Chromosome-level genome sequence of the C4 grass *Themeda triandra* reveals karyotype orthology with sorghum and genetic variation in accessions adapted to diverse environments

Jakob B. Butler[1], Jazmine L. Humphreys[1], Theodore Allnutt[2], Vinod K. Jacob[3,4], Lily Chen[4], Alejandro Correa-Lozano[1], Javier López-Jurado[1], Eloise Foo[1], Ian J. Wright[3,4,5], Steven M. Smith[1], Brian J. Atwell[5*]

[1]ARC Centre of Excellence for Plant Success in Nature and Agriculture, School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia

[2]Royal Botanical Gardens, Melbourne, Victoria, Australia

[3]ARC Centre of Excellence for Plant Success in Nature and Agriculture, Western Sydney University, Sydney, New South Wales, Australia

[4]Hawkesbury Institute for the Environment, Western Sydney University, Sydney, New South Wales, Australia

[5]School of Natural Sciences, Macquarie University, Sydney, New South Wales, Australia

*Corresponding author

ABSTRACT

Themeda triandra (family Andropogoneae) is a pan-continental grass occurring throughout Australia. Recognising its remarkable adaptive capacity, we assembled the nuclear genome sequence of a diploid accession from eastern Australia, finding all ten chromosomes of the 755 Mb assembly highly syntenic with those of *Sorghum bicolor*. Genotyping and cytometry of range-wide accessions suggested several historical radiation events, with polyploids dominating in arid regions, apart from a notable diploid lineage from the arid north-west of Australia (PAN). A detailed comparison of PAN with two diploids from temperate south-eastern Australia through whole-genome resequencing revealed extensive copy number variation and polymorphism, with changes in genes for heat-shock proteins and flowering regulation reflecting their environmental origin. Exploration of the rich genetic diversity in *T. triandra* with respect to environmental adaptation is expected to benefit grassland management programs and enable introgression of novel genes into Andropogoneae crops for climate resilience.

INTRODUCTION

The tribe Andropogoneae is a pan-tropical complex of 90 genera and 1,200 species of C4 grasses, many of which are key species in savanna ecosystems [1]. It also includes major crops such as maize, sorghum and sugarcane. Within the subtribe Anthistiriinae, *Themeda triandra* exemplifies the remarkable geographic breadth of the tribe, occurring naturally across south and south-east Asia, the Middle East, eastern and southern Africa, and Australasia [2-4]. Within Australasia, *T. triandra* (formerly *T. australis*) occurs naturally across every climate zone, extending from the northern tropics to the southern cool-temperate and maritime regions to latitudes of more than 40°S [5, 6].

T. triandra has a generally perennial habit in the field, with vegetative growth in spring culminating in sexual and aposporous flowering during warmer seasons if soil water is available [7]. However, populations in different climate zones exhibit extensive phenotypic variation; for instance, accessions collected from New Guinea to Tasmania have markedly different photoperiod responses [7]. Even within accessions from SE Australia, morphological traits such as tiller number and flowering time vary markedly [6]. Variation in awn length occurs according to climate regime, with cool southern Australian accessions differing from those inland, which are in turn distinct from accessions from the arid north-west [8].

Common garden experiments have shown that certain characters are determined by the site-of-origin of different accessions, indicating that different populations are genetically distinct. For example, the timing of reproductive development among accessions, as estimated by the days from germination to emergence of inflorescences, was found to be related to the climate-of-origin [7]. More recently we have found that in two temperature-controlled treatments, morphological and physiological traits during vegetative development correlated weakly with the temperature and precipitation regimes at the site-of-origin of accessions [9]. When maxima were constrained to 30 °C, accessions originating in the wet sub-tropics took 1 - 2 weeks longer to flower under long days in comparison to temperate accessions. However, at 20 °C (thought to be sub-optimal for C4 grasses), this difference was exacerbated, with the sub-tropical accessions delayed by an average of 5 weeks [9]. Thus, cool growing conditions unmasked a strong 'time to flowering' phenotype among different accessions.

The evolutionary origins of *T. triandra* and specifically, its spread to Australia in the past 1 - 2 million years [3], are critical considerations for understanding divergence within the species. Tight phylogenetic clustering of the Australian accessions relative to others from the Indian Ocean rim has been revealed by plastid DNA sequencing [3, 10]. Nevertheless, genetic fragmentation has been documented across SE Australia, where climate regimes and geographic separation have driven diversity in the ploidy and genotypes of 52 accessions [11]. Diversity between populations of *T. triandra* across SE Australia has been shaped by the strong selective forces of temperature and available water [3], together accounting for a quarter of genetic divergence in these populations [11].

Polyploidy is common in wild plant populations, often thought to influence rates of speciation and adaptation to stress [12]. Polyploids ranging up to octoploids in local populations of *T. triandra* throughout Australia have been reported [5, 11, 13]. This widespread polyploidy is consistent with the species' ecological success across a vast range of climates and soil types [14]. Polyploid *T. triandra* populations are generally better adapted to stressful environments, exhibiting faster vegetative growth rates [15]. Furthermore, tetraploid populations from coastal New South Wales have reproductive advantages compared to diploids [14]. The co-existence of

T. triandra plants with different ploidy levels raises more nuanced questions of adaptation through complex ploidy admixtures within populations [10].

Other mechanisms through which *T. triandra* has adapted to such diverse environments include genome organisation, gene copy number and mutations affecting gene expression and function. Such genome-level phenomena are key targets for the transfer of traits from wild crop relatives to molecular breeding of superior crops [16] but first require extensive genetic resources to be generated. Genome assemblies to the contig level have recently been reported for *T. triandra* accessions from The Philippines and Australia [10, 17]. However, a complete reference genome sequence has been lacking, limiting the further investigations of *T. triandra* evolution.

The aims of this study were to generate a chromosome-scale reference genome for a diploid *T. triandra* accession from eastern Australia, and to use this resource alongside whole-genome resequencing of accessions from contrasting climates to characterise copy number variation and sequence polymorphisms that distinguish divergent lineages. We further aimed to combine SNP genotyping and ploidy assessment across a broader sample set to reconstruct phylogenetic relationships among disparate Australian populations and examine patterns of radiation and genomic differentiation associated with climate.

RESULTS

1. Sequencing and assembly of a reference genome

We assembled a collection of *T. triandra* accessions from established seedbanks, herbarium collections and field sampling, spanning a wide range of Australian climatic regions (Supplementary Table 1, Figure 1). From this collection, a diploid accession from Brisbane Water National Park in Wondabyne (WBW) on the warm-temperate south-eastern coast was selected for reference genome assembly. We generated a total of 96.7 Gbp (114X) of PacBio HiFi reads and 154.1 Gbp of Hi-C sequencing for the assembly of the *T. triandra* WBW genome. The assembled genome was resolved into 10 pseudomolecules representing 10 chromosomes and 12 unassigned contigs, with a total size of 755 Mbp and a contig N50 of 76.5 Mbp (Figure 2, Supplementary Table 2). This assembly contains 75% of the total expected telomeres, suggesting that a near-T2T assembly had been achieved, supported by a BUSCO completeness score of 99.07% (Supplementary Table 2). Chromosomes were numbered and oriented to match *Sorghum bicolor*, reflecting the strong chromosomal synteny between the two genomes described below. A total of 165 Mbp of repetitive sequence motifs covering 22% of the genome were annotated, consisting overwhelmingly of LTR-retrotransposons (Supplementary Table 3). Genome annotation based on *S. bicolor* protein hints and 94 Gbp of Iso-seq data predicted 44,328 protein coding genes, along with 625 ribosomal RNA genes (5S = 163; 5.8S = 204; 18S = 222; 28S = 236) which were localised to two main clusters (chromosomes 3 and 9).

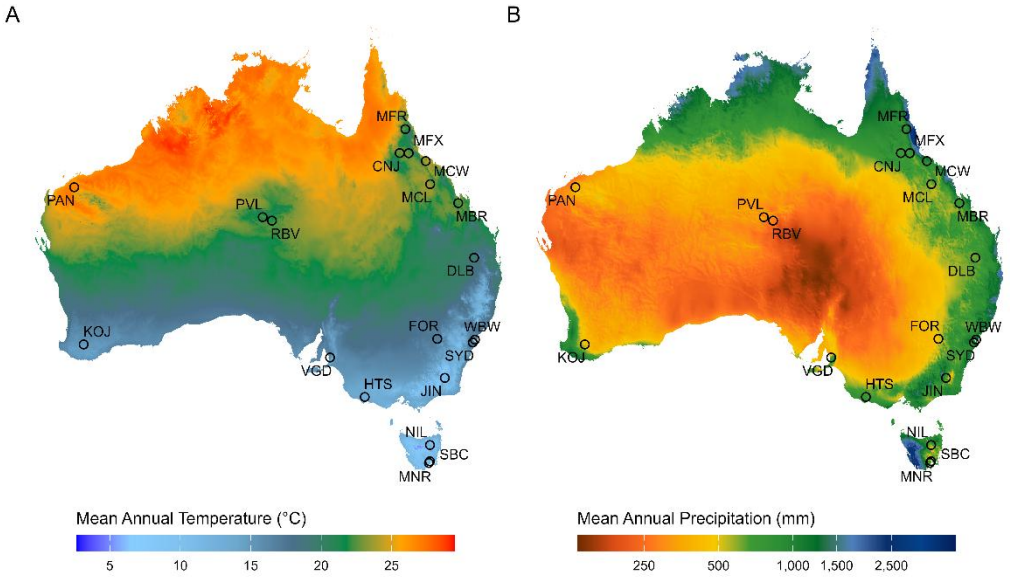

Fig. 1 Geographical locations of *Themeda triandra* accessions used in this study. The maps of Australia are colour-coded according to A) mean annual temperature and B) mean annual precipitation, with variables obtained from gridded datasets from CHELSA [18] at a resolution of ~1 km at the equator and representing the long-term interpolated average of locations between 1981 – 2010. The three letter notations are expanded upon in full in Supplementary Table 1.

To complement the reference genome, whole-genome resequencing was undertaken for six accessions from different climatic regions around Australia. These included the hot arid north-west (PAN), the Mediterranean south-west (KOJ), the cool-temperate south-east (SBC and NIL), the warm-temperate south-east (SYD) and an alpine accession (JIN) (Figure 1). Resequencing enabled high-resolution SNP discovery and the identification of genomic regions with notable patterns of variation (Figure 2). Sliding-window analyses revealed multiple intervals exhibiting elevated or reduced SNP density, consistent with regions of suppressed recombination or localised haplotype divergence among lineages. When combined with both %GC content and gene density, other notable patterns were also distinguishable. The most obvious was the position of centromeric repeat regions, characterized by increased repetitive content and a reduction in %GC, gene density and SNP density. Likewise, the major cluster of rRNA genes on chromosome 3 was large enough to have a visible effect on genome composition (low gene, SNP and repeat density along with increased %GC, Figure 2). Detection of structural variants (SVs) comprising insertions, deletions, and rearrangements further supported this heterogeneity, with clusters of putative SVs concentrated in specific genomic regions (Supplementary Figure 1).

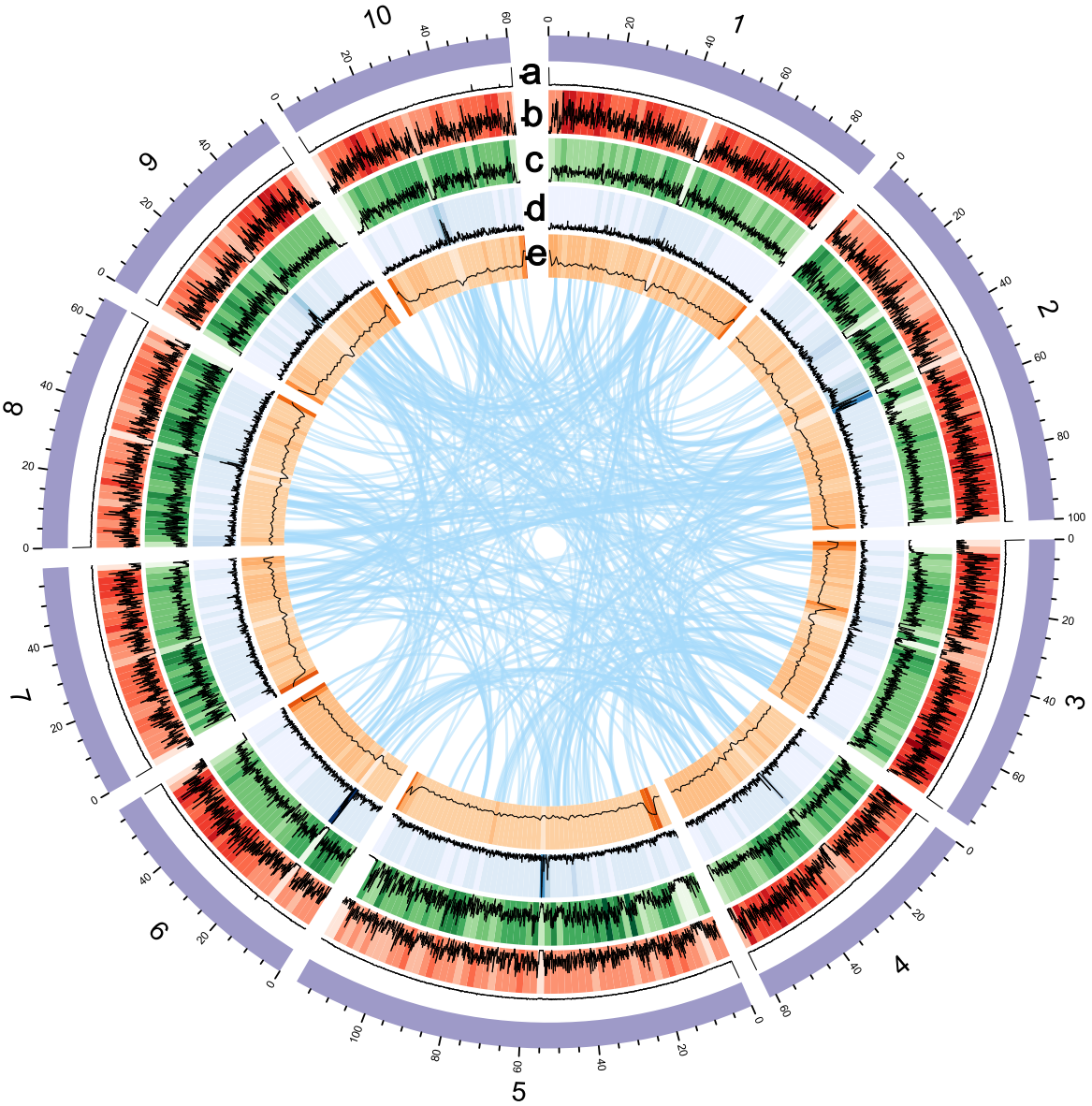

Fig. 2 The genome of *Themeda triandra*. Outer ring of Circos plot shows the length of the 10 pseudomolecules (Mbp). Inner data tracks are quantified across 100kb (line) and 2MB (heatmap) windows: a) Number of telomeric repeats (0 – 500+); b) Gene density (0 - 19 per 100kb, 1 - 207 per 2MB); c) SNP density (0 – 12,309 per 100kb, 239 – 107,360 per 2MB); d) Repeat element density (0 - 187 per 100kb, 17 – 1,437 per 2MB); e) %GC content (0.42 – 0.54% per 100kb, 0.42 – 0.53% per 2MB). Link tracks indicate regions of strong gene synteny between chromosomes (that span ≥ 15 genes and ≥10 gene/Mbp).

2. *Comparisons of* T. triandra, Sorghum bicolor *and* Zea mays

To place the *T. triandra* genome in the context of the Andropogoneae, we compared it to the well-resourced crop genomes of *Sorghum bicolor* and *Zea mays,* with the last common ancestor between these lineages ~15 Mya (and subsequent divergence of *T. triandra* and *Sorghum* of ~12 Mya) [3]. Whole-genome alignment revealed strong chromosomal synteny

between *T. triandra* and *S. bicolor*, with each of the ten *T. triandra* chromosomes aligned predominantly to a corresponding *S. bicolor* chromosome, confirming a largely conserved macrostructure and implying orthologous karyotypes (Supplementary Figure 2). Across the genome, 6,725 high-confidence alignment blocks (>1 kb, ≥90% identity) were detected, with most alignments showing unique 1:1 correspondence, and a minority aligning to multiple locations consistent with segmental duplications. Despite this extensive synteny, many small intrachromosomal inversions were detectable, constituting approximately 94 Mbp (12%) of the genome. These inversions were evident in alignment dotplots as clear diagonal breaks and reversals in local synteny (Supplementary Figure 2). The lack of inter-chromosomal translocations was notable, suggesting that genome evolution has proceeded *via* local structural changes rather than substantial chromosome rearrangement. Orthologous gene synteny also showed strong collinearity between *T. triandra* and *S. bicolor*, although approximately one-third of all syntenic gene blocks were inverted, as would be expected based on the established genome synteny (Figure 3a). Several *T. triandra* regions also show homology to more than one chromosome, particularly in *S. bicolor*, suggesting historical inter-chromosomal duplication of genic regions that have been maintained in both genera (Figure 3a). Some duplicated blocks were also detectable between *T. triandra* and *Z. mays,* although overall synteny was lower, as expected given its more ancient divergence. These patterns of synteny reveal substantial genome reorganisation after the divergence of *Zea* from the *Sorghum*/*Themeda* precursor, and that the broad reorganized state was subsequently maintained between these two genera.

Gene family analysis revealed 14,272 orthologous families (i.e. 49%) conserved across all three lineages. An additional 2,047 families (7%) were unique to *T. triandra* (Figure 3b). These *T. triandra*-specific families were enriched for regulatory and signalling functions, particularly stress- and defence-associated kinase pathways, post-translational modification and transcriptional regulation (Supplementary Table 4a). The number of shared gene families varied between species pairs; while *T. triandra* shared 2,628 families exclusively with *S. bicolor*, it only shared 390 gene families exclusively with *Z. mays*, consistent with its more recent divergence time and greater synteny with the former. Overall, *S. bicolor* resembled maize more closely in terms of gene content (19.5k shared families) than *T. triandra* (14.5k shared families) despite the above shared genome synteny, which suggests stronger or more distinct selective pressures have acted on *Themeda*. A total of 5,282 (18%) gene families were absent in *T. triandra* compared with the other two grasses, but these were not enriched for any particular function. In addition to the presence/absence, variation in the abundance of specific gene-families were also evident in *T. triandra* relative to the other taxa, with expansion in families enriched for enzymes involved in redox chemistry, secondary metabolism, and cell wall modification, and contraction in families dominated by core regulatory, signalling, and protein turnover functions (Supplementary Table 4b,c).

Pairwise synonymous substitution rates (Ks) calculated from 6,345 single-copy orthologues further supported these relationships. The Ks distributions showed peaks at 0.109 for *S. bicolor – T. triandra*, 0.134 for *S. bicolor – Z. mays* and 0.154 for *T. triandra – Z. mays* (Figure 3c). Assuming a *T. triandra – S. bicolor* divergence of ~12 Mya and their split from *Zea* at ~15 Mya [3] this corresponds to an estimated neutral mutation rate of 4.39 - 5.04 × 10^{-9} substitutions per site per year. Notably, all three comparisons also showed a minor shared peak at Ks ≈ 0.33, reflecting a set of genes in the three species that have accumulated more neutral mutations than expected. Using the average mutation rate (4.66 x 10^{-9}), this peak dates to ~35.4 Mya. One explanation for this pattern could be ancient duplication followed by differential gene loss,

consistent with the pattern of historic inter-chromosomal duplication. Significantly, a mere 2% of these genes (11 of 556) were located on different chromosomes in *T. triandra* and *S. bicolor*, instead suggesting that the 35.4 Mya peak reflects genes with deeper ancestral history rather than remnants of genome duplication.

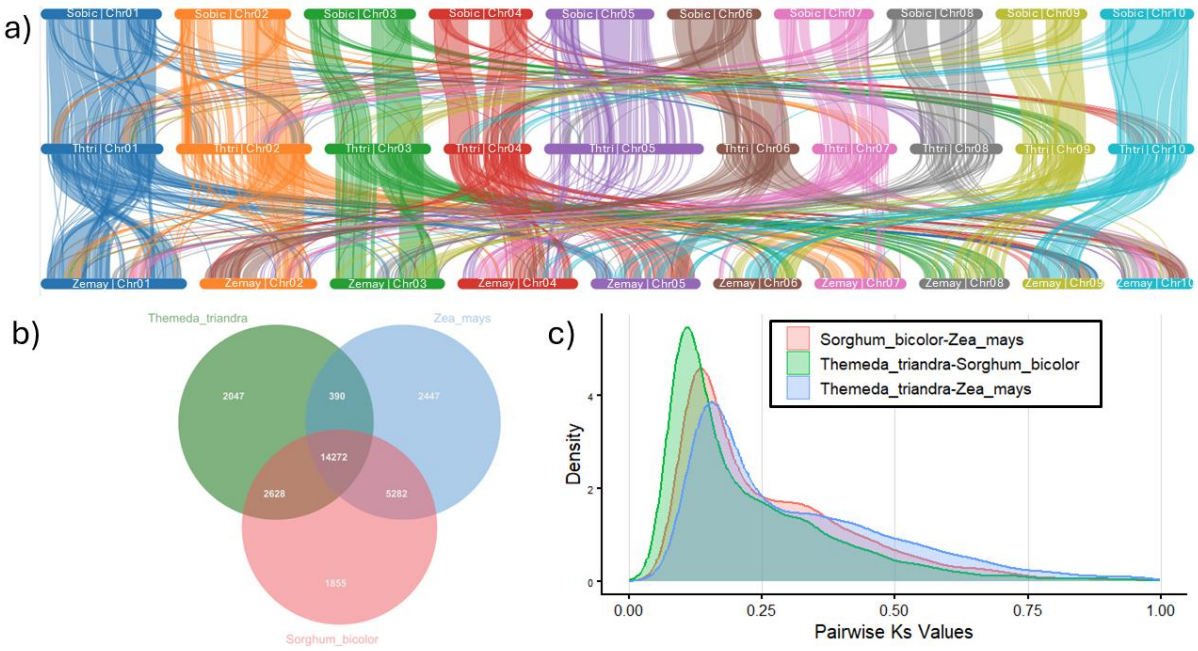

Fig. 3 Comparison of orthologous genes between *T. triandra*, *S. bicolor* and *Z. mays*. A) Synteny of orthologous genes from *T. triandra* (centre) to *S. bicolor* (above) and *Z. mays* (below). Each link represents at least five genes in a collinear arrangement between the genomes compared. The central parts of some *S. bicolor* chromosomes have very few annotated genes, resulting in no visible links. B) Number of unique and orthologous gene families in *T. triandra*, *S. bicolor* and *Z. mays*. C) Distribution of synonymous mutation rate (Ks) values of 6,345 single-copy orthologues when compared pairwise between *T. triandra*, *S. bicolor* and *Z. mays*. A peak in Ks values indicates the most common number of neutral mutations found in these genes between the compared taxa, which can be linked to estimated divergence time and mutation rate.

3. Biogeography, ploidy and phylogeny of Australian T. triandra

Despite their genomic similarity, *S. bicolor* is largely adapted to semi-arid tropical and subtropical regions under irrigation, while *T. triandra* inhabits climates ranging from the hot, arid north-west of Australia to the cool-temperate southern regions. To understand the genetic contribution to this climatic versatility and the relationships between these distinct populations, we sampled disparate *T. triandra* accessions from across the species' Australian range. Specifically, this panel of contrasting accessions originated from cool to hot regimes (mean annual temperatures 10.9 – 26.9 °C) and dry to wet regimes (annual precipitation 281 – 1198 mm) (Figure 1, Supplementary Table 1). Genotyping by ddRAD sequencing generated 50,395 filtered SNPs across the *T. triandra* genome.

Ploidy levels, determined by heterozygous allele ratios and validated with flow cytometry (Supplementary Figure 3; Supplementary Figure 4), ranged from diploid to hexaploid across 19 sampled populations. Phylogenetic reconstruction showed that genetic relationships were

broadly consistent with geographic distance (Figure 4a,b). While most inland arid environments were occupied by polyploids, diploid populations were not restricted to benign environments (SBC, NIL, MNR), also occurring in the hottest and most arid north-west (PAN) of the continent (Figure 4c). These diploid accessions were highly divergent from those sampled from the south-eastern region of the continent (Figure 4a), indicating substantial genetic differentiation independent of ploidy.

Based on the positioning of the *T. triandra* outgroups from China and the Philippines, the phylogeny (Figure 4) supports a diploid origin for the species in Australia, with polyploidy arising as a derived condition. Evidence of mixed-ploidy populations was observed at several sites: for example, genotyping revealed a triploid individual at Pannawonica (Figure 4a), likely arising from hybridisation between local diploids and neighbouring tetraploids (or meiosis error), and cytometric assessment at MFX indicated a mixture of tetraploid and hexaploid plants (Supplementary Figure 4). Some lineages appear to decrease in ploidy relative to the closest branch point, likely reflecting single plant samples from mixed ploidy populations (see Methods) or additional colonisation events [3].

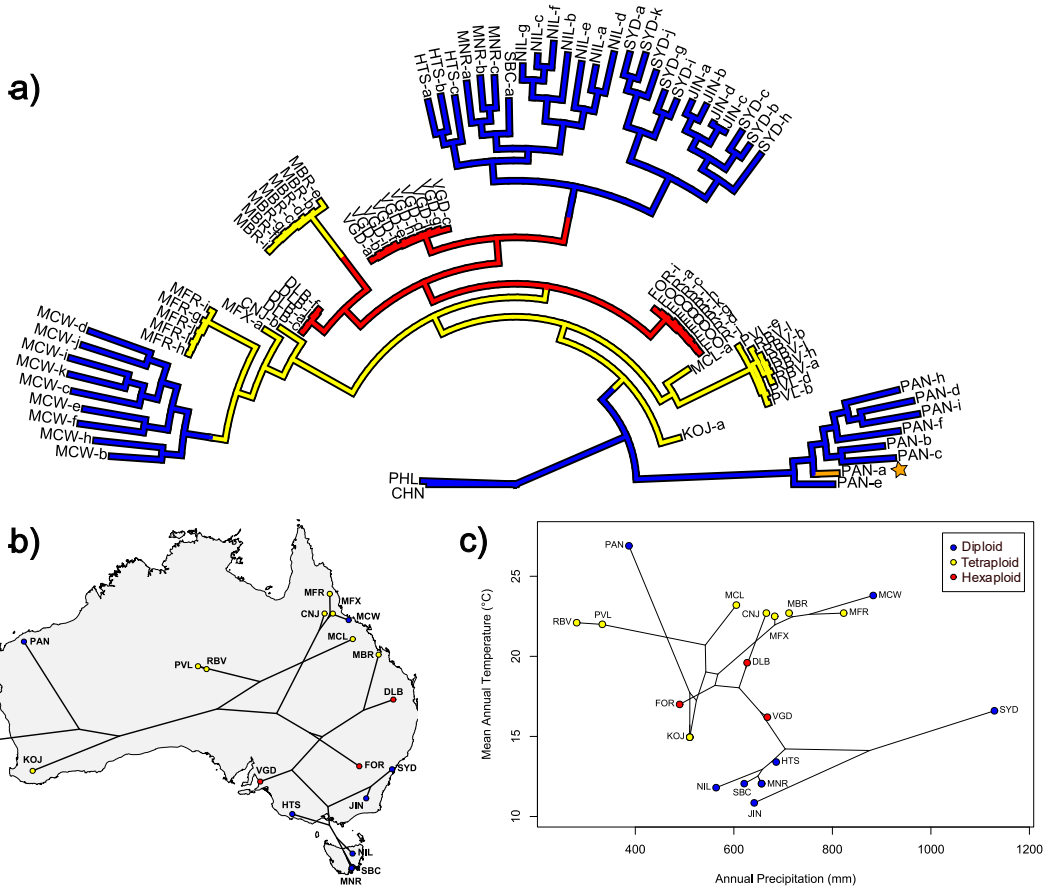

Fig. 4 Nuclear phylogeny of *Themeda triandra* in Australia. A) Maximum likelihood radial phylogram of *T. triandra* populations across its Australian distribution. Ploidy of specific clades is indicated by colour, with 2n=blue, 3n=orange (highlighted with star icon), 4n=yellow, 6n=red, with ancestral ploidy states inferred using stochastic character mapping in *phytools*. B) Maximum likelihood phylogram (with population clades collapsed) projected onto Australia based on coordinates of collection. Accessions PHI (Philippines) and CHN (China) as outgroups are placed outside of Australia. C) Phylomorphospace projection of phylogeny based on mean annual temperature and annual rainfall (mm).

4. Genetic variation in diploids from climate extremes

To characterise the genetic divergence between diploid accessions from contrasting climates, SNPs from the resequencing of three diploid accessions (PAN, SYD, SBC) were annotated for their functional impact using SnpEff [19]. Across three pairwise comparisons, 11,159 unique genes carried high-impact variants predicted to strongly affect protein function, including modification of start/stop codons, premature truncation and frameshifts. When contrasting the PAN accession against both SYD and SBC, genes carrying these mutations were enriched for GO terms related to heat-shock responses (Supplementary Figure 5, Supplementary Table 5). This genome-level response is consistent with the stark contrast in selection pressures on abiotic stress tolerance genes when comparing populations adapted to arid-hot and cool temperate regimes. Gene copy number variation between these accessions was also assessed using sequencing depth profiles with CNVkit [20]. Genes with putative copy number reduction in the PAN accession were strongly enriched for translation-related functions, including translational repression, regulation, initiation, and RNA binding activity, even after false discovery rate correction (Figure 5). Additional functional differences were identified across all three pairwise comparisons, with distinct sets of metabolic and transport-related functions enriched in all three accessions (Supplementary Figure 5, Supplementary Table 5).

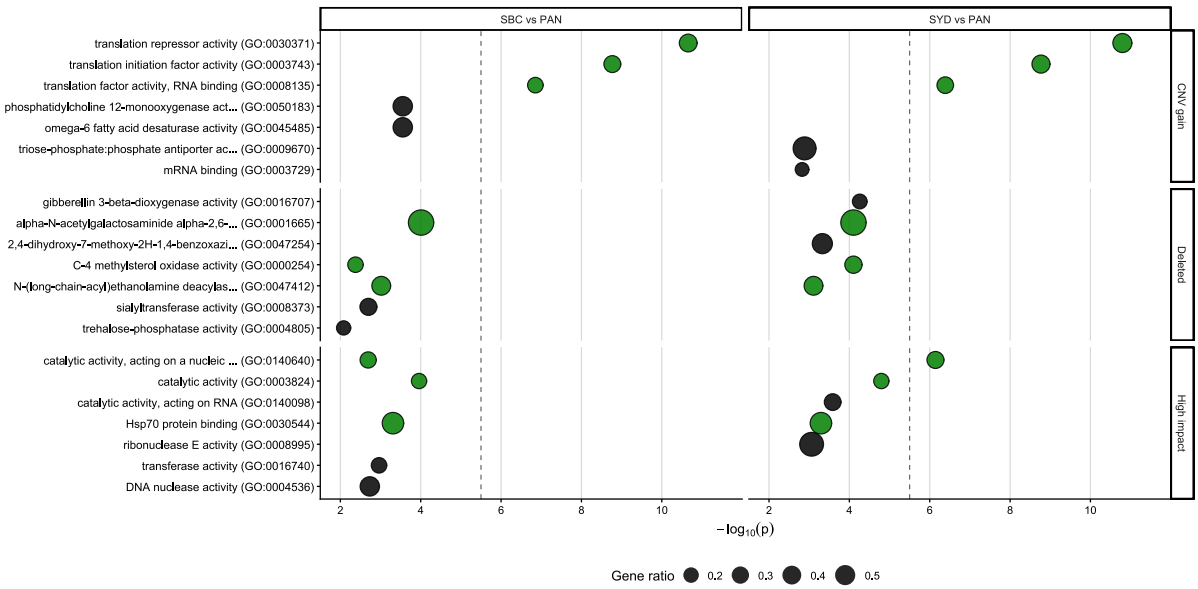

Fig 5. Top five enriched Gene Ontology (GO) terms of genes involved in putative copy number gain, deletions and housing high impact mutations when comparing three diploid accessions of *T. triandra* from different environmental extremes. Accession names are described in Supplementary Table 1. The size of the bubble indicates what proportion of the entire set of genes annotated with that GO term are present in the set associated with putative copy number variation (CNV). Bubbles coloured green indicate GO terms significant in both comparisons. The vertical dotted line indicates the threshold after which P-values remain significant after correction for multiple testing using false discovery rate.

Previous work has shown that *T. triandra* accessions originating from hot environments exhibit substantially delayed flowering at low growth temperatures compared with accessions from colder environments [9]. To investigate the genetic basis of this difference, variation in genes associated with flowering in these three diploid accessions representing environmental extremes was examined. Candidate genes were identified based on homology with known *S. bicolor* flowering genes [21]. Distinct patterns of functional and copy number variation in flowering-related genes were observed among all three accessions (Table 1). In the SBC accession from cool-temperate Tasmania, changes relative to the other accessions were concentrated around putative regulators of *FLOWERING LOCUS C* (*FLC*), a MADS-box transcription factor genes that acts as a potent repressor of flowering, specifically acting on *FLOWERING LOCUS T (FT)* floral promoter genes [22, 23]. These changes included copy number reduction in both a positive and negative regulator of *FLC*, as well as a unique mutation in the vernalisation-responsive gene *SbVRN1*, which also represses *FLC* activity. This pattern is consistent with modulation of flowering in response to cooler environments. The SYD accession showed a partially overlapping but distinct set of changes relative to PAN, also centred on *FLC* regulation, with copy number increases in a positive regulator of *FLC* alongside the florigen *FT* gene and a negative regulator of *FLC*. In contrast, the PAN accession harboured high-impact mutations across a much broader set of flowering genes, spanning signalling, photoperiod response, and floral transition pathways, together with copy number reductions in several genes that promote flowering (Table 1). Collectively, these patterns suggest that populations from the hottest climate regime have undergone more extensive diversification of different genetic pathways interacting with flowering, whereas the colder eastern and south-eastern populations have evolved more targeted genetic changes focused on *FLC*-mediated repression of flowering.

Table 1 Copy number variation and high-impact mutations in sorghum flowering orthologs across three diploid accessions of *T. triandra*. Increase: genes with putative copy number gain ($\log_2$fold change in sequencing depth > 1); Decrease: genes with putative copy number loss ($\log_2$ fold change in sequencing depth < -1); Deleted: genes with < 5 reads coverage; Mutation: genes with SNPs predicted to cause high impact changes to protein structure. Copy number variation is based on pairwise comparison between accessions. Blank cells indicate no difference meeting the above criteria between that accession and the others. Groups of classifications indicated by * highlight relevant accession patterns discussed in the main text. Accession names are detailed in Supplementary Table 1.

T. triandra gene	*S. bicolor* ortholog	*S. bicolor* gene name	SBC	PAN	SYD
ThtriWBW.08G259500	Sobic.008G149100	SbVRN1	Mutation*		
ThtriWBW.08G031500	Sobic.008G007000	SbSEF1	Decrease*	Mutation, Increase	Increase
ThtriWBW.03G493100	Sobic.003G427100	SbFY	Decrease*	Increase	
ThtriWBW.05G107200	Sobic.008G007000	SbSEF1	Decrease*	Increase	Increase
ThtriWBW.05G032600	Sobic.005G038400	SbFT	Deleted*	Increase	Decrease
ThtriWBW.02G459900	Sobic.002G308700	FT (Hd3)	Deleted	*	Deleted
ThtriWBW.03G056800	Sobic.003G031800	Indeterminate growth1		Decrease*	Increase
ThtriWBW.10G178800	Sobic.010G138000	SbCRY2	Increase	Decrease*	Increase
ThtriWBW.11G028900	Sobic.005G038400	SbFT	Increase	Decrease*	Increase
ThtriWBW.02G548500	Sobic.002G388800	P450	Increase	Decrease*	Increase
ThtriWBW.08G158400	Sobic.005G161200	SbGA1		Mutation*	
ThtriWBW.01G088900	Sobic.001G080700	SbHd6		Mutation*	
ThtriWBW.03G479700	Sobic.003G416700	SbLD		Mutation*	
ThtriWBW.03G136500	Sobic.003G101000	SHB1		Mutation*	
ThtriWBW.04G046700	Sobic.002G164200	SbPFT1		Mutation*	
ThtriWBW.06G031300	Sobic.006G015500	SbEMF2		Mutation*	
ThtriWBW.02G019100	Sobic.010G030200	Hox2	Decrease	Mutation, Increase*	Decrease
ThtriWBW.02G011600	Sobic.002G009800	SbVIP4	Decrease	Mutation, Increase*	Decrease
ThtriWBW.03G050900	Sobic.003G029100	SbFRI			Mutation*
ThtriWBW.03G328100	Sobic.003G274400	SbFVE	Decrease	Decrease	Increase*
ThtriWBW.03G134400	Sobic.003G098800	SbFT	Decrease	Decrease	Increase*
ThtriWBW.08G101100	Sobic.008G037300	SbFT	Decrease	Decrease	Increase*
ThtriWBW.08G031400	Sobic.008G007000	SbSEF1	Decrease	Decrease	Increase*
ThtriWBW.08G029600	Sobic.008G007000	SbSEF1	Decrease	Decrease	Increase*
ThtriWBW.08G301100	Sobic.001G248600	SbGA1	Decrease	Decrease	Mutation, Increase*
ThtriWBW.05G107900	Sobic.008G007000	SbSEF1	Mutation	Mutation	Mutation
ThtriWBW.04G207200	Sobic.004G176000	SbEFS	Mutation	Mutation	Mutation
ThtriWBW.02G529900	Sobic.002G367600	P450	Mutation	Mutation	Mutation

DISCUSSION

The assembly of a chromosome-scale reference genome for *T. triandra,* combined with resequencing of accessions from contrasting environments, provides the first comprehensive genomic framework for investigating the success of this species across its remarkably diverse Australian range and beyond to Asia and Africa. The chromosomal synteny with *Sorghum bicolor* opens the potential for findings in *T. triandra* to be translated to major crop genomes. In addition, the genetic variation identified between accessions of *T. triandra* from contrasting climate zones makes it possible to identify genes and pathways underlying climate adaptation. Together, these make *T. triandra* a powerful comparative system for both understanding natural adaptation and informing crop resilience in the Andropogoneae.

The karyotype orthology between *T. triandra* and *S. bicolor,* with largely conserved gene order across all ten chromosomes, means that genes and regulatory regions in *T. triandra* could be mapped directly to their sorghum orthologues. Despite approximately 12 million years of divergence, genome evolution between these lineages has proceeded primarily through

intrachromosomal inversions rather than large-scale chromosome rearrangements, preserving this orthologous relationship. This is particularly relevant because cultivated sorghum has lost an estimated 14 - 35% of its wild progenitor's genetic diversity through domestication [24-26], with modern breeding further narrowing the gene pool [27]. As an undomesticated species spanning diverse environments, *T. triandra* retains adapted variation that has likely been lost in cultivated sorghum. While wild sorghum is clearly the most immediate source of new variation for sorghum breeding, the geographic and climatic distribution of wild *S. bicolor* is narrow compared with *T. triandra* [28], as are the distributions of the 17 native Australian *Sorghum* species, being largely restricted to northern Australia [16]. The chromosomal synteny between these two sentinel *Sorghum* and *Themeda* species provide a route to identify adaptive variation in *T. triandra* and link it to functionally characterised loci in *S. bicolor* (and other well characterised genera), effectively using *T. triandra* as a discovery platform for climate adaptation genes within the Andropogoneae.

Flowering time provides a clear demonstration of this approach. Common-environment (glasshouse) experiments have shown that accessions from warmer environments take significantly longer to flower, with this effect most pronounced under cool growing conditions, while accessions from drier environments flower earlier, likely as a drought escape strategy [9]. The genomic basis of this variation in flowering time can be interpreted through homology with well-characterised *S. bicolor* flowering genes [21]. In the southernmost (Tasmanian) accession of *T. triandra*, genetic differences were concentrated around regulators of the floral repressor *FLC*, including a mutation in the vernalization-responsive gene *SbVRN1*, consistent with adaptation to early flowering in a colder, shorter growing season. The PAN accession, collected from an environment where aridity imposes pressure to reproduce opportunistically within transient, favourable conditions, carried high-impact mutations across a much broader set of flowering genes spanning photoperiod response, floral signalling and transition pathways. These contrasting genetic adaptations in the flowering pathway between southern and northern *T. triandra* accessions suggest distinct evolutionary strategies for adapting flowering time to local environment and climate. Critically, these candidate genes were identified through direct orthology with *Sorghum*, illustrating how the syntenic relationship between these genomes can translate ecological variation in *T. triandra* into candidate genes for crop improvement.

Flowering time represents just one of many trait systems that can now be explored using this resource. The gene family analyses revealed that *T. triandra*-specific and expanded gene families are enriched for stress and defence signalling, redox chemistry, secondary metabolism and cell wall modification. These functions could potentially underpin this species' success across its diverse climate range. Many of these same pathways are targets for crop improvement in sorghum, where narrowed genetic diversity in stress-responsive gene families has been observed in domesticated material [27, 29]. Of particular interest are traits that may have been selected against during domestication due to trade-offs with yield, but which remain intact in wild species. For example, in soybean disease resistance genes have been progressively lost during domestication in multiple crops, with each additional nucleotide-binding leucine-rich repeat (NLR) receptor gene estimated to impose an average 1.6% yield penalty [30, 31]. Similarly, cuticular wax composition affects both drought tolerance and pathogen defence in many plants and has diverged substantially, even within wild species, over short evolutionary timescales [32, 33] and this trait has been further modified during crop improvement [34, 35]. The presence of over 2,000 *T. triandra*-specific gene families enriched for stress and defence signalling, combined with the syntenic framework described here, makes it

possible to identify candidate genes for climate resilience that may have been lost or diminished in the cultivated sorghum gene pool.

This information also has direct applications for conservation and management of *T. triandra* itself. Temperate grasslands dominated by *T. triandra* are among Australia's most endangered ecosystems [36], and restoration efforts are complicated by the species' low seed viability and limited understanding of local adaptation [37]. The phylogeographic structure described here, combined with the ploidy variation across populations, has practical implications for sourcing seed for restoration. Our results show that diploid populations from cool, wet environments diverge strongly at the genome level from those in the hot arid north-west; matching provenance to restoration site is likely to matter for establishment success [38]. Ploidy variation adds a further layer of complexity, with polyploidy substantially reducing predicted genomic vulnerability to future climates [11]. Although our sampling captures only a small subset of each population, polyploids largely dominated the warmer inland environments, yet *T. triandra* plants from the most arid site were diploid, challenging the expectation that polyploidy confers greater tolerance to environmental stress [39].

The reproduction of *T. triandra* combines wind-pollinated outcrossing with self-pollination and apomixis [40]. This has important implications both for natural adaptation and restoration applications. The reproductive flexibility of *T. triandra* enables it to be used for genetic analysis through controlled crosses and segregation studies, making it possible to link adaptive traits to specific genotypes. The reference genome and SNP framework presented here now make it possible to move beyond neutral markers toward identifying the specific genes and variants associated with climate adaptation in different *T. triandra* populations, enabling more informed decisions about seed sourcing, assisted migration, and genetic rescue as conditions shift under climate change.

The chromosome-scale reference genome presented here, together with resequencing from climatic extremes and broad phylogeographic sampling, establishes *T. triandra* as a genomic resource for investigating climate adaptation in a wild Andropogoneae grass. This work opens several directions for further investigation. Resequencing additional accessions, particularly polyploid populations and those from under-sampled regions, would test whether the patterns of adaptive variation hold across the species broader range. A pan-genome incorporating multiple *Themeda* accessions (and species) would capture presence-absence variation not accessible from a single reference, and combining this with existing *Sorghum* and *Zea* pangenomes could establish a tribe-level comparative resource for the Andropogoneae [41]. The tribe Andropogoneae offers vast insights into adaptive diversity, with over 1200 species spanning ~20 million years of divergence [42]. The resources presented here provide both a foundation for conservation genomics and a rich source of candidate genes shaped by millions of years of selection across some of the most contrasting environments occupied by any single grass species.

METHODS

Themeda accession sampling and cultivation

Samples of *T. triandra* were selected from 20 different localities representing diverse climatic origins across the Australian continent, spanning mean annual temps from 10.9 - 26.9 °C and annual precipitation from 281 - 1198 mm (Supplementary Table 1, Figure 1). Accessions were

sourced from suppliers including the Australian Pastures Genebank, Native Seeds Pty Ltd and Nindethana Seed Service Pty Ltd as well as personal collections by BJA and ACL. As personal collections were opportunistic, some populations were represented by multiple seed from a single plant (Supplementary Table 1).

Seeds were germinated in a controlled growth chamber (at 35/20 °C in a 12 h day/night regime, with a peak irradiance of 1000 µmol photons m^{-2} s^{-1} midway through the photoperiod) and transplanted to 4.45-L pots containing a 9:1 mixture of commercial potting mix (<30% fine sand, >70% pine bark; Australian Growing Solutions) and field soil (40:30:30 sandy clay loam). Plants were maintained in climate-controlled glasshouses under a 30/26 °C day/night temperature regime with a photoperiod of 14 h (natural light was supplemented with blue/red LED lighting, adding 600 µmol photons m^{-2} s^{-1}).

For genomic analyses, leaf tissue was sampled from 3-month-old plants and immediately frozen in liquid nitrogen and stored at -80 °C until DNA/RNA extraction.

Library construction and sequencing of reference individual

Leaf tissue samples of the "WBW" accession were sent to Biomolecular Resource Facility (Australian National University) for high molecular weight genomic DNA extraction and long-read library construction using a SMRTbell® prep kit 3.0, with BluePippin size selection used during library preparation. PacBIO-HiFi sequencing was undertaken using the Revio SMRT Cells platform, generating 96.7 Gbp of HiFi long reads with an expected coverage of the *Themeda triandra* genome of 114X. Further leaf tissue samples were also used to prepare a Hi-C library from cross-linked chromatin using an Arima Hi-C v2.0 protocol. This library was sequenced using Illumina NovaSeq 6000, producing 154.1 Gbp of Hi-C reads. RNA was extracted from five different tissues (leaf meristem, mature lamina, root tips, mature root and embryo) of the "WBW" accession using Kinnex full-length RNA kits, preserved in liquid nitrogen and sent to AGRF Australia (Melbourne) for long-read sequencing using PacBio ISOseqX, with 94.0 Gbp of transcript sequence generated.

Genome assembly and annotation

PacBio HiFi reads were assembled using hifiasm v0.19.18 [43]. Contigs belonging to mitochondria and chloroplasts were identified by BLAST search of known sequences, and those contigs were extracted. Haplotigs were removed using purge_haplotigs v1.1.2 [44], and the remaining contigs were then scaffolded using deduplicated Hi-C reads and YaHS v1.2a.1 [45]. The Hi-C contact map was visualized using Juicer v1.6 (https://github.com/aidenlab/juicer). Repetitive elements were discovered using the ENSEMBL repeat annotator [46] and masked in the final assembly, while 74 scaffolds containing >95% repetitive DNA were discarded entirely, along with 4 scaffolds which contained contaminants as determined by NCBI Foreign Contamination Screen [47]. Genome completeness was assessed using BUSCO (poales_odb10 library) [48] *via* compleasm v0.2.5 [49]. The occurrence of telomeric repeats were detected and annotated using tidk v0.2.41 [50].

Structural gene annotation was performed using BRAKER3 v3.0.8 [51], using protein sequences from *Sorghum bicolor* obtained from NCBI (Taxonomy ID 4558) as external hints to guide gene model prediction and the aligned ISOseq transcript data for evidence of model support. Gene models with less than four residues in the translated product were discarded. Predicted protein-coding genes were functionally annotated using InterProScan v5.66-98.0 [52] and eggNOG-

mapper [53], with KEGG orthology determined using BlastKOALA [54]. Annotation of rRNAs was performed using barrnap (https://github.com/tseemann/barrnap).

Genome synteny with other Andropogoneae

Synteny with the *Sorghum bicolor* reference genome (BTx623 v3, [55]) was assessed using *nucmer* [56] (--mumreference -l 40 -b 500 -c 200 -g 1000) with structural variation quantified using *syri* [57]. Given the high level of colinearity detected between these genomes, *T. triandra* chromosomes were named and oriented to maximise synteny with the *S. bicolor* genome.

To characterise patterns of gene family evolution, synteny, and sequence divergence across the Andropogoneae, multiple comparative analyses were performed using algorithms implemented in OrthoVenn3 [58], comparing the gene content of *T. triandra* to that reported in the reference genomes of *S. bicolor* (BTx623 v3.0) and *Zea mays* (B73 v5.0, [59]). Orthologous gene clusters were identified using OrthoMCL [60], with an e-value threshold of 1×10^{-2} applied to define cluster membership. Gene family expansion and contraction along each lineage since divergence were inferred using CAFE5 [61]. Pairwise collinearity of genomic regions containing at least five orthologous genes was assessed among the three species using MCScanX [62]. In addition, single-copy orthologous genes shared among species were identified, and pairwise rates of synonymous substitution (Ks) were calculated for each gene to assess relative rates of sequence mutation and divergence.

Range-wide genotyping using ddRAD sequencing

DNA was extracted from 103 accessions (with 40 samples duplicated) from 18 populations (Supplementary Table 1) using the CTAB method [63] with additional RNAse treatment. For genotyping *via* ddRAD sequencing, DNA was shipped to AGRF Australia for restriction digest, library preparation and sequencing using NovaSeq X Plus (Illumina Inc., San Diego, CA, USA), with 633.5 GB of paired-end sequence generated and an average of 4.4 GB per sample. After quality assessment with fastp [64], reads were demultiplexed using the program stacks [65], specifying the options for cutsite and barcode rescue, removing reads with missing data and low quality, and ignoring the absence of the cut site on reverse reads due to a diversity issue on reverse reads based on the specific library construction [66]. Given this issue the first three bases of the reverse read were also trimmed using seqtk (https://github.com/lh3/seqtk). Adapters were then removed using trimmomatic [67], and all reads (both paired and orphaned) were processed and aligned to the *T. triandra* genome using bwa mem [68] and samtools [69]. As some accessions were duplicated, the samples with greater amounts of sequencing were retained. SNP calling was then performed using bcftools mpileup [69] (-A -d 1000 -q 20 -a AD,DP,SP), and bcftools call. The resulting SNPs were then filtered iteratively; individuals with over 90% missing genotypes were removed, followed by removing SNPs that did not meet the following criteria: maximum missing proportion of 20%, a minimum depth of 5, and a minimum minor allele frequency of 0.05. Subsequently, a SNP pruning step was undertaken to randomly remove a SNP from pairs of SNPs in strong linkage disequilibrium with each other ($r^2 > 0.6$) using bcftools prune.

Whole genome resequencing of six accessions

Full genome resequencing was performed on six selected accessions (Supplementary Table 1). DNA was extracted from leaf samples of these accessions using the CTAB method [63] with RNase treatment and delivered to AGRF Australia for library preparation and sequencing, with 176.6 GB of paired end sequence generated. Reads were quality controlled using the program

fastp [64] (--trim_poly_g –depect_adapter_for_pe --cut right --cut_right_window_size 4 --cut_right_window_quality 20 --length_required 75) and aligned to the *T. triandra* reference genome using bwa mem. Automatic detection of potential DNA structural variation in these samples *via* disjunct alignment of reads was performed using *manta* [70]. SNP calling was then performed against the *T. triandra* reference genome as with the ddRAD data. For subsequent analysis, SNPs called from the resequencing of one accession (KOJ-a) and two outgroup accessions from the Philippines (BioSample SAMN30450876) and China (BioSample SAMN08770639) were intersected with the filtered ddRAD SNP set, with a subsequent missing data filtering of 90% to reduce artefacts arising from combining sequencing technologies [71].

Determining accession ploidy

Ploidy measurements were initially based on flow cytometry and confirmed by analysis of the ddRAD data (below). Approximately 0.5 cm^2 (or a ~2 cm segment) of young leaf tissue was rapidly chopped for 60 s to fineness with a razor blade (replaced after 10 samples) in a petri dish containing 250 µl of Nuclei Extraction buffer supplied in the CyStain UV Precise P kit (Sysmex Partec; Order No. 05-5002). A further 250 µl of extraction buffer was added to yield ~500 µl of suspension (plus debris). The suspension was pipetted through a 30 µm CellTrics® filter (Order No. 04-0042-2317) into a transparent polypropylene 3.5-ml Sysmex sample tube (Order No. 04-2000). Then, 2 ml of CyStain UV Precise P Staining buffer was added and samples were incubated at room temperature for 60 s. Flow measurements were performed by inserting the 3.5-ml tube into a CyFlow® Ploidy Analyser (Sysmex Partec) equipped with a UV LED suitable for DAPI excitation. The instrument was operated with 1× PBS as sheath fluid, and data were acquired using CyView 2.4 software. A sheath fluid flush was performed between each sample.

A DNA control standard supplied by the manufacturer (Sysmex) was run at the beginning of each session as a quality control check. Reference *T. triandra* accessions used to assign ploidy were: 1) a diploid accession from Falmouth, Tasmania (APG 85441) and a 2) a tetraploid accession from Winderlup, near Mylor, South Australia (35.038 S 138.772 E) provided by Marne Dunin (pers. comm.). Reference extracts were used to align G_0/G_1 peaks. Instrument settings, including gain and flow rate, were optimised using the diploid standard to position its G_0/G_1 peak within the linear detection range; these settings were then maintained for all subsequent samples within that run, allowing ploidy levels (diploid, tetraploid or hexaploid) to be assigned based on the relative position of G_0/G_1 peaks compared to the diploid standard. Histograms with a coefficient of variation (CV) of the primary G_0/G_1 peak greater than 6% were excluded, and the corresponding samples were re-analysed.

Ploidy of the samples in the ddRAD data were also assessed by the proportion of heterozygous reads, following the methods of Ahrens et al. [11], with samples shared between methods used for validation. The read depth of each allele was extracted for all SNPs within all samples from the variant call format (VCF) file, and the relative frequency of each allele was calculated. A histogram was made for each sample, showing allele frequencies of all non-homozygous SNPs, and sample ploidy was assigned by visual analysis of the distribution. Assuming consistent read coverage, each heterozygous locus in a diploid individual should have a read-depth ratio of 1:1 for the reference and alternate allele, resulting in a normal distribution of allele frequencies with a single peak at a reference allele proportion of 0.5. Higher ploidy would have different distributions of allele frequencies at heterozygous loci, resulting in one additional peak for each step increase in ploidy (e.g. heterozygous loci in triploid accessions would have expected reference allele read-depth proportions of 0.33 and 0.66, tetraploid accessions 0.25, 0.5 and 0.75, etc). After assessing SNPs of reads depths 5 - 400, only SNPs with a read depth > 200 were

used, as this gave the greatest peak resolution in this data. Determine the ploidy of the resequenced accessions intersected with this data required reducing the depth threshold to 50 to achieve similar resolution.

Phylogenetic analysis

To investigate genetic relationships and diversity among accessions, a maximum likelihood phylogeny was created using a workflow adapted from Severn-Ellis, Scheben [72]. An additional filter to remove SNPs with more than 90% heterozygous calls across all individuals was employed to eliminate sites uninformative to phylogenetic analysis. The filtered VCF file was then converted to PHYLIP format using a custom python script [72]. Maximum likelihood phylogenetic inference was performed using RAxML-NG [73], using the ASC_GTRCAT model and the Lewis ascertainment bias correction, with 100 bootstrap replicates performed to assess node support. Ancestral ploidy states were inferred with stochastic character mapping in phytools [74] (make.simmap, equal rates model, 100 simulations). phytools was also used to project the phylogeny into geographic and morphological space.

Several samples were removed from analysis due to mislabelling of accessions identifiable from the genotyping data/phylogeny. One sample was removed due to unexpected phylogenetic placement combined with unexpected ploidy compared to the rest of the population (Sydney-i). 13 samples were also identified as mislabelled and removed due to unexpected population grouping within single family clades. Specifically, in clades with very short branch lengths (indicating samples of seed from a single plant), mixed populations were identified as problematic, with samples removed based on geographic disagreement with the population in the nearest clade.

Comparison of CNV and SNP impacts in resequenced accessions

To detect putative copy number variations (CNV) among contrasted diploid accessions, changes in resequencing depth were examined using CNVkit [20]. Normalised read depth profiles for accessions from Tasmania (SBC), Sydney, (SYD) and Pannawonica (PAN) were generated relative to a flat reference, providing a common baseline for estimating copy number across the genome. Pairwise comparisons between accessions (PAN–SBC, PAN–SYD and SBC–SYD) were then performed to identify differences in gene-level sequencing depth. Each 5 kbp segment of the chromosomes that intersected with a gene was collated (discarding those genes that spanned two segments with opposite signs in fold change). Log2 fold changes in depth were then calculated for each gene, with values > 1 conservatively classified as copy number gains, values < -1 classified as copy number losses, and genes with fewer than five mapped reads interpreted as putative deletions. To assess for potential functional differences between these accessions, the impact of high-quality SNPs (depth > 5, quality score > 20) on annotated genes was determined using SNPEff [19]. Genes containing SNPs with putative high impact on protein function (modification of start/stop codons, premature truncation and frameshifts) that differed between the three diploid contrasts were identified. All gene lists were tested for enrichment of GO terms using topGO [75] with Fishers exact test.

DATA AVAILABILITY

Scripts used to assemble the *Themeda* genome are available at https://github.com/Royal-Botanic-Gardens-Victoria/Themeda. The genome assembly and sequencing data are available at the NCBI Sequence Read Archive under BioProject accession PRJNA1433027.

ACKNOWLEDGEMENTS

This was a joint project of the Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture (CE200100015) and Bioplatforms Australia. We acknowledge the contribution of the Australian Grasslands Initiative Consortium in the generation of data used in this publication. The Initiative is supported by funding from Bioplatforms Australia, enabled by the Commonwealth Government National Collaborative Research Infrastructure Strategy (NCRIS). We also acknowledge the provision of computing and data resources provided by the Australian BioCommons Leadership Share (ABLeS) program, which is co-funded by Bioplatforms Australia (enabled by NCRIS), the National Computational Infrastructure and Pawsey Supercomputing Research Centre. We also acknowledge the use of the high-performance computing facilities provided by Digital Research Services, IT Services at the University of Tasmania. We thank Juanita Christine Lauer, Marne Marie Dunin and Rachel Burton (University of Adelaide), Caroline Cristofolini (Sydney Botanic Gardens) and Daniel Pascoe (Macquarie University) for assistance with cell cytometry measurements. The reference plant (WBW) was collected by Rachael Gallagher (Western Sydney University). Useful comments on ploidy data by Richard Frankham (Macquarie University) are appreciated.

AUTHOR CONTRIBUTIONS

BJA coordinated funding resources for sequencing, conceptualized the project and assembled the germplasm collection with VKJ, JLH, ACL and EF. VKJ, JLH and BJA collected experimental material for ACL and LC to complete extractions and sample preparation. VKJ, JLH and BJA managed sample dispatch to analytical service laboratories via Bioplatforms facilities. TRA and JBB performed the bioinformatics analysis on all sequencing data and BJA provided the cytometry data in discussion with EF, ACL and JLJ. The manuscript was written by JLH, SMS, JBB and BJA with editing by IJW, EF and VKJ and commentary on a late draft by JLJ and LC.

REFERENCES

1. Arthan, W., et al. Phylogenomics of Andropogoneae (Panicoideae: Poaceae) of Mainland Southeast Asia. *Syst. Bot.* **42**, 418-431 (2017)
2. Holder, F., *A phylogenetic study of the South African representatives of the tribe Andropogoneae (Poaceae)*. 2003, University of the Free State.
3. Dunning, L.T., et al. The recent and rapid spread of Themeda triandra. *Botany Letters* **164**, 327-337 (2017)
4. Arthan, W., et al. Complex evolutionary history of two ecologically significant grass genera, Themeda and Heteropogon (Poaceae: Panicoideae: Andropogoneae). *Bot. J. Linn. Soc.* **196**, 437-455 (2021)
5. Hayman, D. The distribution and cytology of the chromosome races of *Themeda australis* in southern Australia. *Aust. J. Bot.* **8**, 58-68 (1960)
6. Groves, R. Growth and development of five populations of *Themeda australis* in response to temperature. *Aust. J. Bot.* **23**, 951-963 (1975)
7. Evans, L. and R. Knox. Environmental control of reproduction in *Themeda australis*. *Aust. J. Bot.* **17**, 375-389 (1969)
8. Cavanagh, A.M., R.C. Godfree, and J.W. Morgan. Awn length variation in Australia's most widespread grass, *Themeda triandra*, across its distribution. *Aust. J. Bot.* **72** (2024)
9. Jacob, V., et al. Trait–climate relations in *Themeda triandra*: a widely distributed C4 grass and crop wild relative. *bioRxiv* (2026)
10. Dunning, L.T., et al. Hybridisation and chloroplast capture between distinct Themeda triandra lineages in Australia. *Mol. Ecol.* **31**, 5846-5860 (2022)
11. Ahrens, C.W., et al. Spatial, climate and ploidy factors drive genomic diversity and resilience in the widespread grass *Themeda triandra*. *Mol. Ecol.* **29**, 3872-3888 (2020)
12. Estep, M.C., et al. Allopolyploidy, diversification, and the Miocene grassland expansion. *PNAS* **111**, 15149-15154 (2014)
13. Birari, S.P. Polyploidy in species of *Themeda* Forsk. *Caryologia* **34**, 301-310 (1981)
14. Godfree, R.C., et al. Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. *Royal Society Open Science* **4** (2017)
15. Stevens, A.V., et al. Polyploidy affects the seed, dormancy and seedling characteristics of a perennial grass, conferring an advantage in stressful climates. *Plant Biol.* **22**, 500-513 (2020)
16. Ananda, G.K.S., et al. Wild sorghum as a promising resource for crop improvement. *Front Plant Sci.* **Volume 11 - 2020** (2020)
17. Zhu, Y., et al. HEMU: An integrated comparative genomics database and analysis platform for Andropogoneae grasses. *Plant Commun* **5**, 100786 (2024)
18. Karger, D.N., et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017)
19. Cingolani, P., et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80-92 (2012)
20. Talevich, E., et al. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comp. Biol.* **12**, e1004873 (2016)
21. Mace, E.S., C.H. Hunt, and D.R. Jordan. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* **126**, 1377-1395 (2013)
22. Michaels, S.D. and R.M. Amasino. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949-956 (1999)

23. Sheldon, C.C., et al. The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). *PNAS* **97**, 3753-3758 (2000)
24. Burgarella, C., et al. The road to sorghum domestication: Evidence from nucleotide diversity and gene expression patterns. *Front Plant Sci.* **Volume 12 - 2021** (2021)
25. Casa, A.M., et al. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor. Appl. Genet.* **111**, 23-30 (2005)
26. Mace, E.S., et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**, 2320 (2013)
27. Tack, J., J. Lingenfelser, and S.V.K. Jagadish. Disaggregating sorghum yield reductions under warming scenarios exposes narrow genetic diversity in US breeding programs. *PNAS* **114**, 9296-9301 (2017)
28. Mace, E.S., et al. A global resource for exploring and exploiting genetic variation in sorghum crop wild relatives. *Crop Sci.* **61**, 150-162 (2021)
29. Massel, K., et al. Whole genome sequencing reveals potential new targets for improving nitrogen uptake and utilization in *Sorghum bicolor. Front Plant Sci.* **Volume 7 - 2016** (2016)
30. Derbyshire, M.C., et al. The complex relationship between disease resistance and yield in crops. *Plant Biotechnol. J.* **22**, 2612-2623 (2024)
31. Gao, M., et al. Revisiting growth–defence trade-offs and breeding strategies in crops. *Plant Biotechnol. J.* **22**, 1198-1205 (2024)
32. Asadyar, L., et al. Evidence for within-species transition between drought response strategies in *Nicotiana benthamiana*. *New Phytol.* **244**, 464-476 (2024)
33. Lewandowska, M., A. Keyl, and I. Feussner. Wax biosynthesis in response to danger: its regulation upon abiotic and biotic stress. *New Phytol.* **227**, 698-713 (2020)
34. Jordan, W.R., et al. Environmental physiology of Sorghum. II. Epicuticular wax load and cuticular transpiration. *Crop Sci.* **24**, cropsci1984.0011183X002400060038x (1984)
35. Xue, D., et al. Molecular and evolutionary mechanisms of cuticular wax for plant drought tolerance. *Front Plant Sci.* **Volume 8 - 2017** (2017)
36. Cole, B.I. and I.D. Lunt. Restoring Kangaroo Grass (*Themeda triandra*) to grassland and woodland understoreys: a review of establishment requirements and restoration exercises in south-east Australia. *Ecol. Manage. Restor.* **6**, 28-33 (2005)
37. Male, D., et al. Time of seed harvest and sowing determines successful establishment of kangaroo grass (*Themeda triandra*) on Dja Dja Wurrung Country. *Crop & Pasture Science* **76** (2025)
38. Prober, S.M., et al. Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration. *Front. Ecol. Evol.* **Volume 3 - 2015** (2015)
39. Tossi, V.E., et al. Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Front Plant Sci.* **Volume 13 - 2022** (2022)
40. Snyman, H.A., L.J. Ingram, and K.P. Kirkman. *Themeda triandra*: a keystone grass species. *African Journal of Range & Forage Science* **30**, 99-125 (2013)
41. Tao, Y., et al. Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants* **7**, 766-773 (2021)
42. Welker, C.A.D., et al. Phylogenomics enables biogeographic analysis and a new subtribal classification of Andropogoneae (Poaceae—Panicoideae). *Journal of Systematics and Evolution* **58**, 1003-1030 (2020)
43. Cheng, H., et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170-175 (2021)
44. Roach, M.J., S.A. Schmidt, and A.R. Borneman. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018)

45. Zhou, C., S.A. McCarthy, and R. Durbin. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39** (2022)
46. Cunningham, F., et al. Ensembl 2022. *Nucleic Acids Res.* **50**, D988-D995 (2021)
47. Astashyn, A., et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, 60 (2024)
48. Simão, F.A., et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015)
49. Huang, N. and H. Li. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39** (2023)
50. Brown, M.R., P. Manuel Gonzalez de La Rosa, and M. Blaxter. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* **41** (2025)
51. Gabriel, L., et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769-777 (2024)
52. Paysan-Lafosse, T., et al. InterPro in 2022. *Nucleic Acids Res.* **51**, D418-D427 (2023)
53. Cantalapiedra, C.P., et al. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825-5829 (2021)
54. Kanehisa, M., Y. Sato, and K. Morishima. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726-731 (2016)
55. McCormick, R.F., et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* **93**, 338-354 (2018)
56. Marçais, G., et al. MUMmer4: A fast and versatile genome alignment system. *PLoS Comp. Biol.* **14**, e1005944 (2018)
57. Goel, M., et al. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 1-13 (2019)
58. Sun, J., et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* **51**, W397-W403 (2023)
59. Schnable, P.S., et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112-1115 (2009)
60. Li, L., C.J. Stoeckert, Jr., and D.S. Roos. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178-89 (2003)
61. Mendes, F.K., et al. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516-5518 (2020)
62. Wang, Y., et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49-e49 (2012)
63. Doyle, J. and J. Doyle. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11-15 (1987)
64. Chen, S., et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018)
65. Catchen, J., et al. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124-3140 (2013)
66. Krueger, F., S.R. Andrews, and C.S. Osborne. Large scale loss of data in low-diversity Illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS One* **6**, e16607 (2011)
67. Bolger, A.M., M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014)
68. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997** (2013)

69. Danecek, P., et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021)
70. Chen, X., et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2015)
71. Weißbach, S., et al. Reliability of genomic variants across different next-generation sequencing platforms and bioinformatic processing pipelines. *BMC Genom*. **22**, 62 (2021)
72. Severn-Ellis, A.A., et al. Genotyping for species identification and diversity assessment using double-digest restriction site-associated DNA sequencing (ddRAD-Seq). *Methods Mol Biol* **2107**, 159-187 (2020)
73. Kozlov, A.M., et al. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455 (2019)
74. Revell, L.J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024)
75. Alexa, A. and J. Rahnenfuhrer. topGO: Enrichment analysis for gene ontology. *R package version 2.54.0* https://bioconductor.org/packages/topGO (2023)