

RESEARCH

Open Access



New insights into the cold tolerance of upland switchgrass by integrating a haplotype-resolved genome and multi-omics analysis

Bingchao Wu^{1†}, Dan Luo^{1†}, Yuesen Yue², Haidong Yan¹, Min He³, Xixi Ma¹, Bingyu Zhao⁴, Bin Xu⁵, Jie Zhu¹, Jing Wang⁶, Jiyuan Jia¹, Min Sun^{1,7}, Zheni Xie¹, Xiaoshan Wang¹ and Linkai Huang^{1*}

[†]Bingchao Wu and Dan Luo contributed equally to this work.

*Correspondence: huanglinkai@sicau.edu.cn

¹ College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu 611130, China

² Institute of Grassland, Flower and Ecology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

³ State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu 611130, China

⁴ College of Agriculture and Life Sciences, Virginia Tech, Blacksburg, VA 24061, USA

⁵ College of Grassland Science, Nanjing Agricultural University, Nanjing 210095, China

⁶ Key Laboratory for Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610061, China

⁷ Institute of Advanced Study, Chengdu University, Chengdu 610106, China

Abstract

Background: Switchgrass (*Panicum virgatum* L.) is a bioenergy and forage crop. Upland switchgrass exhibits superior cold tolerance compared to the lowland ecotype, but the underlying molecular mechanisms remain unclear.

Results: Here, we present a high-quality haplotype-resolved genome of the upland ecotype “Jingji31.” We then conduct multi-omics analysis to explore the mechanism underlying its cold tolerance. By comparative transcriptome analysis of the upland and lowland ecotypes, we identify many genes with ecotype-specific differential expression, particularly members of the cold-responsive (*COR*) gene family, under cold stress. Notably, *AFB1*, *ATL80*, *HOS10*, and *STRS2* gene families show opposite expression changes between the two ecotypes. Based on the haplotype-resolved genome of “Jingji31,” we detect more cold-induced allele-specific expression genes in the upland ecotype than in the lowland ecotype, and these genes are significantly enriched in the *COR* gene family. By genome-wide association study, we detect an association signal related to the overwintering rate, which overlaps with a selective sweep region containing a cytochrome P450 gene highly expressed under cold stress. Heterologous overexpression of this gene in rice alleviates leaf chlorosis and wilting under cold stress. We also verify that expression of this gene is suppressed by a structural variation in the promoter region.

Conclusions: Based on the high-quality haplotype-resolved genome and multi-omics analysis of upland switchgrass, we characterize candidate genes responsible for cold tolerance. This study advances our understanding of plant cold tolerance, which provides crop breeding for improved cold tolerance.

Keywords: Upland switchgrass, Haplotype-resolved genomes, Allele-specific expression genes, Cold tolerance



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Background

Switchgrass (*Panicum virgatum* L.) is a perennial C4 grass utilized as both a forage crop and a dedicated feedstock for bioenergy production [1, 2]. It comprises two distinct ecotypes, lowland and upland, displaying substantial variations in morphology and environmental adaptability [3, 4]. The lowland ecotype typically thrives in warm, moist environments, featuring greater plant height and broader leaves. In contrast, the upland ecotype predominantly inhabits cold, arid areas and is able to overwinter in colder temperate zones [5]. The upland ecotype likely contains gene resources conferring cold tolerance that differ from those in the lowland ecotype.

Due to the frequent occurrence of extreme weather events caused by global climate change, stable crop production faces significant challenges. Cold stress is one of the most threatening abiotic stresses in the growth and development of plants, impacting the geographical distribution of plants and even leading to plant mortality, thereby causing a decrease in crop yield [6]. Additionally, planting switchgrass on marginal lands is an effective way to increase its cultivation area, but these lands often face various abiotic stresses, especially the threat of low temperature. Upland switchgrass represents an ideal model for understanding how plants respond to cold stress, as it can successfully overwinter in northern cold regions compared to lowland ecotypes. However, few studies have explored the molecular mechanisms underlying the regulation of cold stress responses in the upland ecotype relative to the lowland ecotype, and these mechanisms remain largely uncharacterized.

For most plants with an open-pollination mechanism, the improvement of offspring traits typically results from increased genetic variation and specific expression of allelic genes at certain loci [7]. The time and space-specific expression of different allelic genes can result in significant differences in gene products and lead to distinct phenotypes [8, 9]. Traditional de novo assembly algorithms for most species represent samples as haploid genomes. For highly heterozygous switchgrass, the previously reported collapsed representation may have overlooked half of the heterozygous variants in the genome [10]. It could have introduced assembly errors in regions of divergence between haplotypes. The construction of a haplotype-resolved genome assembly for switchgrass will provide a complete view of the genome and its complex genetic variations, facilitating a comprehensive understanding of allele-specific expression (ASE) genes.

In this study, we uncovered the cold tolerance mechanism of upland switchgrass by constructing a high quality and haplotype-resolved reference genome of upland switchgrass “Jingji31” along with transcriptomic, population genetic, and functional validation assays. We identified a large number of cold-responsive (*COR*) genes with opposite expression trends in upland and lowland switchgrass under cold stress, which may contribute to the cold tolerance differences between the two ecotypes. Additionally, ASE genes were widely present in switchgrass, particularly certain *COR* genes, whose ASE was induced by cold stress. The genome-wide association study (GWAS) and selective sweep analysis identified a large number of candidate genes potentially associated with cold tolerance, among which the overexpression of one candidate gene was found to positively regulate cold tolerance. Our findings not only improve the understanding of cold tolerance in upland switchgrass, which may accelerate genome-assisted breeding of

cold tolerance in this important model energy plant, but also hold promise for promoting comparative genomics studies in other crops.

Results

Assembly and annotation of the upland switchgrass genome

The genome size of the switchgrass upland ecotype “Jingji31” ($2n = 4x = 36$, abbreviated as “JJ31”) was estimated to be ~ 1.19 Gb using k-mer analysis based on 67.8 Gb ($57 \times$ coverage) of Illumina short-read data (Fig. 1a, Additional file 2: Table S1-2). In addition, 53.2 Gb ($44.7 \times$ coverage) of PacBio high-fidelity long read (HiFi) sequences and 145.2 Gb ($122 \times$ coverage) of Illumina-sequenced Hi-C data were generated (Additional file 2: Table S1). Due to the high heterozygosity (1.55%), Hifiasm was used for haplotype-resolved de novo assembly using phased assembly graphs [11]. Two phased haplotypes, “JJ31-A” and “JJ31-B,” were anchored to pseudo-chromosomes using Hi-C reads. The Hi-C interaction map demonstrated that our chromosome-level anchoring was of high quality and reliability (Fig. 1b). Furthermore, through co-linearity analysis with the previously published lowland ecotype switchgrass “AP13” genome, all pseudo-chromosomes of the two haplotypes were successfully assigned to different subgenomes, bearing the same chromosome IDs “K” and “N” [10] (Fig. 1c, Additional file 1: Fig. S1). Ultimately, the genome sizes of “JJ31-A” and “JJ31-B” were determined to be 1.14 Gb and 1.11 Gb, respectively, with 98.47% (“JJ31-A”) and 99.02% (“JJ31-B”) of sequences

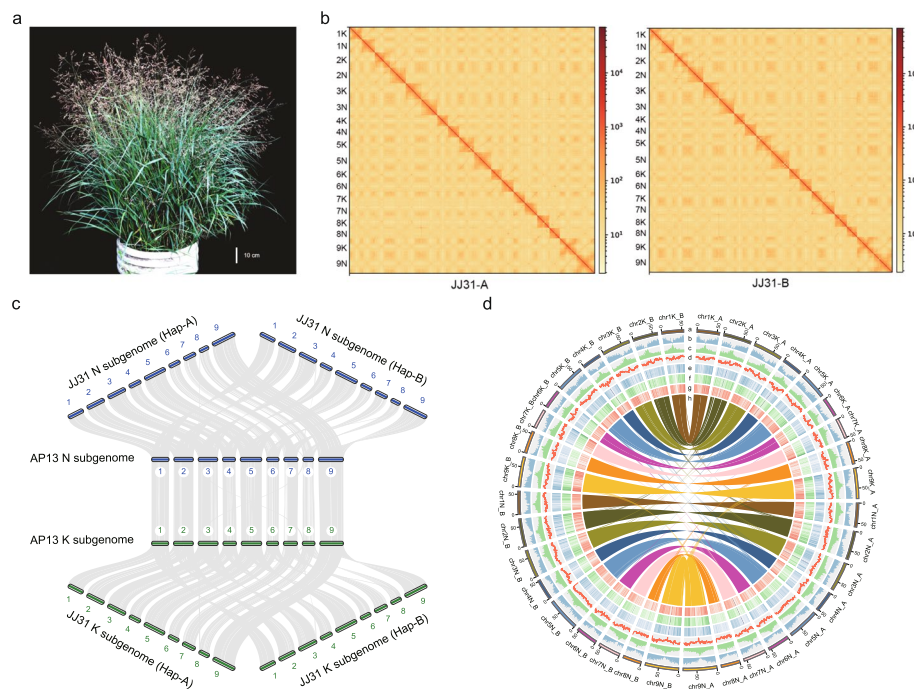


Fig. 1 High-quality haplotype-resolved genome assembly of upland switchgrass JJ31. **a** Flowering morphology of the upland ecotype JJ31. **b** Whole genome Hi-C heat map for “JJ31-A” (left) and “JJ31-B” (right). **c** The chromosome collinearity between the genomes of JJ31 and Alamo. Numbers represent chromosome identifiers. **d** Genome features of the JJ31 genome. Track “a”: chromosome length; track “b”: gene density; track “c”: transposable element (TE) density; track “d”: GC content; track “e–g”: gene expression levels in roots, stems, and leaves; track “h”: collinearity between chromosomes of “JJ31-A” and “JJ31-B”

anchored across 18 chromosomes (Fig. 1d, Table 1, Additional file 2: Table S3-4). Their respective Contig N50 values were 4.9 times and 4.7 times higher than that of the “AP13” genome (5.5 Mb) [10] (Table 1).

We then employed various strategies to assess the quality of the haplotype-resolved genomes. By realigning paired-end reads to the two haploid genomes, we observed alignment rates of 98.07 and 97.93%, respectively (Additional file 2: Table S3). Furthermore, the embryophyta Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis indicated completeness rates of 98.4 and 98.1% for the two haplotypes (Additional file 2: Table S3). Using Merqury to calculate quality values (QV), “JJ31-A” and “JJ31-B” achieved values of 45.16 and 50.27, respectively, exceeding the Vertebrate Genomes Project standard of QV40 [12] (Table 1). The long terminal repeat (LTR) assembly index (LAI) for both phased haplotype genomes approached 20, nearly meeting the golden standard [13] (Additional file 2: Table S3). These results affirm the accuracy, completeness, and contiguity of our two haploid genome assemblies.

Both haplotype-resolved genomes were annotated with a comprehensive strategy that combined homolog prediction, de novo prediction, and other evidence-driven predictions. In the two haplotypes, 79,672 and 79,416 protein-coding genes were predicted, with 98.9% supported by known gene function databases (Table 1 and Additional file 2: Table S3). The gene models of the two haploid genomes had an average coding sequence length of ~1 kb and an average of four exons per gene (Additional file 2: Table S3). We also identified 0.62 and 0.59 Gb repeat sequences, accounting for 54.19 and 53.48% of the two haplotypes, respectively, of which 42.09 and 41.51% were LTRs (Additional file 2: Table S3). Additionally, we identified 2245 and 2206 miRNAs, 1264 and 1291 tRNAs, 9171 and 5221 rRNAs, and 1190 and 1154 snRNAs in the two haplotype genomes, respectively (Additional file 2: Table S3). Considering the overall superior quality of the “JJ31-B” haplotype, it was selected for subsequent analysis unless otherwise stated.

Enhanced cold tolerance and subgenome dominance in upland switchgrass indicated by comparative genomics analysis

To understand the evolutionary relationship between the two ecotypes of switchgrass, we performed a comparative genomic analysis that included several closely related species. The phylogenetic tree showed that the K and N subgenomes of switchgrass diverged approximately 6.6 million years ago (Mya), while the divergence

Table 1 Summary of assembly and annotation of two haplotype genomes in upland switchgrass

| | JJ31-A | JJ31-B |
|-------------------------------|---------|---------|
| Contig N50 (Mb) | 27.05 | 25.98 |
| Contig length (Mb) | 1124.57 | 1136.95 |
| Scaffold N50 (Mb) | 67.96 | 62.02 |
| Scaffold length (Mb) | 1143.31 | 1110.80 |
| Chromosome anchoring rate (%) | 98.47 | 99.02 |
| Gene no | 79,672 | 79,416 |
| Repeat sequence length (Mb) | 619.58 | 594.02 |
| Repeat ratio (%) | 54.19 | 53.48 |
| QV | 45.16 | 50.27 |

between switchgrass and *P. hallii* occurred around 8.8 Mya, consistent with previous studies [10] (Fig. 2a). Notably, the K subgenomes of the two ecotypes diverged at 2.3 Mya, and the N subgenomes at 2.6 Mya (Fig. 2a), suggesting that ecotype divergence occurred during this period. We further determined the divergence time between the two ecotypes to be ~2.2 Mya (Ks peak at 0.03) by calculating the synonymous substitution rate (Ks) of orthologous gene sets between the different subgenomes of the two ecotypes (Fig. 2b, Additional file 2: Table S5). Combining our findings with the previous report on the tetraploidization time of switchgrass (< 4.6 Mya) [10], we propose an evolutionary scenario in which the N and K subgenomes diverged first, followed by tetraploidization and later ecotype differentiation.

It was reported that upland switchgrass exhibits greater cold tolerance than the lowland ecotype [14–16]. Compared to the “AP13,” a total of 4147 expanded, 5030 positively selected, and 4873 specific genes were identified in “JJ31” (likelihood ratio test, $P < 0.05$). These genes were enriched in several stress-responsive pathways and biological processes (Additional file 1: Fig. S2). The expanded and specific genes in “JJ31” were primarily enriched in calcium channel activity, calcium ion transmembrane transporter activity, melatonin receptor activity, and sphingolipid metabolic pathways, which are believed to play crucial roles in protecting plants from cold stress [17–21]. These positively selected specific genes were associated with G-protein coupled receptor activity, which is involved in signal transduction of plant stress

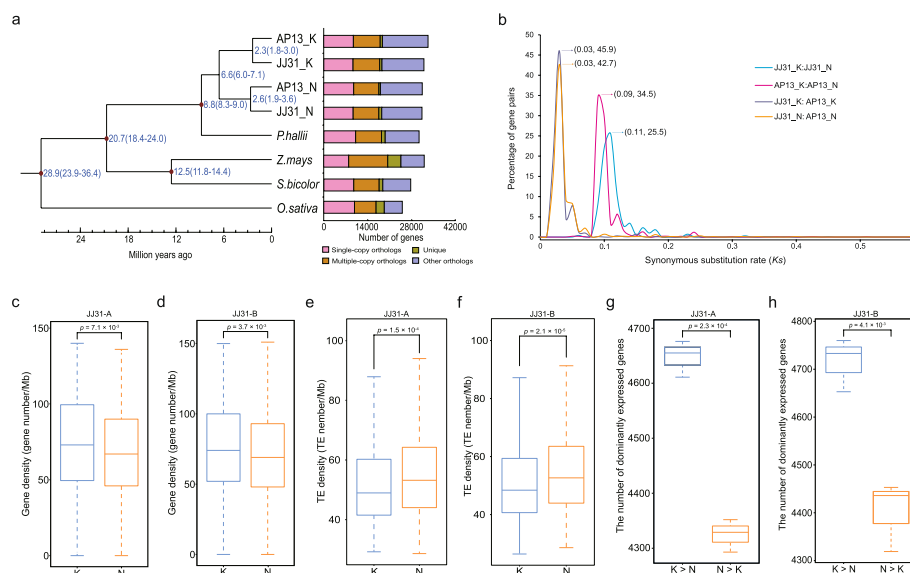


Fig. 2 Comparative genomics and subgenome dominance analysis. **a** Phylogenetic tree and divergence time estimates of “JJ31” and five closely related species. The right panel shows the distribution of single-copy, multi-copy, unique, and other gene orthologs. **b** Evolutionary analysis of the “JJ31” and “AP13.” The Ks distribution is shown for orthologs in the switchgrass genomes. K and N represent two subgenomes, respectively. The parentheses contain the X-axis and Y-axis values corresponding to the peaks. **c–h** Assessment of subgenomic dominance in two haplotype genomes. Gene density of the two subgenomes in the “JJ31-A” (c) and “JJ31-B” (d). TE density of the two subgenomes in the “JJ31-A” (e) and “JJ31-B” (f). Number of dominantly expressed genes in the two subgenomes of the “JJ31-A” (g) and “JJ31-B” (h). The statistical window size is 1 Mb

responses [22]. The above results may help explain the enhanced cold tolerance of upland switchgrass compared to the lowland ecotype.

Subgenome dominance is a common phenomenon in polyploid plants, prompting our investigation into the subgenome characteristics within the two haplotypes of upland switchgrass. We applied a sliding window approach (window size 1 Mb) to evaluate gene and transposable element (TE) densities across the two subgenomes. Compared to the N subgenomes in “JJ31-A” and “JJ31-B,” the K subgenomes exhibited higher gene density (71.2 versus 66.2 genes per Mb and 73 versus 67.6 genes per Mb, $P < 0.01$), more genes with dominant expression (4662 versus 4305 and 4715 versus 4403, $P < 0.01$), and less TE density (53.5 versus 56.4 TEs per Mb and 52.5 versus 55.7 TEs per Mb, $P < 0.01$) (Fig. 2c–h, Additional file 2: Table S6–9). Additionally, the cold stress transcriptome data from both ecotypes revealed slightly more DEGs on the K subgenome (Additional file 2: Table S10). In summary, all metrics indicated that the K subgenome is dominant in both haplotypes, consistent with observations in the “AP13” genome [10].

Contribution of specific differential expression of COR gene families to the cold tolerance differences between the two ecotypes

The upland ecotype “JJ31” exhibited greater cold tolerance than the lowland ecotype “Alamo” (noting that “AP13” was a line/clone selected from “Alamo”) [14] based on the phenotypic and physiological indicators under cold stress (Fig. 3a–c). Leaves of both ecotypes showed significant wilting after 56 days of cold treatment, but only “JJ31” regrew new leaves after returning to room temperature (Fig. 3a). In addition, relative water content (RWC), relative electrical conductivity (REC), and malondialdehyde (MDA) showed significant changes ($P < 0.05$) in “JJ31” after 21 days of cold stress compared to the control, while in “Alamo,” RWC significantly decreased and REC and MDA significantly increased after 14 days of cold stress compared to the control ($P < 0.05$) (Fig. 3b,c). The slower physiological changes under cold stress might indicate better cold tolerance in “JJ31” than in “Alamo.”

To understand the molecular response mechanisms underlying the greater cold tolerance of upland switchgrass, we conducted transcriptome sequencing on the leaves and roots of “JJ31” and “Alamo” at three time points under cold stress (Additional file 2: Table S1). Based on KEGG enrichment analysis, the DEGs identified from each comparison group were mainly enriched in circadian rhythm, MAPK signaling pathway, and plant hormone signal transduction pathway (Fig. 3d, Additional file 1: Fig. S3). Although both ecotypes relied on similar pathways to respond to cold stress, we found that genes related to cold tolerance within these pathways exhibited ecotype-specific expression (Fig. 3e). During the initial exposure of plants to cold stress, Ca^{2+} influx induces the activity of calmodulin (CaM), which activates downstream MEKK1 to positively regulate plant cold tolerance [23, 24]. Our study found that three members of the *CaM* family and one member of the *MEKK1* family were specifically upregulated in “JJ31,” while two and one members of the respective families were specifically downregulated in “Alamo” (Fig. 3e). Two transcription factors, PIF3 and PIF4, are known as negative regulators in plant cold tolerance by inhibiting the expression of *CBF*, while the EIN3-BINDING F-BOX 1/2 (EBF1/2) proteins enhance cold tolerance by degrading PIF3 [25, 26]. We found that nine *PIF3*

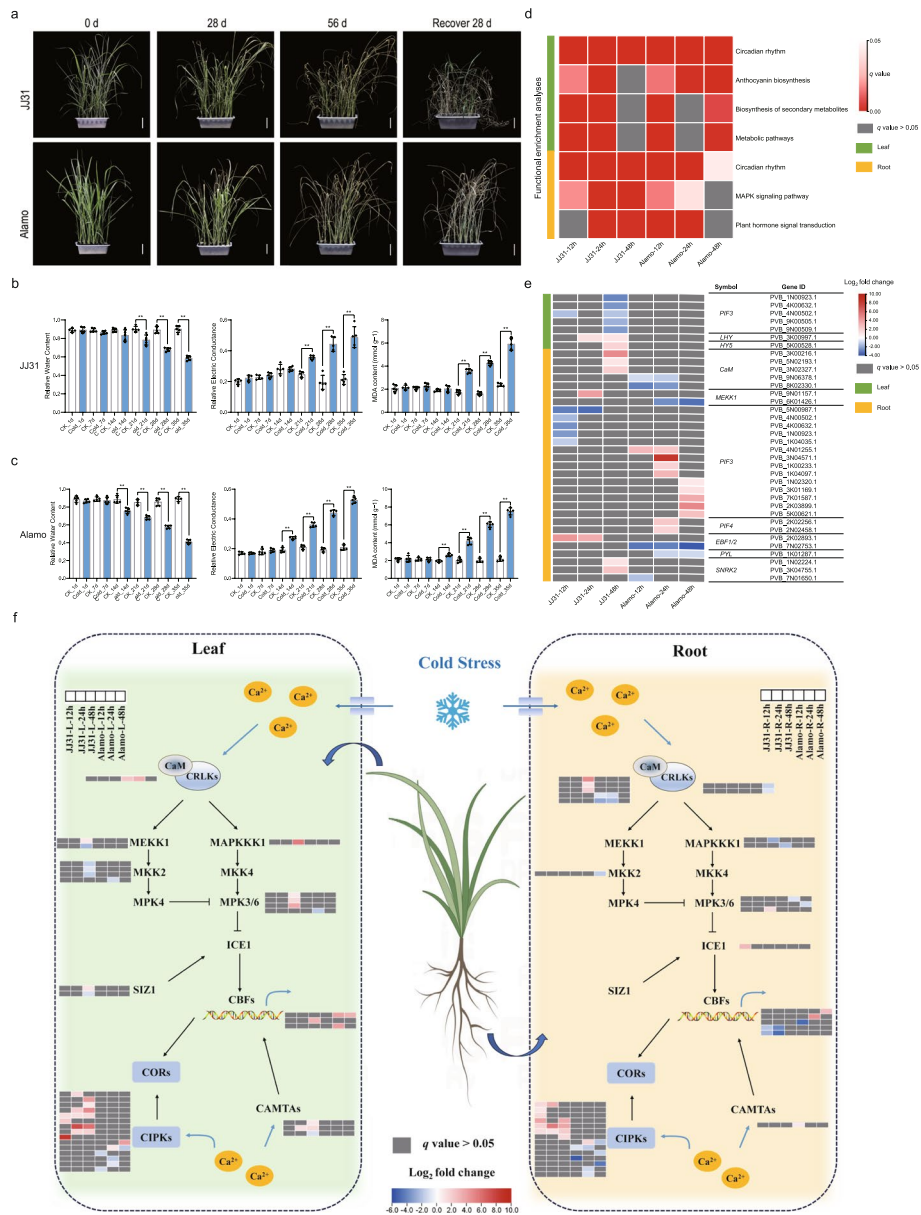


Fig. 3 Transcriptional landscape differences between the two ecotypes under cold stress. **a** Phenotypic changes of JJ31 and Alamo on days 0, 28, and 56 under cold stress at 4 °C and recover at room temperature for 28 days. Scale bar indicates 7 cm. **b** Changes in physiological indicators of JJ31 under control (room temperature) and cold stress (4 °C). From left to right: RWC, REC, and MDA. ** indicates $P < 0.005$. **c** Changes in physiological indicators of Alamo under control (room temperature) and cold stress (4 °C). From left to right: RWC, REC, and MDA. ** indicates $P < 0.005$. **d** KEGG enrichment of all compared DEGs identified in leaves and roots of JJ31 and Alamo under cold stress. **e** Expression changes of cold-tolerance-related genes in the circadian rhythm, MAPK signaling pathway, and plant hormone signaling transduction pathways in two ecotypes. **f** Expression changes of key genes in the CBF-dependent cold response pathway in the two ecotypes

genes were specifically downregulated in “JJ31”, while nine *PIF3* and two *PIF4* genes were specifically upregulated in “Alamo” (Fig. 3e). The two *EBF1/2* genes also exhibited opposite expression changes between the two ecotypes (Fig. 3e). Additionally, one *LHY*, one *HYS*, and two *SNRK2* genes were specifically upregulated in “JJ31,”

while one *SNRK2* and one *PYL* were specifically downregulated in “Alamo” (Fig. 3e). These genes have been reported to positively regulate plant cold tolerance [27–30]. These results suggest that the ecotype-specific expression of these cold stress regulatory genes might explain differences in cold tolerance between the two ecotypes.

The above results imply that there may be more cold-response (*COR*) genes with ecotype-specific expression. We identified 795 *COR* genes in switchgrass based on the 109 *COR* gene families reported in *Arabidopsis* [31] (Additional file 2: Table S11). We identified 182 and 251 specific DEGs significantly enriched in *COR* genes in “JJ31” and “Alamo,” respectively ($P = 0.035$ and $P = 1.0745e - 5$, Additional file 1: Fig. S4, Additional file 2: Table S12-13 and Additional file 3: Note S1). These *COR* genes belonged to 58 and 55 families, respectively, among which members of 25 families showed specifically differential expression in the leaves or roots of only one ecotype (Additional file 1: Fig. S5a). The members of 44 *COR* gene families exhibited differential expression in both ecotypes, where some of these genes showed opposite expression changes between the two ecotypes (Additional file 1: Fig. S5b).

The auxin signaling F-box protein 1 (*AFB1*)-mediated auxin signaling pathway is involved in plant tolerance to abiotic stress, and its overexpression can enhance plant tolerance to salt and cold stresses [32]. Three *AFB1* genes were upregulated in “JJ31,” while four *AFB1* genes were downregulated in “Alamo” (Additional file 1: Fig. S5b). *ATL80*, an E3 ubiquitin ligase and negative regulator in plant cold tolerance [33], had three genes downregulated in “JJ31” and one gene upregulated in “Alamo” (Additional file 1: Fig. S5b). The *Arabidopsis* mutant *hos10-1* completely lost the cold acclimatization response [34]. We found 15 *HOS10* genes upregulated in “JJ31,” while 11 were downregulated in “Alamo” (Additional file 1: Fig. S5b). Interestingly, although *STRESS RESPONSE SUPPRESSOR2* (*STRS2*) was reported to negatively regulate *Arabidopsis* tolerance to salt, osmotic, and heat stress, and not cold stress [35], our study revealed that three *STRS2* genes were downregulated in “JJ31” and two were upregulated in “Alamo” (Additional file 1: Fig. S5b), which may highlight the role of *STRS2* in response to cold stress in switchgrass. In summary, the ecotype-specific expression of the above *COR* genes, as well as their opposite expression changes between the two ecotypes, may contribute to the observed differences in cold tolerance.

The transcriptional regulatory pathway dependent on *CBF* is crucial for the plant response to cold stress [36–38]. Similarly, we found that cold response genes in the *CBF*-dependent pathway were activated to varying degrees in both “JJ31” and “Alamo,” with slight differences between leaves and roots (Fig. 3f). We observed that the specific differential expression of *MEKK1* and *SIZ1* genes occurred only in the leaves. In contrast, the specific differential expression of *CRLKs* and *ICE1* genes occurred only in the roots (Fig. 3f). Additionally, members of the *CaM*, *MPK3/6*, and *CIPKs* families tended to exhibit opposite expression trends between the two ecotypes (Fig. 3f). In conclusion, through the identification of *COR* gene families and comparative transcriptome analysis, we comprehensively revealed the landscape of differential expression of *COR* genes between the two ecotypes, which may contribute to the superior cold tolerance of the upland compared to the lowland ecotype.

The involvement of ASE in response to cold stress revealed by haplotype-resolved genome

The phenomenon of allele-specific expression has been found to be widespread in plants as more genomes are elevated to the level of haplotype resolution, and alleles exhibiting this characteristic are referred to as ASE genes [39]. The haplotype-resolved genome of upland switchgrass enabled us to use RNA-seq data to identify ASE genes, which profoundly impact plant growth and development [40]. According to the correlation between the number of ASE genes and the number of transcriptome samples used in the analysis, we aimed to obtain a complete ASE gene collection in switchgrass by using sufficient data. We found that the number of ASE genes stabilized when the number of transcriptome samples reached 15 by utilizing transcriptome data reported by Zuo et al. (with at least two replicates, Fig. 4a) [41]. A total of 16,801 ASE genes were identified in switchgrass, with the number of alleles biased toward the expression of “JJ31-A” significantly exceeding those biased toward the expression of “JJ31-B” (Fig. 4b, Additional file 2: Table S14). To understand the impact of natural selection on ASE genes and non-ASE genes, we calculated the Ka/Ks ratio between allele pairs. Although most allele pairs exhibited low Ka and Ks values, ASE genes experienced significantly stronger purifying selection pressure compared to non-ASE genes (Fig. 4c, Additional file 2: Table S15). To further investigate potential causes of ASE, we examined the distribution patterns of SNPs surrounding ASE genes and equivalently expressed alleles (EEAs). Compared to EEAs, ASE genes exhibited significantly higher SNP density in the upstream, exonic, intronic, and downstream regions, consistent with previous findings in other plants [42] (Fig. 4d). The SNP density in the upstream region was higher than that in other regions, suggesting that the occurrence of ASE might correlate with the greater variation in the upstream region of the gene.

Similarly, we utilized the transcriptome data obtained in this study under cold stress to identify ASE genes, aiming to explore whether the response of ASE to cold stress differs between the two ecotypes (Additional file 3: Note S2). Compared to the control group, a significant increase in ASE genes were detected in “JJ31” after experiencing cold stress, whereas no significant change was observed in “Alamo,” indicating that more ASE genes were induced by cold stress in “JJ31” (Fig. 4e, Additional file 2: Table S16-17). In total, we identified 10,775 and 8430 cold-induced ASE genes in the two tissues of “JJ31” and “Alamo,” respectively (Additional file 1: Figs. S6 and S7, Additional file 2: Table S18). Among these ASE genes, 5680 were specific to JJ31, 3,335 were specific to Alamo, and 5095 were shared between the two ecotypes (Additional file 1: Fig. S8a). Additionally, cold-induced ASE genes specific to “JJ31” were significantly enriched in *COR* genes ($P=0.02$, Fig. 4f and Additional file 1: Fig. S8b), whereas those specific to “Alamo” showed no significant enrichment ($P=0.63$, Additional file 1: Fig. S8c), supporting the importance of ASE in responding to cold stress in “JJ31.”

We selected the well-known *COR* gene, *CBF*, to validate the accuracy of ASE, as this gene undergoes cold-induced ASE in both “JJ31” and “Alamo.” We observed that the two alleles of the *CBF* gene, *PVA_6 K02793.1* and *PVB_6 K02781.1*, did not show differential expression in the control groups except for the 12 and 24 h control groups of “Alamo,” but *PVB_6 K02781.1* exhibited significantly preferential expression in all cold treatment groups (Fig. 4g). Although the sequence similarity between the two alleles is as high as 98.82%, four SNPs cause changes in three amino acids (Additional file 1: Fig. S9).

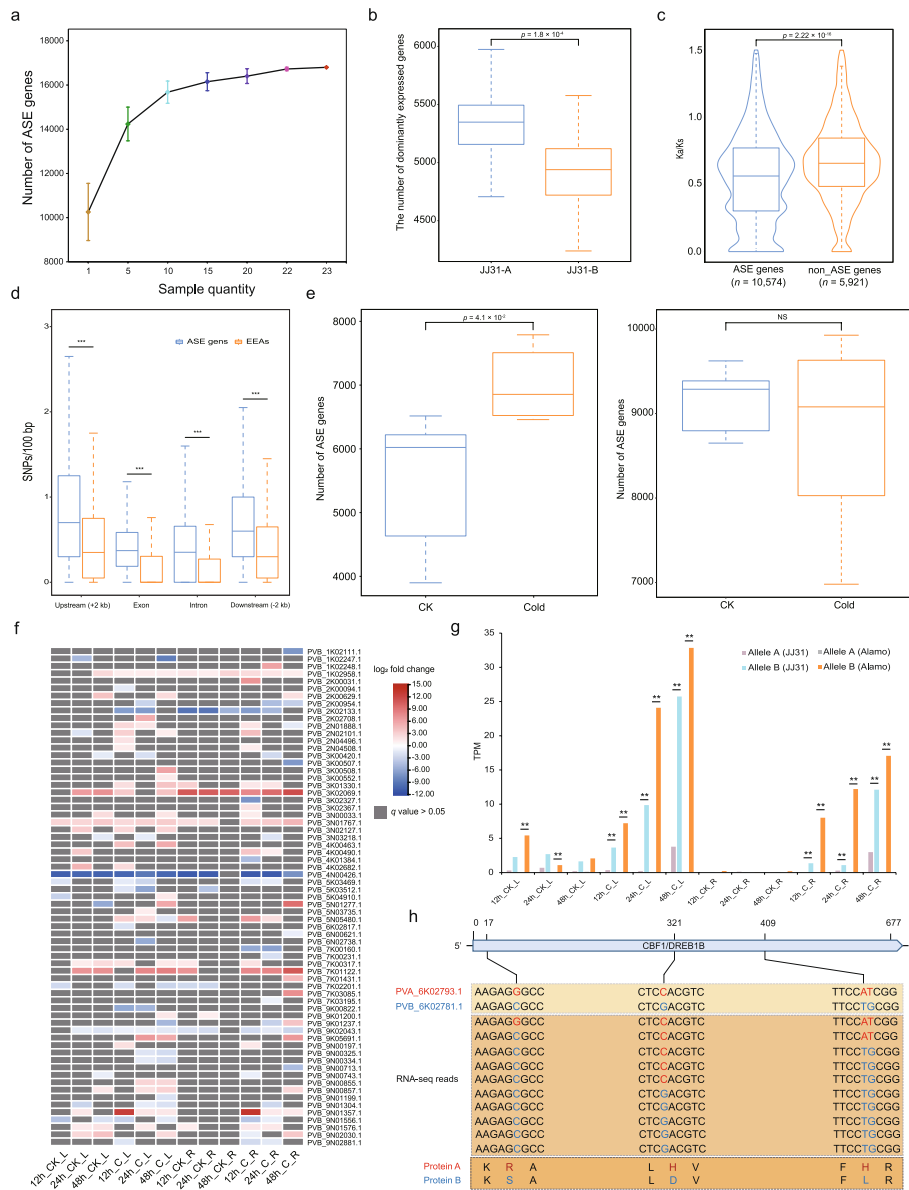


Fig. 4 Characteristics of ASE genes in switchgrass. **a** ASE gene numbers increase with the quantity of RNA-seq samples. The specified number sets were selected randomly from 23 ASE gene sets with three replicates. **b** Dominant expressed alleles in two haplotype genomes. **c** *Ka/Ks* of ASE and non-ASE genes. Minima and maxima are present in the lower and upper bounds of the whiskers, respectively, and the width of the violin are densities of the *Ka/Ks* value. *P* values were calculated with a two-sided Student's *t* test. **d** SNP density in ASE genes and equivalently expressed alleles (EEAs). The y axis represents SNP numbers every 100 bp. *P* values were calculated with two-sided Student's *t* test. *** indicates *P* < 0.0005. **e** Number of ASE genes identified in JJ31 (left) and Alamo (right) under control (room temperature) and cold stress (4 °C). *P* values were calculated with two-sided Student's *t* test. NS indicates not significant. **f** The expression changes of 67 *COR* alleles enriched in JJ31-specific ASE genes under control (room temperature) and cold stress (4 °C) conditions. The suffixes "L" and "R" represent leaves and roots, respectively. **g** The expression levels (TPM) of two alleles of the *CBF* gene (*PVA_6 K02793.1* and *PVB_6 K02781.1*) across different transcriptome samples in JJ31 and Alamo. ** indicates adjusted *P*-value < 0.01. **h** Pattern diagram of *PVB_6 K02781.1* advantage gene expression. Red and blue indicate the allele IDs, corresponding bases, and encoded amino acid types in "JJ31-A" and "JJ31-B," respectively. "RNA-seq reads" represents the proportion of reads containing different SNP types that map to "JJ31-B." From left to right, RNA-seq reads aligned to the SNP sites are as follows: 79 reads (C: 92%, G: 8%), 100 reads (G: 45%, C: 55%), and 67 reads each for the third and fourth sites (T: 81%, A: 19%; G: 85%, T: 13%)

We aligned the RNA-seq data to the reference genome “JJ31-B” and observed a higher proportion of reads containing SNPs corresponding to the B allele type, supporting the dominant expression of *PVB_6 K02781.1* (Fig. 4h). In conclusion, our findings indicate the widespread presence of ASE in switchgrass, with more ASE genes involved in cold stress responses in upland ecotype.

Identification of genes associated with cold tolerance by population genetic analysis

To explore cold tolerance genes in upland switchgrass at the population level, we aligned resequencing data from 340 accessions (242 uplands and 98 lowlands) reported previously to the “JJ31” genome [10] (Additional file 2: Table S19), resulting in 10,654,902 SNPs and 243,831 SVs (Additional file 1: Fig. S10, Additional file 2: Table S20 and Additional file 3: Note S3). A total of 103.7 and 125.9 Mb of genomic sequences covering 5084 and 8428 genes were detected using the sliding window method based on SNPs and SVs, respectively (Additional file 1: Fig. S11). We found that 66 and 111 genes from the two datasets were significantly ($P = 0.0196$ and $P = 0.0018$, respectively) annotated as belonging to the *COR* gene family (Additional file 1: Fig. S11). Among the *COR* genes identified based on SNPs and SVs, approximately 54.5% (36 out of 66) and 62.2% (69 out of 111), respectively, showed differential expression under cold stress (Additional file 2: Table S21). These results suggest that the differential selection of certain *COR* genes in the two ecotypes potentially contributes to the differences in cold tolerance.

To identify candidate genes related to cold tolerance in switchgrass, we conducted a GWAS based on both mixed linear model and logistic model approaches, using two variant datasets and previously reported overwintering rates of 340 switchgrass accessions [10] (Additional file 1: Fig. S12-S15, Additional file 2: Table S22-23). An association signal on Chr3 K was simultaneously detected in both SV-GWAS and SNP-GWAS, including an overlapping region with the selective sweep region (Fig. 5a). We examined the expression of 14 genes around the association signal and found that only *PVB_3 K03605.1* and *PVB_3 K03611.1* were highly expressed under cold stress (Fig. 5b and Additional file 1:

(See figure on next page.)

Fig. 5 Selective sweep analysis between the upland and lowland ecotype populations and GWAS analysis of overwintering rate. **a** Top: Selective sweep detection between two ecotypes on chromosome 3 K using SNPs and SVs (Fst method). The black dashed line indicates the top 5% sweep regions. Middle: GWAS of overwintering rate on chromosome 3 K. The black dashed line shows the significance threshold ($-\log_{10}(P) > 6.69$). Gray bars mark overlapping regions between selective sweeps and GWAS. Bottom: Schematic of *PVB_3 K03611.1* and its upstream 61-bp deletion. **b** The expression levels (TPM) of *PVB_3 K03611.1* under control and cold stress. *** indicates adjusted P -value < 0.001 . The suffixes “L” and “R” represent leaves and roots, respectively. **c** Left: Luciferase (LUC) reporter gene expression observed by soaking in substrate solution; Right: Quantitative detection of the firefly luciferase and renilla luciferase in leaves inoculated with different recombinant vectors using a microplate plate-based fluorescence assay. The error bars indicate the mean \pm s.d.; $n = 3$ biological replicates. * indicates $P < 0.05$ and **** indicates $P < 0.0001$. **d** The distribution proportions of three genotypes with a 61-bp deletion across different ecotypes of germplasm. 0/0 means consistent with the reference genome, 0/1 means heterozygous, and 1/1 means homozygous. **e** Overwintering survival rate of lowland accessions with three genotypes in BRKG (the name of the common garden) area. **f** Phenotypic changes of rice wild type and overexpression lines on days 0, 3, and 6 under 4 °C cold stress, scale bar represents 7 cm. The number in the suffix represents the overexpression line number. **g** Physiological parameters of *PVB_3 K03611.1* overexpression lines and WT under cold stress. Top left: Relative expression levels of *PVB_3 K03611.1* (N.D. indicates not detected). Remaining panels: SOD and POD activities, and MDA content in WT and overexpression lines after 24 h of cold stress at 4 °C. * indicates $P < 0.05$, ** indicates $P < 0.01$, and *** indicates $P < 0.005$

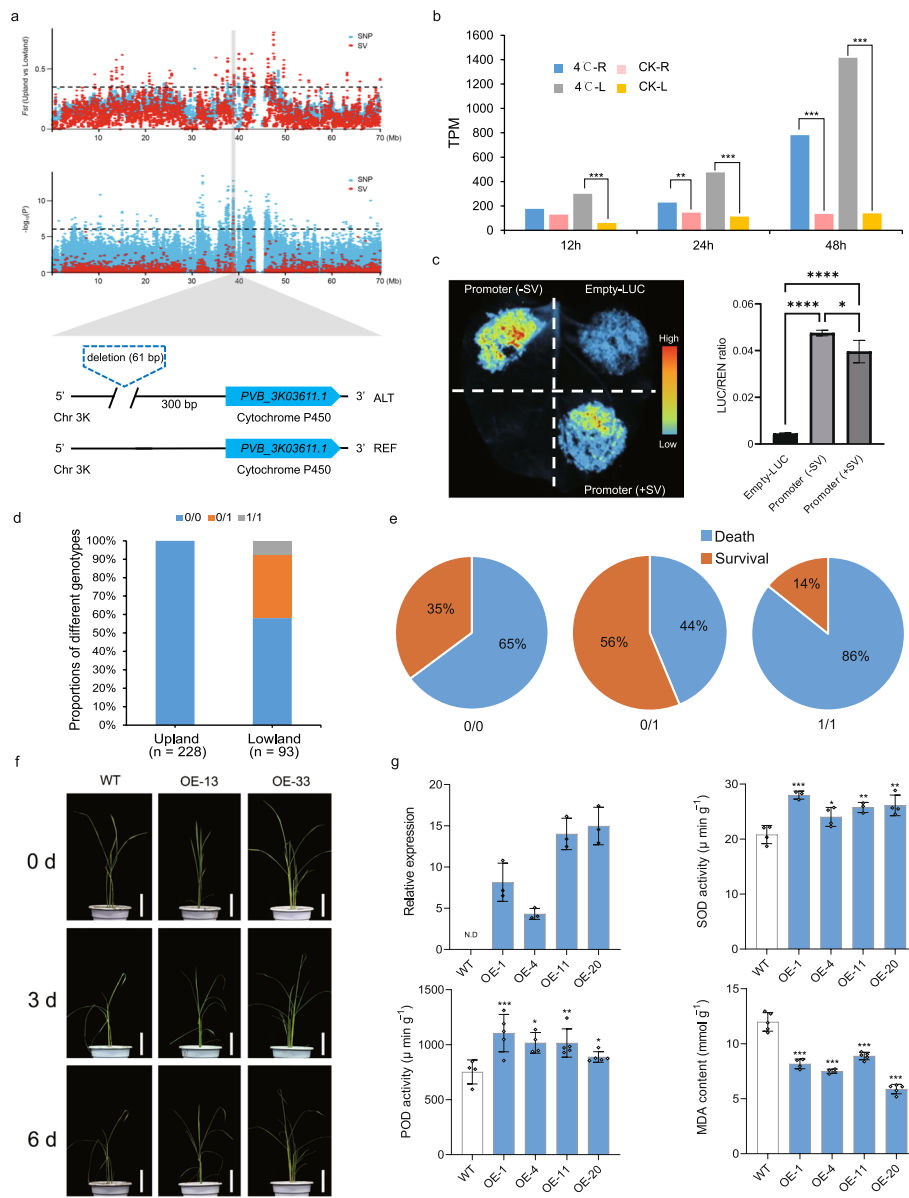


Fig. 5 (See legend on previous page.)

Fig. S16). Interestingly, only *PVB_3 K03611.1* appeared in the overlapping region, which encodes cinnamate-4-hydroxylase belonging to the *CYP450* gene family.

We found a 61-bp deletion located 300 bp upstream of the promoter region of *PVB_3 K03611.1* (Fig. 5a). We confirmed that the presence of this deletion in the promoter region inhibits the expression of the reporter gene using a dual-luciferase reporter system in tobacco (*Nicotiana tabacum*) leaves (Fig. 5c). We further observed this deletion with frequency differences between the two ecotypes, where the 0/1 (heterozygous) and 1/1 (homozygous) genotypes were present in about 40% of the lowland accessions, while the deletion was absent in the upland accessions (Fig. 5d). Analysis of the overwintering survival rate of lowland accessions revealed that 35% and 56% of accessions with 0/0 (same as the reference) and 0/1 genotypes, respectively, could survive the winter, while

this proportion decreased to 14% in accessions with the 1/1 genotype (Fig. 5e). These results indicated that the deletion could reduce the expression of the target gene and might be under purifying selection in the upland accessions.

To validate the role of *PVB_3 K03611.1* in cold tolerance, we overexpressed it in rice. Compared to the wild type (WT) rice, the transgenic lines exhibited less leaf yellowing and withering under cold stress (Fig. 5f and Additional file 1: Fig. S17a). Additionally, the transgenic lines displayed significantly higher activities of superoxide dismutase (SOD) and peroxidase (POD), as well as significantly lower levels of MDA than those in WT plants when exposed to low temperatures, indicating that overexpression of *PVB_3 K03611.1* enhances the cold tolerance of the transgenic lines (Fig. 5g and Additional file 1: Fig. S17b). Collectively, these results support that the deletion in the promoter region may inhibit the expression of *PVB_3 K03611.1*, leading to a reduction in cold tolerance.

Discussion

The application scope and geographical distribution of switchgrass have gradually expanded since 1992, when it was chosen as a model species for bioenergy feedstock [43]. During the formation of modern tetraploid switchgrass, interspecific hybridization between unknown ancestors, polyploidization, whole-genome duplication and diploidization may have occurred [44]. Although the Phytozome database (<https://phytozome-next.jgi.doe.gov/>) has released genome data for two upland switchgrass, these data are subject to “restrictions on dataset usage”, particularly for comparative genomic analysis. We conducted a comparative analysis of the genomes of “JJ31” and the previously published “AP13” and revealed that the divergence between the ecotypes is a relatively recent event that occurred after the subgenome differentiation. These results indicated a high degree of homology between the two ecotypes, which might explain why the upland and lowland ecotypes could easily hybridize at the tetraploid level [45]. Subgenome dominance is a widespread phenomenon in allopolyploids and plays a crucial role in plant evolution and adaptation. A study on cotton (*Gossypium raimondii*) suggested that differences in TE density between subgenomes might be a determining factor of subgenome dominance [46], and our research further supports this perspective. Another study in cotton has shown that, in addition to epigenetic modifications, chromatin dynamics are also important factors influencing subgenome dominance [47]. Whether these factors similarly contribute to subgenome dominance in switchgrass requires further investigation with sufficient data.

The haplotype genomes allowed the identification of ASE genes, addressing the limitations of the previous chimeric assembly genomes. Similar to previous reports [48, 49], we found a positive correlation between the number of ASE genes and the number of transcriptome samples in switchgrass, and these ASE genes were widely present throughout growth and development. ASE genes accounted for 25.88% (16,801 out of 64,928) of all annotated alleles in our study, which is higher than the approximately 7.5% (2676 out of 35,525) identified in the autoploid sugarcane (*Saccharum spontaneum* L.) [50]. We found more ASE genes biased toward “JJ31-A,” indicating that the number of ASE genes biased toward different parents is not balanced, which is consistent with a report in cotton [51]. Additionally, we discovered that there were more cold-induced ASE genes in

the upland ecotype compared to the lowland ecotype. These results indicated that the dominant effects of alleles played a crucial role in the growth, development, and defense against abiotic stresses in switchgrass. Further exploration of the mechanisms underlying ASE can enhance the efficiency and accuracy of genome-assisted breeding strategies.

The two ecotypes of switchgrass are ideal models for studying plant cold tolerance mechanisms. Although upland and lowland ecotypes can coexist in transition zones, the southern regions of the USA are dominated by lowland ecotypes, while the northern regions are dominated by upland ecotypes [52]. This distribution pattern results in differences in cold tolerance between the two ecotypes, with upland ecotypes well adapted to USDA Hardiness Zones 3 to 7, and lowland ecotypes well adapted to USDA Hardiness Zones 5 to 9 [52]. We identified a large number of genes involved in photomorphogenesis and circadian rhythms (e.g., *PIF3*, *PIF4*) that participate in the cold stress response in switchgrass. While their roles in cold tolerance have been demonstrated in *Arabidopsis* [25, 26], the ecotype-specific expression mechanisms regulating cold tolerance in switchgrass require further investigation. Although transcriptomic analysis under 4 °C treatment has to some extent revealed the potential mechanisms underlying the cold tolerance differences between the two ecotypes in this study, whether lower treatment temperatures, particularly freezing stress (< 0 °C), would produce more informative results remains to be further investigated. Future exploration of potential regulatory mechanisms beyond the *COR* gene families at the transcriptional level, as well as epigenetic regulation and post-transcriptional regulation, will provide a more comprehensive understanding of the mechanisms underlying cold tolerance differences between the two ecotypes.

Based on a high-quality upland reference genome, we employed stringent filtering criteria to identify SNPs and SVs in the resequencing data from switchgrass populations planted in cold regions. Based on these variations, we identified a large number of signals associated with overwintering survival rate in switchgrass. It should be noted that the lambda value of GWAS using the mixed linear model in this study showed a certain degree of underestimation, which may be caused by the following reasons. Firstly, as a cross-pollinated allotetraploid, the complex genotype and subgenome interaction effects of switchgrass may lead to systematic bias in statistical analysis [53]. Furthermore, the winter survival rate in this study is a complex trait represented by a binary variable, which may further exacerbate the abnormality of the lambda value. Finally, we incorporated the kinship matrix and population structure matrix (Additional file 1: Fig. S18) into the model analysis to mitigate the lambda value abnormality as much as possible. Additionally, we performed GWAS on the dataset using the SwitchgrassGWAS R package based on a logistic model (Additional file 1: Fig. S13 and S15). The results showed that the lambda values for both SNP-GWAS (0.98) and SV-GWAS (0.87) were improved, indicating more effective control of genomic inflation. Importantly, the association signals identified in this study were consistently detected by both models, further supporting the reliability of our conclusions. When interpreting GWAS results, it is crucial to consider the limitations mentioned above. Selecting signals that pass stringent significance thresholds and integrating multi-omics evidence and validation experiments remains an effective strategy for identifying reliable candidate genes. In the genes identified through selective sweep analysis and GWAS, over 25 and 28% (Additional

file 1: Fig. S19) respectively showed differential expression in at least one tissue after cold stress, indicating that multi-omics integrated analysis is an effective approach for identifying candidate genes. The *CYP450* family has been found to play a crucial role in plant development and stress responses [54, 55]. Some studies have suggested the potential role of certain *CYP450* gene family members in cold tolerance in sorghum [56] and perennial ryegrass [57], but their specific functions remain experimentally unverified. We identified a 61-bp deletion that reduces the expression of the downstream gene *CYP450* (*PVB_3 K03611.1*), thereby influencing cold tolerance and providing new evidence for *CYP450*-mediated cold tolerance in plants. Current breeding strategies for cold tolerance in switchgrass primarily focus on selecting lowland germplasm resources with both cold resistance and high biomass potential [14]. The SV markers developed in this study provide novel molecular tools for cold-tolerant germplasm screening, especially for the rapid identification of cold tolerance at the early seedling stage. Furthermore, the regulatory mechanisms of *CYP450* in the cold stress response need to be further elucidated to design cold-tolerant breeding strategies with multi-target synergistic regulation.

Conclusions

In this study, we constructed a high-quality haplotype-resolved genome for the upland switchgrass “JJ31.” Comparative analysis with the lowland genome revealed that the divergence between the two ecotypes occurred after the differentiation of the two sub-genomes. Through genome-wide identification of the *COR* gene family and comparative transcriptome analysis, we found no significant difference in the number of *COR* gene family members between the two ecotypes. However, a large number of *COR* genes exhibited opposite expression patterns under cold stress, which may contribute to the differences in cold tolerance between the two ecotypes. Based on the haplotype-resolved genome, we identified a large number of ASE genes in switchgrass, with more cold-induced ASE genes present in the upland ecotype. Compared to the lowland ecotype, cold-induced ASE genes in the upland ecotype were significantly enriched in *COR* genes, indicating their important contribution to cold tolerance. Additionally, through the analysis of resequencing data from 340 collected switchgrass accessions, we generated a comprehensive map of genetic variation in switchgrass. Selective sweep analysis and GWAS identified numerous candidate genes associated with cold tolerance. A newly identified cold-tolerance candidate gene, *CYP450* (*PVB_3 K03611.1*), improved cold tolerance when heterologously expressed in rice. A deletion in the promoter region of this gene was confirmed to affect the expression of downstream genes, and this deletion, mainly present in lowland germplasms, was negatively correlated with overwinter survival rates. In summary, our research findings provide valuable resources for future genomic studies of switchgrass and other plant species, advancing genome-assisted breeding for cold tolerance in this important bioenergy crop.

Methods

Sample collection and DNA sequencing

The upland switchgrass cultivar “JJ31” ($2n = 4x = 46$) was derived from the germplasm repository of the Beijing Academy of Agricultural and Forestry Sciences. It is characterized by shorter plant height and earlier flowering time [58]. The plants were propagated

asexually, planted in three pots in the greenhouse, and grown at 26/22 °C (day/night) with a photoperiod of 14/10 h of light/dark. The leaves of plants grown at the E3 stage [59] were collected and immediately frozen in liquid nitrogen, followed by DNA extraction using the DNasecure Plant Kit (TIANGEN). For Illumina short-reads sequencing, ~1.5 µg of genomic DNA was extracted to construct a short insert (350 bp) library using a TruSeq Nano DNA HT Sample Preparation Kit. Sequencing was performed using Illumina HiSeq2500 platforms. The raw reads were trimmed using Trimmomatic (v.0.36) [60] with default parameters. For PacBio HiFi sequencing, SMRTbell libraries were constructed using the SMRTbell Express Template Prep Kit 2.0 (PacBio, CA). Two single-molecule real-time (SMRT) cells were run on the PacBio Sequel II platform. The raw data were processed with the SMRT Link (v.9.0) to obtain HiFi reads, using the parameters `-min-passes = 3` and `-min-rq = 0.99`. For Hi-C sequencing, the library construction method was the same as the protocol previously used in our laboratory [61]. The constructed library was sequenced using the Illumina NovaSeq 6000 platform.

Genome size prospection

To estimate the genome size, k-mer analysis ($K=17$) was performed on Illumina short reads using Jellyfish (v.2.3.0) [62]. The genome size, heterozygosity, and repeat proportion were estimated by GenomeScope (v.2.0) [63] based on the k-mer frequency distribution. The principle for calculating genome size is based on the formula: $G = (N \times (L - k + 1) - B) / D$, where N is the total number of sequence reads, L is the average length of the reads, K is the k-mer length, B is the total number of low-frequency k-mers, D is the estimated total depth based on k-mer distribution, and G is the genome size.

Genome assembly and pseudochromosome construction

HiFi reads were assembled into two haplotype-resolved draft genomes using the Hifiasm software (v.0.15.5) [11]. Initially, an all-vs-all pairwise comparison of HiFi reads was performed for self-correction. After haplotype-aware error correction, the corrected reads were used to construct an assembly graph and generate bubbles within this graph. An initial contig assembly based on the overlap graph was obtained using a modified “best overlap graph” strategy. During the assembly process, optimized parameters suitable for polyploid genomes (`-n-hap 4`) were added to preserve haplotype information as much as possible. Filtered Hi-C reads were aligned to the initial contig assembly using BWA (v.0.7.8) [64], and the alignment results were used as the input in Juicer (v.1.6) [65]. The 3D-DNA workflow selected only uniquely aligned and valid paired-end reads for further assembly [66]. Finally, the order of scaffolds was manually adjusted using Juicebox (v.2.13.07) [67] to obtain the final chromosome assembly. HiCExplore (v.3.7.2) [68] was used to draw heatmaps of the connections between chromosomes.

Genome assessment

To assess the quality of the genome assembly for accuracy, completeness, and continuity, we used BWA (v.0.7.8) [64] to map high-quality Illumina paired-end reads to the genome, evaluating the alignment rate and coverage. BUSCO (v.4.1.2) [69] and the CEGMA (v.2.5) [70] were used to check the completeness of the genome assembly or

annotation. The genome quality was further assessed by calculating the QV values with Merqury (v.1.3) [12] and the LAI with LTR_retriever (v.2.9.8) [13].

Annotation of repetitive sequences

We annotated the repetitive sequences by combining homology-based alignment and de novo prediction. The homology-based alignment method used RepeatMasker (v.4.0.5) [71] and RepeatProteinMask (v.4.0.5) [71] to identify sequences similar to known repetitive sequences based on the RepBase database (<http://www.girinst.org/replib>) [72]. The de novo prediction method utilized LTR_FINDER (v.1.0.7) [73], Piler (v.3.3.0) [74], RepeatScout (v.1.0.5) [75], and RepeatModeler (v.1.0.8) [76] to construct a de novo repeat sequence library, followed by the use of RepeatMasker (v.4.0.5) [71] to predict the repetitive sequences in this library.

Prediction of gene structure

The gene structure was annotated by integrating de novo prediction, homology-based prediction, and transcriptome-based prediction. De novo prediction involved using software such as AUGUSTUS (v.3.2.3) [77], GENSCAN (v.1.0) [78], GlimmerHMM (v.3.0.1) [79], geneid (v.1.4) [80], and NAP (v.2013.11.29) [81] to predict coding regions from the genome with repetitive sequences masked. The homology-based prediction method downloaded protein sequence files of *Arabidopsis*, rice, *Panicum miliaceum*, *Panicum hallii*, and the published genome of a switchgrass line “AP13” selected from the lowland ecotype “Alamo” in the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>). These protein sequences were aligned to the two haplotype genomes of upland switchgrass using tblastN (v.2.2.26) [82] with an e-value threshold of $1e^{-5}$. The Solar (v.0.9.6) [83] software was used to integrate the BLAST results, and GeneWise (v.2.4.1) [84] was employed to predict the precise gene structures in the corresponding genomic regions. The transcriptome-based prediction method used TopHat (v.2.0.13) [85] and Cufflinks (v.2.1.1) [86] to align transcriptome data to the two haplotype genomes. Trinity (v.2.1.1) [87] was utilized to assemble RNA-seq data to create pseudo-expressed sequence tags (pseudo-ESTs), which were then mapped to the two haplotype genomes. Finally, EVidenceModeler (v.1.1.1) [88] was used to integrate the gene sets obtained from the three methods into a non-redundant, more complete gene set (EVM sets). The Program to Assemble Spliced Alignments (PASA) [89] was used to correct the EVM sets, adding information such as UTRs and alternative splicing, to obtain the final gene set.

Annotation of protein-coding genes and non-coding RNA

Six databases were used for the functional annotation of coding genes, including Swiss-Prot (<http://www.uniprot.org/>) [90], InterPro (<https://www.ebi.ac.uk/interpro/>) [91], the Non-Redundant Protein Sequence database (NR, <ftp://ftp.ncbi.nih.gov/blast/db/>), Pfam (<https://pfam-legacy.xfam.org/>) [92], KEGG (<http://www.genome.jp/kegg/>) [93], and the GO database (<http://www.geneontology.org/page/go-database>) [94].

miRNA, rRNA, and snRNA were predicted in the genome using INFERNAL (v.1.1.5) [95] with the Rfam database (<https://rfam.org/>) [96]. For tRNA, tRNAscan-SE (v.2.0.12)

[97] was used to predict tRNA sequences in the two haplotype genomes based on the structural characteristics of tRNA.

Phylogenetic tree construction and divergence time estimation

BLASTP (v.2.7.1) was used to search against the protein sequences of *P. hallii*, *Z. mays*, *S. bicolor*, and *O. sativa*, as well as the two subgenomes of “JJ31-B” and “AP13,” using a default E-value cutoff of $1e^{-5}$. Orthofinder (v.2.3.1) [98] with default parameters was then used to cluster the filtered BLAST results into paralogous and orthologous groups. The sequences of single-copy gene families were aligned using MUSCLE (v.3.8.31) [99], and the alignment results were concatenated to form a super alignment matrix. RAxML (v.8.0.19; <http://sco.h-its.org/exelixis/web/software/raxml/index.html>) [100] was used to construct the phylogenetic tree using the maximum likelihood method, with bootstrap values set to 100. The divergence time of each node on the phylogenetic tree was estimated using the MCMCTree program (v.4.5; <http://abacus.gene.ucl.ac.uk/software/paml.html>) [100] with phylogenetic analysis by maximum likelihood (PAML) with the parameter settings “burn-in = 10,000, sample-number = 100,000, sample-frequency = 2.” The TimeTree database (<http://www.timetree.org/>) [101] provided species divergence times. On the basis of the orthologous genes for the two subgenomes each of “JJ31-B” and “AP13,” the synonymous substitution (K_s) was calculated. The formula $t = K_s/2r$ was used to estimate the divergence time between species, where r is the neutral substitution rate ($r = 6.96 \times 10^{-9}$) [30, 102, 103].

Identification of the COR gene families

The protein sequences of *Arabidopsis* and rice were downloaded from the TAIR (<https://www.arabidopsis.org>) and RGAP (<http://rice.plantbiology.msu.edu>) databases, respectively. Based on the 115 *COR* genes reported in *Arabidopsis*, we used BLASTP (v.2.7.1) to identify the *COR* protein sequences in rice, with an e-value set to $1e^{-10}$ [31]. The top-ranked protein sequences were combined with the *Arabidopsis* protein sequences to create a merged library. Subsequently, we identified the *COR* protein sequences in “JJ31-B” using an e-value of $1e^{-10}$ and identity > 60% [31].

Transcriptomic analyses of switchgrass under low temperature

Seeds of “JJ31” and “Alamo” were sown in plastic pots (10 × 15 × 6 cm) filled with quartz sand and cultivated in a growth chamber (26 °C with 14 h of light, 22 °C with 10 h of darkness). According to a previous study, the 4 °C treatment induces significant physiological changes and the expression of cold tolerance genes in switchgrass, and was therefore selected as the cold stress temperature in this study [104]. Cold stress treatment was then applied to E3 stage [59] seedlings of both ecotypes, with conditions set to 4 °C with 14 h of light and 4 °C with 10 h of darkness, while the control group was maintained under normal conditions. After 12, 24, and 48 h of cold stress treatment, the leaves and roots of JJ31 and Alamo were collected and stored at −80 °C. Three biological replicates were set for each treatment and control, with each replicate consisting of a mixture of three seedlings. RNA was extracted from the mixed samples using the RNeasy Plant Mini Kit (QIAGEN), and the quality of RNA was assessed by RNA gel electrophoresis. High-quality RNA was used to construct cDNA libraries with the NEBNext Ultra

Directional RNA Library Prep Kit. Transcriptome sequencing was performed on the Illumina HiSeq X platform. The raw data were processed to remove adapters and low-quality nucleotide sequences using Trimmomatic (v.0.36) [60]. The quality of the filtered data was assessed using FastQC (v.0.11.9, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). A genome index file was built with Kallisto (v.0.46.0) [105] using “JJ31-B” as the reference genome. Subsequently, the filtered transcriptome clean reads were aligned to the index file to obtain gene count values and transcripts per million (TPM). DESeq2 (v.1.24.0) [106] was employed for identifying DEGs ($|\log_2(\text{fold change})| \geq 0.8$ and adjusted P -value < 0.05) based on gene count values. GO and KEGG enrichment analyses were performed using the OmicShare tools (<http://omicshare.com/tools>).

Physiological index measurement

Leaves of E3 stage [59] seedlings of “JJ31” and “Alamo” were used for physiological measurements, with the cultivation methods and conditions being the same as those used for the seedlings prepared for transcriptome sequencing. The RWC, REC, and MDA content of the leaves were measured on seedlings after 1, 7, 14, 21, 28, and 35 days under both cold treatment and normal conditions. Transgenic rice and WT rice were cultivated for 45 days under 26 °C with 14 h of light and 22 °C with 10 h of darkness, followed by cold stress treatment at 4 °C. After 24 h of cold stress, the MDA content and the activities of POD and SOD enzymes were quantified using rice leaves. The RWC of the leaves was determined using the saturated weighing method [107] based on the formula $RWC = (FW - DW) / (TW - DW)$, where FW refers to the fresh weight of leaves taken from the same part of seedlings, TW is the saturated fresh weight of these leaves after absorbing water, DW refers to the dry weight of leaves after soaking, blanching at 105 °C for 30 min, and then drying at 65 °C until a constant weight is reached. The measurements of REC, MDA, POD, and SOD were based on the methods previously described by our laboratory [61].

Differential expression analysis of allelic genes

Protein sequences from the two haplotype genomes were retrieved using TBtools (v.2.069) [108]. The proteins in “JJ31-A” were compared with those in “JJ31-B” using BLASTP (v.2.7.1), and syntenic blocks within the genomes were identified using MCS-canX [109] with default parameters. Gene pairs with unique alignment relationships between the “JJ31-A” and “JJ31-B” genomes were then identified, with alleles required to originate from the same pair of homologous chromosomes. The sequences of “JJ31-A” and “JJ31-B” were combined into a single file [42]. An index file for the combined sequences was created using Kallisto (v.0.46.0) [105, 110], and the clean RNA-seq data were aligned to the index file to obtain gene count values. Pairwise comparisons of allelic genes (JJ31-A/JJ31-B) were performed using DESeq2 (v.1.24.0) [106] based on the gene count values to identify differentially expressed genes, with criteria set at $|\log_2(\text{fold change})| \geq 1$ and adjusted P -value < 0.05 . Genes meeting the following three conditions were identified as ASE genes: (1) the fold change of one allele compared to the other was > 2 or < 0.5 ; (2) TPM values > 1 in all transcriptome samples; (3) differential expression of alleles in at least one transcriptome sample. All transcriptome samples involved in ASE gene identification had at least two replicates.

SNP calling

SNP calling was performed using GATK (v.4.3.0.0) [111], with detection by HaplotypeCaller and genotyping via GenotypeGVCFs. The SelectVariants tool was used to obtain a collection of SNPs based on the “–select-type-to-include SNP” parameter. This collection was then filtered using the parameters “QD < 2.0 || FS > 60.0 || SOR > 3.0 || MQ < 40.0 || MQRankSum < –12.5.” Finally, VCFtools (v.0.1.16) [112] was employed to further filter the data using parameters with “–max-missing 0.9, –maf 0.05, –minDP 10.”

SV detection

To improve the accuracy of SV identification, we employed three tools: Manta (v.1.6.0) [113], Delly (v.1.1.6) [114], and LUMPY (v.0.3.1) [115]. First, we used LUMPY with the parameters -P -B -S -D to detect SVs, excluding insertions. We filtered results lacking split read support and conducted genotyping with SVTyper (v.0.7.1) [116]. The other two tools were used with their default settings for both SV detection and genotyping. Finally, we merged and filtered the results from these three tools using SURVIVOR (v.1.0.7) [117], with the parameters set to “SURVIVOR merge 1000 3 1 1 0 50.” Only SVs identified by all three tools were retained.

Selective sweep analysis

To identify genomic regions under selection in upland relative to lowland ecotypes, we used VCFtools (v.0.1.16) [112] to perform *Fst* analysis based on a sliding window of 100 kb with a step size of 10 kb [118]. The top 5% windows were identified as selective sweeps [119].

Genome-wide association study

To improve the accuracy of GWAS results, we filtered the SNP and SV variant datasets, removing data with a minor allele frequency (MAF) < 0.05 or missing rate > 0.2. Association analysis was performed using GEMMA (v.0.94.1) [120] based on a mixed linear model. The model was defined as $y = X\alpha + S\beta + K\mu + e$, where y represents the phenotype, X represents the genotype, S is the population structure matrix, and K is the kinship matrix. $X\alpha$ and $S\beta$ represent fixed effects, while $K\mu$ and e represent random effects. To better control for genomic inflation, we also performed GWAS using a logistic model implemented in the SwitchgrassGWAS R package [10]. The significance cutoff was defined as the Bonferroni test threshold, which was set as $0.05/(\text{total number of SNPs})$ and $0.05/(\text{total number of SVs})$ [118], which corresponded to $-\log_{10}(P) = 8.33$ for SNPs and 6.69 for SVs.

Dual-luciferase assays to assess impact of SV on gene expression

To study the impact of SV on gene expression, the activity of dual luciferase was detected by the instantaneous expression of tobacco. Dual-luciferase assays were performed as described in a previous study [61]. The promoter fragment was recombined into the vector pGreenII 0800-LUC to form a reporter vector (Empty-LUC, promoter (-SV)-LUC, and promoter (+ SV) -LUC), and the recombinant plasmid was

transformed into *Agrobacterium tumefaciens* GV3101 (psoup-p19) by freeze–thaw method. The bacterial solution was cultured to an OD₆₀₀ of 0.6 and then injected into the abaxial side of tobacco leaves. Finally, fluorescence in tobacco leaves was observed using a live imaging system (FUION-FX7. EDGE), and the injection areas were collected and fluorescence activity was measured using a luciferase assay kit (Vazyme Biotech Co., Ltd, DL101).

Transgenic rice validation

The CDS sequence of *PVB_3 K03611.1* was synthesized using gene synthesis methods and inserted into the pCAMBIA3300-35S-EGFP vector under the control of the 35S promoter. The wild rice variety used for transgenic verification experiments in this study was Nipponbare (*O. sativa* L. spp. *japonica*). The transformation was performed using the *Agrobacterium*-mediated method as described by Hiei et al. [121]. First, *Agrobacterium* was added to the infection solution to prepare a resuspension with OD₆₀₀ of 0.2. The rice callus was immersed in the *Agrobacterium* resuspension for 10–15 min, and then co-cultivated on medium at 20 °C for 48–72 h. The callus was then transferred to the selection medium and cultured in the dark at 26 °C for 20–30 days. The positive callus tissues screened were inoculated into the secondary screening medium and cultured at 26 °C in the dark for 7–10 days. The positive callus tissues that passed the secondary screening were inoculated into the differentiation medium and cultured at 25–27 °C in the light for 15–20 days. After the 2–5-cm buds appeared, they were inoculated into the rooting medium and cultured at 30 °C in the light for 7–10 days. PCR-positive seedlings were transplanted into soil and grown under conditions of 26 °C with 14 h of light and 22 °C with 10 h of darkness. When the plants reached the four-leaf stage, real-time quantitative PCR was performed, with each sample tested in three technical replicates. Primer information is provided in Additional file 2: Table S24.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03604-8>.

Additional file 1: Fig S1–S19. Fig. S1. Synteny of 18 chromosome-level assemblies between different assemblies. Left: Chromosome synteny between JJ31-A and AP13; Middle: Chromosome synteny between JJ31-B and AP13; Right: Chromosome synteny between JJ31-A and JJ31-B. Fig. S2. Functional enrichment analysis of expanded, positively selected, and specific genes in JJ31. a GO enrichment. b KEGG enrichment. Fig. S3. KEGG enrichment analysis of DEGs in JJ31 and Alamo at three time points under cold stress. a Leaves b, Roots. Fig. S4. Venn diagram between different DEGs sets of two ecotypes. a Venn diagram of DEGs in JJ31 and Alamo leaves. b Venn diagram of DEGs in JJ31 and Alamo roots. c Venn diagram of all specific DEGs and *COR* genes in JJ31. d Venn diagram of all specific DEGs and *COR* genes in Alamo. The letters L and R in the gene set names represent leaves and roots, respectively. Fig. S5. Specifically differentially expressed *COR* gene families in JJ31 and Alamo. a Venn diagram of specifically differentially expressed *COR* gene families in JJ31 and Alamo. b The Log₂expression fold change of *COR* gene families with opposite expression changes in JJ31 and Alamo. Fig. S6. The identification of Cold-Induced ASE genes in JJ31. a Venn diagram of ASE genes in leaves of JJ31 in the control group and cold treatment group at three time points. b Venn diagram of ASE genes in roots of JJ31 in the control group and cold treatment group at three time points. Fig. S7. The identification of Cold-Induced ASE genes in Alamo. a Venn diagram of ASE genes in leaves of Alamo in the control group and cold treatment group at three time points. b Venn diagram of ASE genes in roots of Alamo in the control group and cold treatment group at three time points. Fig. S8. Venn diagram of ASE genes and *COR* genes. a Venn diagram of ASE genes in JJ31 and Alamo. b Venn diagram of JJ31-specific ASE genes and *COR* genes. c Venn diagram of Alamo-specific ASE genes and *COR* genes. d Venn diagram of shared ASE genes and *COR* genes. Fig. S9. Sequence alignment of the two alleles *PVA_6 K02793.1* and *PVB_6 K02781.1* of the *CBF*. a Alignment of CDS sequences. b Alignment of protein sequences. Fig. S10. Characteristics of SVs. a Density of SVs of different lengths. b Distribution of SVs on the genome. Fig. S11. Selective sweep detection between the two ecotype populations using the *Fst* method. a Selective sweep analysis based on SNPs variant set. b Selective sweep analysis based on SVs variant set. The top 5% regions were identified as selective sweep regions, with the dashed line representing the threshold. Fig. S12. Manhattan and QQ plots of genome-wide association studies of BRKG_SRV using SNPs. The dashed line

representing the Bonferroni test threshold. Fig. S13. Manhattan and QQ plots of genome-wide association studies of BRKG_SRV using SNPs based logistic model. The dashed line representing the Bonferroni test threshold. Fig. S14. Manhattan and QQ plots of genome-wide association studies of BRKG_SRV using SVs. The dashed line representing the Bonferroni test threshold. Fig. S15. Manhattan and QQ plots of genome-wide association studies of BRKG_SRV using SVs based logistic model. The dashed line representing the Bonferroni test threshold. Fig. S16. The expression of 14 putative genes around the GWAS signal peak. The letters L and R in the sample names represent leaves and roots, respectively. Fig. S17. Phenotypic and physiological changes of overexpression lines and WT under 4 °C cold stress. a Phenotypic changes of rice wild type and overexpression lines on days 0, 3, and 6 under 4 °C cold stress, scale bar represents 10 cm. b Upper left, relative expression levels of *PVB_3 K03611.1* in WT and overexpression lines. N.D. indicates not detected; the remaining three figures depict the activities of SOD and POD, as well as the MDA content in WT and overexpressing rice lines after 24 h of cold stress at 4 °C. * indicates $P < 0.05$ and *** indicates $P < 0.005$. Fig. S18. Principal component analysis and population structure analysis based on SNPs and SVs. a PCA based on SNPs. b Population structure analysis based on SNPs. c PCA based on SVs. d Population structure analysis based on SVs. Fig. S19. The proportion of DEGs in different tissues under cold stress for candidate genes identified in the selective sweep analysis and GWAS. a The proportion of differential expression of genes in the four cold stress groups based on SNP-based selective sweep analysis. b The proportion of differential expression of genes in the four cold stress groups based on SV-based selective sweep analysis. c The proportion of differential expression of genes identified in SNP-GWAS across the four cold stress groups. d The proportion of differential expression of genes identified in SV-GWAS across the four cold stress groups. The suffix "L" in the group name represents leaves, and "R" represents roots.

Additional file 2: Tables S1-S24. Table S1. Summary of sequencing data from our study. Table S2. Results of Survey analysis. Table S3. Summary of two haplotype genomes. Table S4. Chromosome length of two haplotype genomes. Table S5. Peaks of each Ks distribution of orthologs in JJ31 genomes. Table S6. Gene density in different subgenomes of JJ31. Table S7. TE density in different subgenomes of JJ31. Table S8. Number of upregulated genes in different subgenomes of JJ31. Table S9. Genes with biased expression in the two subgenomes. Table S10. The number of DEGs in the K and N subgenomes across different transcriptome samples. Table S11. Information of cold-responsive gene families. Table S12. The *COR* genes specifically differentially expressed in JJ31. Table S13. The *COR* genes specifically differentially expressed in Alamo. Table S14. The ASE genes identified in the 23 transcriptome samples. Table S15. ASE genes under positive and purifying selection between two haplotypes. Table S16. ASE genes in leaves and roots of JJ31 under cold stress. Table S17. ASE genes in leaves and roots of Alamo under cold stress. Table S18. Cold-induced ASE genes in JJ31 and Alamo. Table S19. Summary of 340 switchgrass accessions from PRJNA622568 dataset. Table S20. Length distribution of SVs in different categories. Table S21. The Log₂ expression fold change of *COR* genes identified in selective sweeps under cold stress conditions. Table S22. Summary of SNP-GWAS results related to BRKG_SRV. Table S23. Summary of SV-GWAS results related to BRKG_SRV. Table S24. Primer design.

Additional file 3: Notes S1-S3. Note S1. Identification of specifically differentially expressed *COR* genes in the two ecotypes. Note S2. ASE analysis under cold stress. Note S3. SV characteristics in switchgrass populations.

Acknowledgements

This research was supported by High-performance Computing Platform of Sichuan Agricultural University.

Peer review information

David Edwards and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

L.H. and B.W. designed and managed the project. B.W., D.L., Y.Y., X.M., J.Z., and J.J. participated in material collecting and processing. B.W. and H.Y. performed bioinformatics analyses. B.W. and D.L. wrote the manuscript. D.L., X.M., and J.Z. contributed to validation works. L.H., M.H., B.Z., B.X., J.W., M.S., Z.X., and X.W. revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the China Agriculture Research System of MOF and MARA, the Sichuan Province Research Grant (2021YFYZ0013), the Modern Agricultural Industry System Sichuan Forage Innovation Team (SCCXTD-2020-16), and the National Natural Science Foundation of China (Nos. 31771866, 32071867 and 32301481).

Data availability

The germplasm resource of JJ31 has been submitted to the Tropical Crop Germplasm Resources Information Platform under the germplasm number 29p0010319000870 (<https://ctcgris.cn/ctbigdata/index.php/Home/detail?id=167,355&from>). The two haplotype genome assemblies, along with the raw resequencing data, Hi-C data, and RNA-seq data generated in this study, have been deposited at the China National Genomics Data Center (NGDC, <https://ngdc.cncb.ac.cn/>) with BioProject ID PRJCA023812 [122]. The published RNA-seq data from different tissues of switchgrass used in this study were downloaded from NCBI under the BioProject ID PRJNA265584 [123]. The published resequencing data used in this study were downloaded from NCBI under the BioProject ID PRJNA622568 [124]. No other scripts and software were used other than those mentioned in the Methods section.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 September 2024 Accepted: 6 May 2025

Published online: 14 May 2025

References

- McLaughlin SB, Kszos LA. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenerg.* 2005;28:515–35.
- Lewandowski I, Scurlock JM, Lindvall E, Christou M. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass Bioenerg.* 2003;25:335–61.
- Wu B, Zhu J, Ma X, Jia J, Luo D, Ding Q, et al. Comparative analysis of switchgrass chloroplast genomes provides insights into identification, phylogenetic relationships and evolution of different ecotypes. *Ind Crops Prod.* 2023;205:117570.
- Porter CL Jr. An analysis of variation between upland and lowland switchgrass, *Panicum virgatum* L., in central Oklahoma. *Ecology.* 1966;47:980–92.
- Childs KL, Nandety A, Hirsch CN, Góngora-Castillo E, Schmutz J, Kaeppler SM, et al. Generation of transcript assemblies and identification of single nucleotide polymorphisms from seven lowland and upland cultivars of switchgrass. *Plant Genome.* 2014;7:plantgenome2013.2012.0041.
- Sinha S, Kukreja B, Arora P, Sharma M, Pandey GK, Agarwal M, et al. The omics of cold stress responses in plants. *Elucidation Abiotic Stress Signal Plants: Funct Genomics Perspect.* 2015;2:143–94.
- Knight JC. Allele-specific gene expression uncovered. *Trends Genet.* 2004;20:113–6.
- Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS. Allelic variation of gene expression in maize hybrids. *Plant Cell.* 2004;16:1707–16.
- Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, et al. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature.* 2009;457:327–31.
- Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, et al. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature.* 2021;590:438–44.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170–5.
- Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020;21:1–27.
- Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 2018;46:e126–e126.
- Casler MD, Tobias CM, Kaeppler SM, Buell CR, Wang ZY, Cao P, et al. The switchgrass genome: tools and strategies. *Plant Genome.* 2011;4:273–82.
- Poudel HP, Sanciangco MD, Kaeppler SM, Buell CR, Casler MD. Genomic prediction for winter survival of lowland switchgrass in the northern USA. *G3: Genes, Genomes, Genetics.* 2019;9:1921–31.
- Adhikari L, Baral R, Paudel D, Min D, Makaju SO, Poudel HP, et al. Cold stress in plants: strategies to improve cold tolerance in forage species. *Plant Stress.* 2022;4:100081.
- Zhu JK. Abiotic stress signaling and responses in plants. *Cell.* 2016;167:313–24.
- Ma Y, Dai X, Xu Y, Luo W, Zheng X, Zeng D, et al. COLD1 confers chilling tolerance in rice. *Cell.* 2015;160:1209–21.
- Bajwa VS, Shukla MR, Sherif SM, Murch SJ, Saxena PK. Role of melatonin in alleviating cold stress in *Arabidopsis thaliana*. *J Pineal Res.* 2014;56:238–45.
- Ali U, Li H, Wang X, Guo L. Emerging roles of sphingolipid signaling in plant response to biotic and abiotic stresses. *Mol Plant.* 2018;11:1328–43.
- Huby E, Napier JA, Baillieul F, Michaelson LV, Dhondt-Cordelier S. Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytol.* 2020;225:659–70.
- Tuteja N, Sopory SK. Plant signaling in stress: G-protein coupled receptors, heterotrimeric G-proteins and signal coupling via phospholipases. *Plant Signal Behav.* 2008;3:79–86.
- Polisensky DH, Braam J. Cold-shock regulation of the *Arabidopsis* TCH genes and the effects of modulating intracellular calcium levels. *Plant Physiol.* 1996;111:1271–9.
- Yang T, Shad Ali G, Yang L, Du L, Reddy A, Poovaiah B. Calcium/calmodulin-regulated receptor-like kinase CRLK1 interacts with MEKK1 in plants. *Plant Signal Behav.* 2010;5:991–4.
- Lee C-M, Thomashow MF. Photoperiodic regulation of the C-repeat binding factor (CBF) cold acclimation pathway and freezing tolerance in *Arabidopsis thaliana*. *Proc Natl Acad Sci.* 2012;109:15054–9.
- Jiang B, Shi Y, Zhang X, Xin X, Qi L, Guo H, et al. PIF3 is a negative regulator of the CBF pathway and freezing tolerance in *Arabidopsis*. *Proc Natl Acad Sci.* 2017;114:E6695–702.
- Dong MA, Farré EM, Thomashow MF. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. *Proc Natl Acad Sci.* 2011;108:7241–6.

28. Catalá R, Medina J, Salinas J. Integration of low temperature and light signaling during cold acclimation response in *Arabidopsis*. *Proc Natl Acad Sci*. 2011;108:16475–80.
29. Raghavendra AS, Gonugunta VK, Christmann A, Grill E. ABA perception and signalling. *Trends Plant Sci*. 2010;15:395–401.
30. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51:739–48.
31. Song XM, Wang JP, Sun PC, Ma X, Yang QH, Hu JJ, et al. Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. *Hortic Res*. 2020;7:20.
32. Zhang X, Wu C, Guo Y, Ren X, Meng Y, Gao Q, et al. Genome-wide analysis elucidates the roles of *GhTIR1/AFB* genes reveals the function of *Gh_D08G0763 (GhTIR1)* in cold stress in *G. hirsutum*. *Plants*. 2024;13:1152.
33. Suh JY, Kim WT. *Arabidopsis* RING E3 ubiquitin ligase AtATL80 is negatively involved in phosphate mobilization and cold stress response in sufficient phosphate growth conditions. *Biochem Biophys Res Commun*. 2015;463:793–9.
34. Zhu J, Verslues PE, Zheng X, Lee BH, Zhan X, Manabe Y, et al. HOS10 encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants. *Proc Natl Acad Sci*. 2005;102:9966–71.
35. Kant P, Kant S, Gordon M, Shaked R, Barak S. STRESS RESPONSE SUPPRESSOR1 and STRESS RESPONSE SUPPRESSOR2, two DEAD-box RNA helicases that attenuate *Arabidopsis* responses to multiple abiotic stresses. *Plant Physiol*. 2007;145:814–30.
36. Kidokoro S, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulatory network of plant cold-stress responses. *Trends Plant Sci*. 2022;27:922–35.
37. Ding Y, Yang S. Surviving and thriving: how plants perceive and respond to temperature stress. *Dev Cell*. 2022;57:947–58.
38. Ding Y, Shi Y, Yang S. Molecular regulation of plant responses to environmental temperatures. *Mol Plant*. 2020;13:544–64.
39. Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020;52:1423–32.
40. Payne JL, Wagner A. The causes of evolvability and their evolution. *Nat Rev Genet*. 2019;20:24–38.
41. Zuo C, Blow M, Sreedasyam A, Kuo RC, Ramamoorthy GK, Torres-Jerez I, et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Biotechnol Biofuels*. 2018;11:1–15.
42. Hu G, Feng J, Xiang X, Wang J, Salojärvi J, Liu C, et al. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nat Genet*. 2022;54:73–83.
43. Sanderson MA, Adler PR, Boateng AA, Casler MD, Sarath G. Switchgrass as a biofuels feedstock in the USA. *Can J Plant Sci*. 2006;86:1315–25.
44. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet*. 2013;9:e1003215.
45. Martinez-Reyna J, Vogel KP, Caha C, Lee DJ. Meiotic stability, chloroplast DNA polymorphisms, and morphological traits of upland × lowland switchgrass reciprocal hybrids. *Crop Sci*. 2001;41:1579–83.
46. Renny-Byfield S, Gong L, Gallagher JP, Wendel JF. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol Biol Evol*. 2015;32:1063–71.
47. Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants*. 2018;4:90–7.
48. Shao L, Xing F, Xu C, Zhang Q, Che J, Wang X, et al. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc Natl Acad Sci*. 2019;116:5653–8.
49. Zhang X, Chen S, Shi L, Gong D, Zhang S, Zhao Q, et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat Genet*. 2021;53:1250–9.
50. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat Genet*. 2018;50:1565–73.
51. Huang C, Cheng Y, Hu Y, Fang L, Si Z, Chen J, et al. Dynamic patterns of gene expressional and regulatory variations in cotton heterosis. *Front Plant Sci*. 2024;15:1450963.
52. Casler MD, Vogel KP, Harrison M. Switchgrass germplasm resources. *Crop Sci*. 2015;55:2463–78.
53. Bourke PM, Van Geest G, Voorrips RE, Jansen J, Kranenburg T, Shahin A, et al. polymapR—linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. *Bioinformatics*. 2018;34:3496–502.
54. Tang M, Zhang W, Lin R, Li L, He L, Yu J, et al. Genome-wide characterization of cytochrome P450 genes reveals the potential roles in fruit ripening and response to cold stress in tomato. *Physiol Plant*. 2024;176:e14332.
55. Chakraborty P, Biswas A, Dey S, Bhattacharjee T, Chakraborty S. Cytochrome P450 gene families: role in plant secondary metabolites production and plant defense. *J Xenobiotics*. 2023;13:402–23.
56. Chopra R, Burrow G, Hayes C, Emendack Y, Xin Z, Burke J. Transcriptome profiling and validation of gene based single nucleotide polymorphisms (SNPs) in sorghum genotypes with contrasting responses to cold stress. *BMC Genomics*. 2015;16:1040.
57. Tao X, Wang MX, Dai Y, Wang Y, Fan YF, Mao P, et al. Identification and expression profile of CYPome in perennial ryegrass and tall fescue in response to temperature stress. *Front Plant Sci*. 2017;8:1519.
58. Wu H, Fan X, Wu J, Sechenbaater, Teng W, Zhang H, et al. Induction of octoploid and its phenotypic and physiological analysis in switchgrass “Jingji31”. *World J For*. 2019;8:10–16. (Chinese).
59. Hardin CF, Fu C, Hisano H, Xiao X, Shen H, Stewart CN, et al. Standardization of switchgrass sample collection for cell wall and biomass trait analysis. *BioEnergy Res*. 2013;6:755–62.
60. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
61. Yan H, Sun M, Zhang Z, Jin Y, Zhang A, Lin C, et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat Genet*. 2023;55:507–18.
62. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–70.

63. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33:2202–4.
64. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
65. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3:95–8.
66. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
67. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3:99–101.
68. Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res*. 2020;48:W177–84.
69. Seppy M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Gene prediction: methods and protocols*. 2019:227–245.
70. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007;23:1061–7.
71. Chen N. Using repeat masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2004;5:4.10. 11–14.10. 14.
72. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005;110:462–7.
73. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*. 2007;35:W265–8.
74. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics*. 2005;21:i152.
75. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. *Bioinformatics*. 2005;21:i351–8.
76. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci*. 2020;117:9451–7.
77. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005;33:W465–7.
78. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997;268:78–94.
79. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene finders. *Bioinformatics*. 2004;20:2878–9.
80. Guigo R. Assembling genes from predicted exons in linear time with dynamic programming. *J Comput Biol*. 1998;5:681–702.
81. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24:2938–9.
82. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
83. Yu XJ, Zheng HK, Wang J, Wang W, Su B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics*. 2006;88:745–51.
84. Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res*. 2004;14:988–95.
85. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
86. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
87. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
88. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9:1–22.
89. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–66.
90. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 2003;31:365–70.
91. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res*. 2017;45:D190–9.
92. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42:D222–30.
93. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 2004;32:D277–80.
94. Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32:D258–61.
95. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29:2933–5.
96. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res*. 2021;49:D192–200.
97. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49:9077–96.
98. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14.
99. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.

100. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
101. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–9.
102. Moniz de Sa M, Drouin G. Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol*. 1996;13:1198–1212.
103. Huang L, Feng G, Yan H, Zhang Z, Bushman BS, Wang J, et al. Genome assembly provides insights into the genome evolution and flowering regulation of orchardgrass. *Plant Biotechnol J*. 2020;18:373–88.
104. Xie Z, Lin W, Yu G, Cheng Q, Xu B, Huang B. Improved cold tolerance in switchgrass by a novel CCCH-type zinc finger transcription factor gene, *PvC3H72*, associated with ICE1–CBF–COR regulon and ABA-responsive genes. *Biotechnol Biofuels*. 2019;12:1–11.
105. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7.
106. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21.
107. Smart RE, Bingham GE. Rapid estimates of relative water content. *Plant Physiol*. 1974;53:258–60.
108. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13:1194–202.
109. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:e49–e49.
110. Hoopes G, Meng X, Hamilton JP, Achakkagari SR, Guesdes FdAF, Bolger ME, et al. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol Plant*. 2022;15:520–536.
111. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
112. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
113. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32:1220–2.
114. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28:i333–9.
115. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15:1–19.
116. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods*. 2015;12:966–8.
117. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 2017;8:14061.
118. Guo J, Cao K, Deng C, Li Y, Zhu G, Fang W, et al. An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol*. 2020;21:1–19.
119. Chao J, Wu S, Shi M, Xu X, Gao Q, Du H, et al. Genomic insight into domestication of rubber tree. *Nat Commun*. 2023;14:4651.
120. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44:821–4.
121. Hiei Y, Ohta S, Komari T, Kumashiro T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J*. 1994;6:271–82.
122. Wu BC, Luo D, Yue YS, Yan HD, He M, Ma XX, Zhao BY, Xu B, Zhu J, Wang J, Jia JY, Sun M, Xie Zn, Wang XS, Huang LK. New insights into the cold tolerance of upland switchgrass by integrating a haplotype-resolved genome and multi-omics analysis. *Datasets*. Genome sequence archive. 2025. <https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA023812>.
123. Zuo CM, Blow M, Sreedasyam A, Kuo RC, Ramamoorthy GK, Torres-Jerez I, et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Datasets*. Genome sequence archive. 2014. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA265584>.
124. Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, et al. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Datasets*. Genome sequence archive. 2020. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA622568>.

Publisher's Note

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