



OPEN ACCESS

EDITED BY

Michael L. Moody,
The University of Texas at El Paso,
United States

REVIEWED BY

Michele Wyler,
MWSchmid GmbH, Switzerland
Jingjing Wang,
Jiangsu Province and Chinese Academy of
Sciences, China

*CORRESPONDENCE

Lijun Yan
✉ yanlijun456@126.com
Xiao Ma
✉ maroar@126.com

†These authors have contributed equally to
this work

RECEIVED 01 July 2024

ACCEPTED 02 September 2024

PUBLISHED 19 September 2024

CITATION

Xiong Y, Li D, Liu T, Xiong Y, Yu Q, Lei X,
Zhao J, Yan L and Ma X (2024) Extensive
transcriptome data providing great efficacy in
genetic research and adaptive gene
discovery: a case study of *Elymus sibiricus* L.
(Poaceae, Triticeae).
Front. Plant Sci. 15:1457980.
doi: 10.3389/fpls.2024.1457980

COPYRIGHT

© 2024 Xiong, Li, Liu, Xiong, Yu, Lei, Zhao, Yan
and Ma. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Extensive transcriptome data providing great efficacy in genetic research and adaptive gene discovery: a case study of *Elymus sibiricus* L. (Poaceae, Triticeae)

Yanli Xiong^{1†}, Daxu Li^{2†}, Tianqi Liu¹, Yi Xiong¹, Qingqing Yu²,
Xiong Lei², Junming Zhao¹, Lijun Yan^{2*} and Xiao Ma^{1*}

¹College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu,
Sichuan, China, ²Sichuan Academy of Grassland Sciences, Chengdu, Sichuan, China

Genetic markers play a central role in understanding genetic diversity, speciation, evolutionary processes, and how species respond to environmental stresses. However, conventional molecular markers are less effective when studying polyploid species with large genomes. In this study, we compared gene expression levels in 101 accessions of *Elymus sibiricus*, a widely distributed allotetraploid forage species across the Eurasian continent. A total of 20,273 high quality transcriptomic SNPs were identified. In addition, 72,344 evolutionary information loci of these accessions of *E. sibiricus* were identified using genome skimming data in conjunction with the assembled composite genome. The population structure results suggest that transcriptome SNPs were more effective than SNPs derived from genome skimming data in revealing the population structure of *E. sibiricus* from different locations, and also outperformed gene expression levels. Compared with transcriptome SNPs, the investigation of population-specifically-expressed genes (PSEGs) using expression levels revealed a larger number of locally adapted genes mainly involved in the ion response process in the Sichuan, Inner Mongolia, and Xizang geographical groups. Furthermore, we performed the weighted gene co-expression network analysis (WGCNA) and successfully identified potential regulators of PSEGs. Therefore, for species lacking genomic information, the use of transcriptome SNPs is an efficient approach to perform population structure analysis. In addition, analyzing genes under selection through nucleotide diversity and genetic differentiation index analysis based on transcriptome SNPs, and exploring PSEG through expression levels is an effective method for analyzing locally adaptive genes.

KEYWORDS

adaptability, transcriptome, genome skimming, SNPs, WGCNA, *Elymus sibiricus*

Introduction

Local adaptation occurs when divergent selection outweighs random genetic drift and the equalizing effect of gene flow between populations (Kawecki and Ebert, 2004). These premises suggest that certain widely distributed wild plants show distinctive local adaptations caused by restricted gene flow between populations due to geographical and environmental isolation. Multiple selection factors interact with geographic distance between populations to shape the evolutionary dynamics of species by restricting gene flow (López-Goldar and Agrawal, 2021). In addition to the molecular perspective and adaptive effects, a comprehensive study of local adaptation in species should also consider ecological factors and the co-evolution of plants and their environment. For example, local adaptation along natural gradients such as latitude and altitude, which have been extensively studied, may be the result of the combined influences of different ecological effects on selection and gene flow (López-Goldar and Agrawal, 2021; Tiffin and Ross-Ibarra, 2014).

Compared to neutral regions, where selection is not expected, regions of genes and linked loci under divergent selection are expected to experience stronger selection pressures (Limborg et al., 2012). The adaptive significance of existing polymorphic sites is often difficult to estimate, as they may not be the direct targets of selection, but rather by-products of hitchhiking effects (Zhang et al., 2024). Large-scale transcriptome sequencing can help us study functional genetic variation, thereby increasing the chances of detecting natural selection, as functional genetic variation is expected to be more directly influenced by natural selection and result in a population-specific expression (Tang et al., 2021). The regulation of gene expression is a complex and intricate process, and studies of the evolution of gene expression have shown significant variability in the variability of regulatory elements both within and between natural populations (Shi et al., 2012). However, studying adaptive evolution at the level of gene expression is only relevant for specific genes or sets of co-expression gene. Given that the environment always influences gene expression, this implies that general evolutionary patterns of gene expression are not necessarily adaptive (Signor and Nuzhdin, 2018). Common garden experiments test how heritable traits are influenced by natural selection (leading to natural variation) and ultimately affect optimal outcomes for population persistence (Schwinning et al., 2022). The natural variation observed could be attributed to the prolonged acclimatization processes that occur in response to the particular environmental conditions of the sites where the species occur.

Elymus sibiricus, the model species of the genus *Elymus*, is widely distributed across the Eurasian continent and exhibits significant genetic diversity and phenotypic variance (Xiong et al., 2021). The extensive habitat diversity has contributed to the unique adaptability of *E. sibiricus* ecotypes across different geographic groups (Han et al., 2022). Due to its remarkable adaptability to the high altitude environment, superior nutritional value and ease of cultivation, *E. sibiricus* is the dominant forage crop and is extensively cultivated in the artificial forage grasslands of the Qinghai-Tibetan Plateau. Nevertheless, the unavailability of the reference genome hampers genetic research on *E. sibiricus*,

resulting in previous studies relying solely on molecular markers such as genomic SSRs (Xiong et al., 2021; Lei et al., 2014), EST-SSRs (Zhang et al., 2019; Zheng et al., 2020), and chloroplast/mitochondrial SSRs (Lei et al., 2014; Xiong et al., 2022), to investigate the population structure and adaptive evolution of the species. Due to their limited coverage, these markers can only reveal specific genetic variations. Considering their larger number and wider distribution, SNP markers are more suitable for population genetic research (Sun et al., 2020). Some SNPs derived from simplified genome sequencing have been used to study the population structure of *E. sibiricus* (Han et al., 2022). However, a significant proportion of these markers are located in non-functional regions. Furthermore, mapping SNPs located in genic regions to specific genes is challenging due to the lack of comprehensive reference genomic information. In this context, the use of transcriptome data allows a thorough investigation of functional regions and genes associated with adaptive traits, and the detection of sequence variation and expression levels within these regions. SNP-based studies of the transcriptome can provide a wealth of SNP markers, and these variants can affect the coding sequence of proteins, thus affecting gene expression and function. In addition, the abundance of SNP markers in the transcriptome allows accurate assessment of genetic variation and population structure (Yang, 1998). However, issues of widespread purifying selection in the transcriptome and concerns about sample size point to the need for comparisons between mRNA and established DNA-based methods (Thorstensen et al., 2020; Pratlong et al., 2015; Thorstensen et al., 2021). Genome skimming sequencing is a cost-effective protocol for obtaining genome-level variation and has shown excellent potential for phylogenetics, species identification, and genetic diversity assessment (Richter et al., 2015; Trevisan et al., 2019). Furthermore, SISRS software allows the identification of phylogenetically informative sites across the genome by skimming sequencing data without reference genome information (Schwartz et al., 2015; Litterman et al., 2023), allowing us to perform population variation calling at the DNA level.

This study included 101 wild *E. sibiricus* accessions from both domestic and international sources, covering the primary distribution regions of *E. sibiricus*. The transcriptomes of these collected accessions were sequenced, and the full-length transcriptome of *E. sibiricus* was assembled using previous research data to construct a reference sequence. The aims of this study included: (1) studying population genetic analysis using transcriptome SNPs and expression levels; (2) analyzing population (geo-group)-specific expression genes (PSEG), investigating genes related to adaptability of *E. sibiricus*; (3) using genome skimming data from 101 accessions as reported by Xiong et al. (2023) to identify homologous sequences at the genome level (Schwartz et al., 2015) and compare their effectiveness in population structure studies with transcriptome data (both transcriptome SNPs and expression levels). This study provides a comprehensive analysis of the population structure of *E. sibiricus*, including transcriptome SNPs, expression levels and genome-level homologous sequences, which could provide a genetic research mode in polyploid species without genomic information. It will also identify the functional genes that contribute to the differentiation of geographic groups of *E. sibiricus*.

Results

Full-length transcriptome assembly, SNPs calling from the transcriptome and genome skimming data

The reference transcriptome of *E. sibiricus* was assembled, comprising 6982 contigs with an N50 of 42,573 bp and a GC content of 48.08% (Table 1). Using this reference transcriptome, 20,273 high quality SNPs were identified from 101 *E. sibiricus* accessions. Furthermore, the genome skimming data were used to assemble the *E. sibiricus* composite genome, which consisted of 33,234 contigs with a total length of more than 3.8 million bp, covering more than 0.05% of the 6.8 Gb nuclear genome. Among these contigs, 1525 had a length of more than 200 bp. A total of 72,344 SNPs were identified based on genome skimming data from 101 *E. sibiricus* accessions.

SNPs derived from the transcriptome data exhibited higher discriminatory ability for various geo-groups of *E. sibiricus*

The PCA plots and phylogenetic trees based on three data sets (transcriptome SNPs, the expression levels, and SNPs-GS) were combined to reveal the population structure of the tested *E. sibiricus* accessions (Figures 1, 2). In general, a greater discriminatory power of SNP data (both transcriptome SNPs and SNPs-GS) to discriminate different geo-groups was observed in PCA plots. Specifically, the PCA plot constructed using SNPs-GS could distinguish several *E. sibiricus* accessions from the QTP (Figure 1B), while the PCA plot constructed using transcriptome SNPs could distinguish more *E. sibiricus* accessions from the QTP and other regions (Figure 1C). However, based on expression levels, the tested *E. sibiricus* accessions had a mixed population structure (Figure 1A). In addition, the phylogenetic analysis also showed a greater discriminatory ability of transcriptome SNPs for distinguishing *E. sibiricus* accessions from different geo-groups (Figure 2), indicating the suitability of SNP data from transcriptome sequencing for further population structure analysis compared to expression levels or SNPs-GS.

Genetic diversity indices using transcriptome SNPs

As transcriptomic SNPs were more effective in revealing the population structure of *E. sibiricus* (Figures 1, 2), we subsequently used transcriptomic SNPs to perform further genetic studies. In

summary, XJ and SC/XZ had the highest (0.3248) and lowest (0.3186) MAF values (Table 2), respectively. For O_het, except for the QH geo-group, all groups had values greater than 0.6, with the MGL geo-group having the highest value (0.625). However, the QH geo-group had the highest value of E_het (0.3894). The O_het values of all the geo-groups were greater than the E_het values.

Gene flow and differentiation among geo-groups based on transcriptome SNPs

The result of the gene flow analysis showed that the optimal K value (migration edges) was eight, indicating the presence of eight gene flow events between the nine geo-groups of *E. sibiricus* (Supplementary Figure S1). This was consistent with the generally low *F*_{st} values between these geo-groups (Supplementary Figure S2). The SC geo-group had the highest number of gene flow events, including gene flows from QH and RUSS to SC, and SC to NM in the early stages (Figure 3). Only a single gene flow event was detected for the Canada, NM, KAZA, and XJ geo-groups. In contrast, no gene flow was observed between MGL and the remaining geo-groups.

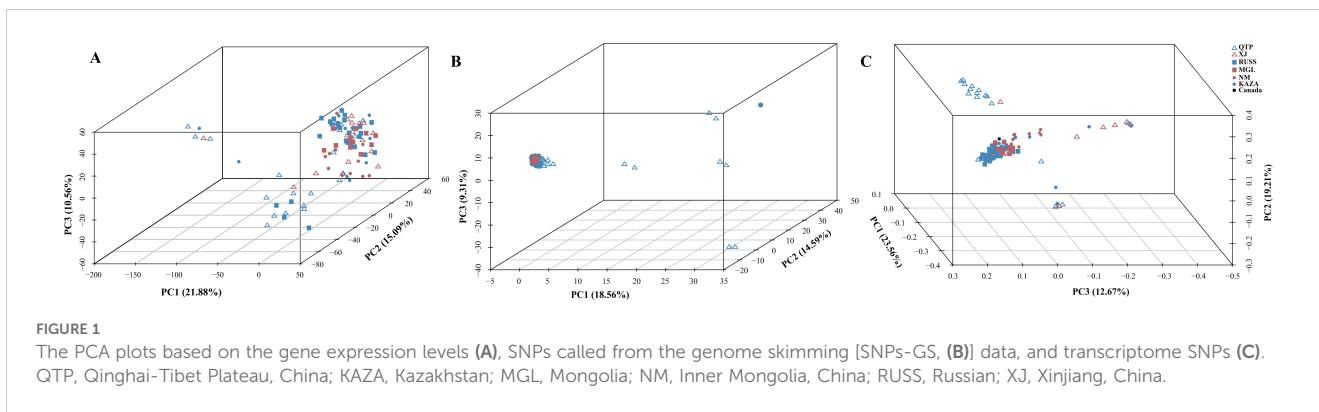
We further integrated the *P*_i values for each geo-group and the pairwise *F*_{st} values to identify the selected regions and associated genes. As the Canada geo-group contained only a single accession, it was excluded from this analysis. A total of 12, 267, 3, 81, 1, 54, 824, and 259 genes were identified in the XZ, XJ, SC, RUSS, QH, NM, MGL, and KAZA geo-groups, respectively. Three and one genes selected in the SC and QH geo-groups, respectively, could not be annotated in the GO database. The selected genes from all six geo-groups were involved in the “development” and “metabolic” processes (Figure 4). With the exception of the NM and KAZA geo-groups, the genes under selection in other geo-groups were also involved in the homeostasis process. In addition, the genes under selection in each geo-group showed responsiveness to specific stresses or stimuli, potentially contributing to the cumulative variation experienced during the long-term adaptation process to environmental factors.

Population differentiation from the perspective of gene expression

Using the expression levels of all genes cannot effectively resolve the population structure of *E. sibiricus* (Figures 1A, 2A). Therefore, we attempted to analyze genes specifically expressed in each geo-group to investigate population differentiation from the standpoint of gene expression. A total of 414, 1899, 1797, 498, 1529, 3295, 556, and 4499 genes were specifically expressed in KAZA, MGL, NM, QH, RUSS, SC, XJ, and XZ geo-groups, respectively. The heatmap

TABLE 1 The quality assessment of the assembled reference transcriptome of *E. sibiricus*.

Number of contig	Largest contig	contigs (>= 50 kbp)	N50	L50	N's per 100 kbp	GC content
6982	170,783	342	42,573	448	0	48.08%



results indicated that the specifically expressed genes in each geo-group showed a significant difference in expression profiles (Supplementary Figure S3).

We also performed GSVA (gene set variation analysis) to identify differentially enriched gene sets in each geo-group compared to the others. Several GO terms related to plant energy metabolism processes were identified, including GO:0052855 and GO:0047453, which are associated with NAD(P)H-hydrate dehydratase activity and were significantly up-regulated in the KAZA geo-group. In addition, GO:0033178 (related to ATPase complex) and GO:0006086, GO:0006637, and GO:0047617 (related to acyl-CoA metabolic process), were significantly upregulated in the MGL geo-group. Furthermore, several GO terms related to carbohydrate metabolism pathways were discovered, including GO:0009052, GO:0061615, and GO:0003872 in the MGL geo-group, GO:0019673, and GO:0008446 in the RUSS geo-group, GO:0006486, GO:0008378, and GO:0019255 in the SC geo-group, GO:0004576 in the XJ geo-group, GO:0019375 and GO:0047714 in the XZ geo-group. Although the plants were grown under controlled growth conditions, some GO terms related to stress tolerance were also detected for some geo-groups. For example, GO:0009408 and GO:0031990, which are associated with heat stress, were up-regulated in the KAZA geo-group, GO:0010225 (response to UV-C) and GO:0009411 (response to UV) were down-regulated in the MGL and SC geo-groups, respectively. GO:1901700

(response to oxygen-containing compound) and GO:0071215 (cellular response to abscisic acid stimulus) were downregulated in the SC geo-group. In particular, six GO terms (GO:0006833, GO:0015793, GO:0015254, GO:0015250, GO:0009992, and GO:0016798) related to water transport were significantly downregulated in the XJ geo-group.

The PSEG were further analyzed between two geo-groups. The analysis revealed that the XZ geo-group had the highest number of shared specifically expressed genes shared with the other geo-group, with a total of 1599 genes (Figure 5). GO annotation results indicated that some of these genes were involved in the response process to UV (Figure 5). In addition, the common specifically expressed genes in the QH, NM and SC geo-groups were found to be responsive to ion, especially in the SC geo-group where the genes were responsive to salt and involved in ion transport process.

WGCNA analysis can help identify potential regulators of PSEG

In this study, we performed a weighted gene co-expression network analysis (WGCNA) to identify the possible core regulatory genes in PSEG. The result showed that all the co-expressed genes could be assigned to 46 modules (Supplementary Figure S4). The regulatory networks in each model identified by WGCNA analysis

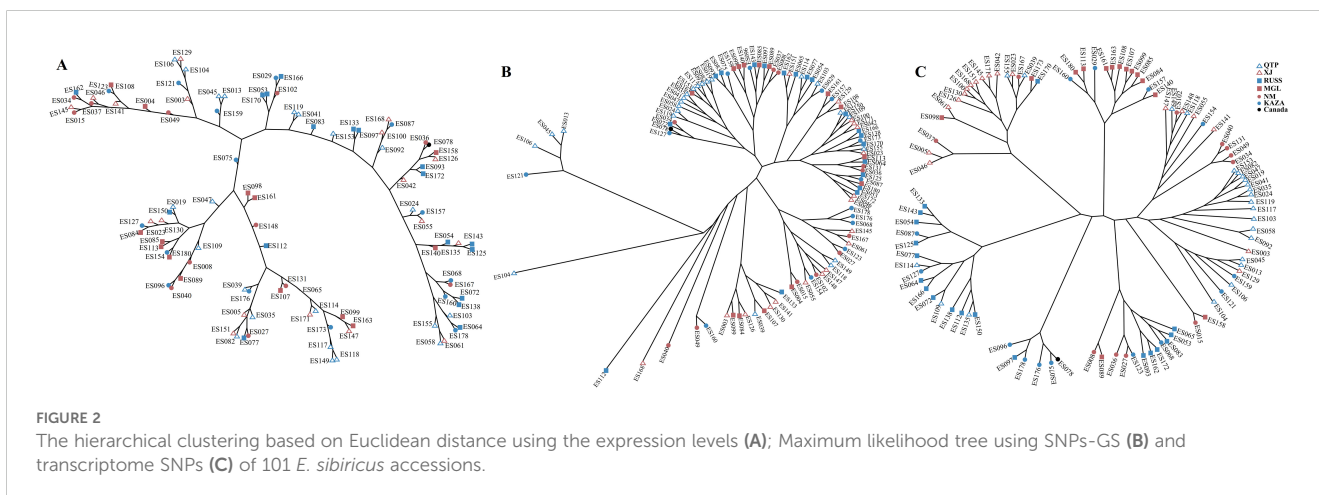
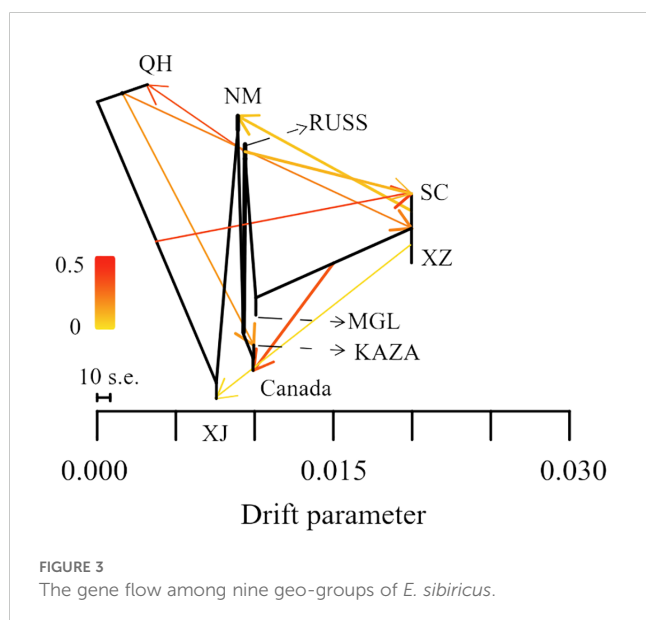


TABLE 2 Genetic diversity indices of each geo-group of *E. sibiricus*.

Geo-groups	MAF	O_het	E_het
Canada	0.3096	0.6191	0.3096
SC	0.3186	0.6188	0.3651
XJ	0.3248	0.6124	0.3824
RUSS	0.3237	0.6189	0.3786
QH	0.3218	0.5881	0.3894
NM	0.3240	0.6207	0.3758
MGL	0.3258	0.6250	0.3779
KAZA	0.3238	0.6012	0.3865
XZ	0.3186	0.6176	0.3605

MAF, minor allele frequency; O_het, observed heterozygosity; E_het, expected heterozygosity.

were constructed and an extensive regulatory relationship between TFs such as NAC, MYB, C2H2, and downstream genes was found in the black model (Figure 6A). Furthermore, we also identified the potential TFs that have the ability to regulate the expression of PSEG in each geo-group (Figure 6B). It is important to note that the genes in the QH, XJ, and NM geo-groups do not function as regulatory genes. Conversely, in the SC geo-group, we discovered two TFs (*MADS-box* and *NAC*) that play a regulatory role in the genes encoding the senescence-associated protease (*SAG39*) and the C2H2-finger domain. In addition, we identified more regulatory TFs in the XZ geo-group, including two genes belonging to the zinc finger family. Notably, our analysis also revealed that some PSEGs are regulated by multiple TFs. For example, one gene, *DNA damage-repair/tolerant (DRT)*, specifically expressed in the XZ geo-group, was regulated by six TFs. These findings highlight the importance of using transcriptome data to effectively identify potential population-specific regulatory genes that may be associated with their own local adaptation.



Discussion

To the best of our knowledge, this study represents the first attempt to assess the effectiveness of three methods - expression levels, transcriptome SNPs, and SNPs-GS - for analyzing population structure in a species lacking a reference genome, with a focus on *E. sibiricus*. The results showed that SNPs, especially transcriptome SNPs, were most effective in analyzing the population structure of *E. sibiricus*. This method identified numerous genes associated with local adaptation in *E. sibiricus*. In addition, although expression levels were less effective in population structure analysis, a significant number of PSEGs were also detected. Their annotations suggested their relevance to local environmental adaptation. Furthermore, by performing WGCNA analysis, some potential regulatory genes of the PSEGs were identified. Thus, we propose that in the absence of a reference genome, the use of transcriptome data (including both expression levels and SNP data) can reveal genes related to local adaptation and evolution of a species.

Transcriptome SNPs is an effective strategy in population analysis

SNP markers have several advantages over other markers, including ease of development and deployment. However, using SNP markers to study the population structure of certain species, particularly polyploidy species in the absence of reference genome information, can be challenging. Most previous studies have relied on genome skimming methods to obtain plastid sequences and nuclear ribosomal genes (Trevisan et al., 2019; Liu et al., 2021b). However, these approaches have only revealed limited genetic variation, which is insufficient for a comprehensive understanding of the true population structure (Del Valle et al., 2020). It is possible to jointly *de novo* assemble and identify evolutionarily informative sites from genome skimming data, as demonstrated for species identification in buffalo fish (Litterman et al., 2023) and phylogenomic reconstruction in *Leishmania* (Harkins et al., 2016). Notably, there are no reports of genome skimming data being used to identify evolutionarily informative sites within species. Given the significant variation among the wild *E. sibiricus* accessions used in our study, we attempted to use genome skimming data to detect evolutionarily informative sites and analyze the population structure of *E. sibiricus*. The assembled composite genome covered approximately 0.05% of the *E. sibiricus* nuclear genome (Xiong et al., 2021). Although the coverage of typical simplified genome sequencing is typically around 5% (Litterman et al., 2023), it remains a challenge to effectively assemble and align genome sequencing data, as well as complicating the annotation of variant sites, in the absence of the reference genome for population structure analysis. In addition, the PCA analysis distinguished only a few *E. sibiricus* accessions from Qinghai that showed significant differences from other accessions (Figure 1), while the phylogenetic tree results show that the accessions from the Qinghai-Tibetan Plateau (in purple) are generally clustered together, while the accessions from the

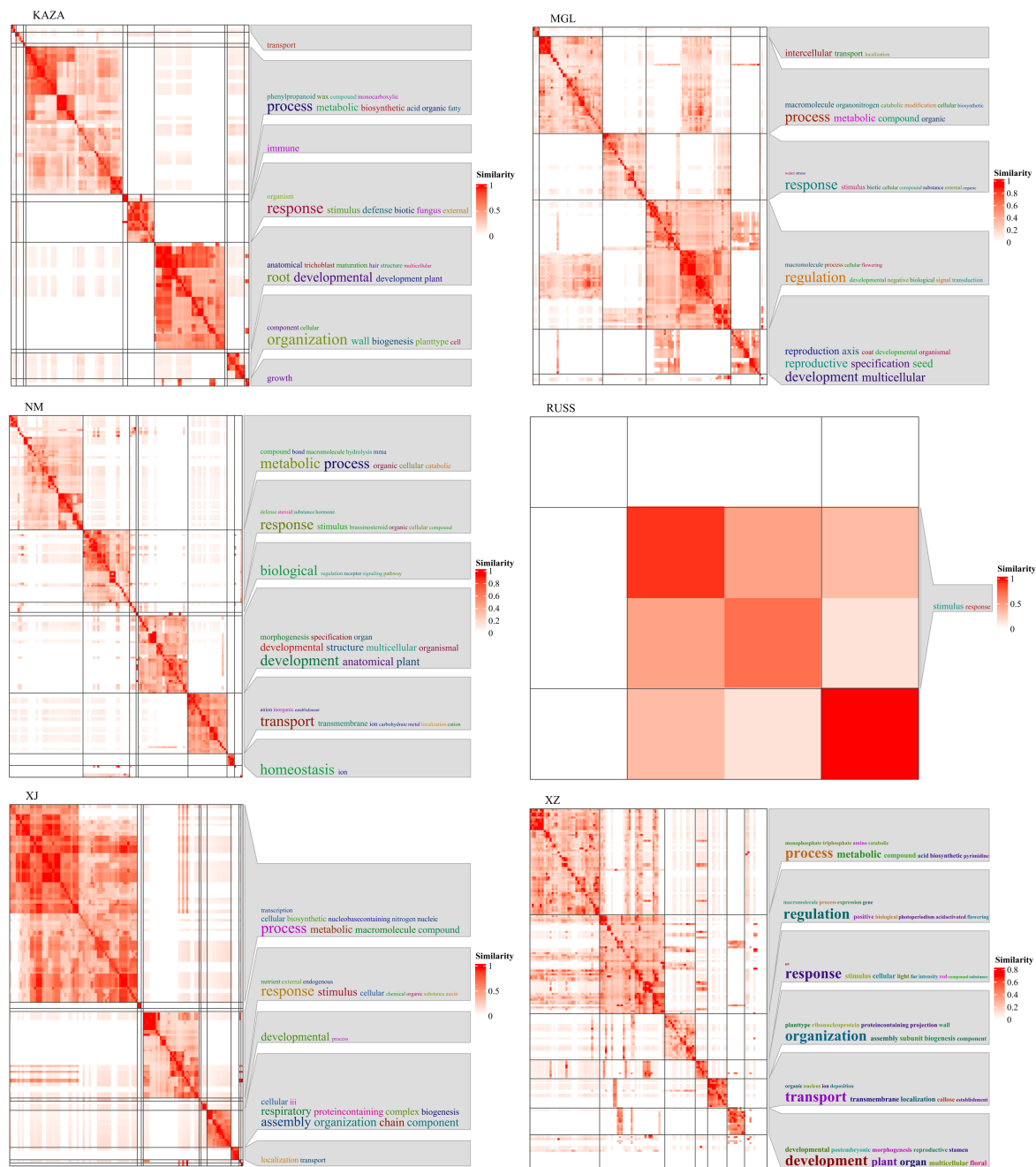


FIGURE 4 The GO annotation result of genes under selection in each geo-group of *E. sibiricus*.

Xinjiang are mainly clustered in three branches in the lower right corner (Figure 2). This suggests that in the absence of genome information, using genome skimming data to identify evolutionarily informative sites for population structure analysis is a viable alternative. However, it is essential to adequately increase the sequencing depth and propose more efficient algorithms for the analysis of short-read sequencing data to accurately infer genetic structure, especially for polyploid species with large genome sizes.

Data from coding genes allow for quality control steps that help to distinguish biological patterns from technical artefacts, enabling SNP analysis data to be tested against neutral expectations at multiple

levels, outperforming typical outlier approaches (De Wit et al., 2015). Many researchers (Wang et al., 2021; Mossion et al., 2022) have successfully developed transcriptome SNPs for the analysis of population structure and genetic diversity. Furthermore, the power of transcriptome SNPs to study intraspecific relationships in *Camellia sinensis* has been shown to exceed that of low-copy nuclear genes (Cheng et al., 2023). In this study, the higher effectiveness of transcriptome SNPs in deciphering the population structure of *E. sibiricus* from different locations compared to expression levels and SNPs-GS (Figures 1, 2) was also found. On the other hand, transcriptomic SNPs may be concentrated in regions of active gene



FIGURE 5
The Venn diagram of population-specifically-expressed genes between two geo-groups, and the GO annotation results of these shared specific-expressed-genes in each geo-group.

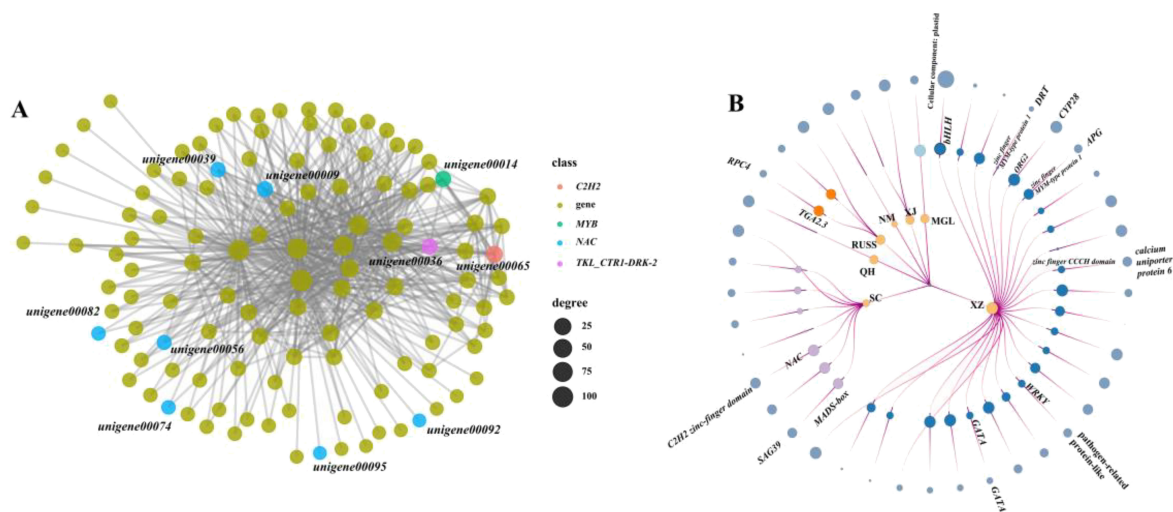


FIGURE 6
The node-link graph showing the regulating relationship between the population-specifically-expressed genes and the TFs. **(A)** regulatory network between TFs and genes in black model; **(B)** the node-link graph of PSEs, the circles from the inside to out represent geo-groups, TFs and population-specific genes, respectively. The circle size indicates the number of times that the genes regulating or be regulated.

expression where variation and recombination may be more frequent, so the present study detected frequent gene flow (Figure 3) and generally small pairwise *F_{st}* coefficients between geo-groups using transcriptomic SNPs (Supplementary Figure S4). This is consistent with a previous research using chloroplast markers (Xiong et al., 2022). However, this seems to differ from the results of previous studies using whole chloroplast genomes (Xiong et al., 2023) and nuclear makers (Han et al., 2022): their results show that gene flow between geo-groups of *E. sibiricus* is relatively low due to the existence of geographic barriers. This may be because transcriptome SNPs capture more recent or small-scale gene flow than chloroplast or nuclear DNA, as these data provide more detailed genetic information (Chen et al., 2021; Tepolt and Palumbi, 2015).

The combination of transcriptome SNPs and expression levels can help understanding adaptive mechanism

Transcriptome SNP can reduce complexity and provide more accurate functional annotation than reduced representation of genomic DNA libraries. This is considered an important method for identifying genes associated with local adaptation for population genomic analyses in non-model plants (Cheng et al., 2023; Liu et al., 2021a). In this study, the *P_i* and pairwise *F_{st}* values based on transcriptome SNPs were used to identify genes under selection within each geo-group. The majority of these genes are involved in the regulation of homeostasis, growth, and developmental processes (Figure 4). While some genes under selection in each geo-group are involved in various stress processes, only the NM geo-group had genes annotated as responding to ion stress processes. This is noteworthy because the NM geo-group is an area with severe saline alkali soils, and *E. sibiricus* from this region shows stronger salt tolerance (Chen et al., 2023). However, no adaptive genes related to local climatic or environmental characteristics (such as the arid climate of Xinjiang and the high altitude of Tibet) were detected in other geo-groups, possibly due to the frequent gene flow events between *E. sibiricus* geo-groups (Figure 3).

In addition to transcriptome SNPs, variation in gene expression levels is also an important source of adaptive traits, as gene transcription levels serve as the medium for gene coding information and final phenotypes (Alonso-Blanco et al., 2009; Bullard et al., 2010). Previous studies have shown that gene expression levels do not support the expected population structure compared to transcriptome variation (Wang et al., 2018), which is consistent with the present study. However, to understand the impact of genetic variation in gene expression levels and its effect on trait adaptation, it is essential to characterize expression level variation at the population scale (Fraser et al., 2011). In this study, transcript abundance was determined at the individual level using 101 wild *E. sibiricus* accessions and PSEGs were identified. In the SC, QH, XZ, and NM geo-groups, the number of PSEGs exceeded the number of selective genes identified by transcriptomic SNPs, and more genes related to local climate adaptation were discovered, such as ion transport (caused by the severe saline-alkaline soils of SC, QH, and NM) and

UV-B radiation (caused by the high altitude of XZ). As the expression levels of many genes controlling quantitative traits are often influenced by regulatory genes (i.e., TFs), this study used WGCNA analysis to explore some transcription factor family members, such as NAC and zinc finger protein, that could potentially regulate PSEGs. Identification of transcriptome SNPs and PSEGs associated with adaptation can shed light on the selective pressures faced by species in their natural environment and their effects on genotype and phenotype. The integration of genotype and regulatory network (especially functional network) is a prominent and inevitable trend in current research on adaptive evolution.

Materials and methods

Materials and transcriptome sequencing

A total of 101 accessions covering the main distribution regions of *E. sibiricus* were included in the present study, divided into seven geographical groups, namely MGL (Mongolia), RUSS (Russia), KAZA (Kazakhstan), Canada (Canada), XJ (Xinjiang, China), NM (Inner Mongolia, China), and QTP (Qinghai, China; Sichuan, China; and Xizang, China) (Supplementary Table S1). Accessions from MGL, RUSS, KAZA and Canada were collected from the National Plant Germplasm System (NPGS). The remaining accessions were collected by the research team according to the following rules: Spikelets were harvested from each individual plant, ensuring a minimum sampling distance of 50 meters between plants. Seeds of all accessions were planted in sand and grown in the greenhouse to minimize the influence of the environment on the expression of the different accessions. At the three-leaf stage, leaves from each accession were collected within one hour and immediately preserved in liquid nitrogen for subsequent RNA extraction. High quality RNA was purified and fragmented after assessment of purity, quantity, and RNA integrity. First-strand cDNA was then synthesized, followed by the addition of A-tailing and adaptor. The amplified products were then purified and paired-end sequenced to a read length of 150 bp on the Illumina platform. Finally, approximately 6 Gb of raw data was obtained for each accession, with each accession having three biological replicates.

The assembly and annotation of the full-length transcriptome, and SNP calling

The full-length *E. sibiricus* transcriptome bam file was obtained from Yu et al. (2023) and converted to a fasta file using the bam2fasta tool (part of the bambamv1.4tool-kit) (Page et al., 2014). This was used for assembly using flye v2.9 with the following parameters: 'flye -pacbio-raw input.fasta -out-dir out_pacbio' (Kolmogorov et al., 2019), which uses repeat graphs for the assembly of long and error-prone reads. The quality of the assembly was assessed using QUAST (Gurevich et al., 2013). The resulting transcript sequences were aligned using the NR (Pruitt et al., 2005), COG (Tatusov et al., 2000), KOG (Koonin et al., 2004),

KEGG (Kanehisa et al., 2004), and GO (Ashburner et al., 2000) databases with an *E*-value of 10⁻⁵. The transferred amino acid sequences were searched against the library of Pfam HMMs using *pfam_scan.pl* (Finn et al., 2014). Coding sequences (CDS) and transcription factors were then identified using the online software TransDecoder v3.0.0 (<https://github.com/TransDecoder/TransDecoder>) and ITAK v1.6 (Zheng et al., 2016), respectively.

Quality control was performed on the reads of the RNA sequencing data using the fastQC v0.12.0 software with the command 'fastqc *.fastq.gz' (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Low quality sequences (Phred < 20 and length < 50 bp) and adaptor sequences were trimmed using *trim_galore*. The clean high-quality reads were compared to the full-length transcriptome data using RSEM v1.1.17 (Li and Dewey, 2011) to calculate gene abundance using default settings, which was further standardized to FPKM (fragment per kb per million reads).

Raw RNA sequencing data from one replicate of each accession were aligned to the assembled full-length transcriptome using the BWA-MEM algorithm of *bwa* v0.7.17-r1188 (Li, 2013; Tarasov et al., 2015) and converted to bam files using *samtools* v1.18 with the command '*samtools view -b*' (Li et al., 2009), followed by the use of *sambamba* v0.5.9 (Li et al., 2009) to sort the resulting bam files and create index files using '*sambamba index*'. Finally, variant detection was performed using the HaplotypeCaller and GenotypeGVCFs with *gatk* v4.0 (McKenna et al., 2010). SNPs near indels were filtered, and the resulting vcf file was further filtered based on specific criteria: minor allele frequency (MAF) < 0.05, maximum missing data < 0.8, minimum depth of sequencing (minDP) < 2, maximum depth of sequencing (maxDP) < 1000, minimum Phred-scale quality score (minQ) < 30, minimum genotype quality score (minGQ) < 10. Finally, only biallelic loci are retained.

Genetic diversity and population structure analysis

SNPs obtained by mapping to the full-length transcriptome (transcriptome SNPs) were used to calculate MAF, observed heterozygosity (*H_o*), and expected heterozygosity (*H_e*) values for each geo-group using PLINK v1.90b4.6 software (Purcell et al., 2007). In addition, transcriptome SNPs were used to perform principal component analysis (PCA) using PLINK v1.90b4.6 (Purcell et al., 2007) with visualization using the *pca_plink_plot* R package. Expression levels were subjected to PCA analysis using the *prcomp* function and then visualized using the *scatterplot3d* R package (Mächler and Ligges, 2003). The vcf file was sorted using *tassel* v5.0 software (Bradbury et al., 2007) and then converted to Phylip format using *run_pipeline.pl*. This was then used to construct the maximum likelihood (ML) tree using *FastTree* v2.1 (Price et al., 2009). The hierarchical clustering tree was constructed based on the Euclidean distance using the *ggraph* R package (Si et al., 2020).

Using the vcf file containing all transcriptome variations, nucleotide diversity (*P_i*) values of each geo-group and pairwise

genetic differentiation index (*F_{st}*) were calculated using *vcftools* v0.1.16 (Danecek et al., 2011) with a window size of 1000 bp and a step size of 100 bp. Regions with the top 5% *F_{st}* values were identified as candidate selected regions and used to extract genes under selection. Gene flow between geo-groups was analyzed using *Treemix* v1.13 software (Pickrell and Pritchard, 2012) and the optimal *m* value was accessed using the *optM* R package (Fitak, 2021). Finally, the *plotting_funcs.R* implemented in *Treemix* was used for visualization.

Population divergence in gene expression

The *rsgcc* R package (Ma and Wang, 2012) was used to identify PSEG with a threshold of 1. GSVA, a method for estimating variation in gene set enrichment through samples of an expression dataset, was used to determine the significantly different Gene Ontology (GO) terms between each geo-group and other accessions using the GSVA R package (Hänzelmann et al., 2013). Furthermore, the functional annotation of common specifically expressed genes in each geo-group was analyzed using the *simplifyEnrichment* R package (Gu and Hübschmann, 2023).

Homologous sequences isolation based on short read sequencing

SISRS (SNP Identification from Short Read Sequences) (Schwartz et al., 2015) allows the extraction of phylogenetically relevant sites from genome sequencing data, even for low coverage sequencing (genome skimming), and is not constrained by the availability of reference genome information. The genome skimming data for the test 101 *E. sibiricus* accessions were obtained from Xiong et al. (2023), with approximately 8 Gb of raw reads (fastq format) for each accession. Given that the genome size of *E. sibiricus* is approximately 6.86 Gb based on survey sequencing (Xiong et al., 2021), the data set of approximately 800 Gb can achieve a minimum coverage of approximately 10X and was found to be sufficient for assembly of a composite genome using *Velvet* v1.2.10 (Zerbino and Birney, 2008). The raw data for each accession were then mapped back to the composite genome using *Bowtie 2* (Langmead and Salzberg, 2012) to extract conserved and variant regions. Alignment file containing the variant information were used to perform the PCA analysis and construct the ML trees using the *adegenet* (Jombart, 2008) and *ggtree* (Yu et al., 2017) R packages, respectively.

WGCNA analysis and network construction

WGCNA is a systems biology approach to describe the correlation patterns between genes (Langfelder and Horvath, 2008). We performed WGCNA using the WGCNA R package (Langfelder and Horvath, 2008) based on the gene expression

matrix of all 101 accessions. The genes with the FPKM values greater than 0.1 in more than 95% of the samples were used for analysis (Li et al., 2020). The weighted score threshold was set to 0.8 (Wan et al., 2018; Liang et al., 2020), and then the regulatory network between transcription factor (TFs) genes and downstream genes was constructed and visualized using the ggraph R package (Si et al., 2020). The node-link graph of PSEGs was connected to show the unique regulatory network of each geo-group.

Data availability statement

The data presented in the study are deposited in the China National GeneBank DataBase repository (<https://db.cngb.org/>), accession number CNP0005048.

Author contributions

YLX: Writing – original draft, Methodology, Formal analysis, Data curation, Writing – review & editing. DL: Supervision, Conceptualization, Writing – review & editing, Writing – original draft. TL: Writing – original draft, Formal analysis, Writing – review & editing. YX: Writing – review & editing, Software, Data curation, Writing – original draft, Formal analysis. QY: Writing – review & editing, Supervision, Resources, Writing – original draft. XL: Writing – original draft, Methodology, Writing – review & editing. JZ: Writing – original draft, Software, Resources, Writing – review & editing. LY: Methodology, Writing – original draft, Writing – review & editing, Investigation. XM: Writing – original draft, Supervision, Funding acquisition, Writing – review & editing.

References

- Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., et al. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21, 1877–1896. doi: 10.1105/tpc.109.068114
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bradbury, P. J., Zhang, Z. W., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Bullard, J. H., Mostovoy, Y., Dudoit, S., and Brem, R. B. (2010). Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *PNAS* 107, 5058–5063. doi: 10.1073/pnas.0912959107
- Chen, C. L., Yang, W. J., Liu, J. Q., Xi, Z. X., Zhang, L., and Hu, Q. (2021). Population transcriptomics reveals gene flow and introgression between two non-sister alpine gentians. *Front. Ecol. Evolution* 9. doi: 10.3389/fevo.2021.638230
- Chen, S. M., Feng, J. J., Xiong, Y., Xiong, Y. L., Liu, Y. J., Zhao, J. M., et al. (2023). Evaluation and screening of wild *Elymus sibiricus* L. germplasm resources under salt stress. *Agronomy* 13, 2675. doi: 10.3390/agronomy13112675
- Cheng, L., Li, M. G., Wang, Y. C., Han, Q. W., Hao, Y. L., Qiao, Z., et al. (2023). Transcriptome-based variations effectively untangling the intraspecific relationships and selection signals in Xinyang Maojian tea population. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1114284
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and vcftools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Del Valle, J. C., Herman, J. A., and Whittall, J. B. (2020). Genome skimming and microsatellite analysis reveal contrasting patterns of genetic diversity in a rare sandhill endemic (*Erysimum teretifolium*, Brassicaceae). *PLoS One* 15, e0227523. doi: 10.1371/journal.pone.0227523
- De Wit, P., Pespeni, M. H., and Palumbi, S. R. (2015). SNP genotyping and population genomics from expressed sequences—current advances and future possibilities. *Mol. Ecol.* 24, 2310–2323. doi: 10.1111/mec.13165
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, 222–230. doi: 10.1093/nar/gkt1223
- Fitak, R. R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* 6, bpab017. doi: 10.1093/biomethods/bpab017
- Fraser, H. B., Babak, T., Tsang, J., Zhou, Y. Q., Zhang, B., Mehrabian, M., et al. (2011). Systematic detection of polygenic cis-regulatory evolution. *PLoS Genet.* 7, e1002023. doi: 10.1371/journal.pgen.1002023
- Gu, Z., and Hübschmann, D. (2023). simplifyEnrichment: a Bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinf.* 21, 190–202. doi: 10.1016/j.gpb.2022.04.008

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We thank the foundation of National Natural Science Foundation of China (32271753), Sichuan Province regional innovation cooperation project (2022TFQ0076), Key R & D project of Sichuan Province (2023YFN0087), and Cooperation project of provincial college, and provincial school (2023YFSY0012).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1457980/full#supplementary-material>

- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 107–115. doi: 10.1093/bioinformatics/btt086
- Han, M. L., Zhang, J. X., Li, D. X., Sun, S. N., Zhang, C. B., Zhang, C. J., et al. (2022). Phylogeographical pattern and population evolution history of indigenous *Elymus sibiricus* L. @ on Qinghai-Tibetan Plateau. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.882601
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 14, 1–15. doi: 10.1186/1471-2105-14-7
- Harkins, K. M., Schwartz, R. S., Cartwright, R. A., and Stone, A. C. (2016). Phylogenomic reconstruction supports supercontinent origins for leishmania. *Infect. Genet. Evol.* 38, 101–109. doi: 10.1016/j.meegid.2015.11.030
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, 277–280. doi: 10.1093/nar/gkh063
- Kawecki, T. J., and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.* 7, 1225–1241. doi: 10.1111/j.1461-0248.2004.00684.x
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5, 1–28. doi: 10.1186/gb-2004-5-2-r7
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559–571. doi: 10.1186/1471-2105-9-559
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lei, Y. T., Zhao, Y. Y., Yu, F., Li, Y., and Dou, Q. W. (2014). Development and characterization of 53 polymorphic genomic-SSR markers in Siberian wildrye (*Elymus sibiricus* L.). *Conserv. Genet. Resour.* 6, 861–864. doi: 10.1007/s12686-014-0225-5
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*. doi: 10.48550/arXiv.1303.3997
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 1–16. doi: 10.1186/1471-2105-12-323
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, Z. H., Wang, P. C., You, C. Y., Yu, J. W., Zhang, X. N., Yan, F. L., et al. (2020). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* 226, 1738–1752. doi: 10.1111/nph.16468
- Liang, W., Sun, F., Zhao, Y., Shan, L., and Lou, H. (2020). Identification of susceptibility modules and genes for cardiovascular disease in diabetic patients using WGCNA analysis. *J. Diabetes Res.* 2020, 4178639. doi: 10.1155/2020/4178639
- Limborg, M. T., Helyar, S. J., De Bruyn, M., Taylor, M. L., Nielsen, E. E., Ogden, R., et al. (2012). Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol. Ecol.* 21, 3686–3703. doi: 10.1111/j.1365-294X.2012.05639.x
- Literman, R., Windsor, A. M., Bart, H. L. Jr., Hunter, E. S., Deeds, J. R., and Handy, S. M. (2023). Using low-coverage whole genome sequencing (genome skimming) to delineate three introgressed species of buffalo fish (*Ictiobus*). *Mol. Phylogenet. Evol.* 182, 107715. doi: 10.1016/j.ympev.2023.107715
- Liu, B. B., Ma, Z. Y., Ren, C., Hodel, R. G., Sun, M., Liu, X. Q., et al. (2021b). Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: A case study in Vitaceae. *J. Systematics Evolution* 59, 1124–1138. doi: 10.1111/jse.12806
- Liu, L., Wang, Z., Su, Y. J., and Wang, T. (2021a). Population transcriptomic sequencing reveals allopatric divergence and local adaptation in *Pseudotaxus chienii* (Taxaceae). *BMC Genomics* 22, 388. doi: 10.1186/s12864-021-07682-3
- López-Goldar, X., and Agrawal, A. A. (2021). Ecological interactions, environmental gradients, and gene flow in local adaptation. *Trends Plant Sci.* 26, 796–809. doi: 10.1016/j.tplants.2021.03.006
- Ma, C., and Wang, X. (2012). Application of the Gini correlation coefficient to infer regulatory relationships in transcriptome analysis. *Plant Physiol.* 160, 192–203. doi: 10.1104/pp.112.201962
- Mächler, M., and Ligges, U. (2003). scatterplot3d - An R package for visualizing multivariate data. *J. Stat. Software* 8, 1–20. Available at: <https://hdl.handle.net/10419/77160>.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Moission, V., Dauphin, B., Grant, J., Kessler, M., Zemp, N., and Croll, D. (2022). Transcriptome-wide snps for *botrychium lunaria* ferns enable fine-grained analysis of ploidy and population structure. *Mol. Ecol. Resour.* 22, 254–271. doi: 10.1111/1755-0998.13478
- Page, J. T., Liechty, Z. S., Huynh, M. D., and Udall, J. A. (2014). BamBam: genome sequence analysis tools for biologists. *BMC Res. Notes* 7, 1–5. doi: 10.1186/1756-0500-7-829
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8 (11), e1002967. doi: 10.1371/journal.pgen.1002967
- Pratlong, M., Haguenaer, A., Chabrol, O., Klopp, C., Pontarotti, P., and Aurelle, D. (2015). The red coral (*Corallium rubrum*) transcriptome: a new resource for population genetics and local adaptation studies. *Mol. Ecol. resources* 15, 1205–1215. doi: 10.1111/1755-0998.12383
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33, 501–504. doi: 10.1093/nar/gki025
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Richter, S., Schwarz, F., Hering, L., Böggemann, M., and Bleidorn, C. (2015). The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biol. Evolution* 7, 3443–3462. doi: 10.1093/gbe/evv224
- Schwartz, R. S., Harkins, K. M., Stone, A. C., and Cartwright, R. A. (2015). A composite genome approach to identify phylogenetically informative data from next-generation sequencing. *BMC Bioinf.* 16, 1–10. doi: 10.1186/s12859-015-0632-y
- Schwinning, S., Lortie, C. J., Esque, T. C., and DeFalco, L. A. (2022). What common-garden experiments tell us about climate responses in plants. *J. Ecol.* 110, 986–996. doi: 10.1111/1365-2745.13887
- Shi, X. L., Ng, D. W. K., Zhang, C. Q., Comai, L., Ye, W. X., and Chen, Z. J. (2012). Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* 3, 950. doi: 10.1038/ncomms1954
- Si, B. B., Liang, Y. X., Zhao, J., Zhang, Y., Liao, X. F., Jin, H., et al. (2020). GGraph: an efficient structure-aware approach for iterative graph processing. *IEEE Trans. Big Data* 8, 1182–1194. doi: 10.1109/TBDATA.2020.3019641
- Signor, S. A., and Nuzhdin, S. V. (2018). The evolution of gene expression in cis and trans. *Trends Genet.* 34, 532–544. doi: 10.1016/j.tig.2018.03.007
- Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., and Chen, F. (2020). The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnol. J.* 18, 1354–1360. doi: 10.1111/pbi.13361
- Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., et al. (2021). Genome-and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol. Plant* 14, 470–487. doi: 10.1016/j.molp.2020.12.003
- Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. doi: 10.1093/bioinformatics/btv098
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Tepolt, C. K., and Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Mol. Ecol.* 24, 4145–4158. doi: 10.1111/mec.13294
- Thorstensen, M. J., Baerwald, M. R., and Jeffries, K. M. (2021). RNA sequencing describes both population structure and plasticity-selection dynamics in a non-model fish. *BMC Genomics* 22, 1–12. doi: 10.1186/s12864-021-07592-4
- Thorstensen, M. J., Jeffrey, J. D., Treberg, J. R., Watkinson, D. A., Enders, E. C., and Jeffries, K. M. (2020). Genomic signals found using RNA sequencing show signatures of selection and subtle population differentiation in walleye (*Sander vitreus*) in a large freshwater ecosystem. *Ecol. Evolution* 10, 7173–7188. doi: 10.1002/ece3.6418
- Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29, 673–680. doi: 10.1016/j.tree.2014.10.004
- Trévisan, B., Alcantara, D. M., MaChado, D. J., Marques, F. P., and Lahr, D. J. (2019). Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. *PeerJ* 7, e7543. doi: 10.7717/peerj.7543
- Wan, Q., Tang, J., Han, Y., and Wang, D. (2018). Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp. eye Res.* 166, 13–20. doi: 10.1016/j.exer.2017.10.007
- Wang, C. Y., Liu, R., Liu, Y. J., Hou, W. W., Wang, X. J., Miao, Y. M., et al. (2021). Development and application of the Faba Bean_130K targeted next-generation sequencing SNP genotyping platform based on transcriptome sequencing. *Theor. Appl. Genet.* 134, 3195–3207. doi: 10.1007/s00122-021-03885-0

- Wang, W., Wang, H., Tang, H., Gan, J., Shi, C. G., Lu, Q., et al. (2018). Genetic structure of six cattle populations revealed by transcriptome-wide SNPs and gene expression. *Genes Genomics* 40, 715–724. doi: 10.1007/s13258-018-0677-1
- Xiong, Y., Lei, X., Bai, S. Q., Xiong, Y. L., Liu, W. H., Wu, W. D., et al. (2021). Genomic survey sequencing, development and characterization of single- and multi-locus genomic SSR markers of *Elymus sibiricus* L. *BMC Plant Biol.* 21, 1–12. doi: 10.1186/s12870-020-02770-0
- Xiong, Y., Xiong, Y. L., Jia, X. J., Zhao, J. M., Yan, L. J., Sha, L. N., et al. (2023). Divergence in *Elymus sibiricus* is related to geography and climate oscillation: a new look from pan-chloroplast genome data. *J. Syst. Evol.* 0, 1–15. doi: 10.1111/jse.13020
- Xiong, Y., Xiong, Y. L., Shu, X., Yu, Q. Q., Lei, X., Li, D. X., et al. (2022). Molecular phylogeography and intraspecific divergences in siberian wildrye (*Elymus sibiricus* L.) wild populations in China, inferred from chloroplast DNA sequence and cpSSR markers. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.862759
- Yang, R. C. (1998). Estimating hierarchical F-statistics. *Evolution* 52, 950–956. doi: 10.1111/j.1558-5646.1998.tb01824.x
- Yu, G. C., Smith, D. K., Zhu, H. C., Guan, Y., and Lam, T. T. Y. (2017). GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Yu, Q. Q., Xiong, Y., Su, X. L., Xiong, Y. L., Dong, Z. X., Zhao, J. M., et al. (2023). Integrating full-length transcriptome and RNA sequencing of Siberian wildrye (*Elymus sibiricus*) to reveal molecular mechanisms in response to drought stress. *Plants* 12, 2719. doi: 10.3390/plants12142719
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zhang, X., Blaxter, M., Wood, J. M., Tracey, A., McCarthy, S., Thorpe, P., et al. (2024). Temporal genomics in Hawaiian crickets reveals compensatory intragenomic coadaptation during adaptive evolution. *Nat. Commun.* 15, 5001. doi: 10.1038/s41467-024-49344-4
- Zhang, Z. Y., Xie, W. W., Zhao, Y. Q., Zhang, J. C., Wang, N., Ntakirutimana, F., et al. (2019). EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC Plant Biol.* 19, 1–18. doi: 10.1186/s12870-019-1825-8
- Zheng, Y., Jiao, C., Sun, H. H., Rosli, H. G., Pombo, M. A., Zhang, P. F., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014
- Zheng, Y. Y., Zhang, Z. Y., Wan, Y. Y., Tian, J. Y., and Xie, W. G. (2020). Development of EST-SSR markers linked to flowering candidate genes in *Elymus sibiricus* L. based on RNA sequencing. *Plants* 9, 1371. doi: 10.3390/plants9101371