

High-quality Chromosome-scale De Novo Assembly of the *Paspalum Notatum* 'Flugge' Genome

Zhenfei Yan

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Huancheng Liu

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Yu Chen

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Juan Sun

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Lichao Ma

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Aihua Wang

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Fuhong Miao

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Lili Cong

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Hui Song

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Xue Yin

College of Grassland Science, Qingdao Agricultural University, Qingdao 266109, China.

Qi Wang

College of Animal Science, Qingdao Agricultural University, Qingdao 266109, China.

Yayun Gong

Berry Genomics Corporation, Beijing, China.

Guofeng Yang (✉ yanggf@qau.edu.cn)

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Zengyu Wang

Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China.

Research Article

Keywords: *Paspalum notatum* 'Flugge', Genome, De novo assembly, Genome annotation

Posted Date: September 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-902562/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on April 11th, 2022. See the published version at <https://doi.org/10.1186/s12864-022-08489-6>.

High-quality chromosome-scale *de novo* assembly of the *Paspalum notatum* ‘Flugge’ genome

Zhenfei Yan¹, Huancheng Liu¹, Yu Chen¹, Juan Sun¹, Lichao Ma¹, Aihua Wang¹, Fuhong Miao¹, Lili Cong¹, Hui Song¹, Xue Yin², Qi Wang³, Yayun Gong⁴, Guofeng Yang^{1*} and Zengyu Wang¹

Abstract

Background: *Paspalum notatum* is a diploid multi-purpose subtropical herb native to South America. Its chromosome length is $2n=20XX$ and has a high ecological significance. It is currently widely planted in tropical and subtropical regions. Despite the gene pool of *P. notatum* being unearthed to a large extent in the past decade, no details about the genomic information of relevant species in Paspalum have been reported. In this study, the complete genome information of *P. notatum* was established and annotated through sequencing and *de novo* assembly of its genome.

Results: The latest PacBio third-generation HiFi assembly and sequencing revealed that the genome size of *P. notatum* is 541M. The assembly result is the highest index among the genomes of the gramineous family published so far, with a contig N50=52Mbp, scaffold N50=49Mbp, and BUSCOs=98.1%, accounting for 98.5% of the estimated genome. Genome annotation revealed 36511 high-confidence gene models, thus providing an important resource for future molecular breeding and evolutionary research. A comparison of the genome annotation results of *P. notatum* with other closely related species revealed that it had a close relationship with *Zea mays*. An analysis of the expansion and contraction of gene families suggested that *P. notatum* contains gene families associated with environmental resistance, increased reproductive ability, and molecular evolution, which explained its excellent agronomic traits.

Conclusion: This study is the first to report the high-quality chromosome-scale-based genome of *P. notatum* assembled using the latest PacBio third-generation HiFi sequencing reads. The study provides an excellent genetic resource bank for gramineous crops and invaluable perspectives regarding the evolution of gramineous plants.

Keywords: *Paspalum notatum* ‘Flugge’, Genome, *De novo* assembly, Genome annotation

Background

Paspalum notatum (*P. notatum*) is a subtropical grass native to South America belonging to the Poaceae family, including diploid and apomictic polyploid biotypes[1, 2]. It has excellent agronomic traits such as fast growth, strong reproductive ability, and resistance to cold, barrenness, high temperature, submergence, and erosion[3-6]. It has been used for water and soil conservation, environmental protection, ecological restoration, and landscape viewing, among other uses, thus greatly improving people's lives amongst other economic and ecological benefits[7, 8]. The grass does not have

strict soil type requirements and possesses a strong ability to grow on sandy soils with lower fertility and aridity[9]. These advantages make it the commonly used turfgrass during the warm seasons[10]. It provides a huge feeding value for the livestock industry[11, 12], thus necessitating increased planting in recent years. The grass is widely planted in tropical and subtropical regions. Notably, different environments provide a new source of genes for *P. notatum*, making it grow rapidly and adapt to various environmental changes[13-15].

P. notatum belongs to the Subtrib. Panicinae Reichb composed of Trib[16, 17]. Andropogoneae Dumort, Trib. Arundinelleae Stapf, Trib. Coleantheae Aschers. et Graebn, Trib. Garnotieae Tateoka, Trib. Isachneae Benth, Trib. Maydeae Dumort, and Trib. Paniceae R. Br[18, 19]. Its exploration degree, especially at the genomic level, is far less than members of Trib[20-22]. Andropogoneae Dumort and Trib. Maydeae Dumort, such as *Sorghum bicolor* and *Zea mays*, and the *Setaria viridis* family. Though it is an important sub-family in the millet tribe, the Subtrib. Paspalinae, there are only a few reports about its genome[23, 24]. Only the reference genome of *E. crusgalli* in the genus *Echinochloa* has been reported[25]. The lack of this information restricts our understanding of the evolutionary history of paspalum and the ability to fully tap the genetic potential of this species for breeding superior varieties, especially in the context of global climate changes[26, 27].

Mapping the genome differences between *P. notatum* and its related species using a robust phylogenetic framework is a basis for a comprehensive understanding of the evolution of its genes and genomes[28, 29]. Moreover, using Hi-C technology to observe the collinearity between the chromosomes of *P. notatum* and its related species significantly improves the accuracy and sensitivity of gene evolution research and enables the prediction of more robust genome structure patterns[30, 31]. Using the complete genomics information of *P. notatum* and that of the closely related species for biological analysis can provide many valuable contributions to the analysis of the differentiation and evolution mechanism of the Poaceae Paspalum. In this study, we obtained the genome information of *P. notatum* by combing Illumina and Nanopore, Pacbio Hifi, and Hi-C to fully understand its genome content and molecular evolution history. The study also aimed to identify the historical events and continuous changes of the geographical environment, the positive selection of genes, and the systematic evolution of *P. notatum*. The study provides a starting point for evolutionary genomics studies and a new research direction for analyzing the evolutionary relationships between *P. notatum* and its related species[32, 33].

Results

Genome-survey, sequencing, and assembly

This study evaluated the genome size, repeatability, heterozygosity, and other genome parameters of *P. notatum* (Figure 1a) [34]. Quality control results of the offline data revealed 57Gbp of Illumina data, with a GC content of 46.08%. A comparison of 10,000 randomly selected clean reads to the NT library through blasting revealed a 96.12% mapping. K-mer analysis performed to estimate the complexity of the genome further predicted a genome size of 549M, with a 1.16% and 58.33% heterozygosity

and repeats, respectively.

The genome sequence of *P. notatum* was predicted using the traditional second-generation sequencing (NGS) data assembly method and the third generation Hifi sequencing (Third-Generation Sequencing, TGS) developed by PacBio[35]. Besides TGS making up for some of the shortcomings of NGS in assembly applications, it also did not require PCR amplification, produced ultra-long read lengths, and had no GC preference. Therefore, using PacBio HiFi for genome assembly is an effective assembly strategy. High-quality HiFi reads were obtained after parameter comparison of the output data. The HiFi reads were 1.9Mbp, with an N50 measure of 1.4kbp.

The contigs were subsequently generated based on the overlap map. The assembled genome (541M) contained 894 contigs, with 52Mbp and 125Mbp contig N50 and contig Max, respectively. The average GC content of the assembled genome was 45.65% (Table 1), which was higher than that of *Oryza sativa* (43.65%)[34] and *Cynodon transvaalensis* (43.6%)[36]. The Illumina reads were subsequently compared with the DNA library to evaluate the quality and completeness of the assembly. The comparison yielded 93.77% of the properly mapped reads. Moreover, the single-copy orthologous gene library used to evaluate the completeness of the genetic space revealed a BUSCO of 98.1% of the assembled genome, highlighting that it had good integrity.

Scaffold construction and curation

Scaffold construction and curation

Hi-C is a high-throughput chromosome conformation capture technology. It utilizes the entire cell nucleus as the research object, fixes and captures the mutual sites in the chromosomes, and then performs high-throughput sequencing to study the spatial distribution of chromatin DNA in the whole genome[37, 38]. A high-resolution chromatin regulatory element interaction map is obtained from the positional relationship. In this study, we generated chromosome-level super scaffolds using the Hi-C data with a data volume of 60G and genome coverage of 110 times. Subsequent analysis of the results of the Hi-C library revealed a genome with a scaffold-Len of 540 Gbp and a scaffold N50 of 49 Gbp. Moreover, 514 Mb genome sequences were mapped to 10 chromosomes after Hi-C assisted assembly, accounting for 95.15% of the sequences. The exchange effect between and within chromosomes was calculated upon completion of the Hi-C assisted assembly to further verify the accuracy of the assembly results. The exchange effect within the chromosomes was much stronger than between the chromosomes. Moreover, the exchange effect of chromosomes in a close physical location was much stronger than in a distant physical location (Figure 1b). These findings suggested that the assembly result was correct. Table 1 summarizes the assembly information.

Table 1 Summary statistic for the *Paspalum notatum* ‘Flugge’ genome

Assembly		
Genome assembly	Estimated genome size	549M
	Total length of assembly	541M
	Number of contig	79
	Contig N50	52Mbp
	Largest contig	126Mbp
	Number of scaffolds	49
	Scaffold N50	49Mbp
	Chromosome length (%)	95.15%
GC content of genome	45.65%	
Annotation		
Transposable elements	Total	328Mbp (60.64%)
	Retrotransposon	263Mbp (49.38%)
	DNA Transposon	20Mbp (3.71%)
		Copies
Noncoding RNAs	rRNAs	828
	tRNAs	846
	miRNAs	133
	snRNAs	1708
Gene models	Number of genes	36,511
	Mean gene length	4,029bp
	Mean coding sequence length	1,503bp

Genome annotation

The gene functions in the genome are inferred by calculating the homology alignments and predicting its repetitive sequences. In this study, we identified MITEs repetitive sequences and LTR transposable elements, accounting for 60.64% and 46.61% of the total sequence, respectively, using structure prediction methods. The LTR-retrotransposons of Copia and Gypsy was 10.03% and 25.31%, respectively. In addition, there were 2,827 single repeats identified in the assembled genome. There were 12 types of ncRNA totaling 3907 ncRNA.

In the same line, 36511 high-confidence gene models were obtained using transcriptome, homologous protein, gene structure, and ab initio prediction strategies after eliminating gene models containing transposable elements. The gene models were unevenly distributed on ten chromosomes.

The average gene length was 4029, with each gene containing an average of five exons. The average lengths of CDS, exons, and introns were 1503bp, 319bp, and 599bp, respectively. We also compared *Paspalum notatum* ‘Flugge’ with five related species, including *Puccinellia tenuiflora*, *Zea mays*, *Sorghum bicolor*, *Echinochloa c*, *Echinochloa h*, and *Brachypodium distachyon*). *Zea mays* had the largest genome (~2.1GB) [39] that was 3.8 times that of *Paspalum notatum* ‘Flugge’. An assembly of

Paspalum notatum 'Flugge' with the largest number of genes (36511), *Puccinellia tenuiflora*, which had the smallest number of genes (22346), and the other four species which had similar numbers of genes, revealed similar average CDS lengths. Annotation comparison analysis of six databases used to annotate the genomes annotated 22,900 genes and predicted the functions of different genes and the number and proportion of genes corresponding to them. The Interpro annotations were subsequently used to fuse multiple databases, revealing a data set of 7976 known common genes (Figure 2).

Gene family and evolution analysis

A comparison of *P. notatum* with the genomes of the six representative species combined with gene family analysis revealed that the 36511 genes of *Paspalum notatum* 'Flugge' clustered with 25,335 gene families. The maximum number of clusters in Arabidopsis was 30,235. However, all the species included in the analysis shared 7,219 gene families (Figure 3a). The analysis suggested that *P. notatum* expanded 146 gene families and contracted 807 gene families in the evolution process. GO analysis showed that the expanded gene family types were related to organic-inorganic compound synthesis, DNA biosynthesis, and nucleosides. Notably, gene families related to acid metabolism were the most enriched. The gene families were potentially involved in plant growth metabolism and stress resistance, thus conferring *P. notatum* with strong resistance, fast growth, and strong reproductive ability[40]. A phylogenetic tree was constructed using 5583 single-copy homologous genes, with Arabidopsis thaliana as the out-group[41]. *P. notatum*, *Zea mays*, *Echinochloa crusgalli*, and *Setaria viridis* clustered together to form a monophyletic group[42]. *Zea mays* (maize) was more closely related to *P. notatum* than the other species, with an estimation that it diverged about 26.1 million years ago (Figure 3b). Collinearity analysis suggested that the chromosomes of *P. notatum* and *Zea mays* showed a certain degree of synchronization. The ten chromosomes of *P. notatum* and ten of *Zea mays* had a good collinear relationship (Figure 3c), indicating that the chromosomes were conserved after the differentiation of the two species.

The WGD events are important indices in plant evolution and also the driving force for plant adaptation to various environments[43, 44]. Changes in the synonymous replacement rate between paralogous genes were used to measure the duplication and loss of genes in the *P. notatum* genome to explore its evolutionary history during the evolution process. The resultant data suggested that the differentiation of *P. notatum* and *Setaria viridis* occurred before the WGD events. Both *P. notatum* and *Setaria viridis* experienced a common WGD event when the KS value was 0.32 (Figure 3d). In addition, the WGD event also occurred when the KS value of *P. notatum* was 0.7 (Figure 3e). The leaf tissues were collected from a plant grown in a stable environment. Thus, the data obtained for the positive selection gene analysis did not reach the threshold designed by the computer. It revealed that no gene was significantly affected by the positive selection gene.

Discussion

The gene and genome data of gramineous plants with excellent agronomic traits are

an important resource for comparative genomics and functional omics. Paspalum is an excellent turfgrass whose high-quality chromosome-scale-based genome was assembled for the first time in this study. These findings improve the genomic resource library of gramineous plants and provide an excellent reference for future research on other Paspalum crops. The latest PacBio third-generation HiFi assembly and sequencing revealed that the genome size of *P. notatum* was 541M. The assembly result is the highest index among the genomes of grasses published so far, with a contig N50=52Mbp, scaffold N50=49Mbp, and BUSCOs=98.1%, accounting for 98.5% of the estimated genome. Notably, the coverage of the assembled genome at the chromosome level was also very high (95.15%) after combining high-throughput sequencing and Hi-C scaffolding. Genome annotation revealed 36511 high-confidence gene models, thus providing an important resource for future molecular breeding and evolutionary research.

P. notatum belongs to the Poaceae family, with limited data regarding its performance in evolutionary history. Genome collinearity analysis revealed that *P. notatum* and *Zea mays* had a good degree of genome collinearity. Both species belong to the Subtrib. Panicinae Reichb and are thus close in phylogeny and genetic relationship. Phylogenetic analyses revealed that *P. notatum* diverged after *Oryza sativa*, *Puccinellia tenuiflora*, *Brachypodium distachyon*, and before *Setaria viridis* and *Echinochloa crusgalli*. These species share the same ancestor with *P. notatum*. The genome information of *P. notatum* will help clarify the evolution process of gramineous species and provide a preliminary understanding of their evolutionary state. *P. notatum* has good resistance to various stresses and can thus provide important genetic resources against biotic and abiotic stresses for Poaceae crops.

Conclusion

This study is the first to report the high-quality chromosome-scale-based genome of *P. notatum* assembled using the latest PacBio third-generation HiFi sequencing reads. The genome has a high coverage rate and the highest completeness index among the gramineous genomes that have been published to date. This study provides an excellent genetic resource bank for gramineous crops and invaluable perspectives regarding the evolution of gramineous plants.

Experimental procedures

For sequencing of genomic DNA, the sample was collected by a qualified postgraduate in vacuutainer tube, from the well-growing *P. notatum*(2n=20,XX), planted in a light incubator in the Grassland Agri-husbandry Research Center. The standard plant ethics norms were and complies with Chinese and international regulations.

followed

DNA isolation and sequencing

P. notatum cv. Crowver was selected as the sampling plant. The plant was grown in an incubator at the Qingdao Agricultural University in Shandong, China. Its leaves were sampled in liquid nitrogen followed by genomic DNA extraction using the Tiangen

DNA secure kit. Sequencing of the DNA was done by Berry Hekang (Beijing, China) using the PacBio third-generation HiFi assembly sequencing platform. Quality and quantity control of the DNA samples were first done, followed by library preparation of the processed DNA, and the libraries were subjected to PE sequencing using Illumina Novaseq. Reads containing adapters, duplicates, and a low sequence quality were first filtered, followed by a random selection of 10,000 of the reads for comparison with the NT library using the BLAST tool. There was no significant external contamination detected. Notably, K-mer analysis was performed to estimate the gene size, heterozygosity, and duplication ratio to have a general understanding of the genome in advance.

Genome assembly and quality evaluation

The NanoDrop 2000 spectrophotometer was used to detect the quality of the genomic DNA[45, 46]. The purified genome was subsequently constructed into a SMRTbell library and then sequenced using the Pacbio SMRT technology. The size of the library was detected using Agilent 2100 bioanalyzer. The obtained data was filtered and then processed using the smrtlink software for ccs processing[47-50]. The hifiasm software was used for assembly, followed by de-hybridization of the contig sequence using the purge-dups software[51, 52]. A single-copy orthologous gene library combined using tblastn, augustus, and hmmer software were finally used to evaluate the integrity of the assembled genome[53-58].

Hi-C data analysis and chromosome construction

Paspalum notatum leaf tissue (100mg) was soaked in paraformaldehyde, a cell cross-linking agent, for 15 minutes to bind DNA. Glycine was then added to the mixture to terminate the chromatin cross-linking reaction, followed by collection and freezing of the treated tissues in liquid nitrogen. The tissues were then ground to powder to extract DNA. Biotin-labeled oligonucleotide ends were added during the end repair, and a covaris breaker was subsequently used to break the extracted DNA recovered into 350bp fragments[59]. The DNA bound to biotin was then captured and purified using avidin magnetic beads, followed by library construction and sequencing using the Illumina PE150 platform[38]. The raw reads were filtered, followed by a random selection of 10,000 sequencing reads for comparison to the NT library using the BLAST tool to check for cell contamination[60, 61]. The JUICER software was then employed to compare the Hi-C data with the draft genome. The 3D-DNA comparison was subsequently used to analyze the Hi-C library results to obtain valid Hi-C data and generate the chromosome level scaffold of the *P. notatum* genome[62-64].

Genome functional annotation

The RepeatMasker, MITE Hunter, LTRharvest, LTR Finder, LTR retriever, and RepeatModeler software were employed to analyze and predict the repetitive sequences to identify the MITEs and LTR transposable elements following the structure prediction method[65, 66]. The software parameters of LTRharvest and LTR Finder were -similar 90 -vic 10 -seed 20 -seqids yes -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis And -D 15000 -d 1000 -L 7000 -l 100 -p 20 -C -M 0.9[67, 68]. The parameters of the RepeatModeler software used to identify

the repetitive sequences in the masked genome from scratch were -engine ncbi -pa 60. In the same line, the parameters of the RepeatMasker software used to mask the repetitive sequences in the genome were -s -nolow -norna -gff -engine ncbi -parallel 20[69].

The tRNAscan-SE software was used to predict the tRNA *ab initio* rRNA. Other types of ncRNA were searched using the Rfam database. Their specific information was obtained through similarity comparison[70-72].

The GeMoMa-1.6.1 software was subsequently used to compare the protein sequences of closely related species with the assembled genome to predict its gene structure[73]. Moreover, RNA comparison data and assembly results were combined to obtain exon and intron boundary information to improve forecast accuracy. The iso-seq standardization process was carried out using the SMRT analysis software to obtain high-quality, full-length transcripts. The Augustus v3.0.3, SNAP, GlimmerHMM, and GeneMark-ET v4.57 software were subsequently used to predict the gene structure[74]. The EVM then integrated the gene prediction results, followed by prediction of the coding gene UTR and alternative splicing using PASA[75, 76]. The predicted genes were annotated using the NR, Swiss-Prot, eggNOG, GO, KEGG, and Interpro databases. Venn analysis of the five major databases was then performed to obtain more accurate gene functional annotation information.

Comparative analysis

The Mummer software set at nucmer -g 1000 -c 90 -l 200 was employed to perform genome collinearity analysis on *P. notatum* and its relative species[77, 78], *Zea mays*, to derive its evolution history. Notably, the OrthoMCL cluster analysis was used to identify the gene protein families[79]. An all-vs-all BLAST alignment of all *P. notatum* gene protein-coding sequences (with 1e-5 as the default e-value) was first performed[80], followed by a calculation of the sequence similarity. The Markov clustering algorithm was then used for cluster analysis (expansion coefficient is 1.5) to obtain the protein family clustering results. Single-copy genes of each species were selected as reference markers, and four-fold degenerate sites were used to construct supergenes because of the imperfect evolution of *P. notatum*. The Mafft software was subsequently used for multiple sequence comparisons of supergenes. A suitable base substitution model was selected, followed by constructing a species-based maximum likelihood (ML) phylogenetic tree and estimating its differentiation time using the RAxML software. The mcmctree tool in the PAML software package (parameters: burn-in=5,000,000, sample-number=1,000,000, sample-frequency=50) was used to estimate differentiation time based on the single-copy gene family[81]. The time calibration point (correction point) was derived from the Timetree website. The Cafe software was subsequently used to analyze the gene families changes between species and then perform a GO functional enrichment analysis on the gene families. The Branch-site model analysis method was employed to detect the positive selection occurring in a specific clade and only affects some sites. An esearch of *P. notatum* and its related species was performed to select one-to-one orthology proteins, which were subsequently aligned using the PRANK software set at default. The Gblocks software set at -t=c -e=.ft -b4= 5 -d=y[82, 83], was then used to filter the alignment results. The

CODEML test in PAML was then used to test the positive selection located in a specific branch and affecting certain sites only. The Chi2 program set at a degree of freedom=2 in PAML was subsequently used to test the correction of multiple hypotheses[84].

The duplicate age distribution method was used to detect WGD events. Blastp was used to compare the longest protein sequence of genes in the genome of *P. notatum*. The MCSscanX software was subsequently used to filter the comparison results[85], and the Yn00 tool in the PAML software package was used to calculate the synonymous replacement rate. A density distribution map based on the Ks values of all paralog gene pairs and Ks values of ortholog gene pairs between the genomes of *P. notatum*, *Setaria viridis*, and other related species was then drawn using Matlab[86].

Abbreviations

NGS: Next-Generation Sequencing; CCS: Circular Consensus Sequencing; BUSCO: Benchmarking Universal Single-Copy Orthologs; Hi-C: High-through chromosome conformation capture; MITEs: miniature inverted repeat transposable elements; LTR: Long terminal repeat; LTR-RT: Long terminal repeat retrotransposons; ncRNA: Non-coding RNA; NR: NCBI nucleotide sequences; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; WGD: Whole genome duplications

Declarations

Acknowledgements

The author would like to thank Professor Guofeng Yang, Professor Zengyu Wang, and Professor Juan Sun (Professor of Grassland Science, Qingdao Agricultural University) for their help in data analysis and article writing. Thank you for the scientific research funding provided by the College of Grassland Science of Qingdao Agricultural University. Thanks for the experimental help provided by Beijing Berry and Kang.

Research involving plants

P. notatum is not endangered or a protected species in China, and it was purchased from Crovo and planted in a light incubator. The seeds are sorted and selected by Professor Guofeng Yang.

Authors' contributions

ZY, ZW and GY conceived and designed this research. ZY analyzed data and wrote the manuscript. ZY, HL, YC and JS executed the data analyses. JS participated in the discussion of the results. LM, AW, FM, QW, XY and LC collected samples. GY, HS, YG contributed to the evaluation and discussion of the results and manuscript revisions. All authors have read and approved the final version.

Funding

This study was supported by the National Nature Science Foundation of China (U1906201), Shandong Forage Research System (SDAIT-23-01), China Agriculture Research System (CARS-34) and the First Class Grassland Science Discipline Program of Shandong Province (1619002), China.

Availability of data and materials

All data generated and analyzed during this current study are available in the Grassland Agri-husbandry Research Center, Qingdao Agricultural University with

permission from the Competent Authority. De novo assembly of genome sequence were submitted in NCBI Database having BioProject ID PRJNA763284 and under accession no. SRR15911435.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Grassland Agri-husbandry Research Center, Qingdao Agricultural University, Qingdao 266109, China. ²College of Grassland Science, Qingdao Agricultural University, Qingdao 266109, China. ³College of Animal Science, Qingdao Agricultural University, Qingdao 266109, China. ⁴Berry Genomics Corporation, Beijing, China.

References

1. Ortiz JPA, Revale S, Siena LA, Podio M, Delgado L, Stein J, Leblanc O, Pessino SC: **A reference floral transcriptome of sexual and apomictic *Paspalum notatum***. *BMC genomics* 2017, **18**(1):318.
2. de Oliveira FA, Vigna BBZ, da Silva CC, Favero AP, de Matta FP, Azevedo ALS, de Souza AP: **Coexpression and Transcriptome analyses identify active Apomixis-related genes in *Paspalum notatum* leaves**. *BMC genomics* 2020, **21**(1):78.
3. Peterson PM, Romaschenko K, Johnson G: **A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees**. *Molecular phylogenetics and evolution* 2010, **55**(2):580-598.
4. Watson L, Dallwitz MJ, Weiller C, Wilson K: **The grass genera of the world**. *New Zealand Journal of Ecology* 1992, **16**(2):151-152.
5. Beard JB-B, Harriet J.: **Beard's turfgrass encyclopedia for golf courses, grounds, lawns, sports fields**. 2005.
6. Peterson P, Columbus T, Pennington S: **Classification and Biogeography of New World Grasses: Chloridoideae**. *Aliso* 2007, **23**(1):580-594.
7. Turgeon AJ: **Turfgrass management**: Turfgrass management.; 1991.
8. Cheng, J. J, Wang: **Bermuda grass as feedstock for biofuel production: A review**.
9. Foster JL, Adesogan AT, Carter JN, Blount AR, Myer RO, Phatak SC: **Intake, digestibility, and nitrogen retention by sheep supplemented with warm-season legume hays or soybean meal**. *Journal of animal science* 2009, **87**(9):2891-2898.
10. Masahiko, Hirata, Wempie, Pakiding: **Spatiotemporal Dynamics in Herbage Mass and Tiller Density in a Bahiagrass (*Paspalum notatum* Flugge) Pasture under Cattle Grazing : Results from 4-year Monitoring in Permanent Quadrats**. *Japanese Journal of Grassland Science* 2004, **50**(2):201-204.
11. **Agriculture; Data on Agriculture Reported by Researchers at Federal University Rio Grande do Sul (Reproductive Analyses of Intraspecific *Paspalum Notatum* Flugge Hybrids)**. *Agriculture Week* 2020.
12. **Agriculture; Findings from Federal University in the Area of Agriculture Described (Nitrogen**

- Use Efficiency and Forage Production In Intraspecific Hybrids of Paspalum Notatum Flugge).** *Chemicals & Chemistry* 2019.
13. Espinoza F, Pessino SC, Quarin CL, Valle EM: **Effect of pollination timing on the rate of apomictic reproduction revealed by RAPD markers in paspalum notatum.** *Annals of botany* 2002, **89**(2):165-170.
 14. Sandhu S, Altpeter F: **Co-integration, co-expression and inheritance of unlinked minimal transgene expression cassettes in an apomictic turf and forage grass (Paspalum notatum Flugge).** *Plant cell reports* 2008, **27**(11):1755-1765.
 15. Gaut BS: **Evolutionary dynamics of grass genomes.** *New Phytologist* 2010, **154**(1):15-28.
 16. Clayton W: **GrassBase-The Online World Grass Flora.** 2008.
 17. Beard JB: **Origins of North American Turfgrasses:** Turfgrass: Biology, Use, and Management; 2015.
 18. Group G, Barker NP, Clark LG, Davis JI, Duvall MR, Guala GF, Hsiao C, Linder EAKP: **Phylogeny and Subfamilial Classification of the Grasses (Poaceae).** *Annals of the Missouri Botanical Garden* 2001, **88**(3):373-457.
 19. Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS: **A worldwide phylogenetic classification of the Poaceae (Gramineae).** *Journal of Systematics and Evolution* 2015, **53**(002):117-137.
 20. Cotton JL, Wysocki WP, Clark LG, Kelchner SA, Pires JC, Edger PP, Mayfield-Jones D, Duvall MR: **Resolving deep relationships of PACMAD grasses: a phylogenomic approach.** *BMC plant biology* 2015, **15**:178.
 21. Grass Phylogeny Working G, II: **New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins.** *The New phytologist* 2012, **193**(2):304-312.
 22. Bocchini M, Galla G, Pupilli F, Bellucci M, Albertini E: **The vesicle trafficking regulator PN_SCD1 is demethylated and overexpressed in florets of apomictic Paspalum notatum genotypes.** *Scientific Reports* 2018, **8**(3030):1-11.
 23. Hittalmani S, Mahesh HB, Shirke MD, Biradar H, Uday G, Aruna YR, Lohithaswa HC, Mohanrao A: **Genome and Transcriptome sequence of Finger millet (Eleusine coracana (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties.** *BMC genomics* 2017, **18**(1):465.
 24. Zhang H, Hall N, Goertzen LR, Bi B, Chen CY, Peatman E, Lowe EK, Patel J, McElroy JS: **Development of a goosegrass (Eleusine indica) draft genome and application to weed science research.** *Pest management science* 2019, **75**(10):2776-2784.
 25. Tanaka H, Hirakawa H, Kosugi S, Nakayama S, Ono A, Watanabe A, Hashiguchi M, Gondo T, Ishigaki G, Muguerza M *et al*: **Sequencing and comparative analyses of the genomes of zoysiagrasses.** *DNA research : an international journal for rapid publication of reports on genes and genomes* 2016, **23**(2):171-180.
 26. Carballo J, Santos BACM, Zappacosta D, Garbus I, Selva JP, Gallo CA, Diaz A, Albertini E, Caccamo M, Echenique V: **A high-quality genome of Eragrostis curvula grass provides insights into Poaceae evolution and supports new strategies to enhance forage quality.** *Scientific Reports* 2019, **9**.
 27. Cannarozzi G, Plaza-Wuthrich S, Esfeld K, Larti S, Wilson YS, Girma D, de Castro E, Chanyalew S, Blosch R, Farinelli L *et al*: **Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef).** *BMC genomics* 2014, **15**.

28. Xu CC, Ge YM, Wang JB: **Molecular basis underlying the successful invasion of hexaploid cytotypes of *Solidago canadensis* L.: Insights from integrated gene and miRNA expression profiling.** *Ecol Evol* 2019, **9**(8):4820-4852.
29. VanBuren R, Wai CM, Keilwagen J, Pardo J: **A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*.** *Plant Direct* 2018, **2**(11).
30. Clayton WD: **Flora of tropical east {Africa}-{Gramineae}.** 1970.
31. Vanburen R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E: **Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*.** *Nature* 2015, **527**(7579):508.
32. Ye CY, Wu D, Mao L, Jia L, Qiu J, Lao S, Chen M, Jiang B, Tang W, Peng Q *et al*: **The Genomes of the Allohexaploid *Echinochloa crus-galli* and Its Progenitors Provide Insights into Polyploidization-Driven Adaptation.** *Mol Plant* 2020, **13**(9):1298-1310.
33. Guo C, Wang YN, Yang AG, He J, Xiao CW, Lv SH, Han FM, Yuan YB, Yuan Y, Dong XL *et al*: **The Coix Genome Provides Insights into Panicoideae Evolution and Papery Hull Domestication.** *Molecular Plant* 2020, **13**(2):309-320.
34. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X *et al*: **Sequencing and de novo assembly of a near complete indica rice genome.** *Nature communications* 2017, **8**:15324.
35. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31**(19):3210-3212.
36. Cheng HY, Concepcion GT, Feng XW, Zhang HW, Li H: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.** *Nat Methods* 2021, **18**(2):170-+.
37. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL: **Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.** *Cell Syst* 2016, **3**(1):95-98.
38. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP *et al*: **De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.** *Science* 2017, **356**(6333):92-95.
39. Gaut BS, Doebley JF: **DNA sequence evidence for the segmental allotetraploid origin of maize.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(13):6809-6814.
40. Feder ME, Hofmann GE: **Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology.** *Annual review of physiology* 1999, **61**:243-282.
41. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*.** *BMC plant biology* 2004, **4**:10.
42. Koonin EV: **Orthologs, paralogs, and evolutionary genomics.** *Annual review of genetics* 2005, **39**:309-338.
43. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N: **Estimating the tempo and mode of gene family evolution from comparative genomic data.** *Genome research* 2005, **15**(8):1153-1160.
44. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, Maiti R, Kodira CD, Neafsey DE, Zeng Q *et al*: **Comparative genomic analyses of the human fungal pathogens**

- Coccidioides and their relatives.** *Genome research* 2009, **19**(10):1722-1731.
45. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM: **De novo assembly of haplotype-resolved genomes with trio binning.** *Nature biotechnology* 2018.
 46. Korf I: **Gene finding in novel genomes.** *BMC bioinformatics* 2004, **5**:59.
 47. Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, Miga KH, Eichler EE, Phillippy AM, Koren S: **HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads.** *Genome research* 2020, **30**(9):1291-1305.
 48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.
 49. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**(14):1754-1760.
 50. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al*: **Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.** *PLoS one* 2014, **9**(11):e112963.
 51. Ramirez F, Bhardwaj V, Arrigoni L, Lam KC, Gruning BA, Villaveces J, Habermann B, Akhtar A, Manke T: **High-resolution TADs reveal DNA sequences underlying genome organization in flies.** *Nature communications* 2018, **9**(1):189.
 52. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics* 2020, **36**(9):2896-2898.
 53. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics* 2018, **34**(18):3094-3100.
 54. Quinlan AR: **BEDTools: The Swiss-Army Tool for Genome Feature Analysis.** *Current protocols in bioinformatics* 2014, **47**:11 12 11-34.
 55. Roach MJ, Schmidt SA, Borneman AR: **Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies.** *BMC bioinformatics* 2018, **19**(1):460.
 56. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS *et al*: **Improved maize reference genome with single-molecule technologies.** *Nature* 2017, **546**(7659):524-527.
 57. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL: **Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.** *Cell Syst* 2016, **3**(1):95-98.
 58. Seppey M, Manni M, Zdobnov EM: **BUSCO: Assessing Genome Assembly and Annotation Completeness.** *Methods in molecular biology* 2019, **1962**:227-245.
 59. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nat Methods* 2015, **12**(4):357-360.
 60. Guan DF, McCarthy SA, Wood J, Howe K, Wang YD, Durbin R: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics* 2020, **36**(9):2896-2898.
 61. Jarvis DE, Ho YS, Lightfoot DJ, Schmockel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Mitchell CT, Saber N *et al*: **The genome of *Chenopodium quinoa* (vol 542, pg 307, 2017).** *Nature* 2017, **545**(7655):510-510.

62. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J *et al*: **A chromosome conformation capture ordered sequence of the barley genome.** *Nature* 2017, **544**(7651):426-+.
63. Teh BT, Lim K, Yong CH, Ng CCY, Rao SR, Rajasegaran V, Lim WK, Ong CK, Chan K, Cheng VKY *et al*: **The draft genome of tropical fruit durian (*Durio zibethinus*).** *Nat Genet* 2017, **49**(11):1633-+.
64. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST *et al*: **Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome.** *Nat Genet* 2017, **49**(4):643-+.
65. Majoros WH, Pertea M, Salzberg SL: **TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders.** *Bioinformatics* 2004, **20**(16):2878-2879.
66. Stanke M, Steinkamp R, Waack S, Morgenstern B: **AUGUSTUS: a web server for gene finding in eukaryotes.** *Nucleic acids research* 2004, **32**(Web Server issue):W309-312.
67. Ter-Hovhannisyanyan V, Lomsadze A, Chernoff YO, Borodovsky M: **Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.** *Genome research* 2008, **18**(12):1979-1990.
68. Haas BJ, Salzberg SL, Wei Z, Pertea... M: **Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments.** *Genome Biology* 2008, **9**(1):R7.
69. McGinnis S, Madden TL: **BLAST: at the core of a powerful and diverse set of sequence analysis tools.** *Nucleic acids research* 2004, **32**(Web Server issue):W20-25.
70. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.** *Nucleic acids research* 2011, **39**(Web Server issue):W316-322.
71. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucleic acids research* 1997, **25**(5):955-964.
72. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR: **Rfam: an RNA family database.** *Nucleic acids research* 2003, **31**(1):439-441.
73. Gremme G, Steinbiss S, Kurtz S: **GenomeTools: a comprehensive software library for efficient processing of structured genome annotations.** *IEEE/ACM transactions on computational biology and bioinformatics* 2013, **10**(3):645-656.
74. Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics* 2013, **29**(22):2933-2935.
75. Storz G: **An expanding universe of noncoding RNAs.** *Science* 2002, **296**(5571):1260-1263.
76. Li L, Stoeckert CJ, Jr., Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome research* 2003, **13**(9):2178-2189.
77. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome research* 2009, **19**(9):1639-1645.
78. Harris RS: **Improved pairwise alignment of genomic DNA.** *Dissertations & Theses - Gradworks* 2007.
79. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.** *Nucleic acids research* 2012, **40**(7):e49.

80. Bailey JA, Eichler EE: **Primate segmental duplications: crucibles of evolution, diversity and disease.** *Nature reviews Genetics* 2006, **7**(7):552-564.
81. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151-1155.
82. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al*: **Ancestral polyploidy in seed plants and angiosperms.** *Nature* 2011, **473**(7345):97-100.
83. Yang JZNRZ: **Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level.** *Molecular Biology & Evolution* 2005(12):2472-2479.
84. Kumar S, Nei M, Dudley J, Tamura K: **MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences.** *Briefings in bioinformatics* 2008, **9**(4):299-306.
85. Hahn MW, Han MV, Han SG: **Gene family evolution across 12 Drosophila genomes.** *PLoS genetics* 2007, **3**(11):e197.
86. Hahn MW, Stajich JE: **Estimating the tempo and mode of gene family evolution from comparative genomic data.** *Genome research* 2005, **15**(8):1153-1160.

Fig. 1 Plant morphology and Hi-C-assisted genome assembly of *P. notatum*. **(a)** Phenotype of the sequenced *P. notatum* plant. **(b)** Hi-C interaction heatmap showing 100-kb resolution super scaffolds

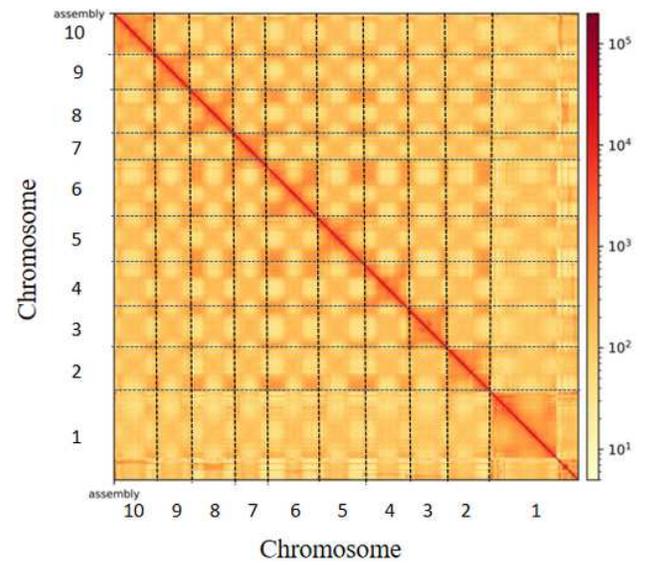
Fig. 2 Venn analysis of five major databases containing gene function annotation information

Fig. 3 Gene family and phylogenetic tree analyses of *P. notatum* and other representative plant genomes. **(a)** Venn diagram of the number of shared gene families. **(b)** A phylogenetic tree based on shared single-copy gene families (left), gene family expansions and contractions among *P. notatum* and seven other species (middle), and Gene family clustering in *P. notatum* and seven other plant genomes (right). **(c)** *P. notatum* and *Zea mays* synteny analysis; the beginning of NC represents the chromosome of *Zea mays*, while the beginning of CHR represents the chromosome of *P. notatum*. **(d)** Genome-wide replication Ks distribution map of *P. notatum* and its related species. **(e)** Genome-wide replication Ks analysis of *P. notatum*.

Figures



a



b

Figure 1

Plant morphology and Hi-C-assisted genome assembly of *P. notatum*. (a) Phenotype of the sequenced *P. notatum* plant. (b) Hi-C interaction heatmap showing 100-kb resolution super scaffolds

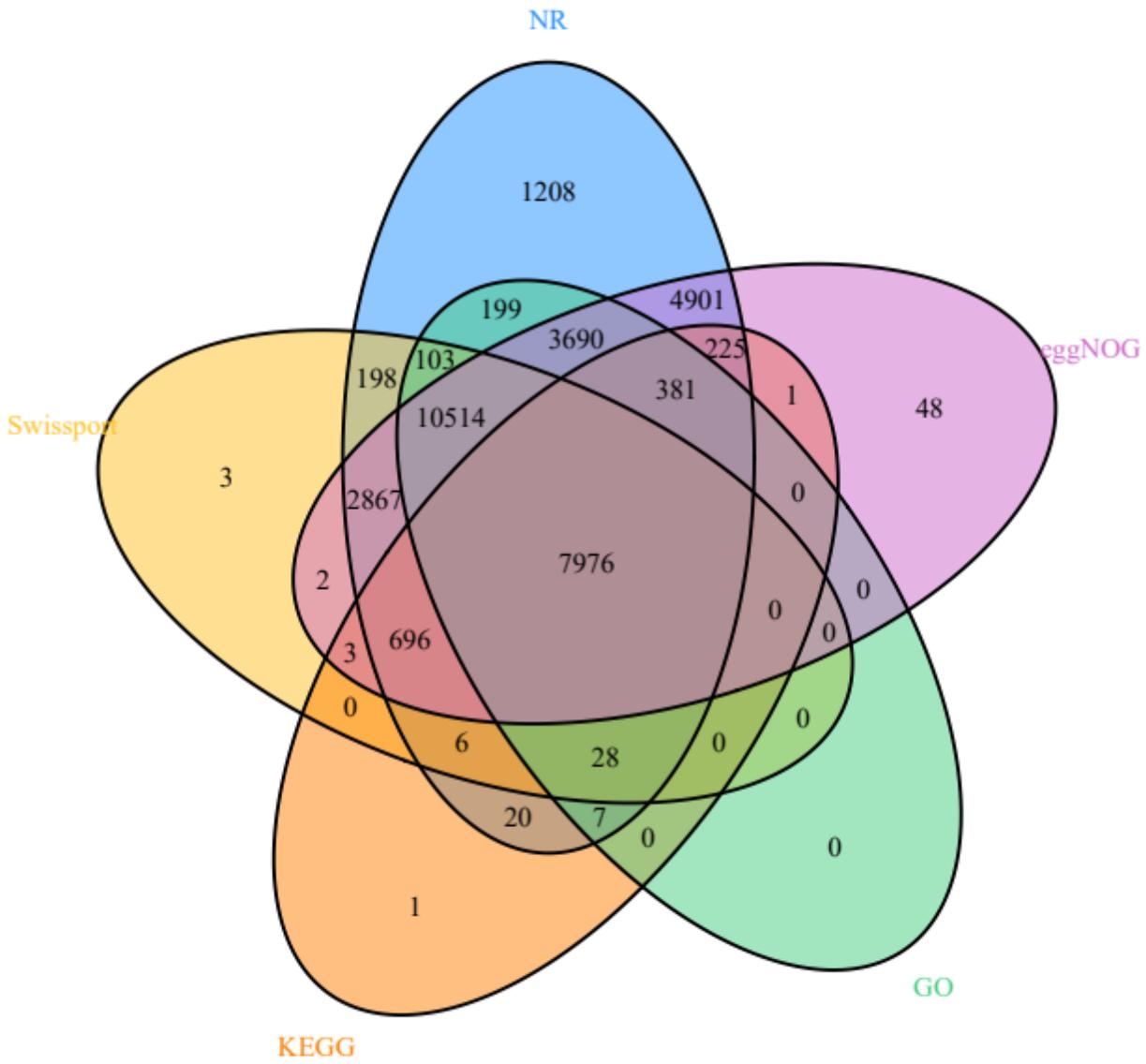


Figure 2

Venn analysis of five major databases containing gene function annotation information

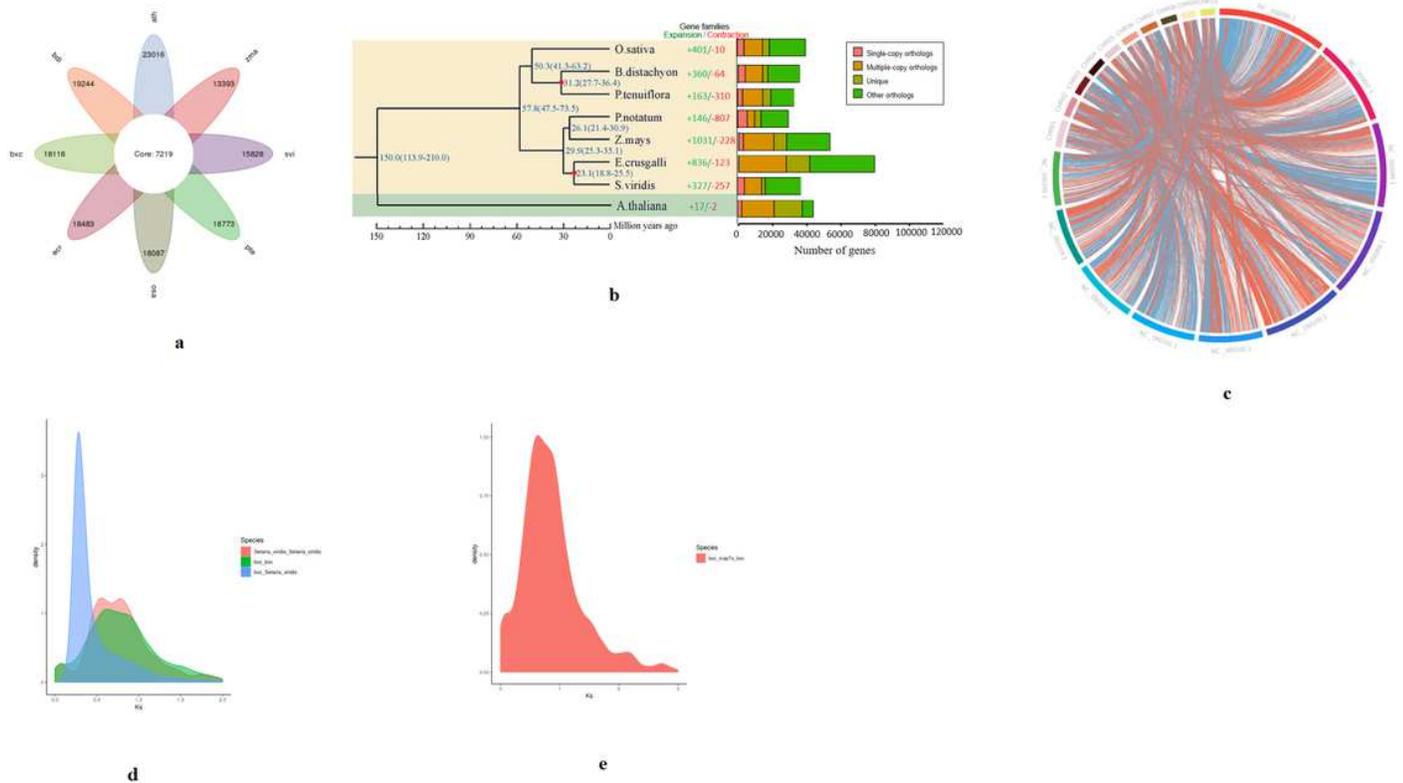


Figure 3

Gene family and phylogenetic tree analyses of *P. notatum* and other representative plant genomes. (a) Venn diagram of the number of shared gene families. (b) A phylogenetic tree based on shared single-copy gene families (left), gene family expansions and contractions among *P. notatum* and seven other species (middle), and Gene family clustering in *P. notatum* and seven other plant genomes (right). (c) *P. notatum* and *Zea mays* synteny analysis; the beginning of NC represents the chromosome of *Zea mays*, while the beginning of CHR represents the chromosome of *P. notatum*. (d) Genome-wide replication K_s distribution map of *P. notatum* and its related species. (e) Genome-wide replication K_s analysis of *P. notatum*.