganda M., Edwards E.J., Nosil P., Osborne C.P., Christin P.-A. 2017a. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. Evolution 71:1541–1555.

Dunning L.T., Liabot A.-L., Olofsson J.K., Smith E.K., Vorontsova M.S., Besnard G., Simpson K.J., Lundgren M.R., Addicott E., Gallagher R. V., Chu Y., Pennington R.T., Christin P.-A., Lehmann C.E.R. 2017b. The recent and rapid spread of *Themeda triandra*. Bot. Lett. 164:327–337.

Dunning L.T., Olofsson J.K., Parisod C., Choudhury R.R., Moreno-Villena J.J., Yang Y., Dionora J., Quick W.P., Park M., Bennetzen J.L., Besnard G., Nosil P., Osborne C.P., Christin P.-A. 2019. Lateral transfers of large DNA fragments spread functional genes among grasses. Proc. Natl. Acad. Sci. USA 116:4416–4425.

Edwards E.J., Smith S.A. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. Proc. Natl. Acad. Sci. USA 107:2532–2537.

Edwards E.J., Osborne C.P., Stromberg C.A.E., Smith S.A., Bond W.J., Christin P.-A., Cousins A.B., Duvall M.R., Fox D.L., Freckleton R.P., Ghannoum O., Hartwell J., Huang Y., Janis C.M., Keeley J.E., Kellogg E.A., Knapp A.K., Leakey A.D.B., Nelson D.M., Saarela J.M., Sage R.F., Sala O.E., Salamin N., Still C.J., Tipple B. 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. Science 328:587–591.

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 C_4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 species. Mol. Biol. Evol. 33:1796–1806.

Endress P.K. 2011. Evolutionary diversification of the flowers in angiosperms. Am. J. Bot. 98:370–396.

Estep M.C., Diaz D.M.V., Zhong J., Kellogg E.A. 2012. Eleven diverse nuclear-encoded phylogenetic markers for the subfamily Panicoideae (Poaceae). Am. J. Bot. 99:e443–e446.

Estep M.C., McKain M.R., Vela Diaz D., Zhong J., Hodge J.G., Hodkinson T.R., Layton D.J., Malcomber S.T., Pasquet R., Kellogg E.A. 2014. Allopolyploidy, diversification, and the Miocene grassland expansion. Proc. Natl. Acad. Sci. USA 111:15149–15154.

Forrestel E.J., Donoghue M.J., Smith M.D. 2014. Convergent phylogenetic and functional responses to altered fire regimes in mesic savanna grasslands of North America and South Africa. New Phytol. 203:1000–1011.

Gallaher T.J., Adams D.C., Attigala L., Burke S. V., Craine J.M., Duvall M.R., Klahs P.C., Sherratt E., Wysocki W.P., Clark L.G. 2019. Leaf shape and size track habitat transitions across forest–grassland boundaries in the grass family (Poaceae). Evolution 73:927–946.

Goodstein D.M., Shu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D.S. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178-D1186.

Grass Phylogeny Working Group (GPWG). 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. MO. Bot. Gard. 88:373–457.

Grass Phylogeny Working Group II (GPWG II). 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. New Phytol. 193:304–312.

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. Syst. Biol. 59:307–321.

Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Gutierrez M., Gracen V.E., Edwards G.E. 1974. Biochemical and cytological relationships in C_4 plants. Planta 119:279–300.

Hackel J., Vorontsova M.S., Nanjarisoa O.P., Hall R.C., Razanatsoa J., Malakasi P., Besnard G. 2018. Grass diversification in Madagascar: In situ radiation of two large C_3 shade clades and support for a Miocene to Pliocene origin of C_4 grassy biomes. J. Biogeogr. 45:750–761.

Hatch M.D. 1987. C$_4$ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta 895:81–106.

Hartley W. 1958. Studies on the origin, evolution and distribution of the Gramineae. I. The tribe Andropogoneae. Aust. J. Bot. 6:116–128.

Hattersley P.W. 1984. Characterization of C$_4$ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. Ann. Bot. 53:163–180.

Heyduk K., Moreno-Villena J.J., Gilman I., Christin P.-A., Edwards E.J. 2019. The genetics of convergent evolution: insights from plant photosynthesis. Nat. Rev. Genet. doi:10.1038/s41576-019-0107-5

Hibberd J.M., Covshoff S. 2010. The regulation of gene expression required for C$_4$ photosynthesis. Annu. Rev. Plant Biol. 61:181–207.

Huang P., Studer A.J., Schnable J.C., Kellogg E.A., Brutnell T.P. 2017. Cross species selection scans identify components of C$_4$ photosynthesis in the grasses. J. Exp. Bot. 68:127–135.

Hulbert L.C. 1988. Causes of fire effects in tallgrass prairie. Ecology 69:46–58.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–589.

Kanai R., Edwards G.E. 1999. The biochemistry of C$_4$ photosynthesis. In: Sage R.F., Monson R.K., editors. C$_4$ Plant Biology. Academic Press. pp. 49–87.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

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

Kellogg E.A. 2015. Flowering Plants. Monocots: Poaceae. Heidelberg: Springer. 416 p.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–

359.

Lehmann C.E.R, Griffith D.M., Simpson K.J., Anderson T.M., Archibald S., Beerling D.J., Bond
W.J., Denton E., Edwards E.J., Forrestel E.J., Fox D.L., Georges D., Hoffmann W.A.,
Kluyver T., Mucina L., Pau S., Ratnam J., Salamin N., Santini B., Smith M.D., Spriggs E.L.,
Westley R., Still C.J., Strömberg C.A.E., Osborne C.P. March 21, 2019 [cited on June 25,
2019]. Functional diversification enabled grassy biomes to fill global climate space. bioRxiv
[Preprint] doi: 10.1101/583625.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.
2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Linder H.P., Lehmann C.E.R., Archibald S., Osborne C.P., Richardson D.M. 2018. Global grass
(Poaceae) success underpinned by traits facilitating colonization, persistence and habitat
transformation. Biol. Rev. 93: 1125–1144.

Long S.P. 1999. Environmental responses. In: Sage R.F., Monson R.K., editors. C_4 plant biology.
Academic Press. pp. 215–249.

Lundgren M.R., Osborne C.P., Christin P.-A. 2014. Deconstructing Kranz anatomy to understand
C_4 evolution. J. Exp. Bot. 65:3357–3369.

Lundgren M.R., Besnard G., Ripley B.S., Lehmann C.E.R., Chatelet D.S., Kynast R.G.,
Namaganda M., Vorontsova M.S., Hall R.C., Elia J., Osborne C.P., Christin P.-A. 2015.
Photosynthetic innovation broadens the niche within a single species. Ecol. Lett. 18:1021–
1029.

Lundgren M.R., Dunning L.T., Olofsson J.K., Moreno-Villena J.J., Bouvier J.W., Sage T.L.,
Khoshravesh R., Sultmanis S., Stata M., Ripley B.S., Vorontsova M.S., Besnard G., Adams
C., Cuff N., Mapaura A., Bianconi M.E., Long C.M., Christin P.A., Osborne C.P. 2019. C_4
anatomy can evolve via a single developmental change. Ecol. Lett. 22:302–312.

Marek P.E., Moore W. 2015. Discovery of a glowing millipede in California and the gradual
evolution of bioluminescence in *Diplopoda*. Proc. Natl. Acad. Sci. USA 112:6419–6424.

McGee M.D., Borstein S.R., Neches R.Y., Buescher H.H., Seehausen O., Wainwright P.C. 2015. A pharyngeal jaw evolutionary innovation facilitated extinction in Lake Victoria cichlids. Science 350:1077–1079.

Mendes F.K., Hahn M.W. 2016. Gene tree discordance causes apparent substitution rate variation. Syst. Biol. 65:711–721.

Metcalfe C.R. 1960. Anatomy of the monocotyledons - I. Gramineae. Oxford, 267 p.

Meyer B.S., Matschiner M., Salzburger W. 2017. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Syst. Biol. 66:531–550.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Moreno-Villena J.J., Dunning L.T., Osborne C.P., Christin P.-A. 2018. Highly expressed genes are preferentially co-opted for C_4 photosynthesis. Mol. Biol. Evol. 35:94–106.

Nair V.J., Ramachandran V.S., Sreekumar P. V. 1982. *Chandrasekharania*: a new genus of Poaceae from Kerala, India. Proc. Indian Acad. Sci. - Sect. B. Part 3, Plant Sci. 91:79–82.

Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Niklaus M., Kelly S. 2019. The molecular evolution of C_4 photosynthesis: opportunities for understanding and improving the world's most productive plants. J. Exp. Bot. 70:795–804.

O'Leary M.H. 1988. Carbon isotopes in photosynthesis. BioScience 38:328–336.

Ogilvie H.A., Heled J., Xie D., Drummond A.J. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Syst. Biol. 65:381–396.

Ogilvie H.A., Bouckaert R.R., Drummond A.J. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34:2101–2114.

Olofsson J.K., Bianconi M., Besnard G., Dunning L.T., Lundgren M.R., Holota H., Vorontsova

M.S., Hidalgo O., Leitch I.J., Nosil P., Osborne C.P., Christin P.-A. 2016. Genome

biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Mol.

Ecol. 25:6107–6123.

Olofsson J.K., Cantera I., Van de Paer C., Hong-Wa C., Zedane L., Dunning L.T., Alberti A.,

Christin P.-A., Besnard G. 2019. Phylogenomics using low-depth whole genome sequencing:

a case study with the olive tribe. Mol. Ecol. Resour. 19:877–892.

Osborne C.P. 2008. Atmosphere, ecology and evolution: what drove the Miocene expansion of C_4

grasslands? J. Ecol. 96:35–45.

Osborne C.P., Freckleton R.P. 2009. Ecological selection pressures for C_4 photosynthesis in the

grasses. Proc. R. Soc. B Biol. Sci. 276:1753–1760.

Osborne C.P., Sack L. 2012. Evolution of C_4 plants: A new hypothesis for an interaction of CO_2

and water relations mediated by plant hydraulics. Philos. Trans. R. Soc. B Biol. Sci. 367:583–

600.

Osborne C.P., Salomaa A., Kluyver T.A., Visser V., Kellogg E.A., Morrone O., Vorontsova M.S.,

Clayton W.D., Simpson D.A. 2014. A global database of C_4 photosynthesis in grasses. New

Phytol. 204:441–446.

Patel R.K., Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation

sequencing data. PLoS One 7:e30619.

Paterson A.H., Bowers J.E., Bruggmann R., Dubchak I., Grimwood J., Gundlach H., Haberer G.,

Hellsten U., Mitros T., Poliakov A., Schmutz J., Spannagl M., Tang H., Wang X., Wicker T.,

Bharti A.K., Chapman J., Feltus F.A., Gowik U., Grigoriev I. V., Lyons E., Maher C.A.,

Martis M., Narechania A., Otillar R.P., Penning B.W., Salamov A.A., Wang Y., Zhang L.,

Carpita N.C., Freeling M., Gingle A.R., Hash C.T., Keller B., Klein P., Kresovich S.,

McCann M.C., Ming R., Peterson D.G., Mehboob-Ur-Rahman, Ware D., Westhoff P., Mayer

K.F.X., Messing J., Rokhsar D.S. 2009. The *Sorghum bicolor* genome and the diversification

of grasses. Nature 457:551–556.

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

Prasad V., Strömberg, C.A.E., Leaché, A.D., Samant B., Patnaik R., Tang L., Mohabey D.M., Ge S., Sahni A. 2011. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. Nat. Commun. 2:480.

Prendergast H.D. V, Hattersley P.W., Stone N.E. 1987. New structural/biochemical associations in leaf blades of C_4 grasses (Poaceae). Funct. Plant Biol. 14:403–420.

Puttick M.N., Thomas G.H., Benton M.J. 2014. High rates of evolution preceded the origin of birds. Evolution 68:1497–1510.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rainford J.L., Hofreiter M., Nicholson D.B., Mayhew P.J. 2014. Phylogenetic distribution of extant richness suggests metamorphosis is a key innovation driving diversification in insects. PLoS One 9:e109085.

Rambaut A., Suchard M.A., Xie W., Drummond A.J. 2013. Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/

Renvoize S.A. 1982a. A survey of leaf-blade anatomy in grasses I. Andropogoneae. Kew Bull. 37:315–321.

Renvoize S.A. 1982b. A survey of leaf-blade anatomy in grasses II. Arundinelleae. Kew Bull. 37:489–495

Renvoize S.A. 1982c. A survey of leaf-blade anatomy in grasses III. Garnotieae. Kew Bull. 37:497.

Renvoize S.A. 1985. A note on *Jansenella* (Gramineae). Kew Bull. 40:470.

Ripley B., Visser V., Christin P.-A., Archibald S., Martin T., Osborne C. 2015. Fire ecology of C_3 and C_4 grasses depends on evolutionary history and frequency of burning but not photosynthetic type. Ecology 96:2679–2691.

Rondeau P., Rouch C., Besnard G. 2005. NADP-malate dehydrogenase gene evolution in

Andropogoneae (Poaceae): gene duplication followed by sub-functionalization. Ann. Bot. 96:1307–1314.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Saarela J.M., Burke S. V., Wysocki W.P., Barrett M.D., Clark L.G., Craine J.M., Peterson P.M., Soreng R.J., Vorontsova M.S., Duvall M.R. 2018. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. PeerJ6:e4299.

Sage R.F. 2004. The evolution of C_4 photosynthesis. New Phytol. 161:341–370.

Sage R.F., Christin P.-A., Edwards E.J. 2011. The C_4 plant lineages of planet Earth. J. Exp. Bot. 62:3155–3169.

Sage R.F., Stata M. 2015. Photosynthetic diversity meets biodiversity: the C_4 plant example. J. Plant Physiol. 172:104–119.

Sage R.F. 2017. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J. Exp. Bot. 68:e11–e28.

Sánchez-García M., Matheny P.B. 2017. Is the switch to an ectomycorrhizal state an evolutionary key innovation in mushroom-forming fungi? A case study in the Tricholomatineae (Agaricales). Evolution 71:51–65.

Silva C., Besnard G., Piot A., Razanatsoa J., Oliveira R.P., Vorontsova M.S. 2017. Museomics resolve the systematics of an endangered grass lineage endemic to north-western Madagascar. Ann. Bot. 119:339–351.

Sinha N.R., Kellogg E.A. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. Am. J. Bot. 83:1458–1470.

Shilla U., Tiwari B.K. 2015. Impact of fire and grazing on plant diversity of a grassland ecosystem of Cherrapunjee. Keanean J. Sci. 4:67–78.

Solbrig O.T. 1996. The diversity of the savanna ecosystem. In: O.T. Solbrig, E. Medina, and J.F.

Silva, editors. Biodiversity and Savanna Ecosystem Processes. Ecological Studies Vol. 121, pp. 1–27. Springer-Verlag, Berlin Heidelberg.

Soreng R.J., Peterson P.M., Romaschenko K., Davidse G., Teisher J.K., Clark L.G., Barberá P., Gillespie L.J., Zuloaga F.O. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. J. Syst. Evol.55:259–290.

Sprent J.I. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytol. 174:11–25.

Spriggs E.L., Christin P.-A., Edwards E.J.. 2014. C_4 photosynthesis promoted species diversification during the Miocene grassland expansion. PLoS One 9:e97722.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stebbins G.L. 1975. The role of polyploid complexes in the evolution of North American grasslands. Taxon 24:91–106.

Strömberg C.A.E., Dunn R.E., Crifò C., Harris E.B. 2018. Phytoliths in paleoecology: Analytical considerations, current use, and future directions. In: Croft D. A. et al. (eds), Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities, Vertebrate Paleobiology and Paleoanthropology, Chap. 12, pp. 235-287. Springer, Cham.

Studer A.J., Schnable J.C., Weissmann S., Kolbe A.R., McKain M.R., Shao Y., Cousins A.B., Kellogg E.A., Brutnell T.P. 2016. The draft genome of the C_3 panicoid grass species *Dichanthelium oligosanthes*. Genome Biol. 17:223.

Tateoka T. 1958. Notes on some grasses. VIII. On leaf structure of *Arundinella* and *Garnotia*. Bot. Gaz. 120:101–109.

Tausta S.L., Miller Coyle H., Rothermel B., Stiefel V., Nelson T. 2002. Maize C_4 and non-C_4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-

localized ancestor. Plant Mol. Biol. 50:635–652.

Thornsberry J.M., Goodman M.M.,Doebley J., Kresovich S., Nielsen D., Buckler E.S. IV. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28:288–289.

Türpe A.M. 1970. Sobre la anatomía foliar de *Jansenella griffithiana* (C. Mueller) Bor (Poaceae:Arundinelleae). Senckenberg. Biol. 51:277–285.

Ueno O. 1995. Occurrence of distinctive cells in leaves of C_4 species in *Arthraxon* and *Microstegium* (Andropogoneae-Poaceae) and the structural and immunocytochemical characterization of these cells. Int. J. Plant Sci. 156:270–289.

Vicentini A., Barber J.C., Aliscioni S.S., Giussani L.M., Kellogg E.A. 2008. The age of the grasses and clusters of origins of C_4 photosynthesis. Glob. Change Biol. 14: 2963–2977.

Visser V., Woodward F.I., Freckleton R.P., Osborne C.P. 2012. Environmental factors determining the phylogenetic structure of C_4 grass communities. J. Biogeogr. 39:232–246.

von Caemmerer S., Quick W.P., Furbank R.T. 2012. The development of C_4 rice: current progress and future challenges. Science 336:1671–1672.

Walker R.P., Acheson R.M., Técsi L.I., Leegood R.C. 1997. Phosphoenolpyruvate carboxykinase in C_4 plants: its role and regulation. Aust. J. Plant Physiol. 24:459–468.

Wang X., Gowik U., Tang H., Bowers J.E., Westhoff P., Paterson A.H. 2009. Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses. Genome Biol. 10:R68.

Wang Y., Bräutigam A., Weber A.P.M., Zhu X.-G. 2014. Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. J. Exp. Bot. 65:3567–3578.

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

Washburn J.D., Schnable J.C., Davidse G., Pires J.C. 2015. Phylogeny and photosynthesis of the grass tribe Paniceae. Am. J. Bot. 102:1493–1505.

Watcharamongkol T., Christin P.-A., Osborne C.P. 2018. C_4 photosynthesis evolved in warm

climates but promoted migration to cooler ones. Ecol. Lett. 21:376–383.

Watson L., Macfarlane T.D., Dallwitz M.J. 1992 onwards. The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. Version: 11th December 2017.

Wingler A., Walker R.P., Chen Z.-H., Leegood R.C. 1999. Phosphoenolpyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. Plant Physiol. 120:539–546.

Xi Z., Liu L., Davis C.C. 2016. The impact of missing data on species tree estimation. Mol. Biol. Evol. 33:838–860.

Yadav S.R., Chivalkar S.A., Gosavi K.V.C. 2010. On the identity of *Jansenella griffithiana* (Poaceae) with a new species from Western Ghats, India. Rheedea 20:38–43.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15:568–573.

Yang Z., Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. 46:409–418.

Yukawa T., Ogura-Tsujita Y., Shefferson R.P., Yokoyama J. 2009. Mycorrhizal diversity in *Apostasia* (Orchidaceae) indicates the origin and evolution of orchid mycorrhiza. Am. J. Bot. 96:1997–2009.

Zuloaga, F.O., Morrone, O. and Giussani, L.M. 2000. A cladistic analysis of the Paniceae: a preliminary approach. In: Jacobs S.W.L., Everett J., editors. Grasses: systematics and evolution. CSIRO. pp. 123–135.

FIGURES

Figure 1. Phylogenetic trees of grasses based on (a) plastomes and (b) genome-wide nuclear data. (a) Bayesian phylogram inferred from coding sequences of plastomes (see Fig. S2 for phylogram based on non-coding sequences). Branch lengths are in expected substitutions per site. Closed circles on nodes indicate Bayesian posterior probability ≥ 0.95. Branches in red lead to C_4 species. (b) Multigene coalescent species tree estimated from 365 nuclear genes . Pie charts on nodes indicate the proportion of quartet trees that support the main topology (in blue), the first alternative (in red), and the second alternative (in orange). Local posterior probabilities are indicated near nodes. Branch lengths are in coalescent units. The major groups of Panicoideae are delimited with shades.

Figure 2. Leaf transverse sections of representatives of Jansenelleae and Andropogoneae. (a) *Jansenella griffithiana*. OS = outer bundle sheath; IS = inner bundle sheath. Scale bar = 50 μm. (b) *Jansenella griffithiana*, *Garnotia stricta*, *Arundinella nepalensis*, *Heteropogon contortus* and *Ischaemum afrum*. The latter three images are from Christin et al. (2013). Arrows with a circle, a dash and a square indicate major veins, distinctive cells and minor veins, respectively. Scale bar = 200 μm.

Figure 3. Phylograms with branch lengths based on amino acid sequences. The tree topologies were fixed to those obtained using 3rd positions of codons for genes encoding five core C_4 enzymes: (a) NADP-malate dehydrogenase (NADP-MDH, gene *nadpmdh-1P1*), (b) NADP-malic enzyme (NADP-ME, gene *nadpme-1P4*), (c) phosphoenolpyruvate carboxykinase (PCK, gene *pck-1P1*), (d) phosphoenolpyruvate carboxylase (PEPC, gene *ppc-1P3*), and (e) pyruvate phosphate dikinase (PPDK, gene *ppdk-1P2*). The major taxonomic groups are indicated with shades, and branches from C_4Andropogoneae are in red. Yellow circles indicate gene duplications. Scale bars = 0.01 expected amino acid substitutions per site (panels a – e are depicted at different scales). C_4 species outside Andropogoneae were pruned from the tree. See Fig. S7 for details, and Fig. S6 for branch lengths based on nucleotides.

Figure 4. C_4-related changes in protein sequences and leaf anatomy in the Andropogoneae grasses.

A time-calibrated phylogenetic tree of Panicoideae is presented, with selected C_3 species outside Andropogoneae (see Fig. S5 for full phylogenetic tree). Branch thickness is proportional to the rate of protein change and colours represent different C_4 enzymes. A simplified transverse section of the leaf is presented on the right, with colours representing the different tissues.

TABLES

Table 1. Divergence time estimates for selected lineages of grasses based on plastome sequences.

Table 2. Leaf anatomical traits in *Jansenella griffithiana* and representatives of Andropogoneae.

Table 3. Summary of branch model comparisons.

SUPPLEMENTARY MATERIAL

Figure S1. Effect of different filtering stringencies and reference species on the set of nuclear genes retained for phylogenetic analyses. (a,f) Number of genes retained; (b,g) taxon occupancy (i.e. proportion of species represented); (c,h) proportion of the alignment that is complete (completeness); (d,i) number of parsimony informative sites; (e,j) proportion of the branches with bootstrap support values $\geq 50\%$. Analyses were performed on the reduced subset of 37 species (a-e) and the complete set of 66 species (f-j), and using either *Setaria italica* or *Sorghum bicolor* as the reference. Each analysis was repeated with a different level of trimming (no trim = no trimming; 50% trim = sites covered by less than 50% of sequences were trimmed; 70% trim = sites covered by less than 70% of sequences were trimmed). Finally, analyses considered either all gene trees ('all') or only gene trees with $\geq 50\%$ of branches with bootstrap support $\geq 50\%$. Dashed rectangles highlight the datasets used to generate the multigene coalescent species tree in Fig. 1 (in black), and Fig. S3 (in grey).

Figure S2. Bayesian phylogram inferred from non-coding sequences of plastomes. Black circles on nodes indicate Bayesian posterior probability ≥ 0.95.

Figure S3. Multigene coalescent species trees inferred from 3,127 nuclear genes using *Sorghum bicolor* as a reference. Pie charts on nodes indicate the proportion of quartet trees that support the main topology (in blue), the first alternative (in red), and the second alternative (in orange). Local posterior probabilities are indicated near nodes. Branch lengths are given in coalescent units. The major groups of Panicoideae are indicated with shades, as in Fig. 1.

Figure S4. Bayesian phylograms inferred from individual nuclear markers. (a) *apo1*, (b) *arodeh*, (c) *dwarf8*, (d) *floricaula*, (e) *kn1*, (f) *phyB*, (g) *rep1* and (h) *waxy*. Bayesian posterior probability values ≥ 70 are indicated near nodes. The main groups of Panicoideae are delimited with shades.

Figure S5. Time-calibrated phylogenetic tree based on coding sequences of plastomes. Bars on nodes indicate the 95% HPD intervals for the ages. The main groups of Panicoideae are delimited with shades.

Figure S6. Bayesian phylograms inferred for C_4-related genes. These phylograms were inferred based on 3rd positions of codons of several core C_4 genes: (a) *nadpmdh-1P1*, (b) *nadpme-1P4*, (c) *pck-1P1*, (d) *ppc-1P3*, (e) *ppdk-1P2*; paralogs of core C_4 genes: (f) *nadpme-1P1*, (g) *nadpme-1P2*, (h) *nadpme-1P3*, (i) *ppc-1P4*, (j) *ppc-1P5*, (k) *ppc-1P7*, (l) *ppdk-1P1*; and other nuclear phylogenetic markers: (m) *arodeh*, (n) *dwarf8*, (o) *kn1*, (p) *phyB* and (q) *waxy*. Branches in red lead to C_4 species. Bayesian posterior probability values ≥ 70 are indicated near nodes. The main groups of Panicoideae are delimited with shades.

Figure S7. Phylograms with branch lengths based on amino acid sequences. The tree topologies were fixed to those obtained using 3rd positions of codons used for branch model tests. Results are shown for core C_4 genes: (a) *nadpmdh-1P1*, (b) *nadpme-1P4*, (c) *pck-1P1*, (d) *ppc-1P3*, (e) *ppdk-1P2*; paralogs of core C_4 genes: (f) *nadpme-1P1*, (g) *nadpme-1P2*, (h) *nadpme-1P3*, (i) *ppc-1P4*, (j) *ppc-1P5*, (k) *ppc-1P7*, (l) *ppdk-1P1*; and other nuclear genes: (m) *arodeh*, (n) *dwarf8*, (o) *kn1*, (p) *phyB* and (q) *waxy*. Branches in red lead to C_4 species. Scale bars = 0.01 expected amino acid substitutions per site.

Table S1. Genomic data information.

Table S2. Read mapping success and nuclear genome coverage in datasets assembled using *Setaria italica* or *Sorghum bicolor* sequences as references.

Table S3. NCBI accession numbers of sequences of nuclear markers assembled in this study.

Table S4. Photosynthetic types and leaf anatomy data of the 66 grass species used in this study.

Table S5. Summary of multigene coalescent species trees inferred from nuclear data under different filtering stringencies.

Table S6. Summary of branch model comparisons performed using the species tree topology.

APPENDIX 1

Jansenelleae Voronts. tribus nov.

Type: *Jansenella* Bor, Kew Bull. 1955: 96. 1955.

Included genera: *Chandrasekharania* V.J. Nair, V.S. Ramach. & Sreek., *Jansenella* Bor

Description:

Annuals with erect culms. Ligule shortly membranous. Leaf blades lanceolate. Inflorescence shortly branched, appearing capitate. Spikelets laterally compressed, 2-flowered. Glumes 2, apically acuminate to shortly awned. Lower floret sterile, staminate, or bisexual. Lower lemma shortly awned, either entire (*Jansenella*) or awn arising between two erose apical lobes (*Chandrasekharania*). Lower palea present. Upper floret bisexual. Upper lemma awned from a bidentate apex, but variable in its shape and indumentum: either with two hair tufts, twisted dehiscent awn arising between long-acuminate lobes (*Jansenella*) or without hair tufts, short straight awn arising between two erose apical lobes (*Chandrasekharania*). Stamens 3. Grain ellipsoid; hilum punctiform.

Leaf anatomy:

Outer and inner bundle sheaths present. More than four cells between consecutive veins. No distinctive cells. Starch storage in the chlorenchyma.

Distribution:

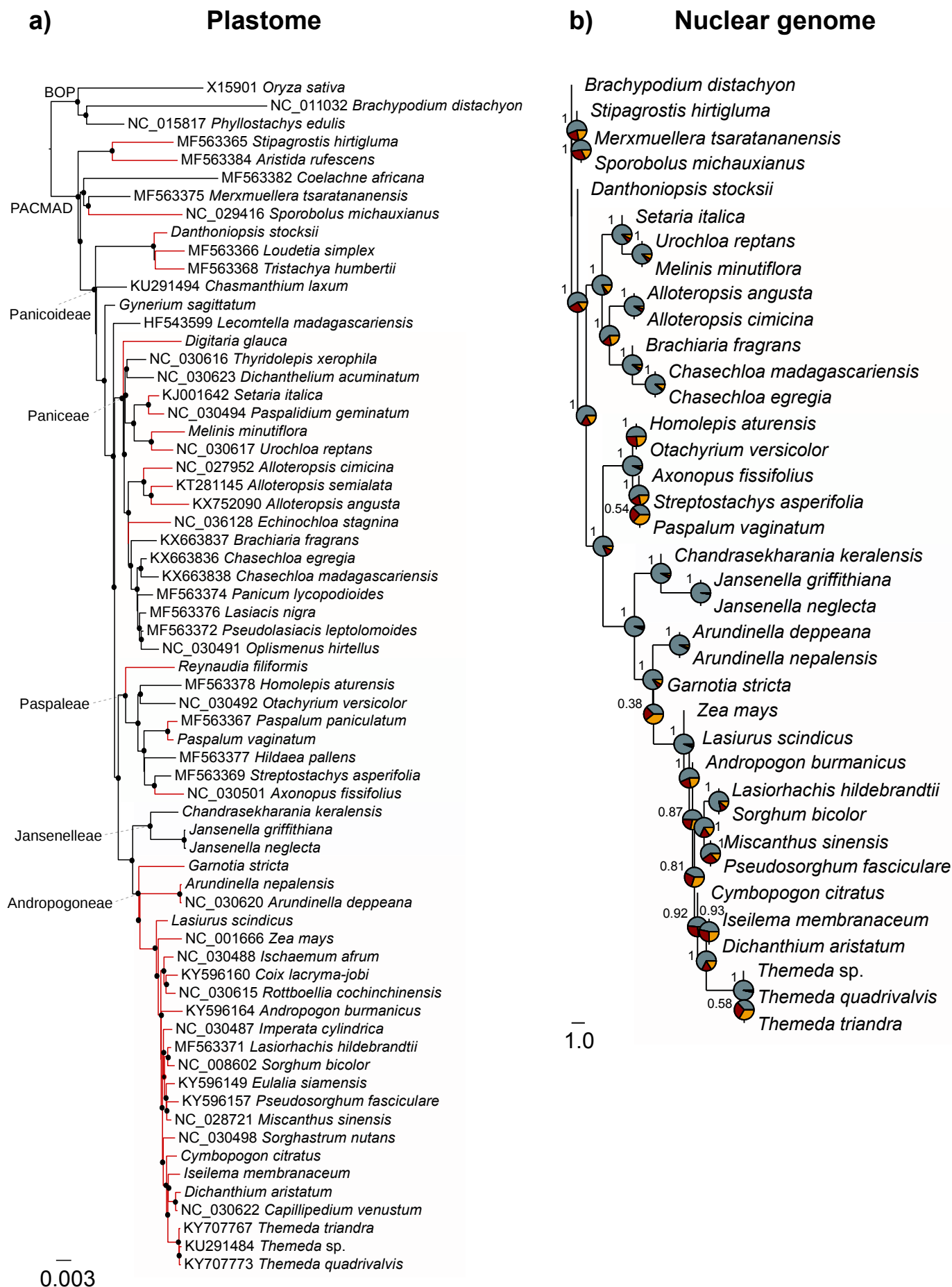India (including Assam and Sri Lanka), Myanmar and Thailand.

Fig. 1

a)   **Plastome**

b)   **Nuclear genome**

Systematic Biology

Fig. 2

Fig. 3

a) BOP — NADP-MDH
Paniceae
Paspaleae
Jansenelleae
Andropogoneae
Arundinellinae
Andropogoneae *s.s.*

b) BOP — NADP-ME
Arundinellinae
Andropogoneae *s.s.*

c) BOP — PCK
Arundinellinae
Andropogoneae *s.s.*

d) PEPC
Arundinellinae
Andropogoneae *s.s.*

e) BOP — PPDK
Arundinellinae
Andropogoneae *s.s.*

Fig. 4

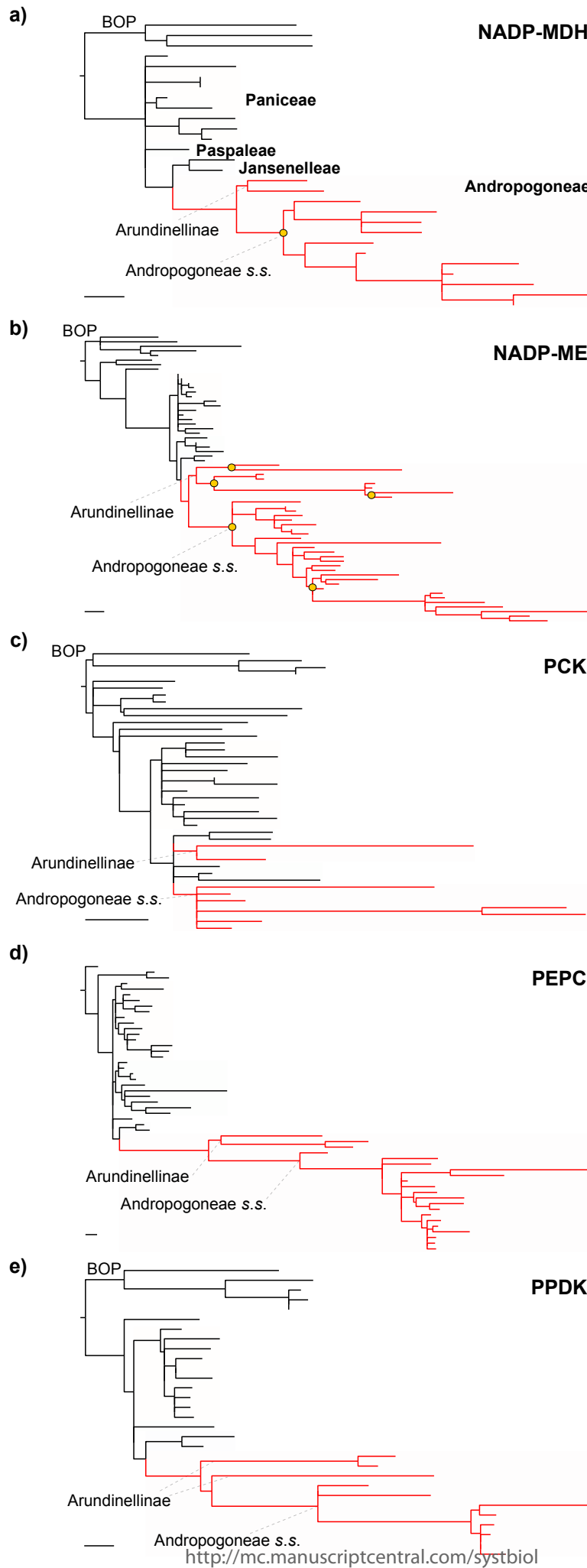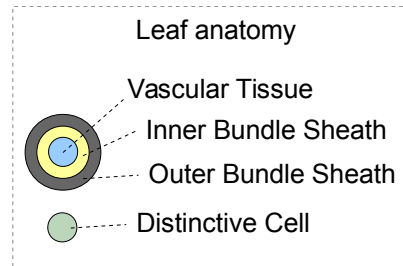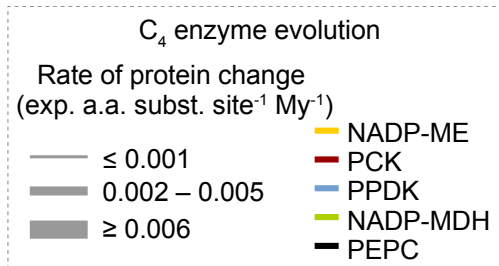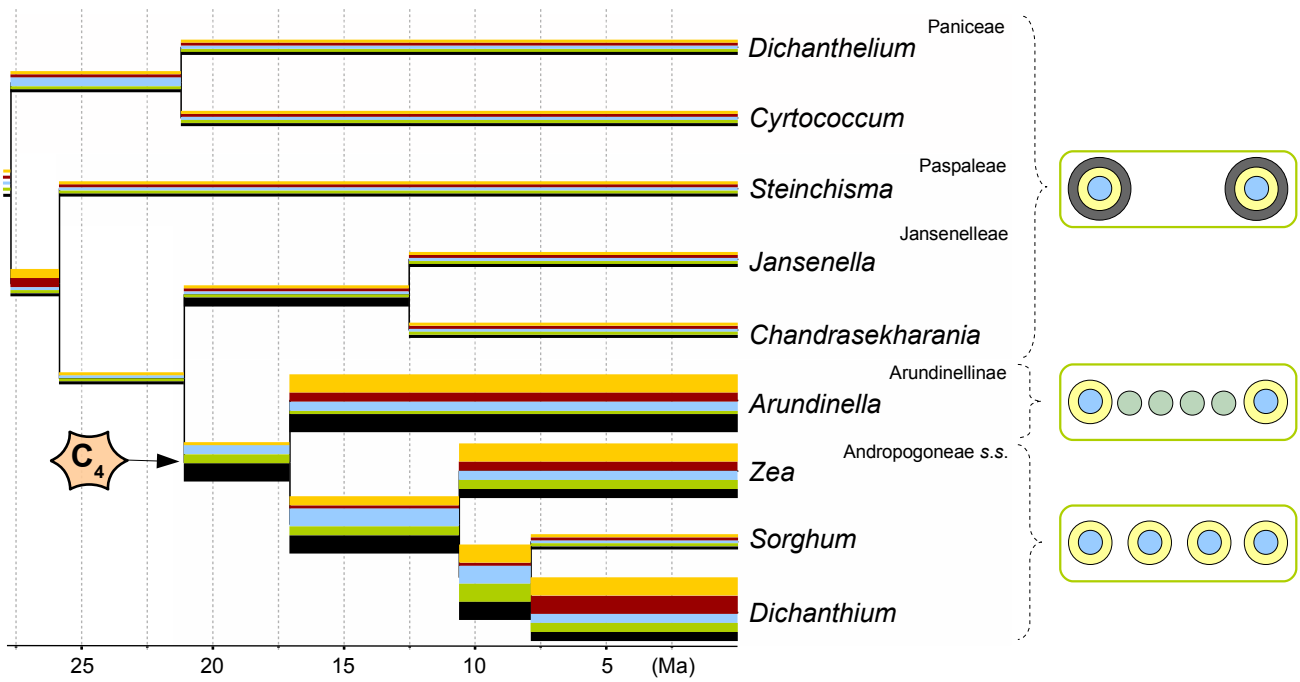**Table 1.** Divergence time estimates for selected lineages of grasses based on plastome sequences[a].

Clade	Macrofossils[b]	Microfossils[c]
BOP crown	34.7 (24.8 – 45.6)	55.8 (39.9 – 73.4)
PACMAD crown	43.4 (34.6 – 51.1)	69.8 (55.7 – 82.2)
Panicoideae crown	36.4 (26.3 – 46.6)	58.6 (42.3 – 75.0)
Jansenelleae / Andropogoneae split	21.1 (14.6 – 27.6)	34.0 (23.5 – 44.4)
Andropogoneae crown	17.9 (12.2 – 23.7)	28.8 (19.6 – 38.1)
Andropogoneae *s.s.* crown	11.9 (8.0 – 16.0)	19.2 (12.9 – 25.8)

[a] Median ages are given in million years ago (Ma), with 95% HPD intervals in parentheses; [b] Secondary calibration using Christin et al. (2014) estimates based only on macrofossils; [c] Secondary calibration using Christin et al. (2014) estimates based on macrofossils plus microfossils.

Table 2. Leaf anatomical traits in *Jansenella griffithiana* and representatives of Andropogoneae.

Lineage/Species	Bundle sheath layers	Interveinal distance (μm)	% Inner Sheath Area	Bundle sheath distance (μm)	Distinctive cells[a]	Starch in BSC[a]
Jansenelleae (C$_3$)						
Jansenella griffithiana	2	471	0.01	403	A	A
Arundinellinae (C$_4$)						
Arundinella nepalensis[b]	1	215	0.27	101	P	P
Garnotia stricta	1	191	0.11	187	P	P
Andropogoneae *s.s.* (C$_4$)						
Arthraxon sp.[c,d]	1	-	-	-	P	P
Chrysopogon pallidus[b]	1	112	0.23	29	A	P
Heteropogon contortus[b]	1	80	0.21	32	A	P
Ischaemum afrum[b]	1	109	0.24	52	A	P
Sorghum halepense[b]	1	119	0.20	53	A	P

[a] A = absent, P = present; [b] Christin et al. (2013); [c] Watson et al. (1992); [d] Ueno (1995).

Table 3. Summary of branch model comparisons.

			Branch models[a]						
Scenarios of adaptive evolution			**Single episode** Preceding C_4 evolution		**Single episode** During C_4 evolution, at the base of Andropogoneae		**Two episodes** During C_4 evolution, at the base of Arundinellinae and Andropogoneae *s.s.*		dN/dS[d]
Gene[b]	N[c]	Null model	Internal branch	Sustained	Internal branch	Sustained	Internal branch	Sustained	
Core C_4 genes									
nadpmdh-1P1 (NADP-MDH)	28	13.53	14.82	5.90	14.74	**0.00***	14.55	1.46*	0.07 – 0.14
nadpme-1P4 (NADP-ME)	63	205.29	202.70	42.20*	205.26	**0.00***	202.73	2.89*	0.08 – 0.30
pck-1P1 (PCK)	41	36.77	38.11	5.40*	-	-	38.77	**0.00***	0.02 – 0.06
ppc-1P3 (PEPC)	51	105.82	107.82	7.83*	82.81*	**0.00***	85.87*	21.78*	0.03 – 0.09
ppdk-1P2 (PPDK)	30	100.33	97.23	51.07*	83.78*	37.69*	**0.00***	53.91*	0.10 – 0.44
Paralogs of core C_4 genes									
nadpme-1P1	30	**1.52**	1.56	0.00	1.28	1.16	3.33	1.73	0.09
nadpme-1P2	21	**0.00**	0.72	1.59	2.00	0.97	0.55	0.93	0.09
nadpme-1P3	24	10.58	**0.00***	4.73	4.77	7.75	9.97	10.05	0.05 – 0.17
ppc-1P4	30	**3.92**	4.08	5.90	5.23	5.30	0.00	3.84	0.07
ppc-1P5	30	**0.00**	2.00	1.84	1.83	1.49	1.84	1.80	0.06
ppc-1P7	19	11.33	13.29	5.59	-	-	12.65	**0.00***	0.09 – 0.04
ppdk-1P1	12	**6.19**	8.17	6.31	7.39	0.37	0.00	0.63	0.18
Other nuclear genes									
arodeh	30	**0.00**	1.99	1.70	-	-	1.85	1.77	0.11
dwarf8	39	46.51	45.75	23.99*	47.23	2.35*	48.34	**0.00***	0.06 – 0.16
knotted1	13	**4.15**	0.00	5.45	-	-	3.60	5.94	0.07
phyB	55	**7.95**	4.67	8.35	0.00	9.86	9.71	9.85	0.09
waxy	55	**0.00**	1.74	1.70	-	-	1.90	1.91	0.05

[a] dAIC values relative to the best-fit model for each gene are shown. The best-fit model is highlighted in bold. Asterisks indicate significant likelihood ratio tests (LRT) against the null model after Bonferroni correction. Two hypotheses of potential enzyme adaptation were tested for each scenario, the first assuming a shift in selective pressure only in the internal branch of the group specified, the second assuming a sustained shift from the internal branch including all descendant branches. Missing values correspond to trees in which Andropogoneae was not monophyletic; [b] C_4 gene annotation following Moreno-Villena et al. (2018); [c] Number of sequences in the alignment; [d] dN/dS ratios of background and foreground branches, respectively, estimated for the best-fit model, except in cases where the null was the best-fit model, for which there was a single dN/dS estimate for all branches.