

RESEARCH ARTICLE

Open Access



Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoideae: Andropogoneae) to sugarcane

Dyfed Lloyd Evans^{1,2,3*} , Shailesh V. Joshi^{1,2} and Jianping Wang^{4,5,6}

Abstract

Background: For over 50 years, attempts have been made to introgress agronomically useful traits from *Erianthus* sect. *Ripidium* (*Tripidium*) species into sugarcane based on both genera being part of the ‘Saccharum Complex’, an interbreeding group of species believed to be involved in the origins of sugarcane. However, recent low copy number gene studies indicate that *Tripidium* and *Saccharum* are more divergent than previously thought. The extent of genus *Tripidium* has not been fully explored and many species that should be included in *Tripidium* are still classified as *Saccharum*. Moreover, *Tripidium* is currently defined as *incertae sedis* within the Andropogoneae, though it has been suggested that members of this genus are related to the Germainiinae.

Results: Eight newly-sequenced chloroplasts from potential *Tripidium* species were combined in a phylogenetic study with 46 members of the Panicoideae, including seven *Saccharum* accessions, two *Miscanthidium* and three *Miscanthus* species. A robust chloroplast phylogeny was generated and comparison with a gene locus phylogeny clearly places a monophyletic *Tripidium* clade outside the bounds of the Saccharinae. A key to the currently identified *Tripidium* species is presented.

Conclusion: For the first time, we have undertaken a large-scale whole plastid study of eight newly assembled *Tripidium* accessions and a gene locus study of five *Tripidium* accessions. Our findings show that *Tripidium* and *Saccharum* are 8 million years divergent, last sharing a common ancestor 12 million years ago. We demonstrate that four species should be removed from *Saccharum/Erianthus* and included in genus *Tripidium*. In a genome context, we show that *Tripidium* evolved from a common ancestor with and extended Germainiinae clade formed from *Germainia*, *Eriochrysis*, *Apocopsis*, *Pogonatherum* and *Imperata*. We re-define the ‘Saccharum complex’ to a group of genera that can interbreed in the wild and extend the Saccharinae to include *Sarga* along with *Sorghastrum*, *Microstegium vimineum* and *Polytrias* (but excluding *Sorghum*). Monophyly of genus *Tripidium* is confirmed and the genus is expanded to include *Tripidium arundinaceum*, *Tripidium procerum*, *Tripidium kanashiroi* and *Tripidium rufipilum*. As a consequence, these species are excluded from genus *Saccharum*. Moreover, we demonstrate that genus *Tripidium* is distinct from the Germainiinae.

Keywords: *Tripidium*, Saccharinae, *saccharum* complex, Plastids, Phylogenetics, Taxonomy, Sorghinae, Assembly, Gene loci

* Correspondence: dyfed121@mail.com

¹South African Sugarcane Research Institute, 170 Flanders Drive, Private Bag X02, Mount Edgecombe, Durban 4300, South Africa

²School of Life Sciences, College of Agriculture, Engineering and Science, University of Kwa-Zulu Natal, Private Bag X54001, Durban 4000, South Africa
Full list of author information is available at the end of the article



Background

Sugarcane (complex hybrids of *Saccharum cultum* Lloyd Evans and Joshi, *Saccharum officinarum* L. and *Saccharum spontaneum* L. [1]) ranks amongst the top-10 crop species worldwide [2]. Sugarcane also provides between 60 and 70% of total world sugar output and is a major source of bioethanol [2]. However, sugarcane production is dependent on rich soil and plentiful water supply. Sugarcane growth is also highly temperature dependent, with an optimum of 27 °C and an optimal humidity of 85% [3]. This means that the sugarcane-growing belt is limited to the tropical and subtropical regions of the globe. As a result, sugarcane breeders have long had an interest in introgressing traits from sugarcane's relatives to alter the growing range of sugarcane, to make the plant more productive, to increase its stress tolerance, and to make it more efficient in terms of resource usage (water, soil conditions, nitrogen) [4]. For optimal breeding success, particularly if using genera outside core *Saccharum* L., an accurate phylogenetic relationship between *Saccharum* and its purported relatives is a paramount necessity.

The concept of the 'Saccharum complex' as a set of interbreeding species and the Saccharinae as a subtribe of sugarcane's close relatives has remained largely unchallenged for 60 years [5] and has been at the core of all molecular and traditional breeding approaches. These relationships can now be re-evaluated in the light of modern molecular taxonomic approaches.

Ever since Mukherjee included *Erianthus* Michx. sect. *Ripidium* Henrard within the 'Saccharum complex' [5], a group of potentially interbreeding genera believed to be involved in the origins of modern sugarcane hybrids (*Saccharum* × *officinarum*/*Saccharum* × *cultum* [1]), considerable effort has been expended in introgressing *Erianthus* species into the sugarcane gene-pool.

Clayton and Renvoize [6] included *Erianthus* in the subgenus Saccharinae Griseb. (along with the genera *Eriochrysis* P. Beauv., *Eulalia* Knuth, *Eulaliopsis* Honda, *Homozeugos* Stapf, *Imperata* Crillo, *Lophopogon* Hack, *Microstegium* Nees, *Miscanthus* Andersson, *Pogonatherum* P. Beauv., *Polliniopsis* Hayata, *Polytrias* Hack, *Saccharum* and *Spodiopogon* Trin), suggesting that *Erianthus* was closely allied to *Saccharum*. Indeed, this paper has had considerable taxonomic influence and Kew's GrassBase defines former *Erianthus* species as part of the *Saccharum* genus [7].

However, the paper of Hodkinson et al. [8], though using relatively few characters for their phylogeny, demonstrated that *Erianthus* was not monophyletic. The Old World *Erianthus* species (sect. *Ripidium*) clustered as sister to *Eulalia* and *Zea* L., whilst the New World species and *Erianthus rockii* Keng (*Saccharum longisetosum* (Andersson) V. Naray.) clustered as sister to *Miscanthidium* Stapf. Hodkinson et al. [8] proposed the name

Ripidium (following Grassl [9] and von Trinius [10]) for the genus corresponding to Old World *Erianthus* species. This name, however, was found to be invalid, as Bernhardt [11] had already proposed *Ripidium* for a fern species within the *Schizaeaceae*, and Valdés and Scholz [12] proposed *Tripidium* as a replacement.

To fully understand the problems associated with the naming and taxonomic positioning of *Erianthus*/*Tripidium* species, we must be fully conversant with the taxonomic origins and derivations of these genera and the potential use of species within them to sugarcane breeders.

The Old World species of *Erianthus* are highly caespitose (forming dense clumps); indeed, Grassl [9] defined them as 'tufted bunch grass types' and it was this dense bunching habit and efficient ratooning that initially led sugarcane breeders to attempt the introgression of *Erianthus* characteristics into sugarcane. The deep rhizomes of *Erianthus* species also mean that they have improved water use and nutrient use efficiency as compared with members of genus *Saccharum* (even some highly rhizomatous *Saccharum spontaneum* accessions) [13].

Genus *Erianthus* was first defined by Michaux [14], with the name being derived from the Greek 'Erión' (wool) and 'anthos' (flower), referring to the woolly glumes possessed by members of the genus. The New World species are found in the Americas, whilst the Old World Species occur from Mediterranean Europe through India, China, South East Asia, New Guinea and Taiwan. As the Old World species (originally placed under the section *Ripidium*) have always been believed to be closest to *Saccharum* (and part of the 'Saccharum complex' and the Saccharinae) they are analysed in detail within this paper. For reference, the nomenclatural history of species within *Erianthus* sect *Ripidium* (*Tripidium*) is given in Table 1.

Following the work of Valdés et al. [12], genus *Tripidium* currently contains three accepted members: *Tripidium ravennae* (L.) H. Scholz, *Tripidium bengalense* (Retz) H. Scholz and *Tripidium strictum* (Host) H. Scholz. NCBI's taxonomy also places all species detailed in Table 1 within genus *Saccharum*, apart from *Tripidium ravennae*, which is defined as the type species for *Tripidium*, and *Tripidium bengalense* [15]. This leaves four additional species (*Saccharum rufipilum*, *Saccharum procerum*, *Saccharum arundinaceum* and *Saccharum kanashiroi*) that could be members of genus *Tripidium*, all of which were analysed in this study.

A recent publication by Welker et al. [16], examining five low copy number gene loci, placed the Old World *Tripidium* (*Erianthus*) species as sister to a clade formed by the core Andropogoneae, the core Saccharinae and the core Sorghinae (though with relatively poor support). The New World species (*Saccharum giganteum* (Walter) Pers. [*Erianthus giganteus* (Walter) P. Beauv.], *Saccharum asperum* (Nees) Steud. [*Erianthus asper* Nees], *Saccharum*

Table 1 History of the nomenclature of species within *Eriarhynchus* sect *Ripidium* (*Tripidium*)

Basionym	Commonly used name	Grassl re-classification [9]	Valdés et al. Re-classification [12]	GrassBase definition [7]	Tropicos definition [47]	Notes
<i>Saccharum arundinaceum</i> Retz.	<i>E. arundinaceus</i> (Retz.) Jesw	<i>Ripidium arundinaceum</i>		<i>Saccharum arundinaceum</i> Retz.	<i>Saccharum arundinaceum</i> Retz.	
<i>Eriarhynchus kanashiroi</i> Ohwi.	<i>E. kanashiroi</i> Ohwi	<i>Ripidium kanashiroi</i>		<i>Saccharum kanashiroi</i> Ohwi	<i>Saccharum kanashiroi</i> Ohwi	
<i>Saccharum procerum</i> Roxb.	<i>E. procerus</i> (Roxb.)	<i>Ripidium procerum</i>		<i>Saccharum procerum</i> Roxb.	<i>Saccharum procerum</i> Roxb.	
<i>Andropogon ravennae</i> L.	<i>E. ravennae</i> (L.) P. Beauv. (syn. <i>R. ravennae</i> (L.) Trin.)		<i>Tripidium ravennae</i> (L.) Scholz	<i>Saccharum ravennae</i> (L.) L. Scholz	<i>Tripidium ravennae</i> (L.) H. Scholz	
<i>E. elephantinus</i> Hook. F.	<i>E. elephantinus</i> Hook. F.	<i>Ripidium elephantinum</i>				Synonymous with <i>S. ravennae</i> [85]
<i>Saccharum bengalense</i> Retz.	<i>E. bengalense</i> (Retz.)	<i>Ripidium bengalense</i>	<i>Tripidium bengalense</i> (Retz.) Scholz	<i>Saccharum bengalense</i> Retz.	<i>Tripidium bengalense</i>	
<i>Andropogon strictus</i> Host	<i>E. strictus</i> (Host) Borbá (syn. <i>R. strictum</i> (Host) Trin.)		<i>Tripidium strictum</i> (Host) Scholz	<i>Saccharum strictum</i> (Host) Spreng.	<i>Saccharum strictum</i> (Host) Spreng.	
<i>Eriarhynchus rufipilus</i> (Steud.) Griseb.	<i>E. hostii</i> Griseb <i>Eriarhynchus rufipilus</i> (Steud.) Griseb.	<i>Miscanthus rufipilus</i>		<i>Saccharum rufipilus</i> Steud.	<i>Saccharum rufipilus</i> Steud.	Synonymous with <i>S. strictum</i> [86]
	<i>E. longesetosus</i> Anderss	<i>Eccolipus longesetosus</i>				Synonymous with <i>Saccharum strictum</i> [87]
	<i>E. hookeri</i> Hack	<i>Eccolipus hookeri</i>				Synonymous with <i>Saccharum strictum</i> [87]
<i>Eulalia japonica</i> Trin.	<i>Ripidium japonicum</i> (Trin.) Trin.;					Synonymous with <i>Miscanthus floridulus</i> (Labill.) Warb. ex K. Schum. & Lauterb. [85]

The above table gives a history of the nomenclature of species within *Eriarhynchus* sect *Ripidium* (*Tripidium*) starting with the Basionym (original name), followed by the treatment of Grassl [9], the re-classification of Valdés et al. [12], the current definition in Tropicos [47] and the current definition in GrassBase [7]. The notes column denotes species that have been re-classified as being synonymous with previously defined species. All species in the table are from the Old World

angustifolium (Nees) Trin. [*Erianthus angustifolius* Nees], and *Saccharum villosum* Steud [*Erianthus trinii* (Hack.) Hack.]) were placed within the *Miscanthus/Saccharum* (Saccharinae s.s.) clade. However, the positioning of *Tripidium* within the phylogeny of Welker et al. [16] differed slightly from that of the earlier paper of Estep et al. [17]. These authors did, however, propose the recognition of genus *Tripidium* as separate from *Saccharum*. Though they included both *Erianthus arundinaceus* and *Erianthus (Tripidium) ravennae*, there was no large-scale treatment of *Tripidium* species and no dating information on the divergence of *Tripidium* in their study. The paper of Soreng et al. [18] attempting to classify the Poaceae as whole, defined *Tripidium* as *incertae sedis*, but they did exclude the genus from *Saccharum* and suggested that *Tripidium* could be allied to the Germainiinae (*Germainia* and related genera). This uncertain placement stems from the reorganization of the Saccharinae and Germainiinae, which left four genera: *Eriochrysis*, *Imperata*, *Pogonatherum* and *Tripidium*, as *incertae sedis* [18]. Thus, whilst there is some agreement that genus *Tripidium* exists, there is no consensus on this finding, and a full circumscription of *Tripidium* has not been undertaken. Moreover, the species most commonly used in sugarcane introgression breeding (*Erianthus arundinaceus*) has never formally been included in genus *Tripidium*.

A larger-scale phylogenetic study with more extensive sampling of complete plastomes and genomic loci was always needed before a recommendation could be proposed for the confident circumscription and placement of *Tripidium* or any of the other genera. All four of the Andropogoneae genera currently defined as *incertae sedis* are included in the present study, with a large-scale sampling of *Tripidium* species. It is especially important to compare across genomic and plastome datasets, particularly as the role of hybridization in driving evolution, most especially in plant species, has recently received resurgent attention [19, 20]. Indeed, molecular studies from a variety of taxa across the tree of life have increasingly acknowledged that hybridization is an important source of evolutionary novelty [21, 22]. Pirie et al. [23] revealed ancient reticulation in the *Danthonioideae* (Poaceae) and ancient reticulation has been revealed in a range of plant genera [19, 24–28]. This is particularly important where there have been ancient rapid radiations [26], such as in the Poaceae [17]. As a result, the comparison of genome-based and plastome-based datasets is needed to examine the possibility of reticulate evolution within a genus, particularly in grasses where hybridization between lineages is especially common [29, 30].

Based on the existing evidence, it can be concluded that genus *Erianthus* is not monophyletic and that the New World *Erianthus* species are separate from the Old World species. Whilst the New World species may be allied to

the *Saccharum* genus, the Old World species are not. Thus, the current inclusion of all former *Erianthus* species into *Saccharum* (as at Kew's GrassBase [7] and the NCBI Taxonomy [31]) is almost certainly erroneous and is in dire need of re-evaluation.

The *Tripidium (Erianthus sect. Ripidium)* genus includes members with chromosome numbers of $2n = 10, 20, 30, 40$ and 60 [12]. Though, in common with other members of the Panicoideae, the base number is $x = 10$, which is the same as for genus *Saccharum* [32]. As a genus, *Tripidium* possesses a number of agronomically useful phenotypic features namely: cold tolerance, drought tolerance, heat tolerance, salt tolerance, disease resistance, improved vigour, dense culm spacing and increased ratooning [13]. As *Erianthus* was included in the 'Saccharum complex' [5, 33] sugarcane breeders have long attempted to introgress the agronomic properties of *Erianthus* into sugarcane. Typically, introgression breeding involves *Erianthus arundinaceus (Tripidium arundinaceum)*, but it has been difficult to both generate fertile progeny and to identify true hybrids [34]. Subsequent experiments have attempted to utilize cytogenetic tools such as GISH and species-specific markers to follow DNA transmission so that true hybrids can be identified at the seedling stage; and to follow *Erianthus* genome regions into subsequent progeny [35–38]. GISH [35] revealed that chromosome elimination occurs in *Saccharum* hybrid $\times E. arundinaceus$ hybrids, indicating that there might be a greater evolutionary distance between *Saccharum* and *Erianthus* than predicted solely on the basis of morphological characteristics. Recent studies have shown that the small percentage of true hybrids obtained in *Saccharum* hybrid $\times E. arundinaceus$ crossings are highly aneuploid, with loss and duplication of *E. arundinaceus* chromosomes as well as frequent interspecific recombinations between sugarcane and *E. arundinaceus* chromosomes [4, 39]. This phenomenon is characteristic of intergeneric hybridizations (wide crosses between evolutionarily very divergent plant species), as exemplified by oat/maize partial hybrids [40].

As considerable effort has been placed in introgressing *Erianthus* into sugarcane (typically with poor success) the question of the evolutionary distance between the two genera has both economic and taxonomic relevance. Particularly as molecular data are casting significant doubts on the veracity of the 'Saccharum complex' as a whole [41]. Clarifying the taxonomic position of the Old World *Erianthus* species (the most commonly used in introgression breeding) within the Andropogoneae and to the *Saccharum sensu stricto* species will help drive forward knowledge-based breeding in sugarcane.

We have sequenced and assembled complete chloroplasts from six *Tripidium* accessions from the South African Sugarcane Research Institute (two of known species and three

re-classified as part of this study) and two accessions of known species from the USDA collection as well as an additional *Saccharum spontaneum* accession [42]. Our assemblies were integrated with 45 previously assembled chloroplast genomes to yield the most comprehensive phylogenomic study of *Tripidium* and allied genera within the Andropogoneae thus far conducted. Moreover, a parallel analysis of five low copy number gene loci (63 species) was also performed.

Comparisons of whole chloroplast and gene loci phylogenies revealed that the *Tripidium* accessions were monophyletic, and sister to a clade formed by the core Andropogoneae and Saccharinae. They diverged from *Saccharum* at least 11 million years ago and are distal to the core Andropogoneae. We also reveal signals of reticulate evolution in genera that are currently defined as *incertae sedis* within the Andropogoneae.

This means that *Tripidium* should be the preferred genus name for these accessions and that they cannot be part of either the Saccharinae subtribe or the ‘Saccharum complex’. Moreover, *Tripidium* is circumscribed, based on both whole plastome and low copy number gene locus phylogenies, as a monophyletic grouping with seven confirmed species that is over 8 million years divergent from sugarcane. In addition, we show that the genera *Eriochrysis* and *Eulaliopsis* should also be excluded from the Saccharinae, as should *Sorghum* (based on low copy number gene phylogenetics). Genera *Saccharum*, *Miscanthidium*, *Miscanthus*, *Sarga*, and *Erianthus* do form a monophyletic group that could be termed the ‘Saccharinae’. However, of these, only *Miscanthus*, *Miscanthidium* and *Erianthus* are sufficiently evolutionarily close to *Saccharum* to allow for hybridization in the wild [1]. Thus, the ‘Saccharum complex’ as a group of interbreeding species should be limited to *Saccharum*, *Miscanthus*, *Miscanthidium* and the New World *Erianthus* species.

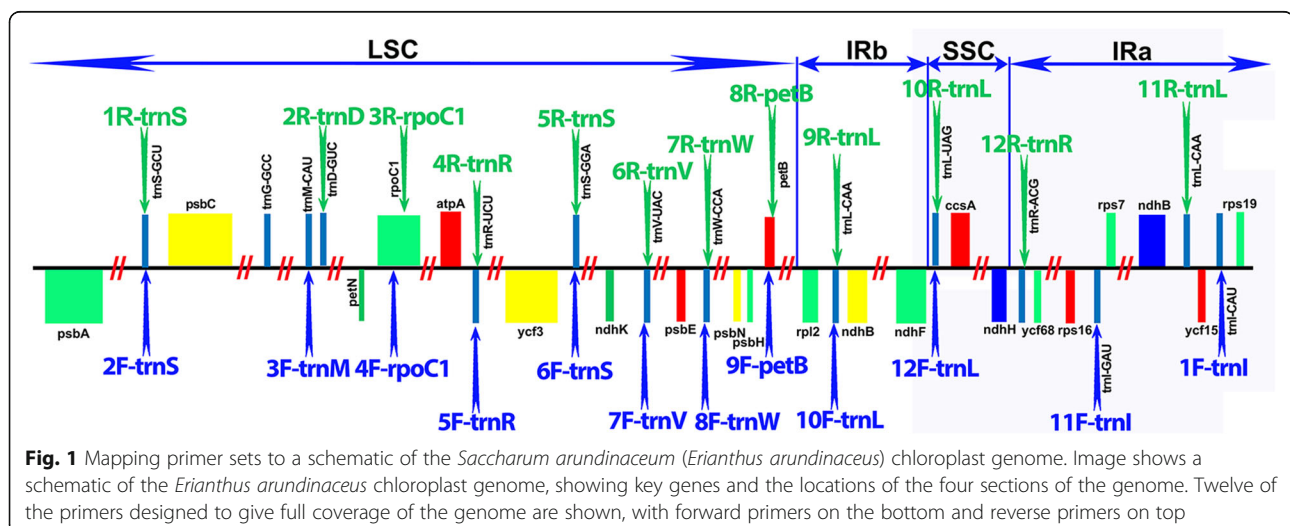
Results

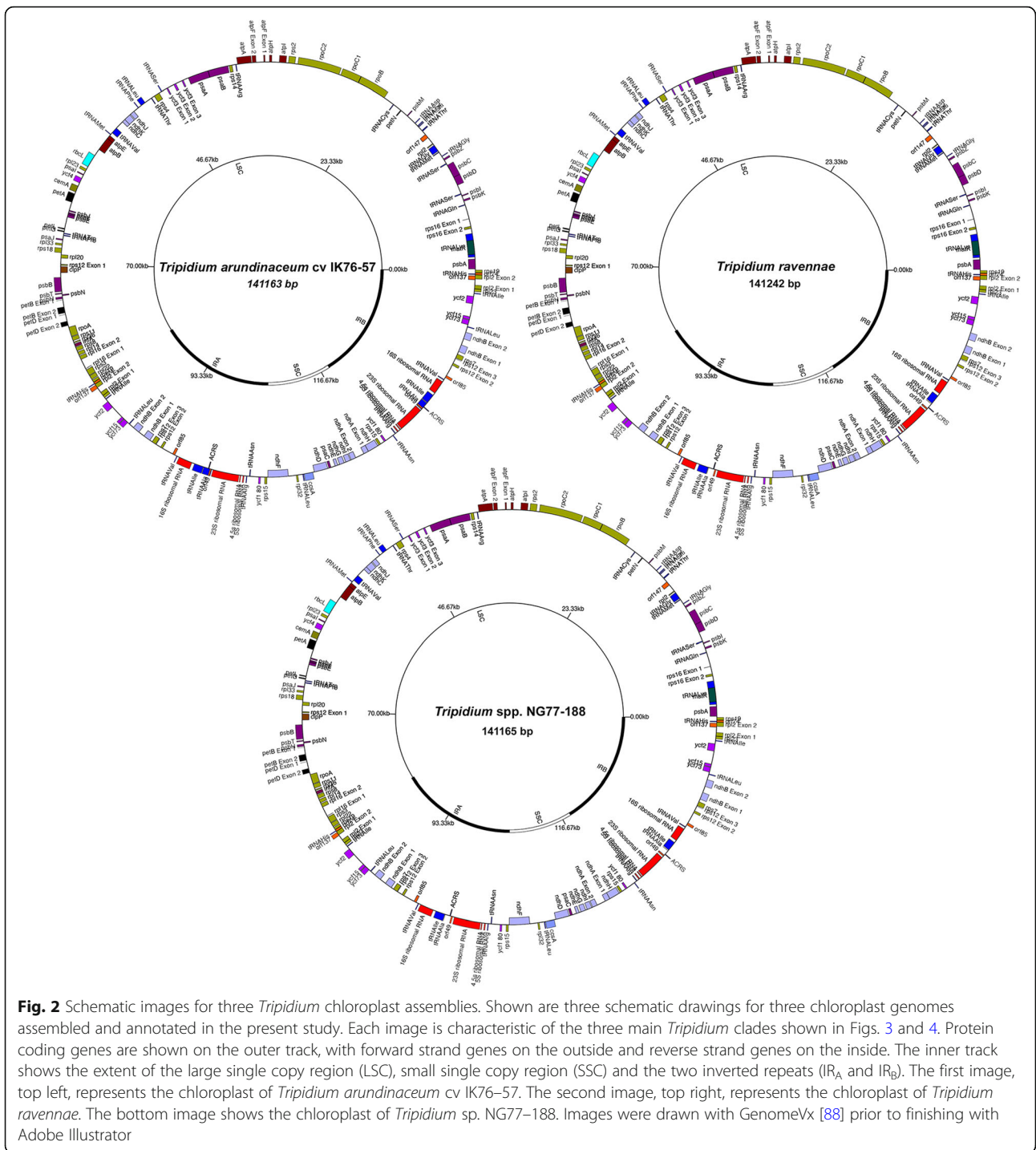
Primer design

Chloroplast amplification primers (an additional PDF file details these Additional file 1) were designed to be universal to the Panicoideae, using *Saccharum* hybrid, *Miscanthus sinensis*, *Zea mays* and *Cenchrus americanus* plastids as references. Wherever possible, tRNAs were used as sites of primer design (these tend to be evolutionarily well conserved). The majority of tRNA-based primers were designed manually with a target melting temperature of 78 °C. For optimal PCR, amplicons were limited to a maximum 20,000 bases. In the few cases where there were no tRNAs available to design primers, the region was input into NCBI’s primer design tool [43]. Final primers were checked with Amplify4 [44] for specificity, primer dimerization and universality. Figure 1 shows a schematic mapping of predicted and subsequently assembled amplicons to the published *E. arundinaceus* chloroplast genome. There are two gaps in the map, within the IR_B region. However, as these two gaps are identical to the sequence on the IR_A inverted repeat, we still have complete coverage of the chloroplast (Fig. 2). Primer pair 13 is not shown on the schematic, as this was designed solely to assist with assembly because SPAdes (our assembler of choice) is sensitive to large troughs and peaks in the assembly graph. As the inverted repeats have double the coverage as compared to the SSC (short single copy) region separating them, an additional primer pair covering the core of the SSC region was designed to increase and even out read coverage in this region. Gel images for the amplicons generated for the *Tripidium* species prior to sequencing are shown as an additional PDF file (Additional file 2).

Phenotypic analyses

Formal phenotypic analyses were performed on the South African Sugarcane Research Institute’s collection





of *Saccharum robustum*, *Saccharum officinarum*, *Saccharum spontaneum* and *Erianthus* collections (excluding probable hybrids). Based on the possession of ciliate auricles and eciliate ligules, deep (rather than lateral) rhizomes, leaf sheaths that are longer than internode to internode distances, a conspicuous midrib on the leaf underside, leaf blades that taper towards both the tip and sheath and caespitose (clump-forming) natures six

accessions were re-classified as Old World *Erianthus* types which probably belonged to genus *Tripidium* (see Kew’s GrassBase for detailed species descriptions, which were strictly adhered to [45]). Accessions showing these features were phenotypically typed in detail and the following accessions were re-classified based on this study: *Saccharum spontaneum* cv Spontaneum 1 to *Tripidium arundinaceum* cv SA-E1; *Saccharum spontaneum* cv

Spontaneum 2 to *Tripidium kanashiroi* cv SA-E2; *Saccharum robustum* cv IK76–417 to *Tripidium arundinaceum* cv IK76–417 and *Saccharum robustum* cv NG77–188 to *Tripidium* sp. cv NG77–188. Of these, *Tripidium* sp. cv NG77–188 was the most phenotypically divergent, and despite being clearly caespitose it was much shorter in stature with thin culms that had waxy bands both above and below the internode and narrow leaves. These accessions, along with one accession known to be *Erianthus arundinaceus*, and one known to be *Tripidium ravennae*, were subject to chloroplast amplification and sequencing.

Based on these studies, and additional phenotypic examination of all species, a new key to the identification of *Tripidium* species was prepared (Table 2). This is the first key to the identification of all *Tripidium* species.

Plastome assemblies

The eight *Tripidium* plastomes assembled in this study were remarkably similar in terms of size, gene and genome

structures. Plastome length varied from 141,105–141,242 bp (*Tripidium rufipilum* and *Tripidium ravennae*, respectively), with a mean of 141,168 bp, the LSC (large single copy region) being the most variable. The *Saccharum spontaneum* SES196 dataset assembled in this study was cognate with the SES205A and SES234B *Saccharum spontaneum* chloroplasts that we had published previously (Additional file 3). All plastome assemblies had 84 protein-coding genes, 33 non-protein coding genes (ribosomal RNA and tRNAs), three pseudogenes and four origins of replication (Fig. 2) [all counts are for unique genes and exclude duplicates on the second, IR_B, inverted repeat].

Plastome sequences derived from PCR amplification (SASRI data) had average read coverage of 330×, whilst plastome sequences derived from reduced representation whole genome amplification sequencing (USDA accessions) had an average of 67× coverage.

Phylogenomic analyses

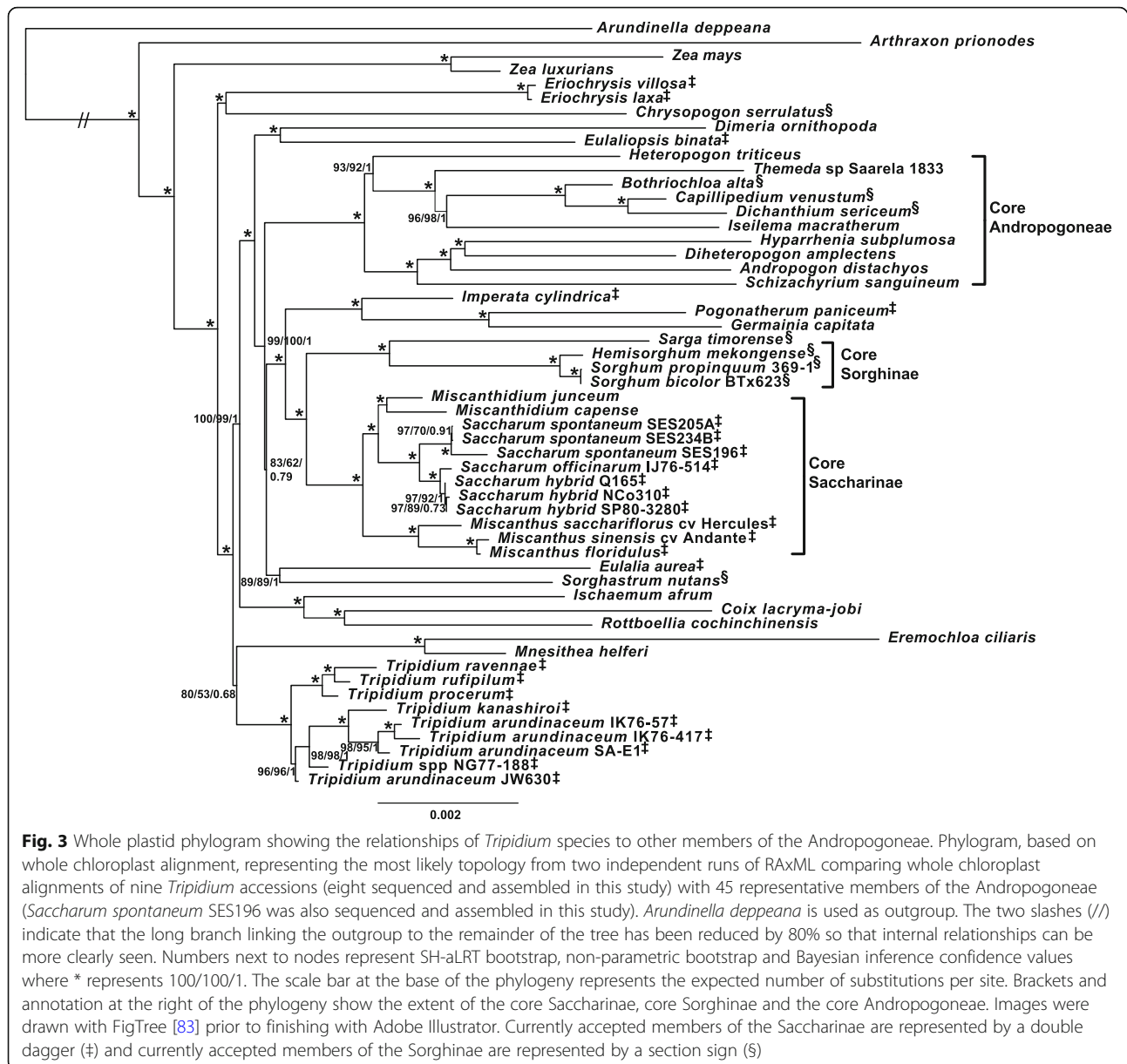
To test the possible influence of partitioning and the two inverted chloroplast repeats on the chloroplast-based phylogenies both maximum likelihood and Bayesian analyses were run with the second inverted repeat IR_B both present and absent from the alignment. Tree topologies in both cases were identical, and branch support showed less than 5% variance. However, backbone branch supports were relatively poor as were some internal branches, particularly within the *Tripidium* genus. Moreover, Neighbor-Joining, Parsimony Ratchet analyses and SH-aLRT analyses presented an alternate tree topology (particularly within *Tripidium* and for the relationship of *Sorghastrum* to the core Andropogoneae and Saccharinae). For an example topology (with supports) using the standard partitioning of the alignment into LSC, IR_A and SSC regions see the PDF document of the additional file (Additional file 4). As such, it was decided to further partition the LSC, IR_A and SSC regions of the chloroplast into protein coding genes, RNA-coding genes and non-coding sequences (8 partitions in all) with all partitions analyzed independently. This is a much finer partitioning than is typically performed for whole chloroplast analyses and it yielded significantly improved branch supports and a more consistent tree topology. For the final partitioning scheme, two independent runs of RAxML (with different ‘-p’ seed numbers and 100 replicates) yielded the same tree topology, with the best log likelihood $-\ln L = 409,657.45$ (Fig. 3). The current phylogeny focuses on the Andropogoneae, with *Arundinella deppeana* (a member of the Arundinelleae) as the outgroup.

Within the SASRI collection, accessions SA-E1 and SA-E2, were previously identified as *Saccharum spontaneum* and accessions NG77–188 and IK76–417 were previously identified as *Saccharum robustum*. This confusion of identification undoubtedly arose due to the

Table 2 Key to the identification of *Tripidium* species

1a. Habitat: Perennial, caespitose
2a. Glumes similar
3a. Leaf blade surface smooth on both sides
4a. Spikelet callus hairs white, 1.5× length of spikelets
5a. Rhachis internode 3.5–4 mm long
<i>Tripidium arundinaceum</i>
5b. Rhachis internodes 6–7 mm long
<i>Tripidium procerum</i>
4b. Spikelet callus hairs yellow or red, 2–3× length of spikelet
<i>Tripidium rufipilum</i>
3b. Leaf blade surface scabrous — rough on both sides
<i>Tripidium strictum</i> [‡]
2b. Glumes dissimilar
6a. Glumes exceeding apex of florets, lower glumes elliptic, upper glumes lanceolate
<i>Tripidium ravennae</i> [‡]
6b. Glumes 1× length of spikelet, lower glume 2 keeled, lower glume hairs 4–9 mm long, upper glume without keels
<i>Tripidium bengalense</i> [‡]
1b. Habitat: Perennial, culms solitary
Key characters: Culms erect, 70–100 cm long, 5–15 mm diameter, rhizomes short, leaf blade hairy at base. Spikelet callus hairs 3–6 mm long.
<i>Tripidium kanashiroi</i>

Presented here is a key to the identification of *Tripidium* species. This key covers all currently accepted *Tripidium* species as well as those species moved into genus *Tripidium* as a consequence of the present study. The characters chosen in this key allow for simple disambiguation amongst the species. The three species marked with a double dagger (‡) superscript are the currently accepted members of *Tripidium*. The four remaining species were added to *Tripidium* as a result of this study

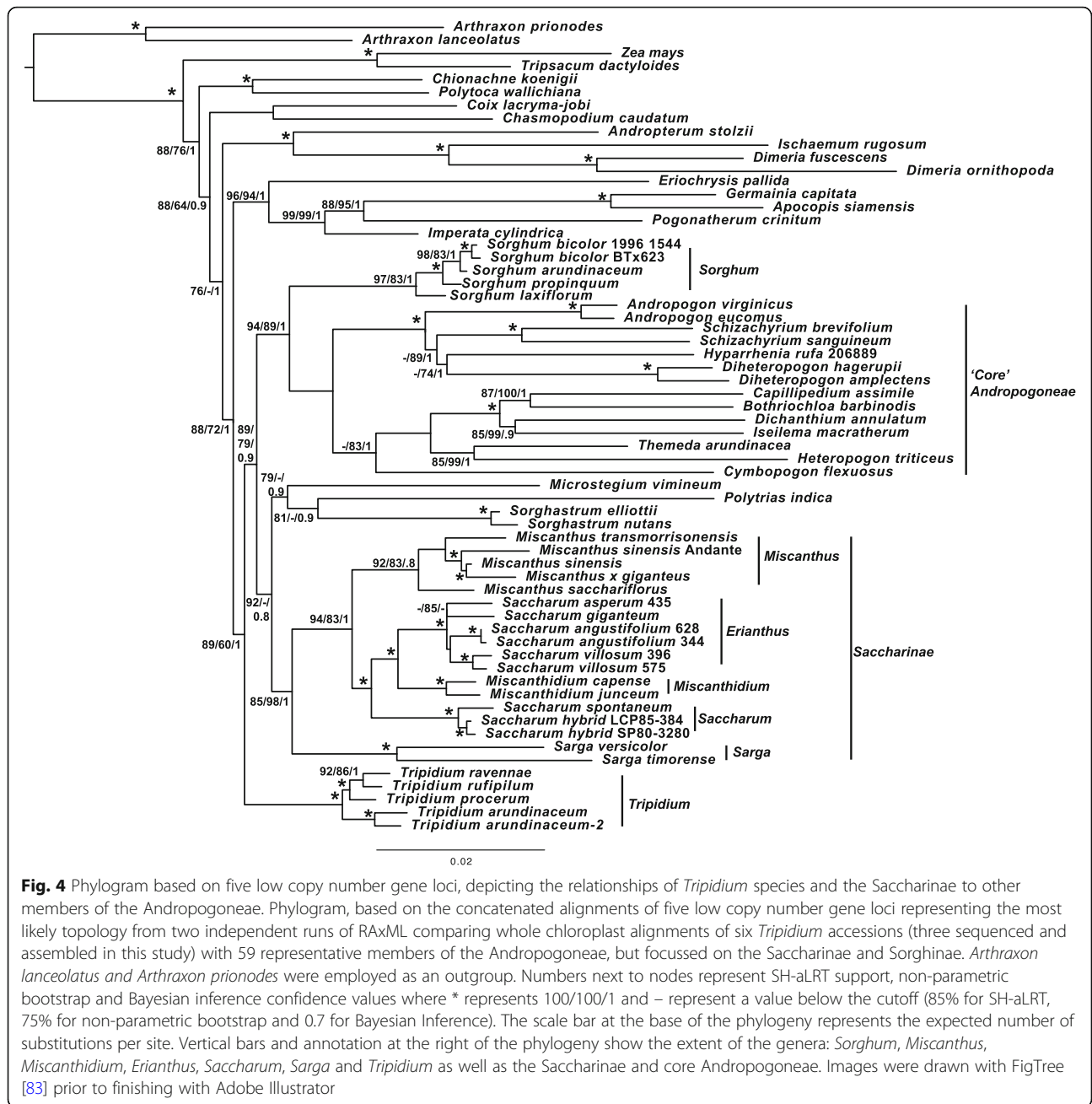


presence of rhizomes and the clump-forming nature of these accessions. However, careful phenotypic analyses revealed that these accessions should more properly be included within genus *Tripidium*, a conclusion that is supported by their monophyly within our phylogeny, particularly if one follows Grassl [9] in excluding the ‘bunch grass’ types from *Saccharum*.

Our gene locus phylogeny (Fig. 4) is based on the work of Estep et al. [17]. As in previous studies, terminal branches have good support, but internal nodes tend to have poorer bootstrap support (but good BI support), an indication of rapid radiation [17]. The topology is almost identical to the phylogeny presented by Welker et al. [16], with the exception that we place *Polytrias indica* as

sister to *Sorghastrum* rather than sister to *Sorghum*. However, our placement of *Polytrias* agrees with the phylogeny of Estep et al. [17]. In addition, comparisons between the chloroplast (Fig. 3) and gene locus (Fig. 4) phylogenies show that both phylogenies are largely congruent.

A side-by-side comparison, presented as a tanglegram (Fig. 5) shows that the gene locus phylogeny and the whole chloroplast genome phylogeny are largely congruent, with only eight genera differing between them. The major differences being the relative positioning of the core *Sorghinae*, a monophyletic clade formed from *Eriochrysis*, *Imperata*, *Pogonatherum*, *Apocopsis* and *Germania* and a monophyletic clade formed from *Ischaemum*,



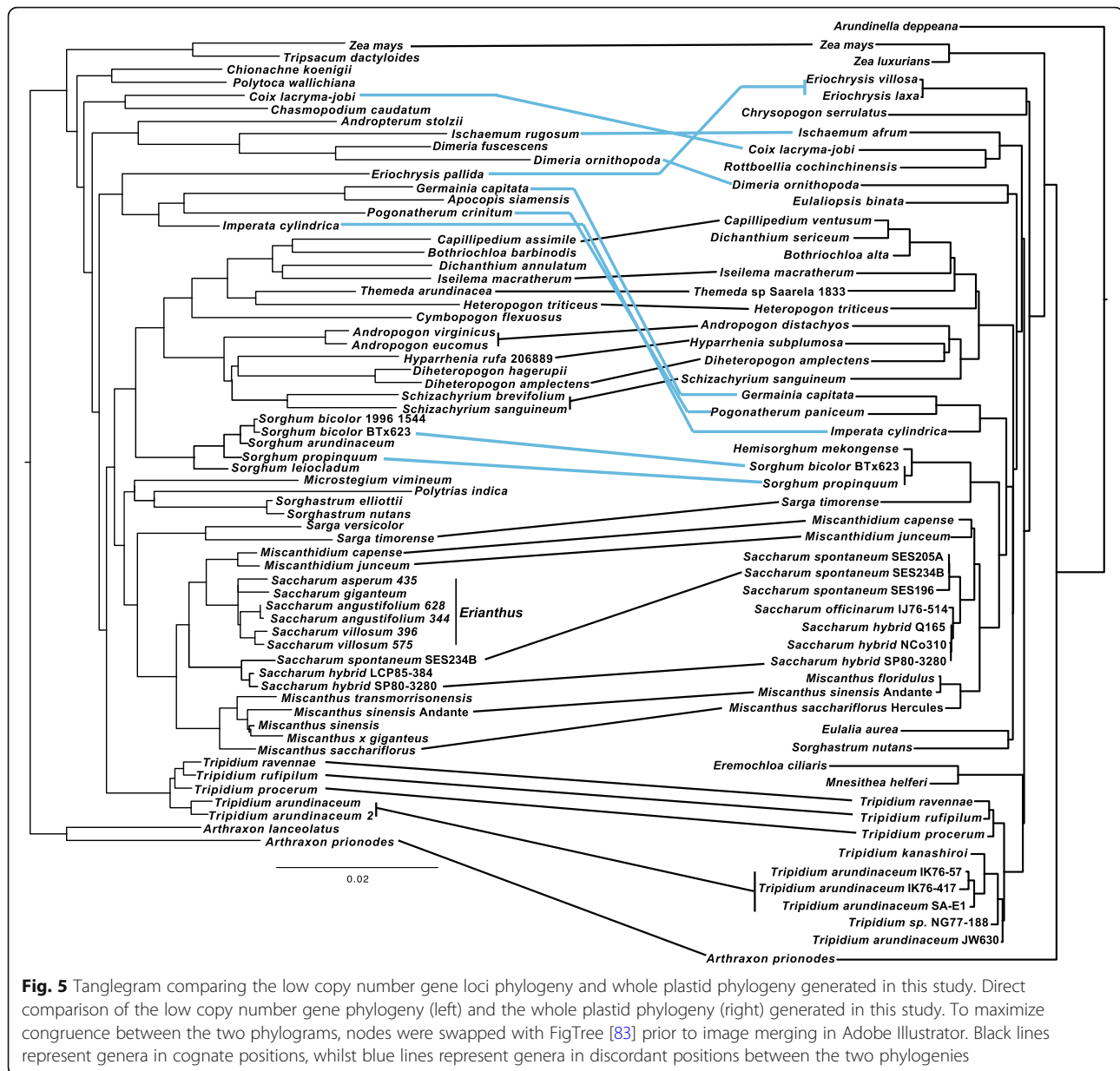
Coix and *Rottboellia* in the chloroplast phylogeny which splits into two more distal clades in the gene locus phylogeny. However, monophyly of *Coix*, *Chasmopodium* and *Rottboellia* is confirmed by the work of Estep et al. [17].

Imperata, *Pogonatherum* and *Germainia* remain monophyletic in both phylogenies, but this clade forms an outgroup to the core Andropogoneae and Saccharinae, with good support, in the chloroplast phylogeny (Figs. 3 and 5) whilst it is sister to *Eriochrysis* with this entire clade being an outgroup to *Tripidium* in the gene locus phylogeny. The core Sorghinae form a monophyletic grouping in both phylogenies, but are placed as sister to *Sarga* in the

chloroplast phylogeny and as sister to the core Andropogoneae in the gene locus phylogeny.

Coix, *Ischaemum* and *Rottboellia* form a monophyletic grouping in the chloroplast phylogeny but monophyly of this group is not supported by the gene locus phylogeny (Figs. 4 and 5) where *Coix* is sister to *Chasmopodium* and this grouping is immediately antecedent to a clade formed by *Andropterum*, *Ischaemum* and *Dimeria*.

Both phylogenies (Fig. 5) support *Tripidium* as a monophyletic grouping which, in the chloroplast phylogeny (Fig. 3), has *Eremochloa ciliaris* and *Mnesithea helferi* as a sister clade (with relatively weak support). Our two

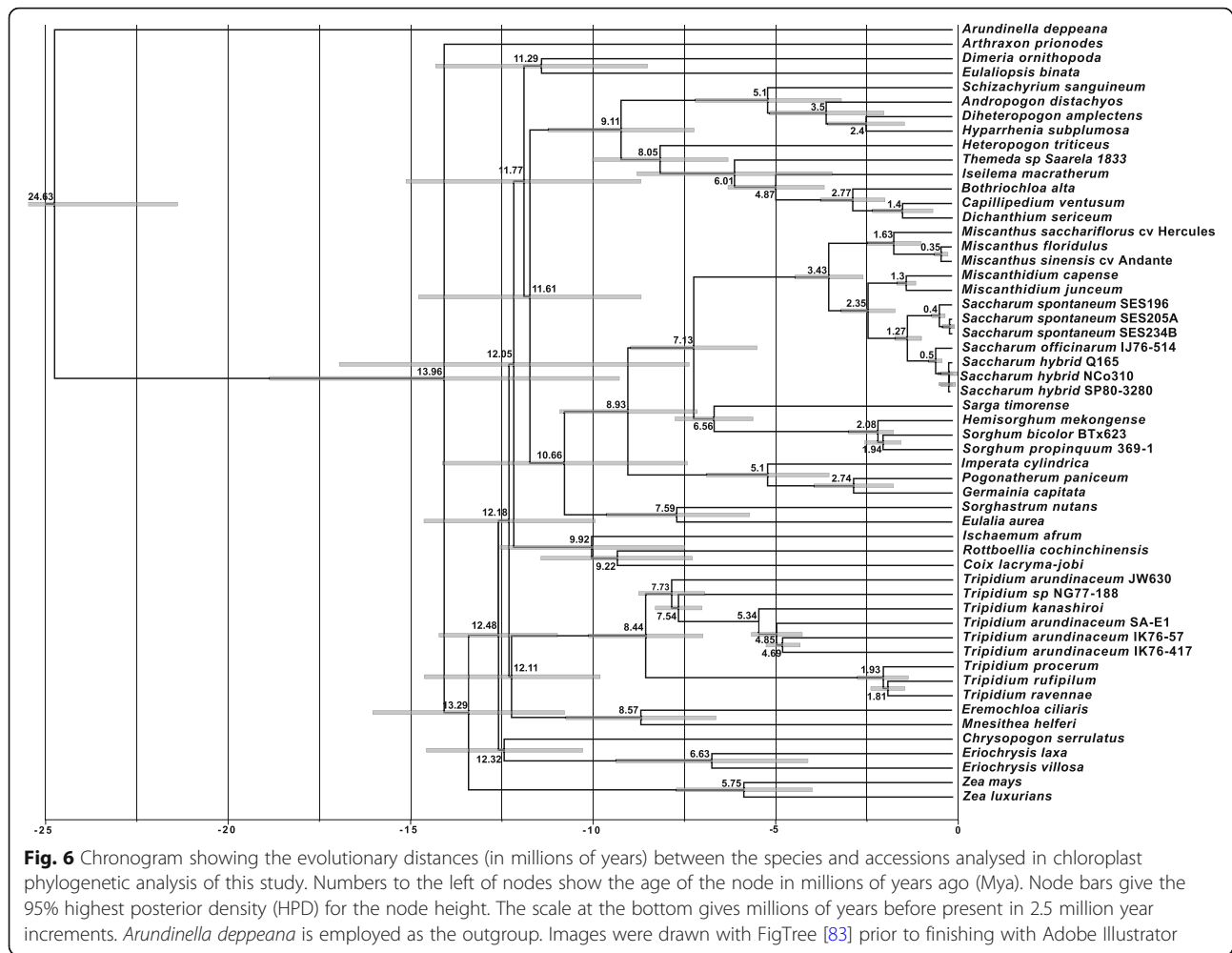


phylogenies support the division of *Tripidium* into a *Tripidium arundinaceum/kanashiroi* clade and a *T. procerum/T. rufipilum* and *T. ravennae* clade, with near 100% branch support in both groups.

Within the core *Andropogoneae*, a division into a clade represented by *Andropogon* and a clade represented by *Bothriochloa* is supported by both phylogenies. Both nuclear locus and plastome phylogenies also support *Sarga* as being sister to the core *Saccharinae*, which is represented by *Miscanthus*, *Miscanthidium* and *Saccharum*. The gene locus phylogeny (Fig. 4) places New World *Erianthus* species (represented by the type species, *Erianthus (Saccharum) giganteus* as sister to *Miscanthidium*. All these genera receive 100% support in both phylogenies as being monophyletic.

The chloroplast phylogeny chronogram (Fig. 6) and gene locus chronogram (Fig. 7) (both chronograms calibrated on the divergence of *Zea mays* at ~13.8 Mya) concur in placing the divergence of *Tripidium* at 12.1–12.2 million years ago, with the crown *Tripidium* species diverging at 7.3–8.4 million years ago (Fig. 5).

Core *Andropogoneae* diverged from the *Saccharinae* at 10.8–11.6 million years ago. The chloroplast-based chronogram (Fig. 6) gives an age for the emergence of the *Andropogoneae* with *Arundinella deppeana* at 24.6 Mya. This is consistent with, but more ancient than, the work of Vicentini et al. [46], placing the origins of *Arundinella* at 19 Mya. Our age of divergence of *Saccharum* and *Miscanthus* at 3.4–3.6 Mya is consistent with previous



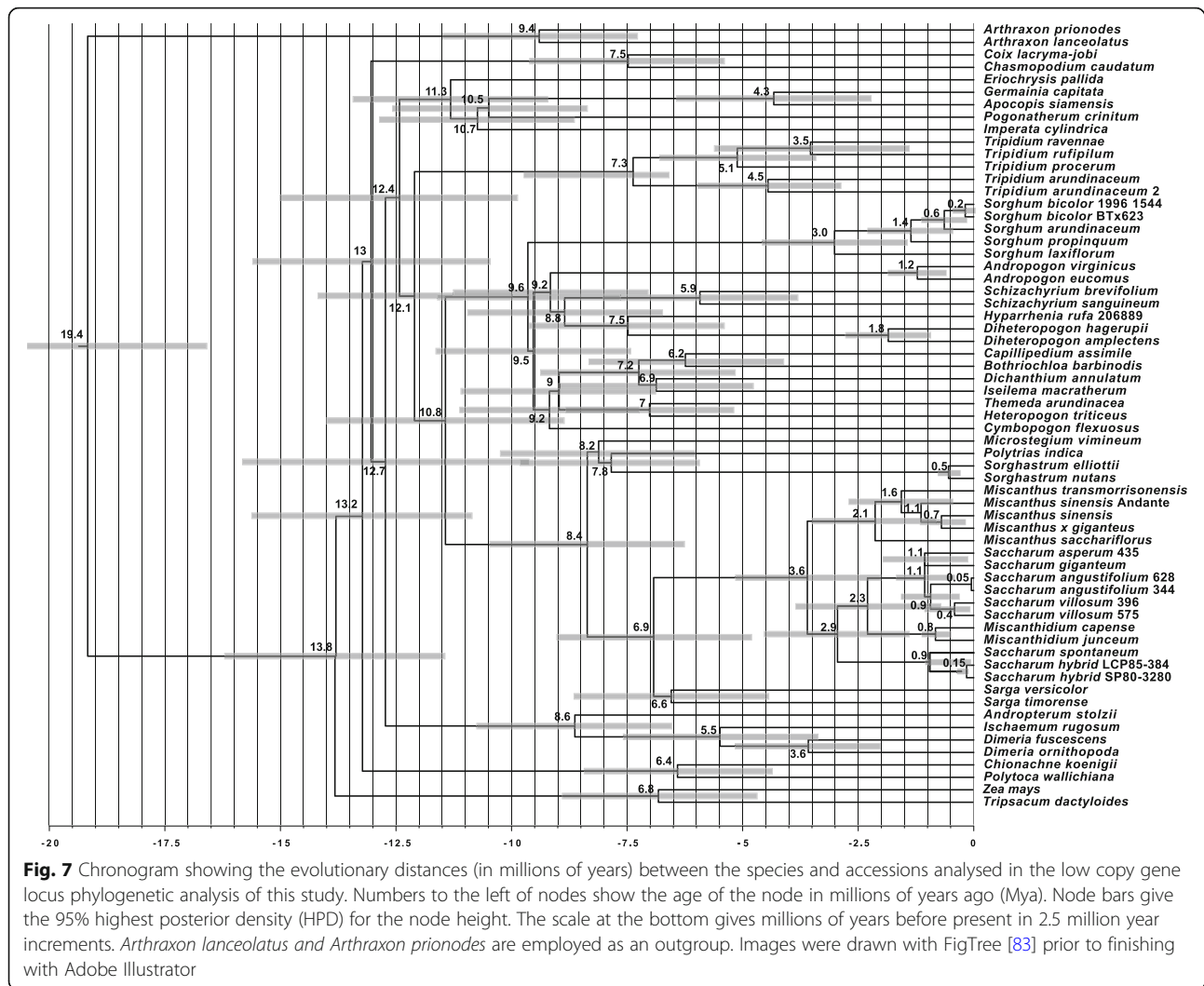
studies that placed this divergence at 3.8 Mya [1, 17]. The split between *Saccharum spontaneum* and the crown *Saccharum* species is consistent with our previous work at 0.9–1.3 Mya [1].

Discussion

The concept of the ‘Saccharum Complex’ as a collection of interbreeding species resulting in the evolution of sugarcane has had a profound, and probably detrimental, effect not only on sugarcane breeding, but also on the taxonomy of the Andropogoneae as a whole. Many of these ‘Saccharum complex’ species have been placed wholesale into the subtribe ‘Saccharinae’. A good case in point is *Erianthus* sect. *Ripidium*. The original definition of this section recognized eight Old World *Erianthus* species, which Grassl [9] placed within genus *Ripidium*. Tropicos, following Scholz [12] places three species within genus *Tripidium* [47] though Kew’s GrassBase places all these species within *Saccharum* [7]. These inconsistencies, and the current definition of genus *Tripidium* as *incertae sedis* within the Andropogoneae necessitates a large-scale

molecular phylogenetic study to verify the validity of the genus. Our data indicates that a more complex partitioning scheme than is typically used for phylogenomic studies (see [Materials and Methods](#) for details) yields improved branch support and improved topological robustness. As a result, this partitioning scheme was employed for all phylogenetic analyses reported in this paper. This finding of improved topological support with finer-grained partitioning is comparable to the study of Folk et al. [19], analysing *Heuchera*.

For the current study, complete chloroplast sequences of eight potential *Tripidium* accessions and a single *Saccharum spontaneum* accession as well as low copy number gene loci (apo1, d8, ep2 exon 7, ep2 exon8 and rep1) were assembled. During assembly and sequencing of the gene loci, we observed no secondary gene copies in *Tripidium*, indicating that there has been no recent hybridization within this genus. Our gene loci assemblies bolstered sampling in *Tripidium*, *Saccharum*, *Miscanthus* and *Sorghum* as compared with earlier studies [16, 17]. By integrating with previously published data



[17, 48] we were able to generate a low copy number gene phylogeny (Fig. 4) to complement our whole chloroplast phylogeny (Fig. 3).

Whilst the chloroplast (Fig. 3) and nuclear gene locus (Fig. 4) phylogenies are broadly congruent, a close comparison, presented as a tanglegram (Fig. 5) demonstrates several points of large-scale phylogenomic discord where the chloroplast and nuclear signals cannot be reconciled. Discordance among genomes has long been observed in plants [49]. Both hybridization and incomplete lineage sampling (ILS) cause phylogenetic discordance, yet expected gene tree distributions under these processes are different.

Manually comparing the two phylogenies, all discordant branches have good support (greater than 85% SH-aLRT, 70% bootstrap and 0.9 BI), which means false identification of conflict could only occur in cases of systematic error [23]. Moreover, the majority of the examples of incongruence presented in Fig. 5 involve branches that are distant in the

tree. Mechanistically, the pattern of discordance observed in Fig. 5 could be explained as resulting either from ancient hybridization or from deep lineage sorting (coalescent stochasticity). However, as hybridization is particularly common in grasses (see [29, 30]) ancient reticulation affords us the most parsimonious explanation for the discordance. This is especially the case for the Andropogoneae, where the chronograms (Figs. 6 and 7) show multiple lineages that diverged rapidly between 12.5 and 11.6 million years ago. This corresponds quite well with the central part of the Middle Miocene Climate Transition (MMCT) [50] and the start of the diversification of C4 grasses [51]. Rapid radiation can often prove problematic for phylogenetic analyses, particularly near the backbone of the phylogenetic tree [23], though our more comprehensive partitioning scheme does generally improve branch support in this region (Additional file 4 and Fig. 3).

Our findings are consistent with reticulate evolution being one of the driving forces behind this radiation. It

is also perhaps unsurprising that three genera currently defined as *incertae sedis* within the Andropogoneae (*Eriochrysis*, *Imperata*, *Pogonatherum*) show signals of ancient reticulation with highly discordant genomic and plastid phylogenetic signals. Two of these genera, *Imperata* and *Pogonatherum* form a monophyletic grouping with *Germainia* in both chloroplast and genomic phylogenies (Figs. 3 and 4). The Germainiinae are currently defined as *Apocopsis*, *Germainia* and *Trachypogon*. Though *Trachypogon* was not represented in our taxon sampling, *Apocopsis* is represented in the low copy number gene locus phylogeny (Fig. 4) and is sister to *Germainia*. The surprise is that the gene locus phylogeny places *Eriochrysis* as the outgroup to the Germainiinae (a finding that is consistent with previous studies [16, 17]). It would seem that reticulation occurred early in the evolutionary history of the Germainiinae, so that genomic and plastid evolutionary histories differ. Based on the low copy number gene phylogeny (Fig. 4), our data support the inclusion of *Imperata* and *Pogonatherum* within the Germainiinae.

The evolutionary histories of *Coix*, *Dimeria* and *Ischaemum* differ greatly between the gene locus and chloroplast phylogenies (Fig. 5) and increased taxon sampling may be required to resolve the true reticulate origins of these species.

Possibly the most surprising discordance between the genomic and plastid phylogenies is the positioning of *Sorghum*. In the chloroplast phylogeny (Fig. 3) it emerges as sister to *Sarga*, with this joint clade being sister to the Saccharinae. However, in the genomic gene locus phylogeny, *Sorghum* is sister to the core Andropogoneae. This new placement of *Sorghum* is consistent with the work of Hawkins et al. [48] and our recent work on extended ITS sequences [52]. This means that *Sorghum* last shared a common ancestor with sugarcane 10.8 million years ago (Fig. 6) rather than the more commonly accepted 7.1 million years ago (Fig. 5). This also has implications for plastid/plastid+genomic datasets, which could lead to misleading phylogenetic interpretations.

Previous to this study, the complete plastid sequence of only a single *Erianthus* species; *Erianthus arundinaceus* has been published [53]. This is an East Asian accession (from the Shizuoka Prefecture of Japan). The SASRI *E. arundinaceus* accession (IK76–57) was collected in Indonesia (the IK designation represents a plant collected on the island of Kalimantan). The South African Sugarcane Research Institute accession SA-E1 was derived from India. Previous morphological studies [9] indicated that Indian and Indonesian accessions of *E. arundinaceus* were quite distinct, with Indian *E. arundinaceus* being closer to *E. ravennae* and *E. procerus*. Our phylogenies (Figs. 3, 4, and 5) show that genus *Tripidium* is monophyletic, with complete support. The genus is divided into two groupings, both with good support. However, *Tripidium*

arundinaceum is clearly not monophyletic. The previously published Japanese accession forms an outgroup to our accession NG77–188, previously miss-identified as *Saccharum robustum*. These two species are basal to the remaining *Tripidium kanashiroi* and *Tripidium arundinaceum* accessions. *T. procerus*, *T. ravennae* and *T. rufipilum* cluster together and are sister to all the other *Tripidium* accessions. Indian *Tripidium arundinaceum* SA-E1 is basal to the crown *Tripidium arundinaceum* accessions, but is separate from *Tripidium (Erianthus) ravennae*. Thus, chloroplast data supports Indian *Tripidium arundinaceum* being phylogenetically more closely related to Indonesian *T. arundinaceum* than to *T. ravennae*.

Both chloroplast-based and low copy number gene locus based chronograms (Figs. 6 and 7) show that *Tripidium* diverged from the other grasses about 12.1 Mya. The two crown *Tripidium* clades diverged about 7.3 million years ago (Fig. 6). The chloroplast analysis (Fig. 3) places *Tripidium* sp. NG77–188 and *T. arundinaceum* JW630 as outgroups to the crown clade formed by *Tripidium kanashiroi* and the *Tripidium arundinaceum* species. Monophyly of this entire grouping has 100% support in both plastome and low copy number gene locus phylogenies (with identical topologies) and monophyly of the two sub-groupings within *Tripidium* has excellent support. Thus, our study provides two independent analyses supporting *Tripidium* as a monophyletic genus that last shared a common ancestor with *Saccharum* species 12.1 Mya. As such, we propose that *Tripidium* should be the preferred genus name for these species and that they should be removed from genus *Saccharum*.

Indeed, it is generally accepted that the demarcation line between species of Andropogoneae that could be members of the Saccharinae, and those which most certainly are not, are the core Andropogoneae. Species and genera that arose before the divergence of the core Andropogoneae (about 11.6 Mya) are not members of the Saccharinae, whilst genera that are sister to the core Andropogoneae could be members of the Saccharinae [54].

Our plastid phylogeny, (Fig. 3) shows that the crown *Tripidium* group formed by *T. kanashiroi* and three *T. arundinaceum* accessions have two basal accessions. The first of these is *T. arundinaceum* JW630, diverging 7.73 Mya (Fig. 5) and the second formed by *Tripidium* sp. NG77–188 diverging some 7.54 Mya. These are evolutionarily distinct from the crown *T. arundinaceum* group and separated from them by over 2 million years. *Tripidium* sp. NG77–188 is also morphologically distinct from the crown *T. arundinaceum* grouping. Despite being characterized as *T. arundinaceum*, accession JW630 is the most basal of the group and it is separated from the remaining *T. arundinaceum* accessions by another species, *T. kanashiroi*. The presence of *T. arundinaceum*

accessions in the basal and crown groups means that *T. arundinaceum* is not monophyletic. Thus, *T. arundinaceum* JW630 is, undoubtedly, a different species from the core *T. arundinaceum* and *T. kanashiroi* species assembled in this study, and the naming and identification of these basal *Tripidium* species requires further investigation.

In our plastid analysis (Fig. 3), the monophyletic *Tripidium* clade is sister to a clade formed by *Eremochloa ciliaris* and *Mnesithea helferi*. It should be noted that Neighbor-Joining analysis supports the placement of *Eulalia* and *Sorghastrum*, but places *Mnesithea* and *Eulalia* as sister to *Tripidium*, but SH-aLRT supports the topology as presented. Indeed, support for *Eremochloa/Mnesithea* being sister to *Tripidium* is only slightly better than the topology where *Eremochloa/Mnesithea* is an immediate antecedent clade to *Tripidium*. This may be partly due to the rapid radiation seen in this portion of the phylogeny and increased sampling in this area may help resolve the uncertainty within the relative position of the *Eremochloa/Mnesithea* clade or uncertainty could be due to ancestral reticulate evolution. However, the uncertainty in the placement of these two clades does not affect our conclusions regarding *Tripidium*.

Our study therefore provides excellent support for *Tripidium* as a monophyletic genus that last shared a common ancestor with *Saccharum* species 12.2 Mya. As such, we propose that *Tripidium* should be the preferred genus name for these species and that they should be removed from genus *Saccharum*.

In addition, our recent study [1] gives a 3.4 million year window in which members of the Poaceae can hybridize in the wild. *Tripidium* species clearly fall outside this window and though they can be crossed by human mediation they cannot naturally hybridize in the wild. Indeed, our findings are entirely consistent with the problems in hybridizing these two very divergent genera that sugarcane breeders have encountered over the past 50 years. In addition, the large evolutionary distance between *Tripidium* and *Saccharum* explains why crosses between the genera result in the duplication and loss of incompatible chromosomes (as hybrids are intergeneric). Thus, in the strictest sense, those rare hybrids that occur between *Tripidium* species and *Saccharum* species should be defined as partial hybrids and not as true hybrids.

As only three species have previously been formally included in genus *Tripidium*, our findings show that the following new combinations are necessary: *Tripidium arundinaceum* (Retz.) Lloyd Evans comb. nov. based on *Saccharum arundinaceum* (Retz.); *Tripidium procerum* (Roxb.) Lloyd Evans comb. nov. based on *Saccharum procerum* (Roxb.); and *Tripidium kanashiroi* (Ohwi) Lloyd Evans comb. nov. based on *Saccharum kanashiroi* (Ohwi). We confirm by molecular phylogenetic data that *Saccharum rufipilum* is not a member of genus

Miscanthus and is sister to *Tripidium ravennae*. As such, it must be brought into genus *Tripidium* as *Tripidium rufipilum* (Steud.) Lloyd Evans comb. nov. based on *Saccharum rufipilum* (Steud.). With the previous combinations, this results in *Tripidium* being a genus with seven confirmed species. To aid identification of these species, a new key to *Tripidium* identification is presented in Table 2. That *Tripidium* species' physical characteristics can be ranked in this fashion is further evidence for the monophyly of *Tripidium* as a genus.

Our low copy number gene phylogeny places *Germainia* and the extended Germainiinae as a clade that is separate from *Tripidium*, but which is sister to *Tripidium*, the core Andropogoneae and the Saccharinae *sensu lato* as a whole. Thus *Tripidium* and *Germainia* are not closely allied as had previously been suggested [18].

As our low copy number gene locus phylogeny (Fig. 5) places *Sorghum* as an outgroup to the core Andropogoneae, we are left with a core Saccharinae subtribe of *Miscanthus*, *Miscanthidium* and *Saccharum*, which all lie within the wild crossing range of about 3.8 million years. Figure 3 clearly shows that the Saccharinae are not monophyletic and that three genera previously included in both the *Saccharum* complex and the Saccharinae (*Tripidium*, *Sorghum* and *Eriochrysis*) can be excluded as they are more distantly related to *Saccharum* than the core Andropogoneae. *Pogonatherum* and *Imperata* can also be excluded due to evolutionary distance, particularly if the low copy gene locus data (Fig. 4) is given precedence. It should be noted that genus *Miscanthidium* is not currently defined as part of the core Saccharinae (Fig. 3). Previously these species were included within *Miscanthus* [8] but our data clearly show that they are far closer to *Saccharum* than to *Miscanthus*. Our low copy number gene locus phylogeny (Fig. 4) also shows that New World *Erianthus* species are sister to *Miscanthidium*, thus they are part of the Saccharinae, but distinct from *Saccharum* (the *Erianthus* + *Miscanthidium* clade being sister to *Saccharum*).

Our phylogenies also demonstrate that the Sorghinae, as currently defined, is not monophyletic, with *Sorghastrum* (currently defined within the Sorghinae) clearly lying outside the Sorghinae, being sister to the core Saccharinae. Three other genera: *Bothriochloa*, *Capillipedium* and *Dichanthium* lie within the core Andropogoneae and can also be excluded from the Sorghinae. *Chrysopogon*, which is also currently included within the Sorghinae [6] can also be excluded as it is distal to the core Andropogoneae. The split between *Sorghum* and *Sarga* species is further emphasized as low copy number gene locus phylogenetics places *Sorghum* as sister to the core Andropogoneae, whilst *Sarga* is sister to the core Saccharinae.

Thus, our data supports a core Sorghinae clade formed only from *Hemisorghum* and *Sorghum* and a Saccharinae clade formed from the genera *Saccharum*, *Miscanthus*,

Miscanthidium and *Erianthus* that has *Sarga* as a sister clade. Phylogenetically, we describe two new subtribes, as there is a natural clade formed from *Eriochrysis pallida*, *Germainia capitata*, *Apocopsis siamensis*, *Pogonatherum crinitum* and *Imperata cylindrical* that can be encompassed within the Germainiinae. An extended clade formed from *Sarga*, *Miscanthus*, *Miscanthidium*, *Erianthus* and *Saccharum* could be given the subtribe name ‘Saccharinae’.

Conclusion

The genera *Tripidium*, *Eulalia* and *Eriochrysis* should be excluded from the *Saccharinae* as they break the subtribe’s monophyly and are too distant from *Saccharum* to hybridize with *Saccharum* species in the wild. By the same argument *Chrysopogon*, *Bothriochloa*, *Capillipedium*, *Dichanthium* and *Sarga* should be excluded from the Sorghinae.

We also clearly demonstrate that genus *Tripidium* is not allied to Germainiinae, as suggested by Soreng et al. [18], though the Germainiinae are sister to *Tripidium*. Here we need to make a clear distinction between the ‘Saccharum complex’, a group of species that can interbreed in the wild (this can only include the genera: *Miscanthus*, *Miscanthidium* and *Saccharum*, as well as the New World *Erianthus* species [16]) and the *Saccharinae*, a taxonomically monophyletic grouping that includes genus *Saccharum* (in totality this subtribe would include *Saccharum*, *Miscanthus*, *Miscanthidium*, the New World *Erianthus* species and *Sarga* possibly with the inclusion of *Microstegium vimineum*, *Polytrias indica* and *Sorghastrum* which forms an outgroup to the ‘Saccharum complex’ (with good support)) based on low copy number gene phylogenetics (Fig. 4).

We also demonstrate, based on low copy gene locus phylogenetics, that genus *Erianthus*, as originally defined, is not monophyletic and should be divided into *Tripidium* and *Erianthus*, with genus *Erianthus* only including the New World species. Moreover *Erianthus* species are sister to *Miscanthidium* and are therefore phylogenetically distinct from *Saccharum*. As *Erianthus giganteus* is the type species we propose that New World *Erianthus* species should be removed from genus *Saccharum* and returned to genus *Erianthus*.

Despite over 50 years’ worth of efforts in introgression breeding, there has been little success in the generation of valid *Saccharum/Erianthus* hybrids. Our phylogeny clearly reveals that, as genera, *Saccharum* and *Tripidium* are 8.6 million years divergent, last sharing a common ancestor 12 million years ago. *Tripidium* is a monophyletic grouping composed of seven species, with a potential for the addition of two additional species, yet to be formally described. The inclusion of *Erianthus* sect *Ripidium* (which we re-classify as *Tripidium* species) within the

‘Saccharum complex’ has led sugarcane breeders down a blind alley for five decades and more, whilst breeders have ignored the potential of the New World *Erianthus* species, which do lie within the wild crossing window with sugarcane. Here we resolve the confusion and re-define the membership of the genus *Tripidium*, the ‘Saccharum complex’ and the *Saccharinae* based on robust molecular systematic studies.

Materials and methods

Plant materials and DNA isolation

Tripidium arundinaceum and *Tripidium ravennae* leaves were collected from the South African Sugarcane Research Institute’s (SASRI) collection, along with four other samples identified as *Tripidium*, but of previously unknown species. *Tripidium procerum*, *Tripidium rufipilum* and *Saccharum spontaneum* SES196 were from the USDA collection and DNA sequencing for these accessions has been reported previously [42]. For the SASRI accessions, total DNA was isolated from liquid nitrogen frozen and ground leaf material using the standard CTAB method according to Wang et al. [55]. Complete chloroplasts were amplified using the 13 primer pairs designed for this project (Additional file 1). Polymerase Chain Reaction (PCR) amplifications were performed using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific). Each reaction contained 50 ng of high quality genomic DNA, 1.5 mM MgCl₂ with 0.2 mM each deoxynucleotide triphosphates (dNTPs), 0.5 M each of forward and reverse primer, 0.4 Units Phusion Hot Start II High-Fidelity DNA Polymerase and reaction buffer as supplied by the manufacturer. Thermal cycling conditions were as follows: Initial denaturation at 98 °C for 1 min, followed by 35 cycles of: 98 °C for 10 s and annealing and extension at 68 °C for 15 min. A final extension step was performed at 72 °C for 10 min followed by a hold at 4 °C. Amplicons were separated by gel electrophoresis (see example gel lanes, which are available in PDF format as an additional document Additional file 2), prior to gel elution and Illumina sequencing (Genotypic Technology Pvt. Ltd).

Phenotypic characterization

Using Kew’s GrassBase [7] descriptions as a foundation, basic phenotypic characters for *T. arundinaceum*, *T. kanashiroi*, *T. bengalense*, *T. ravennae*, *T. procerum*, *T. rufipilum* and *T. strictum* were extracted. Potential *Tripidium* accessions from SASRI’s living collection were phenotypically characterized and matched against the above descriptions. Once identified, further characters were taken from SASRI’s accessions for *T. arundinaceum*, *T. kanashiroi* and *T. ravennae*. Further analysis of *Tripidium strictum* was made against the Missouri Botanical Garden (MBG) specimen of *Saccharum strictum* (MO-2397278). For *T. bengalense*, *T. rufipilum* and *T. procerum* the Herbarium Kewense (K) specimens of *Saccharum bengalense* (K000943381), *S.*

rufipilum (K000309025) and *S. procerum* (K001128357) were analysed. Unique characters were used to generate a key to recognition of *Tripidium* species.

Chloroplast assembly

The sequences derived from three USDA accessions were assembled as described previously [1] with data downloaded from NCBI's sequence read archive (SRA) for the accessions: SRR2899231; SRR2891248 and SRR2891271, using Mirabait [56] and SPAdes v 3.10 [57], with a baiting k-mer of 27 and an assembly k-mer series of 25, 33, 55 and 77. Scaffolds were arranged against the previously published assembly of *Erianthus arundinaceus* JW63 (NCBI accession LC160130.1) and the region corresponding to the second inverted repeat was copied, inverted and stitched into the genome (that this region represented a repeat was confirmed by increased read coverage compared with the remainder of the genome). The SASRI accession raw read data were subject to adapter trimming and cleaning with Trimmomatic [58]. Trimmed reads were assembled, typically generating seven or eight scaffolds, which could be arranged on the *Erianthus arundinaceus* backbone. Any small gaps in the initial assemblies were filled by excising a 2 kb region around the gap and re-assembling this region by baiting reads with Mirabait (k-mer of 31) and assembling the baited reads with SPAdes before inserting the reassembled region to close the gap in the main assembly.

Low copy number locus assembly

Following Estep et al. [17], low copy number regions were assembled for the following genes: *Aberrant panicle organization1* (*apo1*), *Dwarf8* (*d8*), two exons (7 and 8) of *Erect panicle 2* (*ep2*), and *Retarded palea 1* (*rep1*). We assembled these regions from the following GenBank sequence read archive (SRA) accessions: *Sorghum propinquum* (Knuth) Hitchc. (SRR072055); *Sarga versicolor* (Andersson) Spangler (SRR427176); *Sarga timorensis* (Kunth) Spangler Gypsum 9 (SRR424217); *Miscanthidium junceum* (Stapf) Stapf (SRR396848 and SRR396849); *Miscanthus sacchariflorus* (Maxim.) Benth. & Hook. f. ex Franch. var. *Hercules* (SRR486748); *Miscanthus x giganteus* JM Greef & Deuter ex Hodk. & Renvoize (SRR407328 and SRR407325); *Miscanthus trans-morrisonensis* Hayata (SRR396850); *Saccharum* hybrid hort. ex RM Grey LCP85–385 (SRR427145); *Saccharum* hybrid SP80–3280 (SRR1763296), *Saccharum spontaneum* L. SES234B (SRR486146), *Sorghum bicolor* (L.) Moench BTx623 (SRR1945055) and *Sorghum arundinaceum* (Desv.) Stapf PI300119 (SRR999023). *Miscanthus sinensis* Andersson cv. *Andante* and *Andropogon virginicus* L. genomic reads as well as Sanger sequenced low copy number reads for *Tripidium arundinaceum* 2 were kindly donated by BeauSci, Cambridge, UK. For *Tripidium procerum* (ibid) and *Tripidium rufipilum* (ibid), as the data was low

coverage, reads were mapped to the existing *Tripidium ravennae* gene fragments with BWA [59] prior to consensus sequence calling with the Integrative Genomics Viewer (IGV) [60]. The consensus sequences were used as templates for gene assembly in these sequences. Gene loci were assembled as described previously [61] using the closest orthologue for each gene for Mirabaiting and SPAdes for assembly. In *Miscanthus*, *M. sinensis* was used to bait the primary or 'A' genome for each species. All *Miscanthus sinensis* cv *Andante*, *Andropogon virginicus* assemblies along with the *T. rufipilum* and *T. procerum* assemblies and the *Tripidium arundinaceum* 2 sequences have been deposited in EMBL/GenBank under the project identifier PRJEB22229. These and all other gene region assemblies were also deposited in the Dryad digital repository [62].

Assembly finishing

To ensure that the assembly was of high quality, all assembled chloroplasts were finished and polished with a novel pipeline. Raw reads from the SRA pool were mapped back to the assembly with BWA [59], duplicate sequences were tagged with Picard tools [63] prior to optimizing the read alignment with GATK [64] and finally polishing and finishing with Pilon 1.2.0 [65].

Chloroplast annotation

The finished chloroplast alignments were orientated so that they finished with the IR_B region prior to batch upload to the Verdant annotation server [66] for automated annotation. Annotation files were downloaded from Verdant in '.tbl' (Verdant native) format. Genes and pseudogenes identified from our previous analysis [1], and which were not present in the Verdant annotation, were mapped manually with BLAST and were placed in a separate '.tbl' format file. A custom BioPerl script [67] was written that integrated the Verdant output and the additional gene output along with data corresponding to gene annotation and cultivar level taxonomies prior to outputting an EMBL format file and a chromosome file that could be automatically uploaded to the EMBL database. All assembled chloroplasts were submitted to ENA under the project identifier PRJEB20532.

Whole chloroplast alignments

Whole plastid alignments were performed with SATÉ (version 2.2.2) [68], using MAFFT [69] as the aligner, MUSCLE [70] as the sub-alignment joiner and RAxML [71] as the phylogenetic analysis application. Alignment iterations were run for 20 generations past the iterations that yielded the best likelihood score to ensure that the correct global alignment minimum had been reached. This optimal alignment was subsequently edited manually to remove any obvious errors and to trim all gaps of > 20 bp due to only a single sequence to 10 bp (a list of

all chloroplast assemblies with sequence and voucher accessions are given in [Additional file 3]). This alignment was used to generate an initial phylogeny (corresponding to [Additional file 4]), but the raw alignment was used for the alignment finishing steps.

Gene region alignments

Gene regions assembled in this study were merged with the equivalent gene regions from a subset of the data from Welker et al. [16]. All regions were aligned independently using SATÉ, with MAFFT as the aligner and Muscle as the sub-alignment joiner. Alignments were adjusted manually and trimmed. Each individual alignment was input into RAxML and a phylogenetic analysis (100 replicates) was run to identify the most likely tree. If alternate copies of genes were in the same position in the phylogeny these were linked. As we were not analysing close hybrids in this study, only primary copies of each locus were retained. Subsequent to alignment optimization (see below) individual gene/locus alignments were stitched together with a custom Perl script. If the positions of genes were uncertain, all alternate positions were generated. The phylogeny was run with RAxML and the sub-alignment giving the strongest phylogenetic signal was chosen. Using this methodology, we generated the optimal alignment. For the phylogeny generated here, only the alignment yielding the dominant phylogenetic signal was chosen and secondary copies of genes were eliminated. The final, optimized, alignment as well as the phylogram determined from this alignment were deposited in TreeBase (TB2: S23649).

Alignment finishing and optimization

Both the whole chloroplast and individual gene locus alignments were finished and optimized using Prank [72], an indel aware probabilistic multiple alignment program. Terminal taxa representing well-supported groups as defined by SATÉ's RAxML phylogram were constrained using PRANK's 'group' functions. The 'alignment finishing' mode of Prank was initiated with the following command:

```
prank -d=<input_alignment> -t=<input_tree> -o=<output_alignment> -partaligned
```

The output alignment was edited to remove any obviously over-inserted gaps and RAxML (see below) was run for 100 generations to generate a most likely phylogeny. This phylogeny and the manually fixed alignment were used as input for a second round of Prank analysis. After this round of analysis, as the input and output alignments yielded the same tree topology, Prank optimization was deemed complete. Finally, to reduce long-branch issues, all insertions of greater than 20 bp created by just a single sequence were trimmed to 20 bp. This final alignment was used for all subsequent analyses.

Phylogenomic analyses

The whole plastid alignment was divided into LSC, IR_A and SSC partitions. These partitions were further divided into protein-coding gene, RNA-coding gene and non-coding regions. The regions were isolated with the BeforePhylo.pl [73] script and merged into separate partitions. The IR_A region contained only a single tRNA encoding gene, which was added to the SSC RNA-gene partition. This yielded a total of eight partitions. Best-fit evolutionary models for each partition were selected using JModelTest2 [74] and the AICc criterion. The best-fit models were as follows: LSC protein coding: TPM1uf + I + Γ , LSC RNA genes: TVM + Γ ; LSC non-coding: TVM + Γ ; IRA protein coding: TVM + I + Γ ; IRA non-coding: TVM + Γ ; SSC protein coding: TPM1uf + I + Γ ; SSC RNA-gene: TrN + I + Γ and SSC non-coding: TVM + I + Γ . The low copy number gene loci were divided into their five component partitions. Each partition was tested with JModelTest2 and the AICc criterion to determine the best-fit evolutionary models. The optimal models were as follows: apo1: GTR + Γ ; d8: GTR + Γ + I; ep2-exon7: TIM2 + G; ep2-exon8: TrN + Γ + I; rep1: HKY + G. The partitions determined above and their closest model equivalents were used for all subsequent analyses. Neighbor-joining phylogenies were generated with the Ape library in R [75]. Bayesian analyses were run with MrBayes (version 3.1.2) [76], Maximum Likelihood analyses were run with RAxML (version 8.1.17) [72] and SH-aLRT analyses were run with IQ-Tree [77].

For both the chloroplast and low copy number gene datasets Bayesian Markov Chain Monte Carlo (MCMC) analyses were run with MrBayes 3.1.2, using four chains (3 heated and 1 cold) with default priors run for 20,000,000 generations with sampling every 100th tree. Two independent MrBayes analyses, each of two independent runs, were conducted. To avoid any potential over-partitioning of the data, the posterior distributions and associated parameter variables were monitored for each partition using Tracer v 1.6 [78]. High variance and low effective sample sizes were used as signatures of over-sampling. Burn-in was determined by topological convergence and was judged to be sufficient when the average standard deviation of split frequencies was < 0.001 along with the use of the Cumulative and Compare functions of AWTY [79]. The first 50,000 (25%) sampled generations were discarded as burn-in, and the resultant tree samples were mapped onto the reference phylogram (as determined by maximum likelihood analysis) with the SumTrees 4.0.0 script of the Dendropy 4.0.2 package [80].

Maximum Likelihood inference and bootstrapping were performed in RAxML using the same partitioning schemes as detailed above. To obtain the best tree RAxML was run without resampled replicates for 100 generations. The most likely whole plastome tree was obtained in the 70th generation and the most likely

low copy number gene locus tree was obtained on the 53rd generation. These were used as the respective reference tree topologies in all subsequent analyses. To confirm this topology, a second, independent run of RAxML with different seed parameters was also run for both data matrices. In all cases, replicate tree topologies were identical.

To provide support for the Maximum Likelihood phylogeny, a total of 10,000 bootstrap replicates were analysed. Replicate trees were summarized with SumTrees before being mapped onto the best maximally likely tree as determined above.

As an additional measure of branch confidence, SH-aLRT analyses were run for 2500 replicates with IQ-TREE, using the `-bnni` option [77] to reduce the risk of overestimating branch supports due to severe model violations.

Divergence time estimates

Divergence times were estimated using BEAST 2.4.4 [81] optimized for OpenGL graphics running on a MacBook Pro (15-in., 2017 2.9 GHz Intel Core i7) with 16Gib RAM. The concatenated analysis (LSC, IRA and SSC along with their coding gene, RNA gene and non-coding partitions) was run for 20 million generations with sampling every 1000th replication under the BEAST equivalents of the JModelTest2 models (as defined above) with six gamma categories. The tree prior used the Calibrated Yule Model [81] with a relaxed lognormal clock and site models unlinked. Partitions were defined as above. The XML output from BEAUTi was edited to set the starting tree as the most likely tree obtained from RAxML analysis. The site model followed an uncorrelated lognormal relaxed clock [82]. The whole chloroplast analysis was rooted to *Arundinella deppeana*, whilst the low copy number gene locus analysis was rooted to *Arthraxon prionodes* and *Arthraxon lanceolatus*. The age of *Zea mays* divergence was estimated as a normal distribution describing an age of 13.8 ± 2 million years [17] whilst the age of the root was set as 24 ± 4 million years for the chloroplast phylogeny and 19 ± 4 million years for the low copy number gene loci analyses, both based on prior analyses (D Lloyd Evans, personal communication). Convergence statistics were estimated using Tracer v.1.6 [78] after a burn-in of 20,000 sampled generations. Chain convergence was estimated to have been met when the effective sample size was > 200 for all statistics. Tree samples were integrated with SumTrees to generate the maximum clade credibility tree and to determine the 95% highest posterior density (HPD) for each node. The final tree was drawn using FigTree v.1.4.3 [83].

The final alignments and phylogenies are available from TreeBase [84].

Additional files

Additional file 1: List of *Tripidium* chloroplast amplification primers. List of the 13 primers used in amplifying the complete chloroplast sequence of the South African Sugarcane Research Institute *Tripidium* accessions. (PDF 58 kb)

Additional file 2: Gel images of PCR amplicons. Gel images of the 13 PCR amplicons used for *Tripidium* chloroplast isolation and assembly. Example images for the 13 primers are shown, with *Saccharum* hybrid BH10/12 as a positive control. There are images for all six of the *Tripidium* accessions from the South African Sugarcane Research Institute sequenced and assembled in this study. (PDF 311 kb)

Additional file 3: Table of whole chloroplast accessions used for phylogenetics. A table of all the chloroplast sequence accessions (including species, voucher accession and ENA/GenBank accession) that were used for the phylogenetic analyses in this study. Also given are the original references (where applicable) for each sequence. (PDF 150 kb)

Additional file 4: Phylogram with support values for a traditional whole chloroplast analysis. The image depicts the most likely tree topology (with branch support) for an analysis of a whole chloroplast alignment using a standard partition of LSC, IR_A and SSC. Numbers next to nodes give support values (non-parametric bootstrap/Bayesian inference). (PDF 110 kb)

Abbreviations

bp: basepairs; cv.: cultivar; Mya: million years ago; s.s.: *sensu stricto*; sp.: species

Acknowledgments

We thank the South African Sugarcane Research Institute for supporting this work. We also thank Mr. E. Albertse for performing the DNA isolation, PCR amplifications and DNA extractions. We are grateful to the Missouri Botanical Garden (MBG) and the Royal Botanic Gardens, Kew (K) herbaria for access to digital specimens.

Funding

This work was supported by the South African Sugarcane Research Institute.

Availability of data and materials

Sequences assembled and annotated in this study are available from ENA/GenBank under the project accessions PRJEB20532 and PRJEB22229. The sequence datasets and tree topologies generated and/or analysed during the current study are available from the TreeBase repository, <http://purl.org/phylo/treebase/phyloids/study/TB2:523649>. Low copy number gene assemblies, low copy number gene alignments, low copy alignment matrix and maximum likelihood phylogenies along with the chloroplast partition data matrix and chloroplast maximum likelihood phylogeny are available from the Dryad digital repository, <https://doi.org/10.5061/dryad.1k5s048>. All associated computer code has been uploaded to GitHub, <https://github.com/gwydion1/bifo-scripts.git>.

Authors' contributions

SVJ and DLIE conceived the initial experiment, based on plastome sequencing. DLIE conceived the low copy number gene analysis, amended and developed the experimental design, designed the PCR primers, performed the phenotypic analyses, assemblies and phylogenetics, developed the software scripts, analysed and interpreted the data and wrote the paper. SVJ provided the SASRI plant materials and proofread the manuscript. JW provided the sequence data for the USDA accessions. DLIE arranged for the remainder of the plant materials. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

Sampling of plant leaf materials (not destructive to the plant) was conducted according to the South African Sugarcane Research Institute's guidelines.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹South African Sugarcane Research Institute, 170 Flanders Drive, Private Bag X02, Mount Edgecombe, Durban 4300, South Africa. ²School of Life Sciences, College of Agriculture, Engineering and Science, University of Kwa-Zulu Natal, Private Bag X54001, Durban 4000, South Africa. ³BeauSci Ltd., Waterbeach, Cambridge CB25 9TL, UK. ⁴Agronomy Department, University of Florida, Gainesville, FL, USA. ⁵Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, China. ⁶Plant Molecular and Biology Program, Genetics Institute, University of Florida, Gainesville, FL, USA.

Received: 3 December 2017 Accepted: 3 January 2019

Published online: 25 January 2019

References

- Lloyd Evans D, Joshi SV. Complete chloroplast genomes of *Saccharum spontaneum*, *Saccharum officinarum* and *Miscanthus floridulus* (Panicoideae: Andropogoneae) reveal the plastid view on sugarcane origins. *Syst Biodivers*. 2016;14:548–71.
- Reddy BV, Ramesh S, Kumar AA, Wani SS, Ortiz R, Ceballos H, Sreedevi TK. Bio-fuel crops research for energy security and rural development in developing countries. *Bioenergy Res*. 2008;1:48–258.
- Ebrahim MK, Vogg G, Osman MN, Komor E. 1998. Photosynthetic performance and adaptation of sugarcane at suboptimal temperatures. *J Plant Physiol*. 1998;153:587–92.
- Wu J, Huang Y, Lin Y, Fu C, Liu S, Deng Z, et al. Unexpected inheritance pattern of *Erianthus arundinaceus* chromosomes in the intergeneric progeny between *Saccharum* spp. and *Erianthus arundinaceus*. *PLoS One*. 2014;9:e110390.
- Mukherjee SK. Origin and distribution of *Saccharum*. *Bot Gaz*. 1957;119:55–61.
- Clayton B, Renvoize SA. Genera graminum. *Grasses of the World*. 1986;13.
- Kew Gardens GrassBase Entry for *Saccharum*. <http://www.kew.org/data/grasses-db/spindex.htm#S>. Accessed 10 Oct 2017.
- Hodkinson TR, Chase MW, Takahashi C, Leitch IJ, Bennett MD, Renvoize SA. The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *Am J Bot*. 2002;89:279–86.
- Grassl CO. Taxonomy of *Saccharum* relatives: *Sclerostachya*, *Narenga*, and *Erianthus*. *Proceedings of the 14th Congress of the International Society of Sugar Cane Technologists*. 1971. p. 240–248.
- von Trinius CB. *Fundamenta agrostographiae, sive Theoria constructionis Floris graminei; adjecta synopsis generum graminum hucusque cognitorum*. Vienna: JG Heubner; 1820. p. 169.
- Bernhardi JJ. 1801. *J Bot (Schradet)*. 1801;1800:127.
- Valdés B, Scholz H. The euro+ med treatment of Gramineae—a generic synopsis and some new names. *Willdenowia*. 2006;36:657–69.
- Berding JJ, Roach BT. Germplasm collection, maintenance, and use. In: Heinz DJ, editor. *Sugarcane improvement through breeding*, pages: Amsterdam: Elsevier; 1987. p. 143–210.
- Michaux FA. *Flora borealis Americana*. Paris: Caroli Crapelet, Paris; 1803.
- NCBI Taxonomy entries for genus *Tripidium*. <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1612336>. Accessed 12 Oct 2017.
- Welker CA, Souza-Chies TT, Longhi-Wagner HM, Peichoto MT, McKain MR, Kellogg EA. Phylogenetic analysis of *Saccharum* *sl* (Poaceae; Andropogoneae), with emphasis on the circumscription of the south American species. *Am J Bot*. 2015;102:248–63.
- Estep MC, McKain MR, Diaz DV, Zhong J, Hodge JG, Hodkinson TR, et al. Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc Natl Acad Sci U S A*. 2014;111:15149–54.
- Soreng R-J, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Morrone O. A worldwide phylogenetic classification of the Poaceae (Gramineae). *J Syst Evol*. 2015;53:117–37.
- Folk RA, Mandel JR, Freudenstein JV. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst Biol*. 2017;66:320–37.
- Folk RA, Soltis PS, Soltis DE, Guralnick R. New prospects in the detection and comparative analysis of hybridization in the tree of life. *Am J Bot*. 2018 in press. <https://doi.org/10.1002/ajb.21018>.
- Mallet J. Hybrid speciation. *Nature*. 2007;446:279.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, et al. Polyploidy and angiosperm diversification. *Am J Bot*. 2009;2009(96):336–48.
- Pirie MD, Humphreys AM, Barker NP, Linder HP. Reticulation, data combination, and inferring evolutionary history: an example from *Danthonioideae* (Poaceae). *Syst Biol*. 2009;58:612–28.
- Guo X, Thomas DC, Saunders RM. Gene tree discordance and coalescent methods support ancient intergeneric hybridisation between *Dasymschalon* and *Friesodielsia* (Annonaceae). *Molecular Phylogenetics and Evolution*. 2018; 127:14–29. DOI: 2018. <https://doi.org/10.1016/j.ympev.2018.04.009>.
- Záveská E, Fér T, Šída O, Marhold K, Leong-Škomičková J. Hybridization among distantly related species: examples from the polyploid genus *Curcuma* (Zingiberaceae). *Mol Phylogenet Evol*. 2016;100:303–21.
- Whitfield JB, Lockhart PJ. Deciphering ancient rapid radiations. *Trends Ecol Evol*. 2007;22:258–65.
- Hinsinger DD, Gaudeul M, Couloux A, Bousquet J, Frascaria-Lacoste N. 2014. The phylogeography of Eurasian *Fraxinus* species reveals ancient transcontinental reticulation. *Mol Phylogenet Evol*. 2014;77:223–37.
- García N, Folk RA, Meerow AW, Chamala S, Gitzendanner MA, de Oliveira RS, et al. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (*Amaryllidaceae* tribe Hippeastreae). *Mol Phylogenet Evol*. 2017;111:231–47.
- Kellogg EA, Appels R, Mason-Gamer RJ. When genes tell different stories: the diploid genera of Triticaceae (Gramineae). *Syst Bot*. 1996;21:321–47.
- Connor HE. 2004. *Flora of New Zealand — Gramineae supplement I: Danthonioideae*. *N Z J Bot*. 2004;42:771–95.
- NCBI Taxonomy entries for genus *Saccharum*. <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=4546>. Accessed 12 Oct 2017.
- Whalen MD. Taxonomy of *saccharum* (Poaceae). *Baileya*. 1991;23:109–25.
- Daniels J, Smith P, Paton N, Williams CA. The origin of the genus *Saccharum*. *Sugarcane Breed Newsl*. 1975;36:24–39.
- Grassl CO. Problems relating to the origin and evolution of wild and cultivated *Saccharum*. *Indian J Sugarcane Res Dev*. 1964;8:106–16.
- D'Hont A, Rao PS, Feldmann P, Grivet L, Islam-Faridi N, Taylor P, Glaszmann JC. Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* × *Erianthus arundinaceus*, with molecular markers and DNA in situ hybridisation. *Theor Appl Genet*. 1995;91:320–6.
- Alix K, Baurens FC, Paulet F, Glaszmann JC, D'Hont A. Isolation and characterization of a satellite DNA family in the *Saccharum* complex. *Genome*. 1998;41:854–64.
- Alix K, Glaszmann JC, D'Hont A. Inter-Alu-like species-specific sequences in the *Saccharum* complex. *Theor Appl Genet*. 1999;99:962–8.
- Piperidis N, Chen JW, Deng HH, Wang LP, Jackson P, Piperidis G. GISH characterization of *Erianthus arundinaceus* chromosomes in three generations of sugarcane intergeneric hybrids. *Genome*. 2010;53:331–6.
- Huang Y, Wu P, Lin Y, Fu C, Deng Z, Wang Q, et al. Characterization of Chromosome Inheritance of the Intergeneric BC 2 and BC 3 Progeny between *Saccharum* spp. and *Erianthus arundinaceus*. *PLoS One*. 2015;10:e0133722.
- Riera-Lizarazu O, Rines HW, Phillips RL. Cytological and molecular characterization of oat × maize partial hybrids. *Theor Appl Genet*. 1996;93:123–35.
- Burke SV, Wysocki WP, Zuloaga FO, Craine JM, Pires J-C, Edger PP, Mayfield-Jones D, et al. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biol*. 2016;16: 140. <https://doi.org/10.1186/s12870-016-0823-3>.
- Song J, Yang X, Resende MF Jr, Neves LG, Todd J, Zhang J, et al. Natural allelic variations in highly polyploidy *Saccharum* complex. *Frontiers Plant Sci*. 2016;7:804–32.
- NCBI primer design tool. <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>. Accessed 23 Sept 2017.
- Engels B. Amplify4 In Silico PCR. <https://engels.genetics.wisc.edu/amplify/>. Accessed 7 Oct 2017.
- Clayton B, Vorontsova MS, Harman KT, Williamson H. GrassBase — The Online World Grass Flora. <http://www.kew.org/data/grasses-db.html>. 2006 onwards. Accessed 6 Oct 2017.
- Vicentini A, Barber JC, Alicioni SS, Giussani LM, Kellogg EA. The age of the grasses and clusters of origins of C4 photosynthesis. *Glob Chang Biol*. 2008;14:2963–77.
- Tropicos entry for genus *Tripidium*. <http://tropicos.info/NamePage.aspx?nameid=50314867&tab=subordinatetaxa&projectid=48>. Accessed 12 Oct 2017.
- Rieseberg LH, Soltis DE. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*. 1991;5:65–84.

49. Hawkins JS, Ramachandran D, Henderson A, Freeman J, Carlise M, Harris A, Willison-Headley Z. Phylogenetic reconstruction using four low-copy nuclear loci strongly supports a polyphyletic origin of the genus *Sorghum*. *Ann Bot*. 2015;116:291–9.
50. Zachos J, Pagani S, Sloan L, Thomas E, Billups K. Trends, rhythms, and aberrations in global climate 65 ma to present. *Science*. 2001;292:686–93. <https://doi.org/10.1126/science.1059412>.
51. Spriggs EL, Christin PA, Edwards EJ. 2014. C4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS One*. 2014;9:e97722. <https://doi.org/10.1371/journal.pone.0097722>.
52. Snyman SJ, Komape DM, Khanyi H, van den Berg J, Cilliers D, Lloyd Evans D, Barnard S, Siebert SJ. Assessing the likelihood of gene flow from sugarcane (*Saccharum* hybrids) to wild relatives in South Africa. *Frontiers in Bioengineering*. 2018;6:72.
53. Tsuruta SI, Ebina M, Kobayashi M, Takahashi W. Complete Chloroplast Genomes of *Erianthus arundinaceus* and *Miscanthus sinensis*: Comparative Genomics and Evolution of the Saccharum Complex. *PLoS One*. 2017;12:e0169992.
54. Kellogg EA. Phylogenetic relationships of Saccharinae and Sorghinae. In: Paterson AH, editor. *Genomics of the Saccharinae*. New York: Springer; 2013. p. 3–21.
55. Wang J, Roe B, Macmill S, Yu Q, Murray JE, Tang H, et al. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics*. 2010;11:261.
56. Chevreux C, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. *German Conference on Bioinformatics*. 1999;99:45–56.
57. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
58. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*. 2014;btu170.
59. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*. 2009;25:1754–60.
60. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–92.
61. Lloyd Evans D, Joshi SV. Herbicide targets and detoxification proteins in sugarcane: from gene assembly to structure modelling. *Genome*. 2017;60: 601–17.
62. Lloyd Evans D, Joshi SV and Wang J (2017) Data from: whole chloroplast and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoidae: Andropogoneae) to sugarcane. *Dryad*. <https://doi.org/10.5061/dryad.1k5s048>. Last accessed 10 Dec 2018
63. Picard tools. Available from: <http://broadinstitute.github.io/picard/>. Last accessed 10 Nov 2017.
64. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
65. Walker B-J, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9:e112963.
66. McKain MR, Hartsock RH, Wohl MM, Kellogg EA. Verdant: automated annotation, alignment and phylogenetic analysis of whole chloroplast genomes. *Bioinformatics*. 2016;btw583.
67. Lloyd Evans D. Sequence annotation code. Available from: <https://github.com/gwydion1/bifo-scripts.git>. Accessed 10 Nov 2017.
68. Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science*. 2009;324:1561–4.
69. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
70. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
71. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
72. Löytynoja A, Vilella AJ, Goldman N. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics*. 2012;28:1684–91.
73. Zhu, Q. 2014. BeforePhylo.PI version 0.9.0 available from: <https://github.com/qiyunzhu/BeforePhylo>. Accessed 12 Nov 2017.
74. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9:772.
75. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–90.
76. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
77. Nguyen L-T, Schmidt HA, von Haeseler A, Minh QA; IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol and Evol*, 2015; 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
78. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer, Version 1.5. <http://tree.bio.ed.ac.uk/software/tracer/>. Accessed 7 Oct 2017.
79. Nylander JA, Wilgenbusch JC, Warren DL. Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 2008; 24:581–583.
80. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 2010;26:1569–71.
81. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29:1969–73.
82. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4:e88.
83. FigTree, available from <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 10 Nov 2017.
84. Lloyd Evans D, Joshi SV and Wang J (2017) Data from: whole chloroplast and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (Panicoidae: Andropogoneae) to sugarcane. *TreeBase*. <http://purl.org/phylo/treebase/phylogs/study/TB2:S23649>.
85. Filgueiras TS, Peterson PM, Soreng RJ, Judziewicz EJ. Editors. *Catalogue of New World grasses (Poaceae): III. Subfamilies Panicoidae, Aristidoideae, Arundinoideae, and Danthonioideae*, *Contr US Natl herb*. Washington DC: Smithsonian Institution; 2003.
86. Flora of China Editorial Committee. *Flora of China (Poaceae)*. In Wu CY, Raven PH, Hong DY, editors, *Flora of China*. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press. 2006. p. 1–733.
87. Cabi E, Doğan M. *Poaceae*. In: Güner A, Aslan S, Ekim T, Vural M, Babaç MC editors. *Türkiye Bitkileri Listesi*. İstanbul: Nezahat Gökyiğit Botanik Bahçesi ve Flora Araştırmaları Derneği Yayını. 2012. p. 690–756.
88. Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 2008;24:861–2.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

