

RESEARCH

Open Access

Rapidly evolving genetic features for desert adaptations in *Stipagrostis pennata*



Xixu Ding, Tingting Zhang* and Lei Ma*

Abstract

Background: *Stipagrostis pennata* is distributed in the mobile and semi-mobile sand dunes which can adapt well to extreme environments such as drought and high temperature. It is a pioneer plant species with potential for stabilizing sand dunes and ecological restoration. It can settle on moving sand dunes earlier than other desert plants. It can effectively improve the stability of sand dunes and help more plants settle down and increase plant diversity. However, despite its important ecological value, the genetic resources available for this species are limited.

Results: We used single-molecule real-time sequencing technology to obtain the complete full-length transcriptome of *Stipagrostis pennata*, including 90,204 unigenes with an average length of 2624 bp. In addition, the 5436 transcription factors identified in these unigenes are rich in stress resistance genes, such as MYB-related, C3H, bHLH, GRAS and HSF, etc., which may play a role in adapting to desert drought and strong wind stress. Intron retention events are abundant alternative splicing events. *Stipagrostis pennata* has experienced stronger positive selection, accelerating the fixation of advantageous variants. Thirty-eight genes, such as CPP/TSO1-like gene, have evolved rapidly and may play a role in material transportation, flowering and seed formation.

Conclusions: The present study captures the complete full-length transcriptome of *Stipagrostis pennata* and reveals its rapid evolution. The desert adaptation in *Stipagrostis pennata* is reflected in the regulation of gene expression and the adaptability of gene function. Our findings provide a wealth of knowledge for the evolutionary adaptability of desert grass species.

Background

Stipagrostis pennata is a pioneer plant species in the desert. It grows on mobile and semi-mobile sand dunes and has excellent sand fixation ability. It can colonize mobile dunes earlier than other plants. It is reported to be diploid ($2n = 2x = 22$) or tetraploid ($2n = 4x = 44$) [1]. Compared with other desert plants, it has excellent restoration properties such as drought resistance, sand burial resistance, and wind erosion resistance. The colonization of *Stipagrostis pennata* can effectively improve the stability of sand dunes and help more plants settle and increase plant diversity.

Stipagrostis pennata has evolved various mechanisms to adapt to desert environment. It has a developed root system and sheath structure. It can settle in flowing sand dunes and grow well. Besides, the reproduction of

Stipagrostis pennata is guaranteed in arid environments through various characteristics. For example, its curled leaves reduce water evaporation, the flowering time is short, and the number of seeds is large [2–5].

However, the lack of genetic information on *Stipagrostis pennata* has hindered research on its adaptation to stress. In the present study, we combined single-molecule real-time (SMRT) and second-generation sequencing (SGS) to generate the full-length *Stipagrostis pennata* transcriptome. We analyzed the gene function and structure of the *Stipagrostis pennata*, and identified alternative splicing. In order to identify the genes necessary for the adaptive evolution of *Stipagrostis pennata*, selection pressure analysis was performed. Based on the functional annotations of rapidly evolving genes, the adaptive evolution law of specific gene functions are initially revealed. The present study will provide important genetic resources for elucidating the adaptive evolution

* Correspondence: ztting@shzu.edu.cn; mlei@shzu.edu.cn
College of Life Sciences, Shihezi University, Shihezi City, Xinjiang, China



© 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

mechanism of *Stipagrostis pennata*, and will also help to better understand the evolutionary adaptability of desert grass species.

Results

Transcriptome sequencing

We obtained 6,128,860 subreads from raw data by removing joints and low-quality reads. The average subreads length is 2172 bp. After screening, 382,837 circular consensus sequences and 294,509 full-length non-chimeric read sequences (FLNC) were generated, with an average length of 2509 bp.

After error correction and removal of redundant transcripts, a total of 90,204 non-redundant transcripts were produced. Each represents a unique full-length transcript with an average length of 2624 bp and an N50 of 2788 bp (Additional file 1: Table S1).

Annotation of transcripts

The 84,610 (93.80%) sequences are annotated in at least 1 of the 7 databases [6–8] (Nr, Nt, Pfam, KOG, Swiss-prot, KEGG, and GO) (Additional file 1: Figure S1). Most genes (91.28%) are annotated in the Nr database. Moreover, 33,204 unigenes are annotated in all 5 databases. Relatively small number of unigenes (347) are only annotated in the Nr database (Additional file 1: Figure S2). By searching the Nr database, 34.4%, 23.5%, 14.01% and <10% of the unigene sequences of *Stipagrostis pennata* match the corresponding sequences of *Setaria italica*, *Sorghum bicolor*, *Dichanthium oligosanthes* and *Zea mays*, respectively (Additional file 1: Figure S3).

For Gene Ontology analysis (GO), in the category of biological process (BP), the main subcategories of the classified genes include metabolic processes, cellular processes, and single-organism processes. Cell and cell parts were the most abundant terms of the cellular component (CC). In terms of the molecular function (MF) category, binding and catalytic activity are main subcategories (Additional file 1: Figure S4).

EuKaryotic Ortholog Groups (KOG) annotation showed that the most abundant genes are related to post-translational modification, protein turnover, and chaperones. This indicates that post-translational modification of proteins and regulation of protein-level modifications are very active in *Stipagrostis pennata*. The “general function prediction only” group has more abundant genes. This group mainly corresponds to conventional gene functions (Additional file 1: Figure S5).

KEGG annotation analysis was performed using unigenes. In the cellular process category, the most abundant genes are related to transport and catabolism (Additional file 1: Figure S6). In terms of environmental information processing, the most abundant genes are related to signal transduction. The number of genes

related to transcription and translation is highly enriched in information processing, while the number of genes related to gene replication and repair is less. There are many genes enriched in metabolism-related pathways, and carbohydrate metabolism is dominant.

Identification of putative transcription factors (TFs)

We identified 5436 TFs in *Stipagrostis pennata*. The most abundant families of TF are: MYB-related (321 genes), C3H (290 genes), bHLH (282 genes), GRAS (277 genes), HSF (267 genes), C2H2 (234 genes), and SNF2 (228 genes) (Fig. 1). By comparing the number of TFs in *Stipagrostis pennata* and their proportion in three species of Gramineae, it is found that the proportion of main TFs in *Stipagrostis pennata* is special (Fig. 2). The proportion of C3H, GRAS and HSF in *Stipagrostis pennata* are greater than the other three gramineous crops. These TF families play a role in stress resistance. For example, a key role of the C3H gene family is lignin synthesis [9], and the HSF gene family is known to be the key regulatory genes for heat stress adaptation [10].

Identification of alternative splicing (AS) in *Stipagrostis pennata*

Intron retention (IR) are the most AS event in *Stipagrostis pennata*. It accounts for approximately 85% of the total number of AS events (Fig. 3). In addition, exon skipping (ES) and alternative donor sites (AltD) each accounts for 6% of the total number of AS events. Alternative acceptor sites (AltA) are the least AS events.

Intron retention (IR) and exon skipping (ES) are functionally different (Fig. 4). IR events are rich in transportation-related genes. For example, there are many IR genes for protein binding and transporter activity. However, the ES genes are not annotated in the above GO terms. In contrast, ES genes are rich in oxidoreductase activity and biological process regulation. These findings may provide insight into the

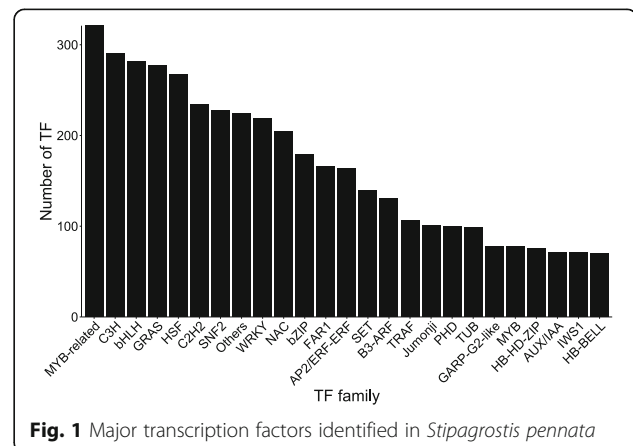


Fig. 1 Major transcription factors identified in *Stipagrostis pennata*

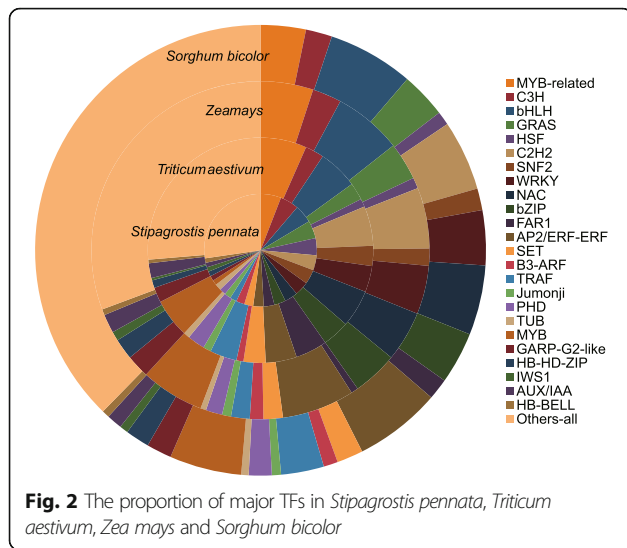


Fig. 2 The proportion of major TFs in *Stipagrostis pennata*, *Triticum aestivum*, *Zea mays* and *Sorghum bicolor*

regulatory mechanisms that support *Stipagrostis pennata*'s adaptation to adverse environments.

Identification of orthologous genes

Compared with other gramineous crops, *Stipagrostis pennata* gradually formed a unique desert adaptation mechanism during the evolution process. We analyzed the homology between *Stipagrostis pennata*, *Arabidopsis thaliana*, *Oryza sativa*, *Setaria italica*, *Triticum aestivum*, *Sorghum bicolor* and *Zea mays*. For *Stipagrostis pennata*, there are fewer genes shared (7708 sequences) with *Arabidopsis thaliana* than with the other five cereal crops. *Stipagrostis pennata* and *Sorghum bicolor* have the highest number of orthologs. The number of homologous sequences unique to a single species is 14,296. The number of orthologous genes shared among all species is 8839. There are 6416 orthologous genes found only in *Stipagrostis pennata* (Fig. 5).

In order to analyze the composition and function of the orthologous genes shared by above-mentioned gramineous plants, GO function enrichment was performed on orthologous genes. *Stipagrostis pennata*'s unique genes are

mainly enriched in the single biological process (Fig. 6). The 8839 orthologous genes shared by all species were significantly enriched to 22 GO entries. Four of those entries belong to the categories of molecular functions and cellular components. The remaining entries belong to the category of biological processes. This result indicates that the orthologous genes are mainly involved in biological processes (Fig. 7).

In the category of biological processes, most orthologous genes are related to RNA modification. RNA modification is important for regulating of RNA biological functions. In addition, there are many genes related to RNA splicing, RNA shearing and rRNA processing. Furthermore, orthologous genes are also related to protein transport, defense response, lipid catabolism, oxidative stress and osmotic stress. In the category of cellular components, orthologous genes are related to mitochondrion and plasma membrane. Finally, in the molecular function category, the orthologous genes are mainly related to oxidoreductase activity and nutrient reservoirs.

Selection pressure analysis

Stipagrostis pennata is a pioneer Gramineae and can adapt to harsh desert environments. An in-depth understanding of the evolutionary pressures of *Stipagrostis pennata* can provide insight for subsequent genetic mining. To gain further insight, *Oryza sativa* was selected as a reference species to analyze the selection pressure of orthologous genes of *Stipagrostis pennata*, *Sorghum bicolor*, *Zea mays*, *Triticum aestivum*, and *Setaria italica*. The ratio of non-synonymous substitutions to synonymous substitutions ω (Ka/Ks^{-1}) values were calculated in the gene pairs of the target species and *Oryza sativa*. Figure 8 shows that when the ω value is greater than 1, the ω value distribution of *Stipagrostis pennata* is different from the other four crops. High ω is a typical feature of rapid evolution. This indicates that *Stipagrostis pennata* has undergone species-specific genetic variation while adapting to the harsh desert environment.

As the value of ω increases, the difference between *Stipagrostis pennata* and the other four grasses become more obvious. According to the ω values of *Stipagrostis pennata*, homologous gene pairs were divided into three groups. These groups were: 0 ~ 1, 1 ~ 2, and 2 ~ max of the value of ω . The Spearman rank correlation coefficients of ω were calculated in those groups, respectively. As shown in Fig. 8, in the total area of ω , all four species are highly significantly positively correlated with each other ($P < 0.01$). However, when the ω value is greater than 1, the correlation between *Stipagrostis pennata* and *Sorghum bicolor*, *Zea mays*, *Triticum aestivum* and *Setaria italica* is not significant. This indicates that positive selection in *Stipagrostis pennata* is different to the other grasses. These genes with

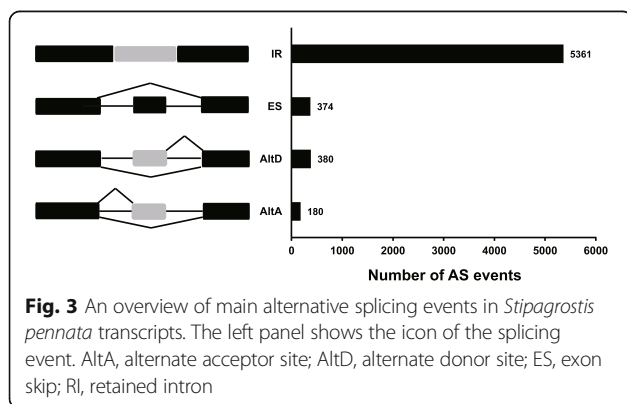
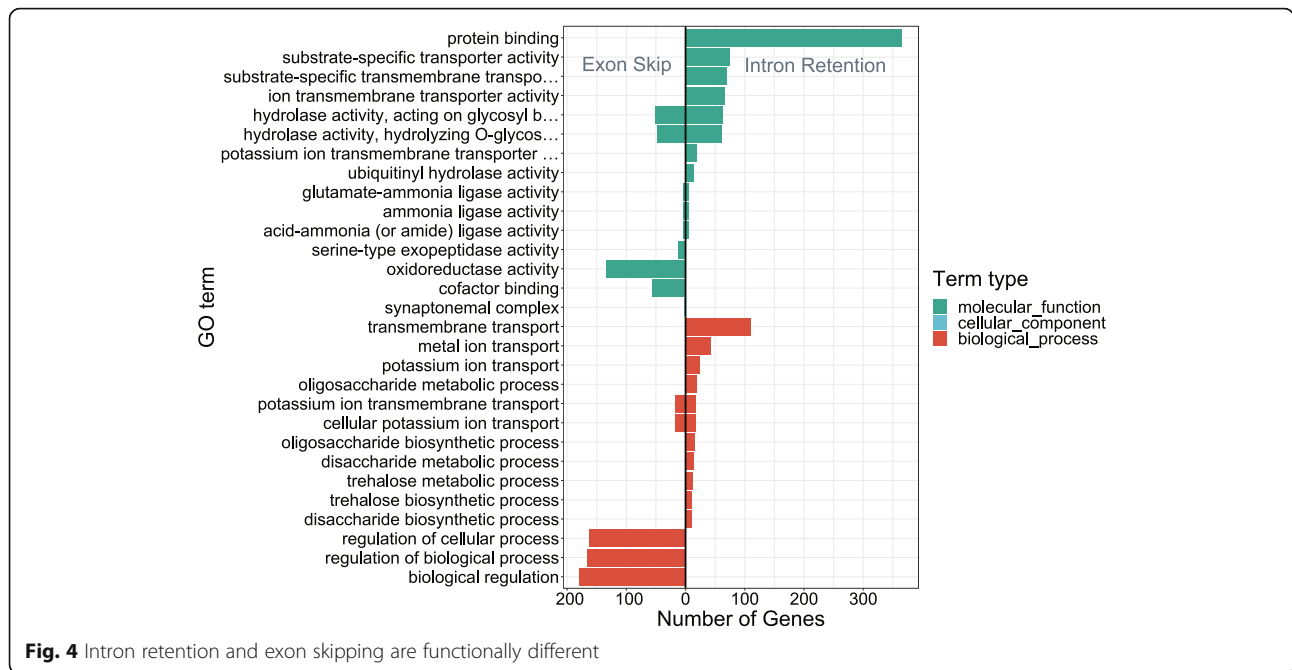


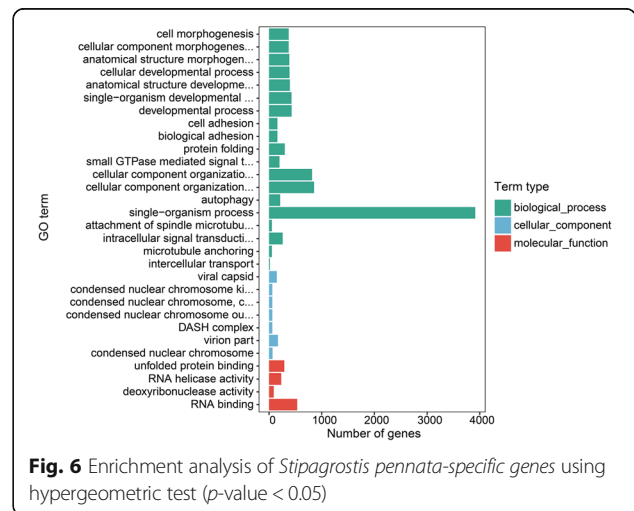
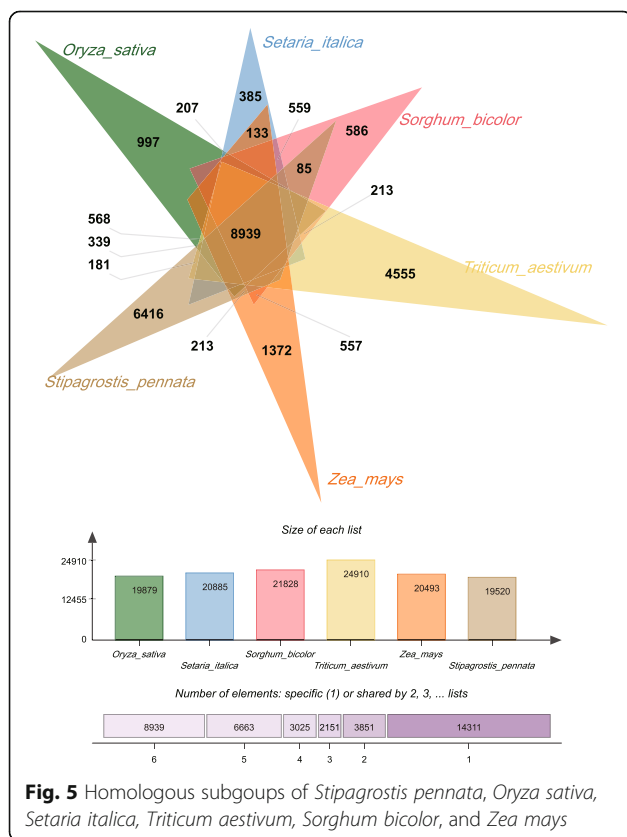
Fig. 3 An overview of main alternative splicing events in *Stipagrostis pennata* transcripts. The left panel shows the icon of the splicing event. AltA, alternate acceptor site; AltD, alternate donor site; ES, exon skip; RI, retained intron

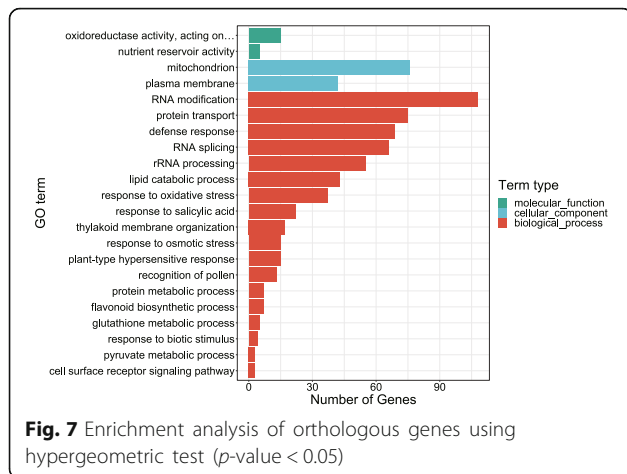


increased molecular evolutionary rate (ω) may be the key to understanding the adaptation of *Stipagrostis pennata* to the desert.

Although most genes are shared among all plants and are under purification selection, there are 38 *Stipagrostis pennata*-specific genes under positive selection (Fig. 9, Additional file 2). Some of these species-specific genes are related to transporter activity in the category of molecular function. They are involved in ion binding, heterocyclic compound binding, organic cyclic compound binding and protein binding. Furthermore, some genes are related to post-transcriptional modification and protein modification.

Rapidly evolving genes may play a role in the desert adaptation process. For example, the gene *i2_LQ_SP*





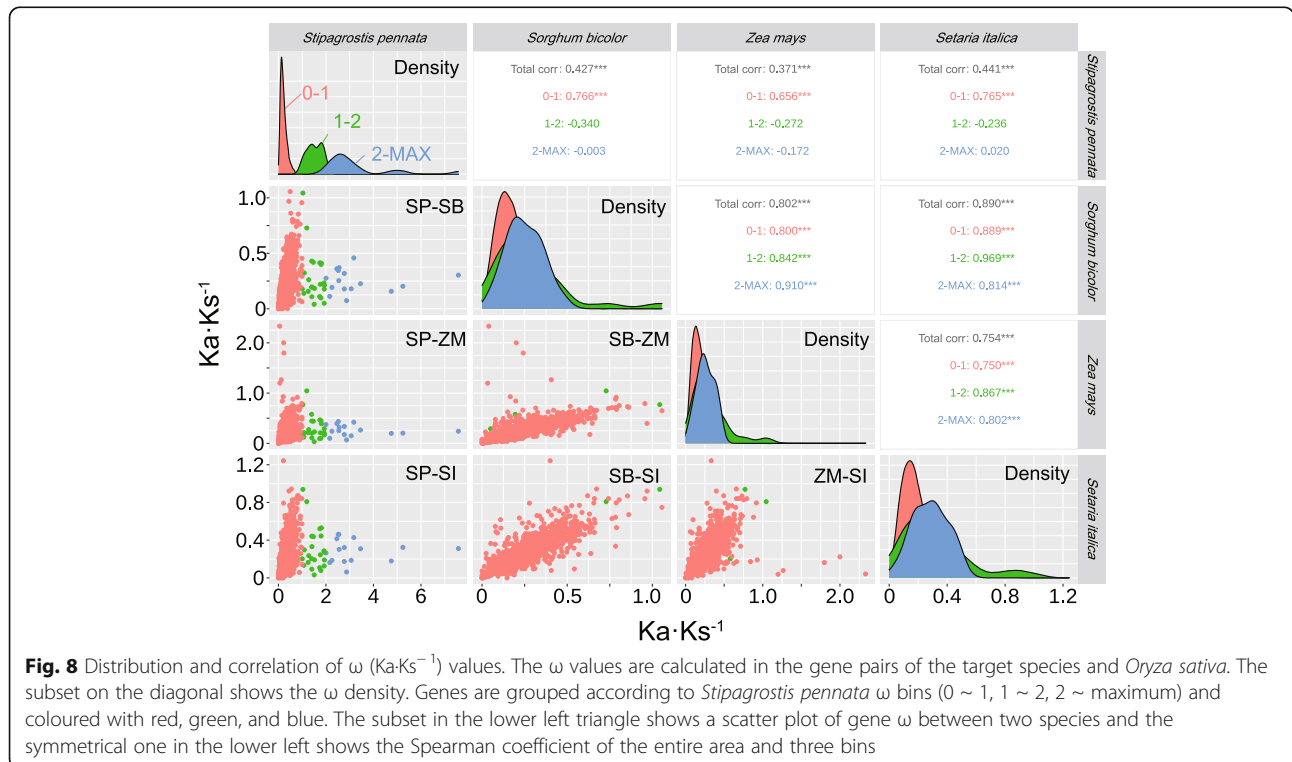
c5507/f1p1/2479 is a rapidly evolving gene of *Stipagrostis pennata*. This gene is a *CPP/TSO1-like* gene with two cysteine-rich CXC motifs. It plays a key role in plant reproductive tissue formation (including plant flower formation). We compared this gene with the genes of five other Gramineae species. In the CXC motif region, the sequences of the six species were generally consistent. However, in the region where the two CXC motifs join, *Stipagrostis pennata* has some differences in amino acid positions. As shown in Fig. 10A, in *Stipagrostis pennata*, arginine is replaced by proline and aspartic acid is replaced by glutamic acid. Due to the variation of these two loci, the structure of *Stipagrostis pennata* CPP/TSO1-like

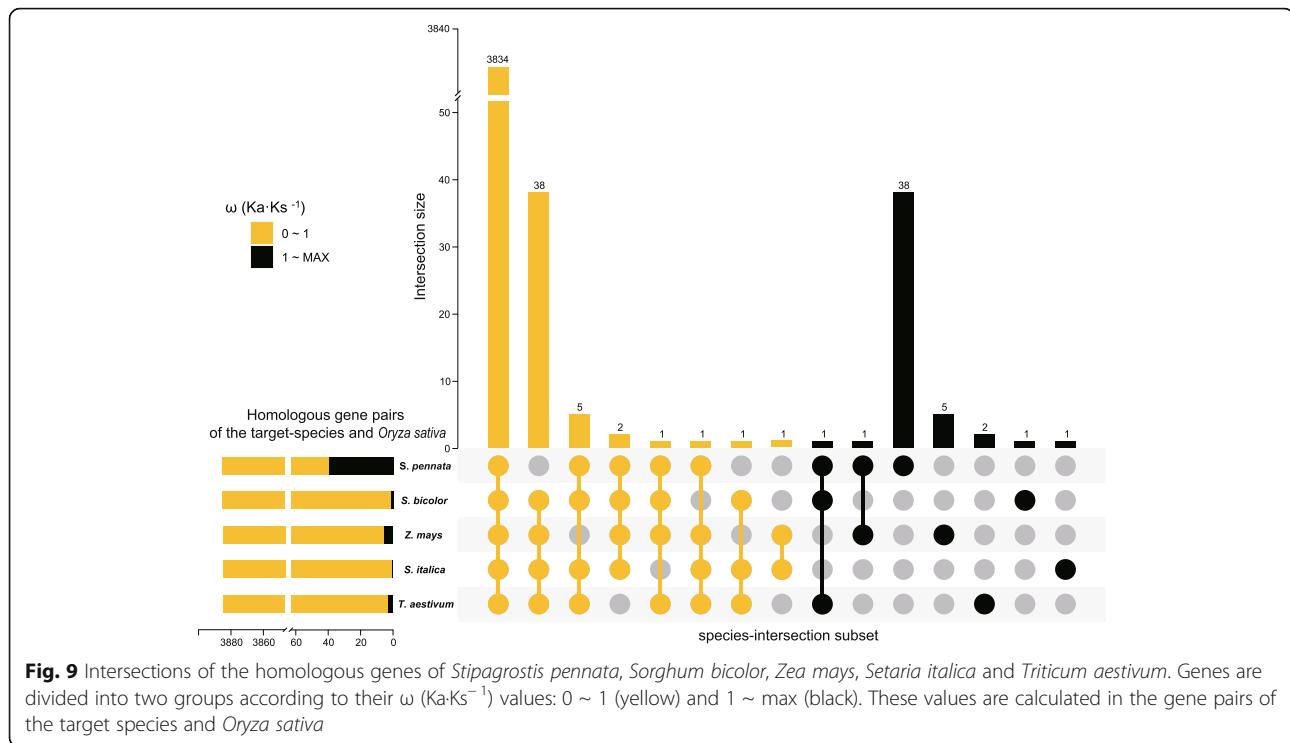
protein may be affected. Compared with the other 5 species, the three-dimensional structure of the *Stipagrostis pennata* protein is looser in the junction area between the CXC motifs. We hypothesize that the variant three-dimensional structure of *Stipagrostis pennata* CPP/TSO1-like protein plays a role in its adaptation to flowering and seed formation.

Discussion

Plants living in harsh environments have undergone significant adaptations during their long evolutionary history. Research on adaptive evolution can not only have a basic understanding of the evolutionary history of this plant, but also lay the foundation for genetic mining of features related to stress resistance [11–15]. *Stipagrostis pennata* is a pioneer plant that grows in the desert. It has formed a variety of resistance mechanisms, such as drought tolerance. However, the genetic adaptation mechanism of *Stipagrostis pennata* has not yet been elucidated. In the present study, we used PacBio SMRT technology and Illumina short-read sequencing to obtain the most complete full-length transcript of *Stipagrostis pennata* to date. These findings laid the foundation for future research on *Stipagrostis pennata*.

Plant adaptation is the result of a combination of gene expression regulation and function adaptive evolution [16, 17]. Gene expression regulation helps to improve the ability of plants to adapt to changes in the external environment [18–20]. Desert environments are harsh





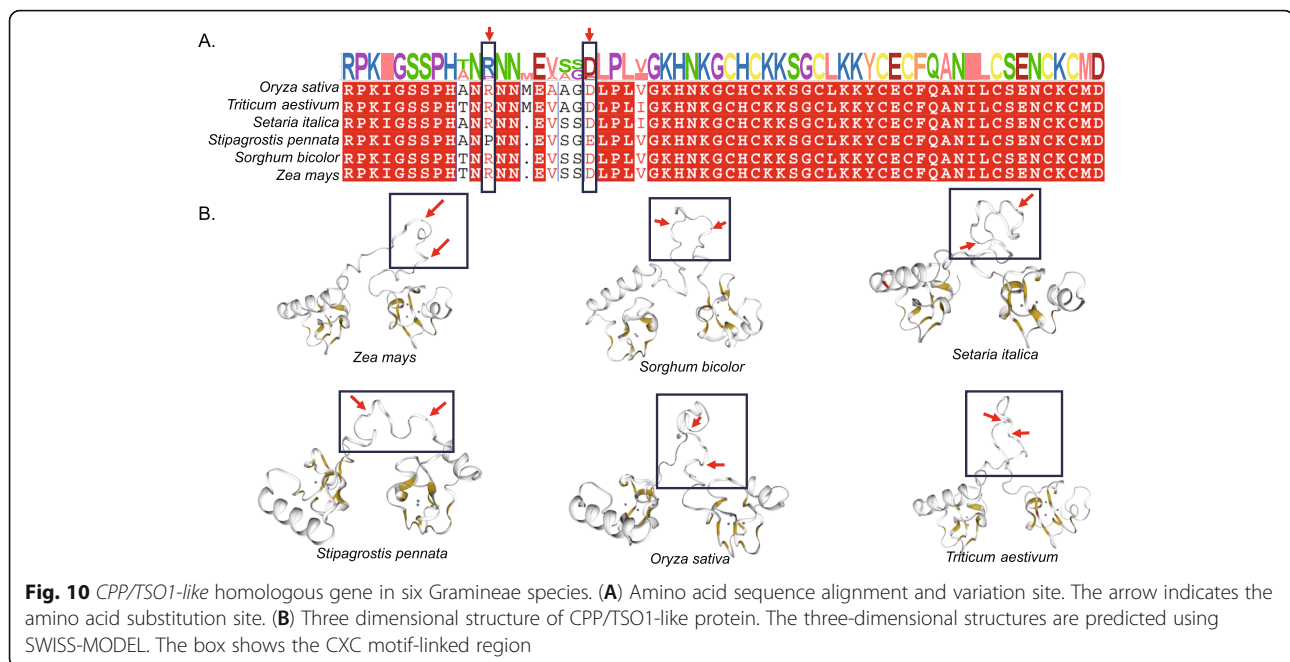
and varied. Any plant that survives in desert environments faces complex environmental stresses. Dynamic regulation of specific gene expression and biochemical processes is particularly important. In such dynamic process, transcription factors are one of the important regulatory components [21].

The 24 TFs, such as MYB-related, C3H, bHLH, and HSF genes, are abundant transcription factors in *Stipagrostis pennata*. These transcription factor families are closely related to abiotic stress responses [22–24]. *Stipagrostis pennata*, as a pioneer plant living in mobile dunes, has tough stem tissue that can withstand strong desert winds. This tissue may be the result of changes in gene expression related to lignin and cellulose synthesis [25]. The C3H is an important regulatory gene in lignin synthesis. It is the main transcription factor that regulates the formation of plant secondary walls [26, 27]. We propose that *Stipagrostis pennata* can improve wind erosion resistance by regulating the synthesis of cellulose and lignin. Interestingly, bHLH is an important transcription factor that can regulate drought resistance of plants such as *Arabidopsis thaliana* and *Triticum aestivum*. In addition, HSF is involved in the adaptation of plants to heat stress [28–32]. Furthermore, alternative splicing is another important mechanism for regulating plant gene expression. Intron retention events are abundant alternative splicing events in *Stipagrostis pennata*. Intron retention helps diversify plant gene functions, thereby enhancing the ability of plants to adapt to

the environment [33]. In *Stipagrostis pennata*, intron retention events may improve stress tolerance through flexible transcript splicing.

The adaptive evolution of gene expression is an important mechanism for plants to adapt to changes in the external environment. However, the range of gene expression is limited. Therefore, the adaptability of gene function has a greater impact in the evolution process [34]. The homologous genes, identified in *Stipagrostis pennata* and major cereal crops, are mainly enriched in GO terms for RNA modification and processing. *Stipagrostis pennata* has species-specific genes under positive selection pressure. The 38 rapidly evolving genes unique to *Stipagrostis pennata* may be involved in molecular translocation, binding and localization. In addition, the molecular functions, ion binding, heterocyclic compound binding, organic cyclic compound binding and protein binding of *Stipagrostis pennata* are all positively selected.

CPP/TSO1-like genes are involved in plant flowering and seed formation. In *Arabidopsis thaliana*, mutations in this gene lead to changes in flower structure and reduced pollen levels, and affect the development of ovules and carpels. The 3-D structure shows that the junction region between the two CXC motifs of the gene in *Stipagrostis pennata* is looser due to the variation of the two sites. This variant 3D structure may help *Stipagrostis pennata* adapt to the harsh desert environment.



Conclusions

Stipagrostis pennata grows in the desert and can adapt to harsh and extreme environments. It has a unique mechanism in the regulation of gene expression and the adaptive evolution of gene function. Most of its TFs is related to plant resistance regulation. IR events are main AS events in *Stipagrostis pennata*. The rapidly evolving genes may be related to molecular transport, flowering and seed formation. Finally, our findings may lay the foundation for future research on *Stipagrostis pennata*.

Methods

Plant materials

Fresh *Stipagrostis pennata* tissues were collected from the southern margin of the Gurbantungut Desert in Xinjiang in China. Botanist Ping Yan from Shihezi University identified the *Stipagrostis pennata*, and the voucher specimens (No. 6704) have been deposited in the the College of Life Sciences, Shihezi University China. In order to obtain more complete transcriptome information, we collected the roots, stems, leaves, flowers, and seedlings of *Stipagrostis pennata*. The research on the *Stipagrostis pennata* complies with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. To obtain as many expressed genes as possible, five different tissues (including leaves, stems, flowers, roots, and root sheaths) were sampled. Tissue samples were frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Arabidopsis thaliana, *Triticum aestivum*, *Oryza sativa*, *Setaria italica*, *Sorghum bicolor* and *Zea mays* genomes were downloaded from the Ensembl Plants database.

PacBio Iso-Seq library preparation

Total RNA were extracted from the samples. RNA degradation and contamination were analyzed using 1% agarose gel electrophoresis. RNA purity was measured using a Nanodrop ND-1000 spectrophotometer (Nano Drop Technologies, Wilmington, DE, USA). Precise quantification of RNA concentrations was carried out using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). RNA integrity was measured using Agilent 2100 technology (Agilent Technologies, Palo Alto, CA, USA).

The Iso-Seq library was prepared according to the Iso-form Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit. The BluePippin Size Selection System protocol was performed as described by Pacific Biosciences (PN 100-092-800-03). Libraries were subsequently sequenced using a Pacific Biosciences sequencing machine.

Illumina cDNA library preparation

A total of 1.5 μg RNA per sample was used for cDNA library preparation. Sequencing libraries were generated using a NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] according to the manufacturer's recommendations. Index codes were added to attribute identification sequences to each sample. Library preparations were sequenced on an Illumina HiSeq 2500 platform (Illumina,

San Diego, CA, USA). Sequencing was performed in a paired-end manner. All sequencing was performed in a high-throughput manner at the Novogene Bioinformatics Institute (Novogene, Beijing, China).

Functional annotation and selection pressure analysis

The Blast2GO program [35] (<http://www.blast2go.com>) was used to annotate GO terms (<http://www.geneontology.org>) based on the NR annotations. A cutoff E-value of $\leq 1e-6$ was employed. All enrichment analyses were performed using hypergeometric test (p -value < 0.05). Plant transcription factors were predicted using iTAK [36] software and TFDB 2.0 database [37]. The transcription factor annotations for *Triticum aestivum*, *Zea mays* and *Sorghum bicolor* were downloaded from iTAK database. We used CNCI [38], CPC [39], Pfam-scan [40], and PLEK [41] to predict coding sequences [42]. AStrap [43] was used to identify alternative splicing events in *Stipagrostis pennata*. Homologous genes were screened using OrthoVene2 [44]. Synonymous substitutions (Ka) and non-synonymous substitutions (Ks) were calculated using KaKs_calculator 2.0 [45].

Abbreviations

AS: Alternative splicing; BP: Biological process; CC: Cellular component; ES: Exon skipping; FLNC: Full-length non-chimeric read sequences; GO: Gene Ontology; IR: Intron retention; Ka: The number of nonsynonymous substitutions per non-synonymous site; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: Eukaryotic Ortholog Groups; Ks: The number of synonymous substitutions per synonymous site; MF: Molecular function; NR: NCBI non-redundant protein; NT: NCBI nucleotide sequences; Pfam: Protein family; SGS: Second-generation sequencing; SMRT: Single-molecule real-time; TF: Transcription factor; TGS: Third-generation sequencing

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08124-w>.

Additional file 1

Additional file 2

Acknowledgements

We thank the anonymous reviewers for helpful suggestions on the manuscript.

Authors' contributions

LM and TZ conceived the project and research plans. XD performed the experiment and analysis, and wrote the article with contributions of all the authors. LM and TZ supervised and complemented the writing. All authors read and approved the final manuscript.

Funding

The research was supported by the National Natural Science Foundation of China (31760302, 32060145, 32060300 and 31560310) and the Science Foundation of Shihezi University (RCZK201953). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The RNA sequencing data used in this study have been uploaded to the National Center for Biotechnology Information (NCBI) as Bioproject ID: PRJNA763703.

Declarations

Ethics approval and consent to participate

The research on the *Stipagrostis pennata* comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 19 March 2021 Accepted: 26 October 2021

Published online: 23 November 2021

References

- Shariat A, Mirzaei Nadoushan H, Madah A: Karyotypic Investigation No Several Species of *Stipagrostis*. *Iranian Journal of Rangelands and Forests Plants Breeding and Genetic Research*. 2003;9:25–38.
- Jha UC, Bohra A, Singh NP: Heat stress in crop plants: its nature, impacts and integrated breeding strategies to improve heat tolerance. *Plant Breeding*. 2014;133(6):679–701.
- Wahid A, Gelani S, Ashraf M, Fool Ad MR: Heat tolerance in plants: An overview. *Environ Exper Botany*. 2007;61(3):199–223.
- Barnabás B, Jäger K, Fehér A: The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ*. 2010;31(1):11–38.
- Golldack, Luking, Yang: Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep*. 2011;30(8):1383–91.
- Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M: KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*. 2021;49(D1):D545–51.
- Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
- Kanehisa M: Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. 2019;28(11):1947–51.
- Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P: Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC bioinformatics*: 2009: BioMed Central; 2009;1–15.
- Scharf KD, Berberich T, Ebersberger I, Nover L: The plant heat stress transcription factor (HSF) family: structure, function and evolution. *Biochim Biophys Acta*. 2012;1819(2):104–19.
- Ma J, Lu J, Xu J, Duan B, He X, Liu J: Genome-wide identification of WRKY genes in the desert poplar *Populus euphratica* and adaptive evolution of the genes in response to salt stress. *Evolu Bioinform*. 2015;11:EBO S22067.
- Shi Y, Yan X, Zhao P, Yin H, Zhao X, Xiao H, Li X, Chen G, Ma X-F: Transcriptomic analysis of a tertiary relict plant, extreme xerophyte *Reaumuria soongorica* to identify genes related to drought adaptation. *PLoS One*. 2013;8(5):e63993.
- Zhang C-H, Ma T, Luo W-C, Xu J-M, Liu J-Q, Wan D-S: Identification of 4CL genes in desert poplars and their changes in expression in response to salt stress. *Genes*. 2015;6(3):901–17.
- Zhang J, Feng J, Lu J, Yang Y, Zhang X, Wan D, Liu J: Transcriptome differences between two sister desert poplar species under salt stress. *BMC Genom*. 2014;15(1):337.
- Zhang J, Xie P, Lascoux M, Meagher TR, Liu J: Rapidly evolving genes and stress adaptation of two desert poplars, *Populus euphratica* and *P. pruinosa*. *PLoS One*. 2013;8(6):e66370.
- Wright SI, Gaut BS: Molecular Population Genetics and the Search for Adaptive Evolution in Plants. *Mol Biol Evol*. 2004;22(3):506–19.
- Hough J, Williamson RJ, Wright SI: Patterns of Selection in Plant Genomes. *Annual Review of Ecology, Evolution, and Systematics*. 2013;44(1):31–49.
- Cassone BJ, Kamdem C, Cheng C, Tan JC, Hahn MW, Costantini C, Besansky NJ: Gene expression divergence between malaria vector sibling species *Anopheles gambiae* and *An. coluzzii* from rural and urban Yaoundé Cameroon. *Mol Ecol*. 2014;23(9):2242–59.
- Davidson EH: The regulatory genome: gene regulatory networks in development and evolution. London: Elsevier; 2010.

20. Jiang S-Y, Ma A, Ramamoorthy R, Ramachandran S. Genome-wide survey on genomic variation, expression divergence, and evolution in two contrasting rice genotypes under high salinity stress. *Genom Biol Evol.* 2013;5(11):2032–50.
21. Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al. *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science.* 2000; 290(5499):2105–10.
22. Du H, Wang Y-B, Xie Y, Liang Z, Jiang S-J, Zhang S-S, Huang Y-B, Tang Y-X. Genome-Wide Identification and Evolutionary and Expression Analyses of MYB-Related Genes in Land Plants. *DNA Res.* 2013;20(5):437–48.
23. Gomez E, Royo J, Muniz LM, Sellam O, Paul W, Gerentes D, Barrero C, Lopez M, Perez P, Hueros G. The Maize Transcription Factor Myb-Related Protein-1 Is a Key Regulator of the Differentiation of Transfer Cells. *Plant Cell.* 2009; 21(7):2022–35.
24. Xiong H, Li J, Liu P, Duan J, Zhao Y, Guo X, Li Y, Zhang H, Ali J, Li Z: Overexpression of OsMYB48-1, a Novel MYB-Related Transcription Factor, Enhances Drought and Salinity Tolerance in Rice. *PLOS ONE.* 2014;9(3): e92913.
25. Miri A, Dragovich D, Dong Z. Vegetation morphologic and aerodynamic characteristics reduce aeolian erosion. *Sci Rep.* 2017;7(1):12831–831.
26. Liu C, Xu X, Kan J, ming Cheng Z, Chang Y, Lin J, Li H: Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of *Pyrus betulaefolia*. *PeerJ.* 2020;8:e9328.
27. Zhou X, Yang S, Lu M, Zhao S, Cai L, Zhang Y, Zhao R, Lv J. Structure and Monomer Ratio of Lignin in C3H and HCT RNAi Transgenic Poplar Saplings. *Chemistryselect.* 2020;5(24):7164–9.
28. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell.* 2003;15(1):63–78.
29. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 2011;66(1):94–116.
30. Never L, Scharf KD, Gagliardi D, Vergne P, CzarneckaVerner E, Gurley WB. The HSF world: Classification and properties of plant heat stress transcription factors. *Cell Stress Chaperones.* 1996;1(4):215–23.
31. Scharf KD, Rose S, Zott W, Schoffl F, Nover L. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J.* 1990;9(13):4495–501.
32. Scharf K-D, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (HSF) family: Structure, function and evolution. *Biochim Biophys Acta.* 2012;1819(2):104–19.
33. Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R. Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J.* 2004;39(6):877–85.
34. Zhang J, Wu F, Yan Q, John UP, Cao M, Xu P, Zhang Z, Ma T, Zong X, Li J: The genome of *Cleistogenes songorica* provides a blueprint for functional dissection of dimorphic flower differentiation and drought adaptability. *Plant biotechnology journal.* 2021;19(3):532.
35. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–6.
36. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant.* 2016;9(12):1667–70.
37. He Z, Jin J, Liang T, Yi Z, Gu X, Ge G, Luo J. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res.* 2011;39(Database issue):D1114–7.
38. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 2013;41(17):e166.
39. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 2007;35(suppl_2):W345–9.
40. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44(D1):D279–85.
41. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics.* 2014;15(1):311.
42. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–5.
43. Ji G, Ye W, Su Y, Chen M, Huang G, Wu X. AStrap: identification of alternative splicing from transcript sequences without a reference genome. *Bioinformatics.* 2019;35(15):2654–6.
44. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 2019;47(W1):W52–8.
45. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinform.* 2010;8(1):77–80.

Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

