Cui et al. *Horticulture Research* (2021)8:93
https://doi.org/10.1038/s41438-021-00519-w

Horticulture Research
www.nature.com/hortres

ARTICLE Open Access

The genome of the warm-season turfgrass African bermudagrass (*Cynodon transvaalensis*)

Fengchao Cui[1], Geli Taier[1], Manli Li[2], Xiaoxia Dai[1], Nan Hang[1], Xunzhong Zhang[3], Xiangfeng Wang[4] and Kehua Wang[1]

Abstract

Cynodon species can be used for multiple purposes and have high economic and ecological significance. However, the genetic basis of the favorable agronomic traits of *Cynodon* species is poorly understood, partially due to the limited availability of genomic resources. In this study, we report a chromosome-scale genome assembly of a diploid *Cynodon* species, C. transvaalensis, obtained by combining Illumina and Nanopore sequencing, BioNano, and Hi-C. The assembly contains 282 scaffolds (~423.42 Mb, N50 = 5.37 Mb), which cover ~93.2% of the estimated genome of C. transvaalensis (~454.4 Mb). Furthermore, 90.48% of the scaffolds (~383.08 Mb) were anchored to nine pseudomolecules, of which the largest was 60.78 Mb in length. Evolutionary analysis along with transcriptome comparison provided a preliminary genomic basis for the adaptation of this species to tropical and/or subtropical climates, typically with dry summers. The genomic resources generated in this study will not only facilitate evolutionary studies of the Chloridoideae subfamily, in particular, the Cynodonteae tribe, but also facilitate functional genomic research and genetic breeding in *Cynodon* species for new leading turfgrass cultivars in the future.

Introduction

Bermudagrasses (*Cynodon* spp.) are warm-season (C_4) perennial grass species that originate primarily from open areas in southeastern Africa[1]. Following the early diversification associated with the grazing of herbivorous African hoofed animals, their secondary/postancestral centers of origin/diversification occurred in multiple locations, including in South Africa, India, Australia, Afghanistan, and China[2,3]. The *Cynodon* genus (*Cynodon* Rich.) is classified in the *Cynodonteae* tribe, Chloridoideae subfamily, and Poaceae family, with limited diversification of ~9–11 species, including *Cynodon aethiopicus* Clayton & Harlan, *C. barberi* Rang. & Tadul., *C. coursii* A. Camus, *C. dactylon* (L.) Pers., *C. incompletus* Nees, *C.* × *magennisii* Hurcombe, *C. nlemfuensis* Vanderyst, *C. parviglumis* Ohwi, *C. plectostachyus* (K. Schum.) Pilg., *C. radiatus* Roth, and *C. transvaalensis* Burtt Davy[4–7] (The Plant List, www.theplantlist.org, 2012). Bermudagrasses are widespread and troublesome weeds, but more importantly, owing to a number of important characteristics of *Cynodon* species, they can be used for multiple purposes with enormous economic and ecological significance[8,9]. In particular, the good tolerance to abiotic stresses (i.e., traffic, heat, salinity, and drought), ability to grow in nearly all types of soil conditions, rapid recovery potential, low growing nature and aggressive sod-forming growth habit with rhizomes and solons all make bermudagrasses the most used warm-season turfgrasses worldwide in the turf industry, and these grasses are ideally suited for any kind of application, including for ground covers, lawns, sports fields, golf course tees, fairways, and putting greens[6,10]. They are also important forages in the southern United States[11] and have attracted research interest as biofuel grasses[12], phytoremediation plants for soil reclamation[13,14], and medicinal plants for their pharmacognostic properties (e.g., as anti-inflammatory, diuretic,

Correspondence: Xiangfeng Wang (xwang@cau.edu.cn) or Kehua Wang (kehwang@cau.edu.cn)
[1]Department of Turfgrass Science and Engineering, College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China
[2]Department of Breeding and Seed Science, College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China
Full list of author information is available at the end of the article

© The Author(s) 2021
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

antiemetic, antidiabetic, and blood-purifying agents) in Asia and Australia[15,16]. However, the genetic basis of the favorable agronomic traits of *Cynodon* species is poorly understood, partially due to the limited availability of genetic and genomic resources.

The genus *Cynodon* belongs to the "PACC" clade that contains the subfamilies Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae[17] or the current "PACMAD" clade, which regroups six subfamilies, namely, Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, and Danthonioideae[18,19]. The "PACMAD" lineage is the only grass group within which the C_4 photosynthesis pathway has evolved, and Panicoideae and Chloridoideae are the two major groups in the clade, with ~3600 and 1600 species, respectively[7,20,21]. Compared to the Panicoideae subfamily, which includes some important agricultural crops that have been extensively studied, such as sorghum, maize, sugarcane, foxtail millet, and switchgrass, the Chloridoideae subfamily largely lags behind in terms of scientific exploration, particularly at the whole-genome level. Moreover, among the four most important genera within Chloridoideae, namely, *Cynodon*, *Eleusine*, *Zoysia*, and *Eragrostris*, *Cynodon* currently lags behind the other three in terms of genomics and molecular breeding applications[9,22]. For instance, several reference genomes have been reported for each of the three genera, such as two for *Eleusine* (i.e., finger millet, *E. coracana*; goosegrass, *E. indica*)[23,24], three for *Zoysia* (i.e., *Z. japonica*, *Z. matrella*, and *Z. pacifica*)[25], and three for *Eragrostris* (i.e., *E. nindensis*; weeping love grass, *E. curvula*; tef, *E. tef*)[26,27], but so far, there is no reference genome available for the *Cynodon* genus, which significantly limits our understanding of its evolutionary history and our ability to exploit the full genetic potential of these species for molecular breeding of superior cultivars, particularly in the face of global climate change and worldwide water shortage.

Cynodon plant species can range from diploid to hexaploid and have 2n chromosome numbers from 18 to 54[1,4,28,29]. Among all the *Cynodon* species, the two with the highest scientific, industrial and economic importance are C. dactylon and C. transvaalensis, namely, common bermudagrass and African bermudagrass, respectively. Common bermudagrass, also known as dog's tooth grass, Bahama grass, devil's grass, couch grass, Indian doab, wiregrass, and scutch grass, is a cosmopolitan grass that is native to most of the eastern hemisphere and is now found throughout the tropical, subtropical, and warm-temperate climatic regions of the world[30]. In contrast, African bermudagrass is more confined; this species is native to only South Africa and has now been introduced to Madagascar, Central and Northwest Africa, Greece, Iran, Southeast Australia, and part of the United States, with relatively limited dispersal and naturalization[30,31]. Common bermudagrass is usually tetraploid with 36 chromosomes ($2n = 4x = 36$), while African bermudagrass is diploid with 18 chromosomes ($2n = 2x = 18$)[2,28]. A number of leading varieties/cultivars used for turfgrass are sterile triploid hybrids from natural (*C.* × *magennisii*) or human-made crosses between common bermudagrass and African bermudagrass, such as Tifgreen, Tifway, Tifdwrf, MS Supreme, Champion, and MiniVerde[10].

Here, we report a chromosome-scale genome assembly of C. transvaalensis obtained by combining Illumina and Nanopore sequencing, BioNano, and Hi-C. The assembly contains 282 scaffolds (~423.42 Mb, N50 = 5.37 Mb), which cover ~93.2% of the estimated genome of C. transvaalensis (~454.4 Mb). We annotated 28,444 genes in C. *transvaalensis*, which is comparable to the gene number of another Cynodonteae tribe species, *Oropetium thomaeum* (28,446–28,835 genes, genome size of ~236–245 Mb)[32,33]. The expanded genes of C. transvaalensis are primarily involved in disease resistance and abiotic stress responses. Transcriptomic analysis of the plants under various abiotic stresses (i.e., high temperature, cold, drought, and salinity) provided further information for understanding the mechanism underlying its responses/adaptation to changing environments. As the first de novo genome assembly of *Cynodon* species in the genus, the genome resources generated in this study will not only facilitate evolutionary studies of the Chloridoideae subfamily, in particular, the Cynodonteae tribe, but also facilitate functional genomic research and genetic breeding in *Cynodon* species for new leading turfgrass cultivars in the future.

Results

Genome sequencing, assembly, and evaluation

In this study, we used *C. transvaalensis* (Fig. 1a) for whole-genome sequencing[34]. Based on Illumina data (~18.8 Gb), *k*-mer frequency analysis revealed that the genome size was ~454.4 Mb (Supplementary Fig. 1), similar to the size determined by flow cytometry (~444.3 Mb) (Supplementary Fig. 2). The heterozygosity was also evaluated to be 1.40%, and the high heterozygosity rate led to an obvious secondary peak at the 1/2 position from the main peak (Supplementary Fig. 1).

In order to overcome the difficulty of assembly with high levels of heterozygosity, a series of technologies, including Illumina sequencing, Nanopore sequencing, Bionano, and Hi-C, were utilized to assemble the genome, and the sequencing and assembly workflow is shown in Supplementary Fig. 3. In this process, a total of 80.13-Gb Nanopore long reads (~176.3× coverage of the genome, N50 = 28 kb) were de novo assembled into contigs, and the contigs were polished by using Illumina short reads (~41.4× coverage of the genome, Supplementary Table 1).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

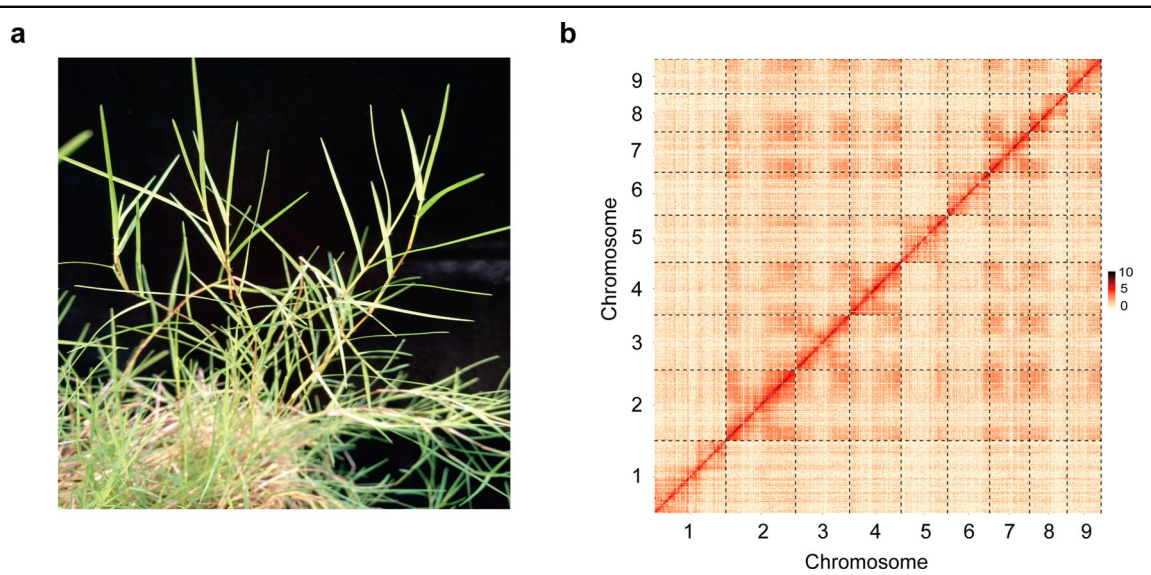

Fig. 1 Plant morphology and Hi-C-assisted genome assembly of *C. transvaalensis*. a Phenotype of the sequenced *C. transvaalensis* plant. **b** Hi-C interaction heatmap showing 100-kb resolution superscaffolds

Table 1 Summary statistics for the *C. transvaalensis* genome

Estimated genome size	444.30 Mb
Total length of assembly	423.42 Mb
Number of contigs	327
Contig N50	2582,667 bp
Largest contig	7992,975 bp
Number of scaffolds	282
Scaffold N50	5371,758 bp
Largest scaffold	35,132,054 bp
Chromosome length	383.08 Mb
GC content of the genome	43.4%
Number of gene models	28,444
Mean gene length	5350 bp
Mean coding sequence length	1280 bp
Total size of transposable elements	128.14 Mb

The assembled genome (~420.45 Mb) contained 327 contigs, in which the contig N50 and the longest contig were 2.58 and 7.99 Mb, respectively (Table 1). Subsequently, ~52.6 Gb of Bionano optical maps (~115.8× coverage of the genome, N50 = 3.39 Mb) were applied to integrate contig sequences into scaffolds and to improve the assembly accuracy (Supplementary Fig. 4), and then, gap sequences longer than 200 kb were deleted (Supplementary Table 2). The remaining 282 scaffolds (~423.42 Mb) accounted for 95.30% of the genome determined by flow cytometry, and the scaffold N50 and the longest scaffold were 5.37 and 35.13 Mb, respectively. Hi-C data (~77.3 Gb, ~170.10× coverage of the genome) were then used to correct, sort, and orient the scaffolds into chromosome-level superscaffolds (pseudomolecules) (Fig. 1b). In total, 90.48% of the scaffolds (~383.08 Mb) were anchored to nine pseudomolecules, of which the longest was 60.78 Mb. The average GC content of the assembled genome was 43.4%, which was close to that of rice (43.6%)[35] and lower than that of zoysiagrass (44.1%)[25]. The assembly information was summarized in Table 1.

To evaluate the quality and integrity of the assembly, Illumina reads were aligned to the genome, and the mapping ratio was 97.62%. In addition, the completeness of the genetic space was assessed by using 1375 Benchmarking Universal Single-Copy Orthologs (BUSCO) genes from Embryophyta[36], and the results indicated that 97.9% of the genes could be annotated and 97% of them were complete (Supplementary Table 3).

Gene annotation

In the assembled genome, repetitive sequences and transposable elements (TEs) accounted for 38.11% and 33.45% of the total sequence, respectively. Among the TEs, long terminal repeat (LTR) retrotransposons were predominant, accounting for 22.14% of the genome. The percentages of *Gypsy* and *Copia* LTR retrotransposons were 14.85% and 2.15% (Supplementary Table 4), respectively. Moreover, long interspersed nuclear elements, short interspersed nuclear elements, and DNA transposons covered 2.9%, 0.01% and 7.49% of

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

the genome, respectively. We also identified 117,768 simple sequence repeats (SSRs) for the assembled genome (Supplementary Table 5).

By combining the transcriptome, homolog and ab initio prediction strategies (Supplementary Table 6), a total of 28,444 high-confidence gene models were obtained after removing gene models containing TEs, which were unevenly distributed on nine pseudomolecules. The average length of the genes was 5350 bp, and each gene contained 5.3 exons. The average CDS, exon, and intron lengths were 1281, 241, and 944 bp, respectively. The gene density, GC content, and *Gypsy* and *Copia* distributions on each pseudomolecule are shown in a circular diagram (Fig. 2). We also compared the *C. transvaalensis* genome with five related C_4 species (*Zoysia japonica*, *Zea mays*, *Sorghum bicolor*, *Setaria italica*, and *Setaria viridis*). Among them, the maize genome was the largest (~2.1 Gb), being ~5.5 times

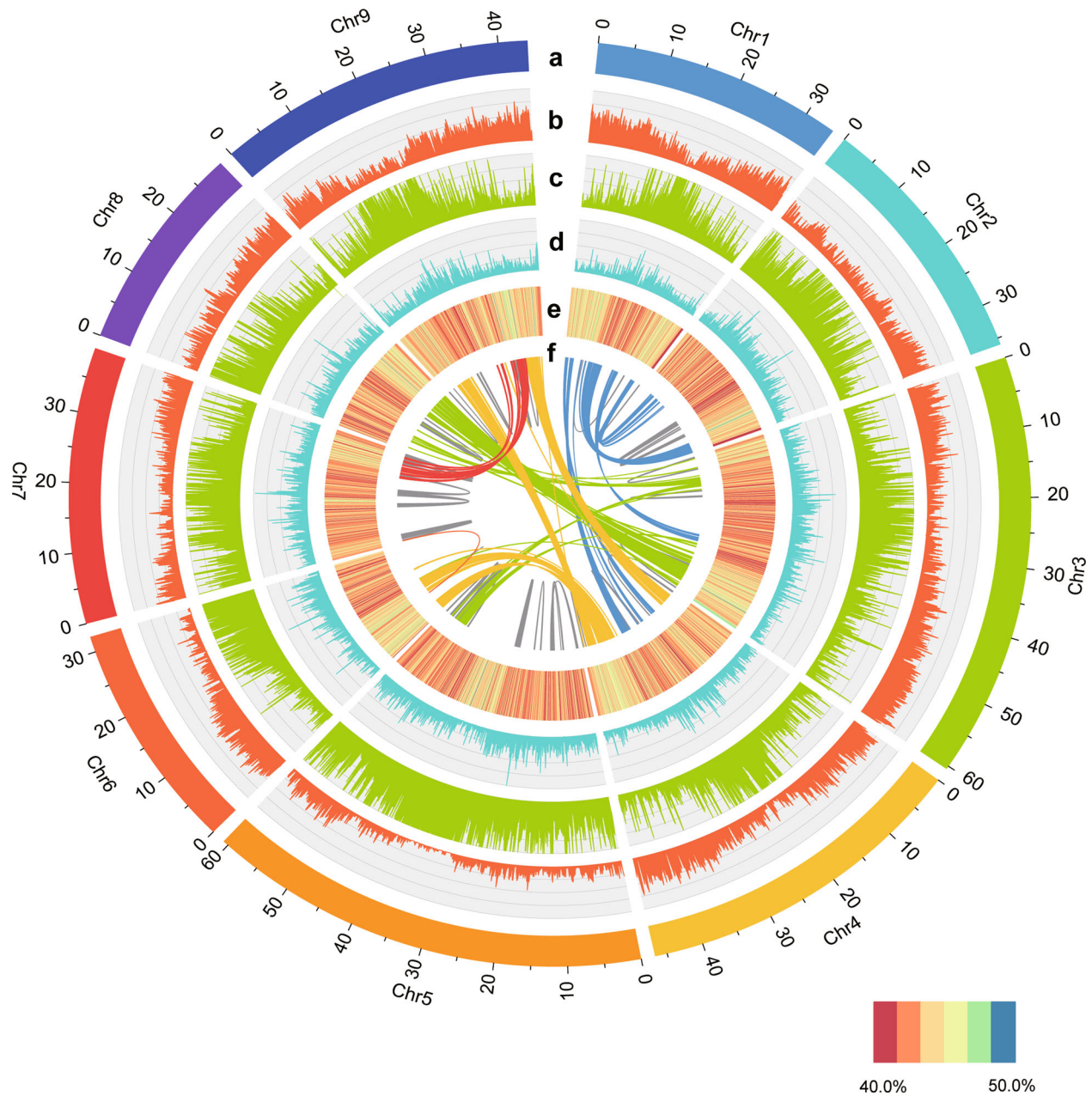

Fig. 2 Features of the *C. transvaalensis* genome. (a) Length of each pseudomolecule. (b) Density of genes. (c) Distribution of *Gypsy*-type retrotransposons. (d) Distribution of *Copia*-type retrotransposons. (e) Heatmap of the GC content. (f) Whole-genome duplication (WGD) events shown by syntenic relationships, with blocks containing more than ten paralogous gene pairs

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

larger than the *C. transvaalensis* genome (Supplementary Fig. 5a). Zoysiagrass had the largest number of genes (49,103), while *C. transvaalensis* had the fewest genes (28,444). The other four species had similar gene numbers (Supplementary Fig. 5b). We found that the distribution of CDSs and gene lengths of the six species showed similar patterns (Supplementary Fig. 5c, d). Gene functions were also validated by searching the predicted gene models in the databases (KEGG, GO, KOG, Swiss-Prot, Nr). In total, 27,146 genes (95.44%) could be annotated, 75.38% and 95.20% of which were found in Swiss-Prot and Nr (Supplementary Fig. 6a), respectively, and 4458 gene models could be annotated across the five databases (Supplementary Fig. 6b). Moreover, 0.06% of the genome was annotated as non-coding RNAs, which included 803 miRNAs, 760 tRNAs, and 62 rRNAs (Supplementary Table 7).

Gene family and evolution analysis

Gene family analysis was performed by comparing the African bermudagrass (*C. transvaalensis*) genome with those of twelve other representative plant species (*Aegilops tauschii*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine max*, *Hordeum vulgare*, *Oryza sativa*, *Phyllostachys edulis*, *S. bicolor*, *S. italica*, *S. viridis*, *Z. japonica* and *Z. mays*). The results revealed that 25,260 genes in *C. transvaalensis* were clustered into 16,595 gene families, and 3184 genes were not clustered into any gene family. The clustered gene family number for *A. thaliana* was the lowest (12,775), while that of *S. viridis* (20,874) was the highest (Supplementary Table 8). Among the five species (*C. transvaalensis*, *Z. mays*, *Z. japonica*, *S. italica*, and *S. viridis*), the number of shared orthologous gene families was 8005, which was higher than that in *C. transvaalensis* and other species (Fig. 3b and

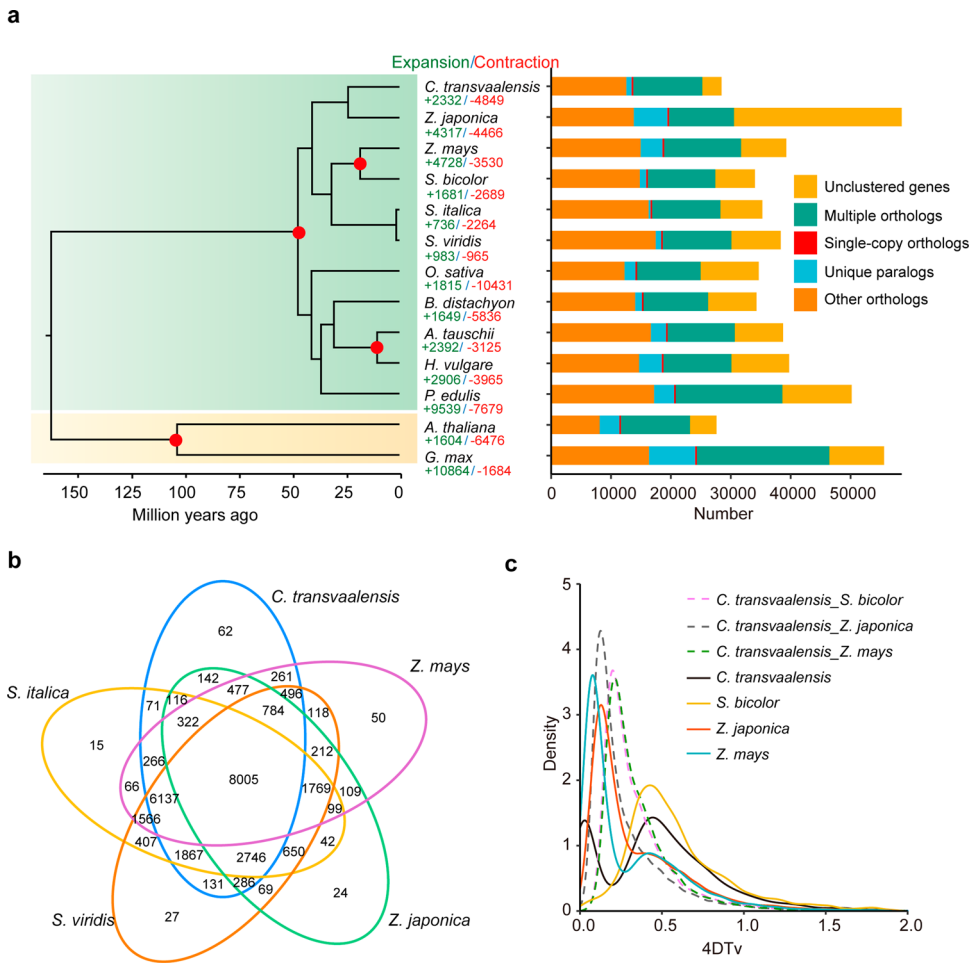

Fig. 3 Gene family and phylogenetic tree analyses of *C. transvaalensis* and other representative plant genomes. a Gene family clustering in *C. transvaalensis* and twelve other plant genomes (right), gene family expansions and contractions among *C. transvaalensis* and twelve other species (middle) and a phylogenetic tree based on shared single-copy gene families (left). **b** Venn diagram of the number of shared gene families within *C. transvaalensis*, *S. italica*, *S. viridis*, *Z. japonica*, and *Z. mays*. **c** Distribution of fourfold degenerate sites of the third codons (4DTv) for syntenic genes within *C. transvaalensis*, *S. bicolor*, *Z. japonica*, and *Z. mays*

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

Supplementary Fig. 7a, b). Then, 278 single-copy orthologous genes were utilized to construct the phylogenetic tree, with *A. thaliana* and *G. max* as outgroup species (Fig. 3a), and the tree indicated that the evolutionary relationship between *C. transvaalensis* and zoysiagrass was closer than that among other species. Based on the published fossilization time, the divergence time between *C. transvaalensis* and zoysiagrass was estimated to be ~23.6 million years ago (Mya).

Whole-genome duplication (WGD) events are an important driving force for plant evolution and adaptation to the environment. To explore the evolutionary history of the *C. transvaalensis* genome, fourfold degenerate sites of the third codons (4DTv) and synonymous substitutions per site (*Ks*) were calculated according to homologous and paralogous gene pairs. When the 4DTv peaks were 0.34–0.45, *C. transvaalensis*, maize, and sorghum experienced the shared WGD event. In addition, *C. transvaalensis* had a recent WGD event when the 4DTv was 0.03 (Fig. 3c), and the WGD event is shown in the circular diagram above (Fig. 2). During the evolution and speciation process, *C. transvaalensis* and zoysiagrass had a closer divergence time than *C. transvaalensis* and other species (*S. bicolor* and *Z. mays*), consistent with the findings from the phylogenetic tree. As expected, *Ks* distributions displayed trends similar to those of the 4DTv results (Supplementary Fig. 8). Positive selection analysis detected 28 genes ($Ka/Ks > 1$), most of which could be annotated to the Swiss-Prot database, such as genes involved in UV sensing, signal transduction, and oxidation-reduction (Supplementary Table 9).

Genome synteny

Based on the phylogenetic tree, genomic synteny analyses of *C. transvaalensis* and the other five species (*Z. mays*, *Z. japonica, S. bicolor*, *S. italica*, and *S. viridis*) were performed to explore their evolutionary relationships. Both *C. transvaalensis* and zoysiagrass were perennial warm-season grasses, and genomic dot plots showed good collinearity (Fig. 4a). The synteny pattern of these two species was 1-to-2 (Fig. 4b), and then, we defined two subgenomes (*Z. japonica* A and *Z. japonica* B) according to the synteny relationships displayed in Fig. 4c. Interestingly, compared with the most recent common ancestor (MRCA), *C. transvaalensis* and zoysiagrass might have experienced chromosome breakage-fusion events (*C. transvaalensis* Chr3 vs. zoysiagrass 8A and zoysiagrass Chr3 vs. zoysiagrass 8B), although most of the chromosomes had 1-to-1 collinearity. After speciation, intergenomic linear rearrangement events also caused some inversions (*C. transvaalensis* Chr6 vs. zoysiagrass 6A and zoysiagrass Chr6 vs. zoysiagrass 6B), which might have contributed to species specificity and environmental adaptation.

To examine the expression patterns, we identified 8816 syntenic genes in *C. transvaalensis* and two zoysiagrass subgenomes, and 6500 genes were retained after filtering low transcripts per million values according to leaf transcriptome sequencing data. Overall, the expression of *C. transvaalensis* was significantly higher than that in the two zoysiagrass subgenomes (Fig. 5b, c). Standardized gene expression (z-score) could be grouped into four clusters, each of which included a different number of genes (Cluster 1: 3158; Cluster 2: 1445; Cluster 3: 751; Cluster 4: 1146) and displayed a different expression pattern (Fig. 5a). From Cluster 1 to Cluster 4, the dominant species were *C. transvaalensis*, *Z. japonica* A, *Z. japonica* B, and *Z. japonica* B, respectively, which might suggest that although *C. transvaalensis* and zoysiagrass had a 1-to-2 synteny pattern, their expression levels were basically the same. The two subgenomes of zoysiagrass often exhibited complementary expression patterns, contributing to the balanced expression in the plants[37]. GO analyses were also performed for different clusters, and terms such as response to hormone, cytokinin (CK) metabolic process and catalytic activity were significantly enriched in clusters 1, 2, and 3, respectively (Supplementary Table 10). Here, we further compared the expression of several key genes involved in CK metabolism and signaling pathways[38,39] among *C. transvaalensis*, *Z. japonica* A, and *Z. japonica* B, such as the adenosine phosphate-isopentenyltransferase (IPT), Arabidopsis histidine phosphotransferases (AHP), Arabidopsis type-B response regulators (Type-B ARR), and cytokinin oxidase/dehydrogenase (CKX) pathways. Notably, the overall CK pathway of *C. transvaalensis* was more active than that of *Z. japonica*, as demonstrated by the higher expression levels of positively regulated genes in the CK pathway (e.g., IPT, LONELY GUY (LOG), AHP, and Type-B ARR) and lower expression of negatively regulated genes (e.g., CKX, Type-A ARR) in *C. transvaalensis* (Fig. 5d). More interestingly, bermudagrasses are aggressive turfgrass species and can recover from foot traffic and other damage quickly, whereas zoysiagrasses are known for their slow growth and establishment rates[40]. Therefore, the growth rate difference between the two grasses might be partly attributable to the different regulation levels of CK metabolism and signaling pathways, considering the important role of CK in plant growth and development[38,39]. In general, CK-deficient plants develop stunted shoots, while accelerated plant growth rates are accompanied by high elevation of active CKs[41,42].

For *C. transvaalensis* and the other four species (*Z. mays*, *S. bicolor*, *S. italica*, and *S. viridis*), we also performed synteny comparisons, and the results all showed high collinearity. The synteny pattern was 1-to-2 in *C. transvaalensis* vs. maize, which was likely because maize originated from allotetraploid ancestors[43], and the

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

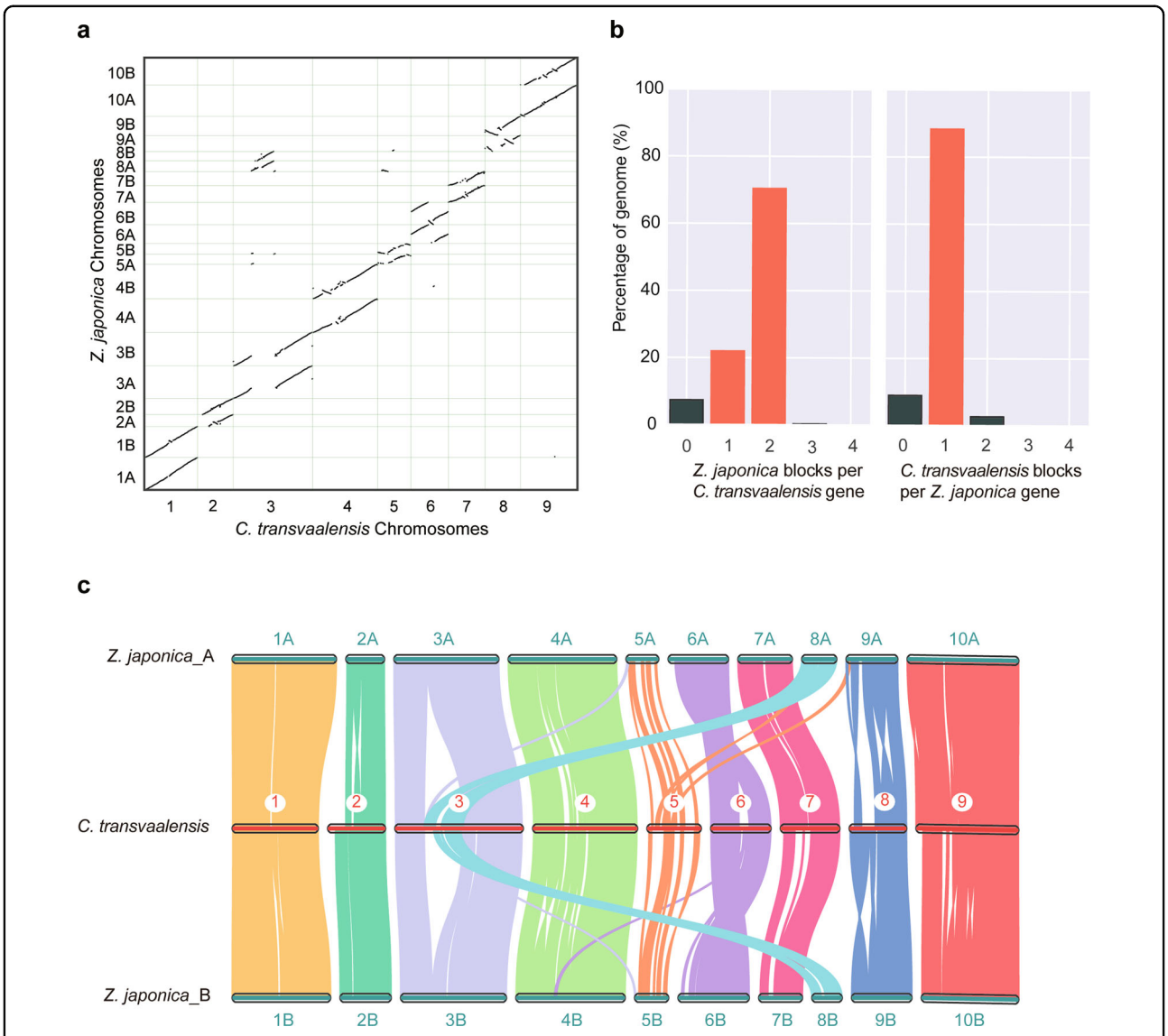

Fig. 4 Genomic synteny analysis of *C. transvaalensis* and *Z. japonica*. a Dot plots for syntenic genes between *C. transvaalensis* and *Z. japonica*. **b** Synteny pattern in *C. transvaalensis* vs. *Z. japonica*. **c** Subgenome classification of *Z. japonica* (*Z. japonica* A and *Z. japonica* B) and synteny analysis

patterns for *S. bicolor*, *S. italica*, and *S. viridis* were all 1-to-1 (Supplementary Fig. 9). Notably, zoysiagrass and sorghum had the same chromosome order in dot plots, indicating that there was no chromosomal rearrangement after speciation. The dot plots between zoysiagrass vs. *S. italica* and zoysiagrass vs. *S. viridis* were very similar because *S. italica* and *S. viridis* were closely related species with high genome consistency (Supplementary Fig. 10).

Genome expansion and contraction

Genome expansion and contraction of *C. transvaalensis* were identified based on gene families. In total, 2332 expanded and 4849 contracted gene families were obtained, and the number of expanded genes was more than nine times the number of contracted genes (Fig. 3a). We also performed GO and KEGG analyses, and the GO results were mainly enriched in terms such as ADP-binding, ATPase activity, and oxidoreductase activity (Supplementary Table 11). KEGG pathways were mainly enriched in pathways such as MAPK signaling and glucose metabolism (amino sugar and nucleotide sugar metabolism, galactose metabolism, tryptophan metabolism, and glycolysis/gluconeogenesis) (Supplementary Table 12 and Supplementary Figs. 11 and 12). Overall, most of the expanded genes were involved in stress resistance and energy metabolism, which could help to increase the environmental adaptability of plants.

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

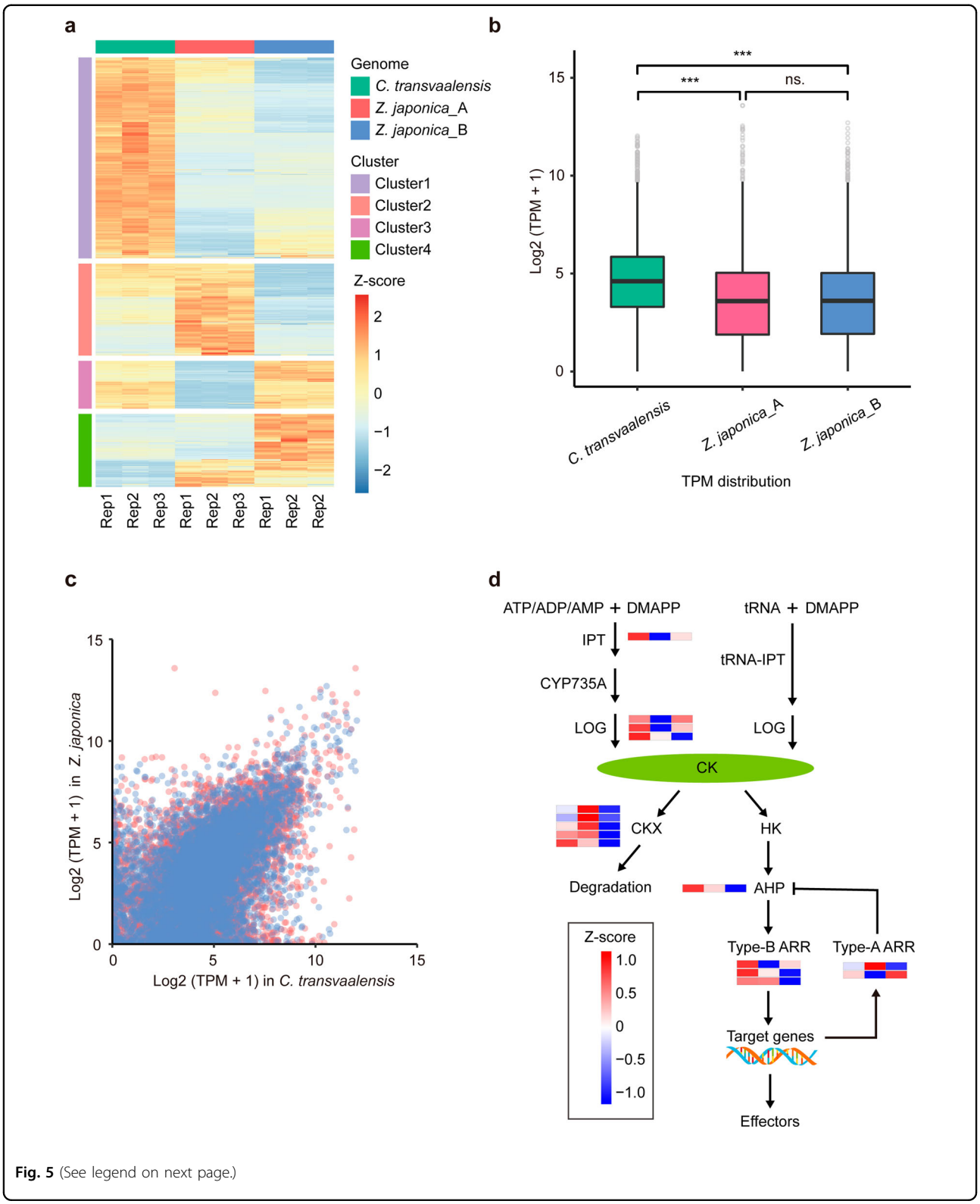

Fig. 5 (See legend on next page.)

Heat shock protein (HSP)-encoding gene families were among the significantly expanded gene families and contained 80 different *HSP* genes, and the majority of the *HSPs* exhibited a different expression pattern in response to abiotic stresses (Fig. 6a and Supplementary Table 13). Subsequently, the HSP70 family was chosen from the six

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

(see figure on previous page)

Fig. 5 Gene expression pattern identification in *C. transvaalensis* and two subgenomes of zoysiagrass (*Z. japonica A* and *Z. japonica B*). **a** Gene expression heatmap of four clustered expression patterns. **b** Syntenic gene expression comparison within *C. transvaalensis*, *Z. japonica* A, and *Z. japonica* B. A nonparametric Wilcoxon rank-sum test was performed to evaluate significant differences for three pairs of expression levels. n.s. not significant; ***$P < 0.01$. **c** Syntenic gene dot plots for *C. transvaalensis* vs. *Z. japonica* A (red dots) and *C. transvaalensis* vs. *Z. japonica* B (blue dots). **d** Overview of cytokinin (CK) metabolism and signaling pathways showing the core genes enriched and their expression in *C. transvaalensis*, *Z. japonica* A, and *Z. japonica* B. Briefly, the first step of CK synthesis is that isopentenyladenine (iP) nucleotides are formed by IPT enzymes and then converted to the respective trans-zeatin (tZ) nucleotides by CYP735As. Activation of the inactive form of CKs is catalyzed by LOG enzymes, while the inactivation of all metabolites is catalyzed by CKX enzymes. Active CKs bind to HK receptors, thereby initiating phosphorelay signaling via HPTs to Type-B ARRs, which activate cytokinin response genes. Among these, Type-A ARRs are negative feedback regulators. DMAPP dimethylallyl diphosphate, IPT adenosine phosphate-isopentenyltransferase, tRNA-IPT tRNA-isopentenyltransferase, CYP735A cytochrome P450 monooxygenases, LOG cytokinin nucleoside 5′-monophosphate phosphoribohydrolase, CKX cytokinin oxidase/dehydrogenase, HK histidine kinase, AHP Arabidopsis histidine phosphotransferase, Type-B ARR Arabidopsis type-B response regulator, Type-A ARR Arabidopsis type-A response regulator

HSP families as an example for phylogenetic and protein motif analysis, and 193 *HSP70* genes were identified among seven species, namely, C. transvaalensis (32), *A. thaliana* (18), *B. distachyon* (30), *O. sativa* (32), *S. bicolor* (33), *S. viridis* (29), and *Z. japonica* (19), with conserved HSP70 domains. According to the predicted protein subcellular locations[44], the phylogenetic tree classified the HSP70s into six groups (I–VI) (Fig. 6b and Supplementary Table 14). The tree also revealed orthologous and paralogous relationships among the HSP70 members of the seven species. Notably, in addition to the 18 orthologous *HSP70* gene pairs between *C. transvaalensis* and zoysiagrass, there were 24 and 7 paralogous pairs in *C. transvaalensis* (4 tandem and 20 segmental duplicated pairs) and zoysiagrass (2 tandem and 5 segmental duplicated pairs), respectively (Fig. 6c and Supplementary Tables 15, 16). These results supported not only the relatively close relationship between the two Chloridoideae species, namely, *C. transvaalensis* and zoysiagrass, but also the respective occurrence of species-specific *HSP70* gene duplication events in both species, particularly the speculated genome-wide duplication event in *C. transvaalensis* after the divergence of the two species (Fig. 3c). Similar expansion mechanisms of the HSP70 family, such as tandem and segmental gene duplication, were also found in other plant species[45,46], which is among the major forces eventually driving the evolution of plant genomes[47,48].

In addition, several fatty acid biosynthetic process-encoding gene families were also expanded, such as the 3-ketoacyl-CoA synthase (KCS), palmitoyl-acyl-carrier protein thioesterase, and 3-oxoacyl-[acyl-carrier-protein] synthase (Supplementary Table 17) families. A total of 13 out of the 19 *KCS* genes were upregulated by either drought or heat stress, but nearly none of them were upregulated under low temperature (Supplementary Fig. 13). For contracted gene families, KEGG pathways were significantly enriched in phenylpropanoid biosynthesis (Supplementary Table 18), which could affect lignin content via both syringyl and guaiacyl monomers according to the pathway map (Supplementary Fig. 14)[49]. A previous study found that the lignin content in C. transvaalensis was less than that in zoysiagrass; therefore, C. transvaalensis could potentially regulate the lignin content to provide more readily digestible forage for animals[50].

Discussion

Notably, the estimated genome sizes of C. transvaalensis determined by k-mer analysis and flow cytometry in this study were 454.4 and 444.3 Mb (Supplementary Figs. 1 and 2), respectively, which were ~8–17% smaller than the previously reported estimation (~495–540 Mb)[8,28,51]. We also estimated the genome size of common bermudagrass ($2n = 4x = 36$) using flow cytometry analysis. As expected, its genome (~905.3 Mb, Supplementary Fig. 2) was nearly double the size of the diploid African bermudagrass used in the study[28]. We postulated that the smaller genome size might be due to internal standard differences, experimental errors, or other differences among studies.

Before this study, zoysiagrass (*Z. japonica*) was the first reported turfgrass species with a relatively high-quality draft genome (e.g., contig N50 value = 2.4 Mb, covering 85.6% of the estimated genome), but its coverage at the chromosome scale was poor (70.2%) due to the limitations of short-read Illumina sequencing, the scaffolding method used (genetic linkage map) and its genome properties (e.g., allotetraploid, high proportion of heterozygous alleles)[25]. Here, we report the first high-quality chromosome-scale genome assembly (contig N50 value = 2.6 Mb, covering 93.2% and 84.3% of the estimated genome at the contig and scaffold levels, respectively, and annotated gene BUSCO value = 96.73%) of a turfgrass species (Supplementary Fig. 15), C. transvaalensis, using a combination of Nanopore sequencing, BioNano optical mapping, and Hi-C scaffolding. It is also the first reference genome of the genus *Cynodon*; this genome is superior to the recently published genomes of several other Chloridoideae subfamily species, such as goosegrass (N50 value = 32.5 kb, BUSCO value = 95%)[24],

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

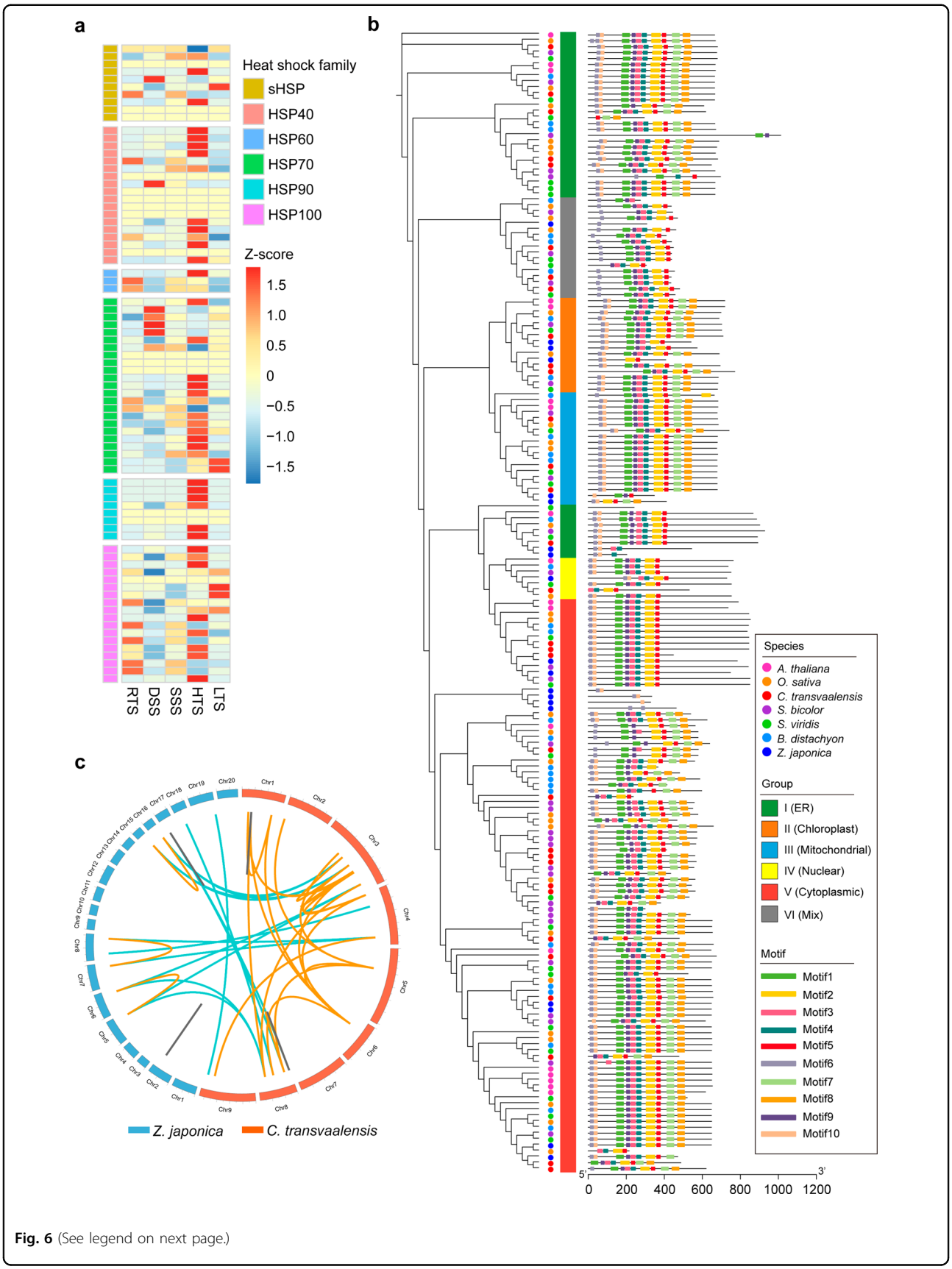

Fig. 6 (See legend on next page.)

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

(see figure on previous page)
Fig. 6 HSP family expression and expansion analysis. a Heatmap for expanded heat shock protein-encoding gene families. RTS, HTS, DSS, SSS, and LTS represent the shoot samples of *C. transvaalensis* under optimum temperature (25/30 °C, day/night, control), high temperature (45 °C for 6 h), drought stress (relative leaf water content of ~60%), salinity stress (200-mmol/L NaCl), and low temperature (4 °C for 6 h), respectively. **b** HSP70 family phylogenetic tree and motif patterns among seven species (*A. thaliana*, *B. distachyon*, *C. transvaalensis*, *O. sativa*, *S. bicolor*, *S. viridis*, *Z. japonica*). The unrooted tree was generated with IQ-TREE 2 software (http://www.iqtree.org/) using the full-length amino acid sequences of the 193 HSP70 proteins. The maximum-likelihood method was selected with 1000 bootstrap replicates. Subfamilies of HSP70s (I–VI) are highlighted with different colored vertical bars next to the gene names in the tree visualized with ITOL (https://itol.embl.de/). Motifs were identified by MEME (https://meme-suite.org/). The sizes of the HSP70 proteins and motifs can be estimated using the scale at the bottom. **c** Orthologous and paralogous *HSP70* gene pair links between *C. transvaalensis* and *Z. japonica*. Orange and gray links indicate intraspecies segmental and tandem duplications, and sky-blue links indicate interspecies orthologous gene pairs

finger millet (N50 value = 23.7 kb)[23], tef (N50 value = 85 kb)[27], *Eragrostis curvula* grass (N50 value = 380 kb, BUSCO value = 96%)[26], and *Oropetium thomaeum* (V1 assembly, N50 value = 2.4 Mb, BUSCO value = 72%)[32,33]. As the 3rd chromosome-scale assembly among the Chloridoideae grasses, the pseudomolecules in the C. transvaalensis reference genome would facilitate further comparative genomic and quantitative genetic studies, such as more accurate identification of syntenic orthologs in other grasses and more detailed comparisons of chromosome evolution in grasses, particularly Chloridoideae species[33].

Genomic collinearity is very common in grass families[52]. In this study, we compared the genome of C. transvaalensis with that of Z. japonica, and as expected, C. transvaalensis showed a high genomic syntenic relationship with Z. japonica due to their close phylogenetic relationship within the Chloridoideae subfamily (Figs. 3 and 4). Similarly, another Cynodon species, C. dactylon, was reported to have a high degree of chromosome-level collinearity with two Chloridoideae grasses, including Z. japonica[53]. According to the phylogenetic analysis, C. transvaalensis and Z. japonica diverged from the MRCA ~23.6 Mya, and they shared an MRCA with the Panicoideae subfamily species (*S. bicolor*, *S. italica*, *S. viridis*, and *Z. mays*) belonging to the "PACMAD" clade ~40.6 Mya, whereas ~47.3 Mya, they shared an MRCS with Bambusoideae (*P. edulis*), Oryzoideae (*O. sativa*), and Pooideae subfamily species (*A. tauschii*, *B. distachyon, H. vulgare*) belonging to the "BOP" clade (Bambusoideae, Oryzoideae, and Pooideae), which is consistent with previous studies[17,53,54].

As a C_4 perennial grass species, unlike its ubiquitous counterpart C. dactylon, the native distribution of C. transvaalensis is nearly completely confined to South Africa, which makes it of particular interest for investigation of its evolution and adaptation to the environment, such as the tropical and/or subtropical climate, typically with dry summers, since its origination[30]. Gene family expansion and contraction due to selection pressure play important roles in the adaptation and evolution of species[55–57]. All six families of HSPs were expanded (Fig. 6), including the small HSPs, HSP40/DnaJ, HSP60/GroEl, HSP70/DnaK, HSP90/HtpG, and HSP100/Clp, which are generally grouped based on their approximate molecular weight[58,59]. HSPs usually protect cells against various abiotic stresses by functioning mainly as molecular chaperones to control the proper folding and conformation of proteins; in particular, numerous studies have reported their important roles in plant heat stress tolerance[60–63]. For instance, HSP101 is a key component for thermotolerance acquisition in plants[64]. The expansion of the HSP families of *C. transvaalensis* indicated possible evolution and adaptation to warm/high temperatures. Moreover, the transcriptome data here showed that the majority of the expanded HSPs were induced under high temperature, and some of them were induced by drought, salinity, or even low temperature (Fig. 6a), which further indicated their major roles in the response to heat stress and possibly some roles in the responses to other abiotic stresses[65]. Several genomic studies have also reported the expansion features of some HSP families, which are thought to be related to differences in environmental adaptation[57,66,67].

In addition to the HSP families, several families of genes involved in fatty acid biosynthetic processes were also expanded, including the KCS family (Supplementary Table 17). KCS genes are known to contribute to very-long-chain fatty acid (VLCFA) synthesis and to play an important role in wax biosynthesis[68,69]; thus, they are involved in plant drought tolerance[70–72]. More gene copies of KCS might help *C. transvaalensis* reduce water loss and better confront dry climates by altering VLCFA biosynthesis. Some bermudagrass species are able to form a layer of wax on the leaf surface to adapt to osmotic stress[73]. The leaf wax layer can also protect plants from strong UV irradiation[74,75], such as the condition faced by *C. transvaalensis* during the dry and hot summer in Africa. In addition, UVR8, a photoreceptor for ultraviolet B that triggers a signaling pathway for ultraviolet protection[76], was among the 28 positively selected genes (Supplementary Table 9), indicating the adaptation of *C. transvaalensis* to a high-ultraviolet environment. Similar gene selection and UV adaptation

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

mechanisms were also reported in other plant species, such as *Juglans sigillata*[77] and maca (*Lepidium meyenii*)[78] growing at high altitudes.

In summary, here, we report the first chromosome-level reference genome in the genus *Cynodon*, a high-quality genome assembly of *C. transvaalensis* determined using several types of sequencing data and assembly methods. Evolutionary analysis and transcriptome comparison showed a preliminary genomic basis for its adaptation to a warm climate with a dry summer environment. The genomic resources generated in this study will not only facilitate further evolutionary studies of the Chloridoideae subfamily but also provide valuable resources for functional genomic research and genetic breeding of new bermudagrass cultivars to be used as turfgrass in the future.

Materials and methods

DNA isolation and sequencing

Leaf samples from single stolon-propagated C. transvaalensis plants grown in the greenhouse of China Agricultural University, Beijing, China, were collected and frozen in liquid nitrogen. Genomic DNA was extracted by the QIAGEN® Genomic DNA Extraction Kit (Qiagen, Hilden, Germany), and the size was selected using the BluePippin system (Sage Science, USA). The selected and purified DNA was then prepared for sequencing following the protocol provided with the genomic sequencing kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK). Single-molecule real-time sequencing was conducted on a Nanopore GridION X5 platform by NextOmics Biosciences (Wuhan, China). Before assembly, reads of low quality, reads shorter than 2000 nt, and reads with adapters were filtered out. Then, 2000 randomly selected reads were subjected to BLAST+ 2.9.0 searches against the NT database, and no obvious external contamination was found. For Illumina sequencing, a separate paired-end library with an insert size of 350 bp was constructed and sequenced by NextOmics Biosciences in accordance with the manufacturer's protocol using the Illumina NovaSeq platform (Illumina, Inc., San Diego, USA) with 150-bp paired-end reads. A total of 19.44 Gb of Illumina raw sequencing data was generated for the survey of C. transvaalensis, and the data were also used for correction and accuracy evaluation of the genome of C. transvaalensis.

Genome survey of African bermudagrass and common bermudagrass

The genome size of C. transvaalensis was estimated using both flow cytometry (EPICS XL, Beckman Coulter, Inc., CA, USA), with sorghum (*Sorghum bicolor*) "BTx623," maize (*Zea mays*) "B73," and rice (*Oryza sativa*) "Nipponbare" as the internal standards[28,79], and the *k*-mer method[80] using Illumina genomic DNA sequencing data. Quality-filtered reads were subjected to 17-mer frequency distribution analysis using the Jellyfish 2.0 program (www.genome.umd.edu/jellyfish.html). The genome size (G) of C. transvaalensis was estimated using the following formula: $G = K_{num}/K_{depth} = b_{num}/b_{depth}$, where K_{num} is the number of k-mers, K_{depth} is the expected depth of k-mers, b_{num} is the base number, and b_{depth} is the expected base depth. The count distribution of the 17-mer followed a Poisson distribution, with two peaks occurring at depths of 17 and 36 (Supplementary Fig. 1).

Genome assembly, polishing, and quality evaluation

Several genome assembly software programs were applied to assemble the genome of C. transvaalensis preliminarily, including wtdbg2 v2.5 (-g 450m -x ont)[81], smartdenovo v1.0 (-k 21)[82], and Nextdenovo (read_cutoff = 1k; seed_cutoff = 25k) (https://github.com/Nextomics/NextDenovo). The quality-controlled Nanopore reads were corrected with Nextdenovo and assembled with Smartdenovo for further genome assembly with the same parameters as those described above. Illumina short reads were mapped to the preliminary genome assembly using minimap2 (v0.7.12-r1039, default parameters)[83], and the reads were then used to polish the assembly four times with NextPolish (v1.0.5, -max_depth 100)[84]. Finally, BWA MEM (http://bio-bwa.sourceforge.net/) and BUSCO v3.0.1[36] were used to assess the quality and completeness of the assembly.

BioNano library construction, sequencing, and scaffolding

High-molecular-mass genomic DNA was isolated as previously described[32,85,86]. Briefly, ~3 g of fresh young leaves of C. transvaalensis was ground to a fine powder in liquid nitrogen and transferred to a 50-ml Falcon tube. The sample was filtered through one layer of Miracloth and two layers of cheesecloth. Nuclei were purified with Percoll gradients and washed extensively before being embedded in 60 μl of low-melting agarose. The DNA plugs were then treated overnight with 500 μl of lysis buffer containing proteinase K, detergent, and β-mercaptoethanol, followed by digestion in 100 μl of RNase A for 1 h. DNA was extracted from the samples, purified, and quantified with a Qubit 3.0 fluorometer. Finally, ~600 ng of genomic DNA was subjected to labeling with the SaphyrPrep Reagent Kit (BioNano Genomics, San Diego, USA). The labeled library was loaded onto a Saphyr Chip and imaged on a Saphyr imaging instrument (BioNano Genomics). Molecules collected from BioNano chips were filtered using a molecule length cutoff of <150 kb and/or a molecule minSites cutoff of <9, and 52.6 Gb of clean data was aligned to the Nanopore genome assembly to generate a molecular quality report, yielding a mapping rate of 45%. To further obtain a longer scaffold, the de novo assembly of Nanopore reads was mapped to the BioNano single-molecule genomic map as previously described[77,87] using the Bionano

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

Access 1.1.2 and Bionano Solve 3.2 hybrid-scaffolding pipeline (hybrid-scaffolding parameters: nonhaplotype without extension and splitting).

Hi-C data analysis and pseudochromosome construction

The Hi-C library was prepared according to Belton et al.[88] and Shi et al.[89] with a modification. In brief, freshly harvested young leaves of C. transvaalensis plants were cut into ~2-cm pieces and fixed in nucleus isolation buffer with 2% formaldehyde after vacuum infiltration for 15 min. The chromatin crosslinking reaction was stopped by adding glycine, and the fixed tissue was then collected, frozen in liquid nitrogen, and ground to powder for nucleus isolation. DNA was extracted from the isolated nuclei, purified, and digested with 100 units of HindIII, and the crosslinked fragments were labeled at the ends with biotin-14-dCTP. The ligated DNA was sheared into 300- to 600-bp fragments and then blunt-end repaired and A-tailed, followed by purification through biotin-streptavidin-mediated pull-down. The purified DNA fragments were used for Hi-C library construction. The libraries were quantified and sequenced on the Illumina NovaSeq platform using the PE 150 layout, which yielded ~93.6 Gb of data with 624 million paired-end reads. Then, quality control of Hi-C raw data was performed using HiC-Pro (v2.8.0) as described by Burton et al.[90]. First, low-quality sequences (quality scores <20), sequences with adapters and sequences shorter than 30 bp were filtered out using fastp v0.12.6[91], and then, the 265 million clean paired-end reads were mapped to the assembled draft sequence using Bowtie2 (v2.3.2) to obtain the unique mapped paired-end reads[92]. As a result, 87 million uniquely mapped paired-end reads were generated, of which 70.34% (61.5 million) were valid interaction pairs. In combination with the valid Hi-C data, we subsequently used the LACHESIS (ligating adjacent chromatin enables scaffolding in situ) de novo assembly pipeline[93] to generate chromosome-level scaffolds of the C. transvaalensis genome.

Genome repeat element identification and gene function annotation

In order to annotate the repeat sequences in C. transvaalensis, based on the principle of repeat sequence-specific structure and de novo prediction, we first used the software LTR_FINDER[94], MITE-Hunter[95], and RepeatModeler (www.repeatmasker.org/RepeatModeler/) to build a repeat sequence database and then combined it with the Repbase database[96] to form the final repeat sequence database. Finally, repeat sequence prediction of the C. transvaalensis genome was performed by using the software RepeatMasker (www.repeatmasker.org) with the final repetitive sequence data as an index file.

For prediction and annotation of the protein-coding genes, the automated eukaryotic gene structure annotation tool EVidenceModeler (EVM)[97] was used by combining three approaches: de novo gene prediction with Augustus[98], de novo assembly of transcripts from RNA-seq data with Transdecoder (https://github.com/TransDecoder/TransDecoder/releases), and prediction of homologous proteins with GeMoMa[99]. Then, TransposonPSI (http://transposonpsi.sourceforge.net) was used to identify potential transposon ORFs within the gene set with PSI-Blast, and the genes were filtered out from the final gene set. Following annotation, singleton and duplicated genes were identified with MCscan[100], and WGD events were visualized with Circos (http://circos.ca/)[101]. The longest protein sequences for each gene were subjected to BLAST analysis against the Swiss-Prot, Nr, KEGG, and KOG/COG databases to identify homologous proteins using Blastp with default parameters, except E value ≤ 1e − 5. The functional domains and possible GO terms in the protein sequences were identified with InterProScan (http://www.ebi.ac.uk/interpro/) with -goterms -t p -f GFF3 -pa -cpu 100. BUSCO evaluation was used to further verify the gene annotation results by searching against the embryophyta_odb10 database.

OrthoMCL[102] (http://orthomcl.org/orthomcl/, v2.09) with the parameters -abc -I 1.5 was used to cluster the gene families for the thirteen species (*A. tauschii*, *A. thaliana*, *B. distachyon*, *C. transvaalensis*, *G. max*, *H. vulgare*, *O. sativa*, *P. edulis*, *S. bicolor*, *S. italica*, *S. viridis*, *Z. japonica*, and *Z. mays*) herein. The protein sequences of the genes from 278 single-copy gene families for the 13 species described above were aligned using MAFFT[103] and then converted into a coding sequence (CDS) alignment. The poorly aligned CDS regions were filtered out with Gblocks[104], and the RaxML method[105] was used to construct a phylogenetic tree with the GTRGAMMA model (Bootstrap = 100). Among the 13 species, *A. thaliana* and *G. max* were used as the outgroup species.

Based on the gene family alignment data, the divergence times of C. transvaalensis from the other plants were estimated using MCMCTREE in the PAML software package[106], and the times were further calibrated using the predicted divergence times of *Z. mays* and *S. bicolor* (11.73–33.32 Mya), *A. thaliana* and *G. max* (97.08–109.04 Mya), and *A. tauschii* and *H. vulgare* (7.75–13.54 Mya) based on available Timetree (http://timetree.org) fossil records.

The 278 common single-copy gene families in all 13 species were used to calculate the expanded and contracted gene families for each lineage using CAFÉ with the default parameters[107]. The orthologous genes among all 13 species were identified using Blastp with an E value ≤ 1e − 5, and then, genes under positive selection in the C. transvaalensis lineage were further identified based on the orthologous gene information using Codeml implemented in the PAML software package[106]. WGD

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

was displayed using 4DTv (fourfold synonymous third-codon transversion). MCscan[100] was used to identify syntenic regions and generate a synteny plot between C. transvaalensis and *Z. japonica*. The 4DTv value was calculated to predict WGD events, and the ratio of non-synonymous (*Ka*) to synonymous (*Ks*) nucleotide substitution rates was calculated using Kaks_calculator[108].

RNA-seq analysis and KEGG pathway enrichment

Total RNA was extracted from leaves of C. transvaalensis under optimum growth chamber conditions [30/25 °C day/light, photosynthetically active radiation 800 μmol/m^2/s], drought stress (after water withholding for 5 days with a relative leaf water content of ~60%), heat stress (45 °C for 6 h), cold stress (4 °C for 6 h), and salinity stress (200-mM NaCl for 24 h, soil salinity level was increased by 50 mM daily) with three biological replicates for Illumina RNA-seq. A total amount of 1 μg of RNA per sample was used for RNA-seq library preparation using TruSeq RNA Sample Prep Kits (Illumina) according to the manufacturer's instructions. The 15 libraries were sequenced on an Illumina NovaSeq platform, and 150-bp paired-end reads were generated. Raw Illumina RNA-seq data/reads were trimmed for quality using fastp (https://github.com/OpenGene/fastp), and the quality of the resulting trimmed reads was further evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). For PacBio full-length transcriptome sequencing, equal amounts of the total RNA extracted with TRIzol reagent (Tiangen Biotech, Beijing, China) from each sample of leaves, stems, and roots were pooled together and used for library preparation. The Iso-Seq library was prepared using the Isoform Sequencing protocol with the SMARTer PCR cDNA Synthesis Kit (Takara, Dalian, China) and the BluePippin Size Selection System protocol (Sage Science, Beverly, USA) with some modifications by NextOmic Biosciences (Wuhan, China). PacBio sequencing was then performed on the PacBio Sequel platform according to the manufacturer's protocol. HISAT2 v2.1.0[109] was applied to map the quality-filtered data to the genome of C. transvaalensis assembled here, and StringTie v1.3.6[110] was used to calculate gene expression. GO and KEGG pathway enrichment analysis was performed by clusterProfiler[111].

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32071887, 31472140) and the Beijing Municipal Natural Science Foundation (6182025). We thank NextOmic Biosciences (Wuhan, China) for their assistance with the PacBio sequencing data, Illumina sequencing data, Hi-C data, and BioNano data. We also thank Dr. Xiqing Ma from China Agricultural University for his generosity in sharing unpublished transcriptome data for *Z. japonica* grown under control conditions.

Author details

[1]Department of Turfgrass Science and Engineering, College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China. [2]Department of Breeding and Seed Science, College of Grassland Science and Technology, China Agricultural University, Beijing 100193, China. [3]School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. [4]National Maize Improvement Center, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100913, China

Author contributions

X.W. and K.W. designed and coordinated the study. F.C., X.W., M.L., and K.W. wrote the manuscript and performed the data analysis. G.T., M.L., X.D., and N.H. collected and grew the plant material and conducted abiotic stress treatments. X.Z. gave helpful suggestions for the whole project and further revised the manuscript. All authors approved the manuscript.

Data availability

The Nanopore long reads and Illumina short reads were uploaded to the China National Center for Bioinformation GSA (Genome Sequence Archive) database under BioProject PRJCA003581, and the submission ID is subPRO005221.

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41438-021-00519-w.

Received: 30 September 2020 Revised: 2 February 2021 Accepted: 6 February 2021
Published online: 01 May 2021

References

1. Forbes, I.Jr, & Burton, G. W. Chromosome numbers and meiosis in some *Cynodon* species and hybrids. *Crop Sci.* **3**, 75–79 (1963).
2. Harlan, J. R., & de Wet, J. M. J. Sources of variation in Cynodon dactylon (L). *Pers. Crop Sci.* **9**, 774–778 (1969).
3. Peterson, P., Columbus, T. & Pennington, S. Classification and biogeography of new world grasses: chloridoideae. *Aliso* **23**, 580–594 (2007).
4. Harlan, J. R., de Wet, J. M. J., Rawal, K. M., Felder, M. R. & Richardson, W. L. Cytogenetic studies in *Cynodon* L. C. Rich. (Gramineae). *Crop Sci.* **10**, 288–291 (1970).
5. Watson, L. & Dallwitz, M. J. *The Grass Genera of the World* (CABI, 1992).
6. Beard, J. B. & Beard, H. J. *Beard's Turfgrass Encyclopedia for Golf Courses, Grounds, Lawns, Sports Fields* (Michigan State University Press, 2005).
7. Peterson, P. M., Romaschenko, K. & Johnson, G. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Mol. Phylogenet. Evol.* **55**, 580–598 (2010).
8. Bethel, C. M. et al. A framework linkage map of bermudagrass (*Cynodon dactylon* × *transvaalensis*) based on single-dose restriction fragments. *Theor. Appl. Genet.* **112**, 727–737 (2006).
9. Khanal, S. et al. SSR-enriched genetic linkage maps of bermudagrass (*Cynodon dactylon* × *transvaalensis*), and their comparison with allied plant genomes. *Theor. Appl. Genet.* **130**, 819–839 (2017).
10. Turgeon, A. J. (ed.) *Turfgrass Management* (Prentice-Hall/Pearson Higher Education, 2011).
11. Ball, D. M., Hoveland, C. S. & Lacefield, G. D. *Southern Forages: Modern Concepts for Forage Crop Management* (International Plant Nutrition Institute, 2015).
12. Xu, J., Wang, Z. & Cheng, J. J. Bermuda grass as feedstock for biofuel production: a review. *Bioresour. Technol.* **102**, 7613–7620 (2011).
13. Garcia, A. & Mangaroo, A. S. Residual concentration of selected heavy metals in a sewage sludge-amended soil and uptake by coastal bermudagrass. in *Sewage Sludge: Land Utilization and the Environment* 187–192 (1994).
14. Razmjoo, K. & Adavi, Z. Assessment of bermudagrass cultivars for phytoremediation of petroleum contaminated soils. *Int. J. Phytoremediat.* **14**, 14–23 (2012).
15. Rai, P. K., Dolly, J., Rai, D. K., Bechan, S. & Geeta, W. Antioxidant potential of oral feeding of *Cynodon dactylon* extract on diabetes-induced oxidative stress. *J. Food Biochem.* **34**, 78–92 (2010).
16. Al-Snafi, A. Chemical constituents and pharmacological effects of *Cynodon dactylon*-A Review. *IOSR J. Pharm.* **6**, 17–31 (2016).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

17. Gaut, B. S. Evolutionary dynamics of grass genomes. *N. Phytol.* **154**, 15–28 (2002).
18. Cotton, J. L. et al. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC Plant Biol.* **15**, 178 (2015).
19. II, G.P.W.G. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *N. Phytol.* **193**, 304–312 (2012).
20. Clayton, W. D., Vorontsova, M. S., Harman, K. T. & Williamson, H. GrassBase—the online world grass flora. http://www.kew.org/data/grasses-db.html (2006).
21. Group, G. et al. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Mo. Bot. Gard.* **88**, 373 (2002).
22. Soreng, R. J. et al. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *J. Syst. Evol.* **55**, 259–290 (2017).
23. Hittalmani, S. et al. Genome and transcriptome sequence of Finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genom.* **18**, 465 (2017).
24. Zhang, H. et al. Development of a goosegrass (*Eleusine indica*) draft genome and application to weed science research. *Pest Manag. Sci.* **75**, 2776–2784 (2019).
25. Tanaka, H. et al. Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Res.* **23**, 171–180 (2016).
26. Carballo, J. et al. A high-quality genome of *Eragrostis curvula* grass provides insights into Poaceae evolution and supports new strategies to enhance forage quality. *Sci. Rep.* **9**, 10250 (2019).
27. Cannarozzi, G. et al. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genom.* **15**, 581 (2014).
28. Wu, Y. Q. et al. Genetic analyses of Chinese *Cynodon* accessions by flow cytometry and AFLP markers. *Crop Sci.* **46**, 917–926 (2006).
29. Gulsen, O. & Ceylan, A. Elucidating polyploidization of bermudagrasses as assessed by organelle and nuclear DNA markers. *OMICS* **15**, 903–912 (2011).
30. Beard, J. B. *Origins of North American Turfgrasses in Turfgrass: Biology, Use, and Management (Agronomy Monographs)* 1st edn, (eds Stier, J. C., Horgan, B. P. & Bonos, S. A.) (American Society of Agronomy, Soil Science Society of America, Crop Science Society of America, Inc., 2013).
31. Clayton, W. D., Renvoize, S. A. & Phillips, S. M. *Flora of Tropical East Africa: Gramineae* (Crown Agents for Overseas Governments & Administrations, 1974).
32. VanBuren, R. et al. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* **527**, 508–511 (2015).
33. VanBuren, R., Wai, C. M., Keilwagen, J. & Pardo, J. A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* **2**, e00096 (2018).
34. Reasor, E. H. et al. Genotypic and phenotypic evaluation of off-type grasses in hybrid Bermudagrass [*Cynodon dactylon* (L.) Pers. x *C. transvaalensis* Burtt-Davy] putting greens using genotyping-by-sequencing and morphological characterization.*Hereditas* **155**, 8 (2017).
35. Du, H. et al. Sequencing and *de novo* assembly of a near complete *indica* rice genome. *Nat. Commun.* **8**, 15324–15324 (2017).
36. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
37. Ye, C. Y. et al. The genomes of the allohexaploid *Echinochloa crus-galli* and its progenitors provide insights into polyploidization-driven adaptation. *Mol. Plant* **13**, 1298–1310 (2020).
38. Werner, T. & Schmulling, T. Cytokinin action in plant development. *Curr. Opin. Plant Biol.* **12**, 527–538 (2009).
39. Kieber, J. J. & Schaller, G. E. Cytokinin signaling in plant development. *Development* **145**, dev149344 (2018).
40. Turgeon, A. J. *Turfgrass Management*, 9th edn (Pearson Prentice Hall, 2010).
41. Werner, T., Motyka, V., Strnad, M. & Schmülling, T. Regulation of plant growth by cytokinin. *Proc. Natl Acad. Sci. U. S. A.* **98**, 10487–10492 (2001).
42. Prerostova, S. et al. Cytokinins: their impact on molecular and growth responses to drought stress and recovery in *Arabidopsis*. *Front. Plant Sci.* **9**, 655 (2018).
43. Gaut, B. S. & Doebley, J. F. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl Acad. Sci. U. S. A.* **94**, 6809–6814 (1997).
44. Yu, C. S., Chen, Y. C., Lu, C. H. & Hwang, J. K. Prediction of protein subcellular localization. *Proteins* **64**, 643–651 (2006).
45. Sarkar, N. K., Kundnani, P. & Grover, A. Functional analysis of Hsp70 superfamily proteins of rice (*Oryza sativa*). *Cell Stress Chaperones* **18**, 427–437 (2013).
46. Zhang, L. et al. Genome-wide analysis and expression profiling under heat and drought treatments of HSP70 gene family in soybean (*Glycine max* L.). *Front. Plant Sci.* **6**, 773 (2015).
47. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **4**, 10 (2004).
48. Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338 (2005).
49. Moura, J. C., Bonine, C. A., de Oliveira Fernandes Viana, J., Dornelas, M. C. & Mazzafera, P. Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J. Integr. Plant Biol.* **52**, 360–376 (2010).
50. Hamido, S. A., Guertal, E. A. & Wood, W. Seasonal variation of carbon and nitrogen emissions from turfgrass. *Am. J. Clim. Change* **5**, 448–463 (2016).
51. Karaca, M., Lang, D. J., Yerk-Davis, G. L., Sukumar, S. & Ainsworth, A. J. Determination of DNA content and genome in *Cynodon* species by flow cytometry. *Crop Res.* **20**, 1–12 (2000).
52. Soreng, R. et al. A world-wide phylogenetic classification of Poaceae (Gramineae). http://www.tropicos.org/projectwebportal.aspx?pagename=ClassificationNWG&projectid=10 (2014).
53. Fang, T., Dong, H., Yu, S. & Moss, J. Q. Sequence-based genetic mapping of *Cynodon dactylon* Pers. reveals new insights into genome evolution in Poaceae. *Commun. Biol.* **3**, 358 (2020).
54. Guo, C. et al. The Coix genome provides insights into panicoideae evolution and papery hull domestication. *Mol. Plant* **13**, 309–320 (2020).
55. Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C. & Cristianini, N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* **15**, 1153–1160 (2005).
56. Sharpton, T. J. et al. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res.* **19**, 1722–1731 (2009).
57. Brawley, S. H. et al. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc. Natl Acad. Sci. U. S. A.* **114**, E6361–E6370 (2017).
58. Feder, M. E. & Hofmann, G. E. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* **61**, 243–282 (1999).
59. Schulze, E.-D., Beck, E. & Muller-Hohenstein, K. *Plant Ecology* (Springer, 2005).
60. Wang, W., Vinocur, B., Shoseyov, O. & Altman, A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci.* **9**, 244–252 (2004).
61. Wahid, A., Gelani, S., Ashraf, M. & Foolad, M. R. Heat tolerance in plants: an overview. *Environ. Exp. Bot.* **61**, 199–223 (2007).
62. Kim, S.-R. & An, G. Rice chloroplast-localized heat shock protein 70, OsHsp70CP1, is essential for chloroplast development under high-temperature conditions. *J. Plant Physiol.* **170**, 854–863 (2013).
63. Sanmiya, K., Suzuki, K., Egawa, Y. & Shono, M. Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants. *FEBS Lett.* **557**, 265–268 (2004).
64. Hong, S. W. & Vierling, E. Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proc. Natl Acad. Sci. U. S. A.* **97**, 4392–4397 (2000).
65. Hýsková, V. & Ryšlavá, H. Roles of HSP70 in plant abiotic stress. *Molecular Approaches in Plant Abiotic Stress* 1st edn, (eds Gaur, R. K. & Sharma, P.) 44–66 (CRC Press, Boca Raton, 2014).
66. Ma, T. et al. Genomic insights into salt adaptation in a desert poplar. *Nat. Commun.* **4**, 2797 (2013).
67. Zhang, Z. et al. Adaptation to extreme Antarctic environments revealed by the genome of a sea ice green alga. *Curr. Biol.* **30**, 3330–3341.e7 (2020).
68. Leng, L. et al. A subclass of HSP70s regulate development and abiotic stress responses in Arabidopsis thaliana. *J. Plant Res.* **130**, 349–363 (2017).
69. Samuels, L., Kunst, L. & Jetter, R. Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annu. Rev. Plant Biol.* **59**, 683–707 (2008).
70. Xue, D., Zhang, X., Lu, X., Chen, G. & Chen, Z.-H. Molecular and evolutionary mechanisms of cuticular wax for plant drought tolerance. *Front. Plant Sci.* **8**, 621 (2017).
71. Zhang, G. et al. The KCS gene is involved in the formation of chloroplast stromules and other physiological processes in jute (*Corchorus capsularis* L.). *Ind. Crop Prod.* **141**, 111781 (2019).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026

72. Lokesh, U. et al. Role of plant fatty acid elongase (3 keto acyl-CoA Synthase) gene in cuticular wax biosynthesis. *J. Agric. Allied Sci.* **2**, 35–42 (2013).
73. Zhou, Y., Lambrides, C. & Fukai, S. Drought resistance and soil water extraction of a perennial C4 grass: Contributions of root and rhizome traits. *Funct. Plant Biol.* **41**, 505 (2014).
74. Long, L. M., Patel, H. P., Cory, W. C. & Stapleton, A. E. The maize epicuticular wax layer provides UV protection. *Funct. Plant Biol.* **30**, 75–81 (2003).
75. Reicosky, D. A. & Hanover, J. W. Physiological effects of surface waxes: I. light reflectance for glaucous and nonglaucous Picea pungens. *Plant Physiol.* **62**, 101–104 (1978).
76. Christie, J. M. et al. Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. *Science* **335**, 1492–1496 (2012).
77. Ning, D.-L. et al. Chromosomal-level assembly of *Juglans sigillata* genome using Nanopore, BioNano, and Hi-C analysis. *Gigascience* **9**, giaa006 (2020).
78. Zhang, J. et al. Genome of plant maca (*Lepidium meyenii*) Illuminates genomic basis for high-altitude adaptation in the central Andes. *Mol. Plant* **9**, 1066–1077 (2016).
79. Dolezel, J. & Bartos, J. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* **95**, 99–110 (2005).
80. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
81. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
82. de Lannoy, C., de Ridder, D. & Risse, J. A sequencer coming of age: *de novo* genome assembly using MinION reads. *bioRxiv* 142711 (2017).
83. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
84. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* **36**, 2253–2255 (2019).
85. Zhang, H.-B., Zhao, X., Ding, X., Paterson, A. H. & Wing, R. A. Preparation of megabase-size DNA from plant nuclei. *Plant J.* **7**, 175–184 (1995).
86. Zhang, M. et al. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protoc.* **7**, 467–478 (2012).
87. Weissensteiner, M. H. et al. Discovery and population genomics of structural variation in a songbird genus. *Nat. Commun.* **11**, 3403 (2020).
88. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
89. Shi, J. et al. Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat. Commun.* **10**, 464 (2019).
90. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
91. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
92. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
93. Korbel, J. O. & Lee, C. Genome assembly and haplotyping with Hi-C. *Nat. Biotechnol.* **31**, 1099–1101 (2013).
94. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
95. Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).
96. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
97. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
98. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **7**, 62 (2006).
99. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
100. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
101. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
102. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
103. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
104. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
105. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
106. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
107. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
108. Zhang, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom, Proteom. Bioinform.* **4**, 259–263 (2006).
109. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
110. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
111. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-021-00519-w/6446620 by guest on 05 August 2026