Revealing the full-length transcriptome of caucasian clover rhizome development

Xiujie Yin (yinxiujie@126.com)
Northeast Agricultural University
Kun Yi
Northeast Agricultural University
Yihang Zhao
Northeast Agricultural University
Yao Hu
Northeast Agricultural University
Xu Li
Northeast Agricultural University
Taotao He
Northeast Agricultural University
Jiaxue Liu
Northeast Agricultural University
Guowen Cui
Northeast Agricultural University

Research article

Keywords: Caucasian clover (Trifolium ambiguum. Bieb.), full-length transcriptome, RNA-Seq, rhizome, plant hormone, TFs

Posted Date: March 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-18210/v1

Version of Record: A version of this preprint was published on September 16th, 2020. See the published version at https://doi.org/10.1186/s12870-020-02637-4.

Abstract

Background: Caucasian clover (*Trifolium ambiguum* M.Bieb.) is a strongly rhizomatous, low-crowned perennial leguminous and ground-covering grass. The species may be used as an ornamental plant and is resistant to cold, arid temperatures and grazing due to a well-developed underground rhizome system and a strong clonal reproduction capacity. However, the posttranscriptional mechanism of the development of the rhizome system in caucasian clover has not been comprehensively studied. Additionally, a reference genome for this species has not yet been published, which limits further exploration of many important biological processes in this plant.

Result: We adopted PacBio Sequencing and Illumina Sequencing to identify differentially expressed transcripts in five tissues taproot (T1), horizontal rhizome (T2), swelling of taproot (T3), rhizome bud (T4) and rhizome bud tip (T5) of the caucasian clover rhizome. In total, we obtained 19.82 GB clean data and 80,654 nonredundant transcripts were analyzed. Additionally, we identified 78,209 open reading frames (ORFs), 65,227 coding sequences (CDSs), 58,276 simple sequence repeats (SSRs), 6,821 alternative splicing (AS) sites, 24,29 long noncoding RNAs (lncRNAs) and 4,501 putative transcription factors (TFs)from 64 different families. Compared with other tissues, T5 exhibited more differentially expressed genes, and co-upregulated genes in T5 are mainly annotated as involved in phenylpropanoid biosynthesis. We also identified betaine aldehyde dehydrogenase (BADH) as a highly expressed gene specific to T5. WGCNA cluster analysis of transcription factors and physiological indicators were combined to reveal 11 candidate genes (MEgreen-GA3), three of which belong to the HB-KNOX family, that are up-regulated in T3. We analyzed 276 differential transcripts involved in hormone signaling and transduction, and the largest number of transcripts are associated with the IAA signaling pathway, with significant upregulation in T2 and T5.

Conclusions: Taken together, this study contributes to our understanding of gene expression across five different tissues and provides preliminary insight into rhizome growth and development in caucasian clover.

Background

Caucasian clover (*Trifolium ambiguum* M.Bieb.) also known as Kura clover, is low-crowned perennial legume that is strongly rhizomatous[1]; the species originates from the region encompassing Caucasian Russia, eastern Turkey and northern Iran[2]. Caucasian clover has deep, semi-woody, usually branched main roots, and many branched roots grow new plantlets, either at the ends or nodes[3-5]. The species can tolerate continuous grazing[3], extreme winter temperatures[4], seasonal moisture deficit and many serious diseases that affect other types of clover[5, 6]. These features are attributed to its prominent primary roots, low-spread crowns and well-developed rhizome systems[7].

Recently, high-throughput RNA sequencing (RNA-seq) technology has become a powerful and cost-efficient means to facilitate an understanding of differential gene expression and regulatory mechanisms,

especially for plant species without a reference genome[8, 9]. RNA-seq has been widely used in the study of rhizome transcriptomics, including in various rhizomatous species such as sorghum (*Sorghum halepense* and *Sorghum propinquum*)[10, 11], bamboo (*Phyllostachys praecox*)[12], *Oryza longIstaminata*[13-15], *Equisetum hyemale* L.[16], *Panax ginseng*[17], reed (*Phragm ites- australis*)[18], tropical lotus (*Nelumbo nucifera*)[19], CangZhu (*Atractylodes Lace- a*)[20], *Ligusticum chuanxiong*[21], ginger (*Zingiber officinale*)[22], and *Miscanthus lutarioriparius*[8]. Identification of energy, metabolism, hormones and protein genes associated with rhizome development has revealed that plant rhizomes are also rich in growth-related regulatory factors. For example, 48 important transcription factors (TFs) belonging to the bHLH families YABBY, NAM, TCP, TALE, and AP2 are expressed specifically or in abundance in the shoot tip and elongation regions of wild rice[13-15].

In the last few years, an increasing number of PacBio full-length transcriptomes have been generated. The PacBio sequencing platform is a single-molecule sequencing technology with a longer read length than second-generation sequencing and an average read length of up to 15 kb. It does not require assembly and can completely retain the entire sequence from the 3' to 5' ends of an RNA, but with a higher error rate; moreover, the second-generation stepwise approach can correct mistakes[23-26]. PacBio has been utilized to detect alternative splicing and substitution polyadenylation in *Phyllostachys praecox* rhizomes, and more than 42,280 different splicing isoforms and a large number of AS events were found to be associated with the rhizome system. The results indicate that posttranscriptional regulation plays an important role in the rhizome system[27]. Moreover, a combination of Illumina and PacBio sequencing applied to various root tissues, particularly the periderm, has provided a more complete view of the Danshen *(Salvia miltiorrhiza)* transcriptome[28]. PacBio sequencing and RNA-seq analysis together have also been used to identify differentially expressed transcripts along a developmental gradient from the shoot apex to the fifth internode of *Populus* Nanlin- 895, showing 15,838 differentially expressed transcripts, of which 1,216 are TFs[29]. Compared with traditional herbage legumes, such as white clover and red clover, the rhizome is one of the most distinctive characteristics of caucasian clover[6, 30]. The rhizome system has important functions regarding energy storage, transport and vegetative reproduction[31-34]. Elucidating the molecular mechanisms underlying rhizome initiation and development will not only contribute to a better understanding of this important biological process but will also serve as a theoretical basis for efficiently identifying finding important rhizome genes. In this study, we combined Illumina and PacBio sequencing to generate a more complete caucasian clover full-length transcriptome, with analysis of gene expression in five different tissues of caucasian clover rhizomes. To the best of our knowledge, this is the first report on transcriptome profiling of candidate genes related to rhizome development in caucasian clover. These genes are excellent candidates for further functional characterization to elucidate their roles in rhizome differentiation, growth and development.

Results

Analysis of PacBio sequencing datasets

Transcriptome sequencing of the caucasian clover rhizome was completed, and 19.82 GB clean data were obtained using one cell. We identified 658,323 reads of inserts (ROIs) with a mean length of 2,286 bp, quality of 0.94; 12 passes from 720,832 polymerase reads with full passes>=0 and a predicted consensus accuracy>0.8 (Table 1). As the length of the cDNA library included the length distribution of the all ROIs (Additional file 1: Figure S1a), the cDNA library (1-6K) was accurate. In total, the ROIs included and 62.87% (449,460) full-length (FL) reads and 29.4% (193,513) non full-length reads of the entire transcriptome sequence from the 5’ to the 3’ end and polyA tail. Additionally, the number of full-length nonchimeric (FLNC) reads was 441,885, with an average full-length nonchimeric read length of 1,969 bp (Table 1). The main number distribution of ROIs is shown in additional file 1: Figure S1b.

As PacBio sequencing results have a high error rate, FLNC reads were clustered using the iterative isoform-clustering (ICE) algorithm and corrected with the Illumina Hiseq2500 platform to correct errors. We generated 227,516 consensus isoforms with an average consensus isoform length of 2,086 bp, including 148,836 high-quality isoforms (Table 1). We successfully obtained 80,654 nonredundant transcripts using CD-HIT for caucasian clover rhizomes analysis.

cDNA library	1-6K
PacBio sequencing	
polymerase reads	720,832
Reads of Insert	658,323
Read Bases of Insert	1,505,553,284
Mean Read Length of Insert	2,286
Mean Read Quality of Insert	0.94
Mean Number of Passes	12
Date size(G)	19.82
ROI databases	
Number of five prime reads	506,205
Number of three prime reads	521,059
Number of poly-A reads	514,174
Number of filtered short reads	15,350
Number of non-full-length reads	193,513
Number of full-length reads	449,460
Number of full-length non-chimeric reads	441,885
Average full-length non-chimeric read length	1,969
ICE clustering	
Number of consensus isoforms	227,516
Average consensus isoforms read length	2,086
Number of polished high-quality isoforms	148,836
Number of polished low-quality isoforms	78,366
Percent of polished high-quality isoforms(%)	65,42
Non-redundant transcripts	80,654

Table 1 Statistics of the PacBio sequencing data.

Prediction of ORFs, SSRs, and lncRNAs and identification of AS events

To identify putative protein-coding sequences, we predicted 78,209 open reading frames (ORFs) using TransDecoder. In total, 65,227 coding sequences (CDSs) were identified with start and stop codons, and the distribution of the numbers and lengths of complete CDSs is shown in Fig.2a. Among them, 12,630 transcripts were distributed in the 100-200 bp range.

A total of 79,424 sequences (167,351,883 bp) were examined, including 58,276 simple sequence repeats (SSRs) and 36,110 SSR-containing sequences (Additional file 2: Table S1). The number of sequences containing more than one SSR was 13,856, and the number of SSRs present in compound form was 10,041. In addition, most are mononucleotides (33,533), dinucleotides (8,610) and trinucleotides (14,026).

In this study, 2,429 long noncoding RNAs (lncRNA) transcripts were predicted by coding potential calculator (CPC), coding-non-coding index (CNCI), pfam protein structure domain analysis and coding potential assessment tool (CPAT) (Fig.1b), revealing candidate lncRNAs for future research.

Regarding alternative splicing (AS) events, 6,821 were detected. Because no reference genome is available for caucasian clover, we could not identify the types of AS. Nonetheless, as AS is an important mechanism for regulating gene expression and producing proteome diversity, we show the results of these AS events in the KEGG enrichment (Fig.1c), and these genes were found to be highly enriched “Glycolysis/Gluconeogenesis”(147), “Spliceosome”(129), “Carbon metabolism”(129) , “Protein processing in endoplasmic reticulum”(109) and “Biosynthesis of amino acids”(96).

Transcripts annotation

We annotated 77,927 (96.61%) transcripts in at least one of seven databases, including NCBI non-redundant protein sequences (NR), SwissProt (a manually annotated and reviewed protein sequence database), Gene Ontology (GO), Clusters of orthologous groups (COG), Eukaryotic ortholog groups (KOG), Protein family (Pfam) and Kyoto encyclopedia of genes and genomes (KEGG). The number of detailed annotations for five of the databases (GO, COG, NR, KEGG and SwissProt) is shown in a Venn diagram (Additional file 3: Figure S2).

Through homologous species analysis comparing transcriptome sequences in the NR database, 77,721 transcripts were annotated. Approximately 65.23% (50,689) of the sequences were aligned to *Medicago truncatula* sequences, followed by *Cicer arietinum* (23.72%, 18,435) (Fig.2a). All the assembled transcripts were subjected to searches against the KOG database to evaluate the effectiveness and completeness of the transcriptome annotation, and the results were divided into 26 main categories (Fig.2b). The clusters “General function predicted only (8,552)”, “Signal transduction mechanisms (6,551)”, and “Posttranslational modification, protein turnover, chaperone (5,374)” represented three of the largest groups, followed by “Carbohydrate transport and metabolism (3,146)” and “Transcription (2,769)”.

18,529 KEGG pathway analysis was performed to identify associated biochemical pathways (Fig.2c). A total of 33,383 (42.84%) transcripts were matched in the KEGG database and further classified into 128 KEGG pathways. “Biosynthesis of amino acids” (1396), “Carbon metabolism” (1389), “Protein processing in endoplasmic reticulum” (1232) ,“Starch and sucrose metabolism” (1189) and “Spliceosome” (1170) were the most represented pathways.

Based on GO analysis, 57,583 transcripts were enriched in the three ontologies (Fig.2d). “biological process”, “molecular function”, and “cellular component”. Transcripts involved in biological processes mainly included “metabolic process” (39,010), “cellular process” (33,383), and “single-organism process” (26,933). In molecular function, transcripts were mainly enriched in “binding” (32,350) “catalytic activity” (32,206), and “transporter activity” (3,404). Regarding the “cellular component” category, the major classes of transcripts were related to “cell part” (24,076) “cell” (23,984) and “organelle” (16,648).

Statistics of differentially expressed genes (DEGs) and expression of specific genes

To identify genes expression differences in the development of the caucasian clover rhizome, we focused on the identification of differentially expressed genes (DEGs). As shown in the diagram of the DEG distribution in different rhizome tissues (Fig.3a), T3 and T4 had the largest number (3,3612) of DEGs, with 18,372 up-regulated and 15,240 down-regulated genes. Moreover, T5 and other tissues (T1, T2, T3 and T4) had more DEGs, as shown in a Venn diagram of the distribution of DEGs in different tissues compared to T5; 4,585 co-upregulated genes and 4,196 co-down regulated genes were found in T5 (Fig.3b and c). With regard to up-regulated genes in T5, only 65 were enriched in KEGG analysis: phenylpropanoid biosynthesis (9), protein processing in endoplasmic reticulum (8), phenylalanine metabolism (7), carbon metabolism (7) and isoflavonoid biosynthesis (6) (Fig.3d). For down-regulated genes in T5, 231 genes were annotated, mainly involving carbon metabolism (35), biosynthesis of amino acids (34), glycolysis/gluconeogenesis (29), protein processing in endoplasmic reticulum (25) and spliceosome (21) (Fig.3e).

We investigated transcript expression levels in the five tissues, and the T1 had the highest number of expressed genes (76,124), followed by T4 (75.978), T2 (75,885), T3 (74,396) and T2 (74,327) (Additional file 4: Figure S3a). The number of genes coexpressed in each tissue was 68,241. FPKM values were used to represent the expression levels of genes, and the distributions in each tissue are shown in Additional file 4: Figure S3b. To determine genes specifically expressed in tissues and provide insight into the specialized developmental process, specific genes (at least two repeats and FPKM>0.1) with the top 5 expression levels were selected for investigation(Table 2). Among them, FPKMs were higher in T2 and T5: F01_cb16574_c994/f1p0/1592 (mitogen-activated protein kinase, MAPK3) and F01_cb71 58_ c 94/f1p0/1000 (betaine aldehyde dehydrogenases, BADH).

	Gene	FPKM	Description
T1	F01_cb14545_c33/f1p0/902	1.04	Diphosphoinositol-polyphosphate
	F01_cb17480_c9/f1p0/696	0.90	ADP-ribosylation factor 1
	F01_cb15376_c1/f1p0/1605	0.66	Protein of unknown function
	F01_cb14185_c1/f2p0/1231	0.50	Putative DNA-binding protein
	F01_cb8972_c16/f1p0/1728	0.42	Cytochrome P450
T2	F01_cb16574_c994/f1p0/1592	18.39	mitogen-activated protein kinase
	F01_cb17761_c3939/f1p0/3262	1.57	SNF2 family N-terminal domain;
	F01_cb10989_c11/f1p0/1938	1.21	Cytochrome P450
	F01_cb16704_c9/f2p2/873	0.60	Plant invertase/pectin methylesterase inhibitor
	F01_cb8297_c8/f1p0/2003	0.56	EH-domain-containing protein;
T3	F01_cb8987_c27/f1p0/456	4.07	ADP-ribosylation factor 1-like
	F01_cb8782_c40/f44p1/1360	3.55	peroxidase
	F01_cb7280_c24/f1p1/2016	2.57	NAC domain-containing protein
	F01_cb7489_c3/f1p0/2140	2.10	PHD-finger
	F01_cb8820_c25/f1p0/1072	2.08	ribosomal protein
T4	F01_cb17761_c89877/f1p0/2957	11.28	Auxin response factor
	F01_cb17761_c104162/f2p1/2284	11.02	BURP domain
	F01_cb16338_c23/f1p0/313	5.44	Calmodulin-like protein
	F01_cb17761_c21336/f1p0/3638	3.97	Leucine Rich Repeat
	F01_cb1066_c94/f1p0/2823	3.35	probable galactinol—sucrose galactosyltransferase 2-like
T5	F01_cb7158_c94/f1p0/1000	146.47	betaine aldehyde dehydrogenase 1
	F01_cb5102_c38/f1p0/684	38.70	probable protein phosphatase
	F01_cb16574_c20256/f2p0/1154	36.04	60S ribosomal protein L2
	F01_cb9053_c18/f1p0/304	30.32	-
	F01_cb11585_c6/f2p0/314	20.12	-

Table 2 Specific genes of the top five FPKMs for each tissue

TF prediction and WGCNA analysis

Transcription factors (TFs) play critical roles in plant growth and development. We examined 4,501 putative TFs from 64 different families((Additional file 5: Table S2), and the top 20 TF families are shown in Fig.4a, with the AP2/ERF-ERF (374) , C3H (372), BHLH (324), WRKY (305), GRAS (302), NAC (270), BZIP (246), C2H2 (239), and MYB-related (180) families having the most. These TFs are widely involved in plant growth and responses to stress and are related to rhizome development.

We used weighted gene co-expression network analysis (WGCNA) to further explore the relationship between TFs (filtering TF with FPKM value <1 and K-ME<0.7) and physiological characteristics in the rhizome of caucasian clover (Additional file 6: TableS3). Highly correlated gene clusters are defined as modules, in which genes within the same cluster correlated highly. WGCNA analysis identified eight distinct modules (labeled with different colors in Fig.4b). The correlation coefficient between the characteristic genes of each module of 10 different modules and each different sample (trait) is presented in Fig.4c. Notably, two module-trait relationships (MEgreenyellow-IAA, and MEgreen-GA3) were highly significant ($r^2 > 0.8$, $p < 10^{-4}$). In addition, the majority genes of the green module were up-regulated for six traits, except for ABA; most of these TFs were mainly up-regulated in T3 (Additional file 7: FigureS4). We identified 11 hub genes based on the criteria of KME (eigengene connectivity) ≥0.99 and edge weight value ≥0.5 in the green module based on the regulatory network (Fig.4d). These hub genes mainly belong to the HB-KNOK, AP2/ERF-ERF, GRAS, C2H2, C3H and NAC families (MEgreen-GA3); moreover, these TF families were upregulated in T1, T2 and T3, particularly in T3 (Fig. 4e), and may be related to the formation of nodules in the rhizome.

Identification of hormone signaling-related genes in rhizome development

Plant hormones play an important role in all aspects of development. Accordingly, we mapped DEGs to hormone signaling and transduction pathways for caucasian clover and analyzed their expression in different tissues. In total, 276 transcripts are related to the synthesis and metabolism of eight hormones, including auxin (IAA), abscisic acid (ABA), ethylene (ETH), cytokinin (CTK), gibberellic acid (GA), brassinosteroid (BR), jasmonic acid (JA) and salicylic acid (SA) (Fig.5). The maximum number of transcripts (62) is related to IAA synthesis and metabolism, followed by ABA at 60, JA at 52, SA at 24, ETH at 28, BR at 20, CTK at 18 and GA at 10. All these significant genes related to hormones synthesis and metabolism exhibited different expression in the different tissues. In addition, most genes associated with the IAA pathway were up-regulated in T2 and T5. Regarding SA signaling, almost all genes belong to the TGA family, with up-regulation only in T4. Most transcripts associated with BR signaling showed higher levels in T4 and T5. For CTK transduction, all crucial genes associated with CTK signaling pathway were identified as DEGs. Only three genes showed no up-regulation trend in T3; which may be tissues in which cells divide in large numbers. Only ten DEGs are associated with GA signaling; five transcripts are classified as GID1 and significantly up-regulated in T2. Genes related to ABA signaling and transduction displayed no significant change. For JA signaling, 52 transcripts all were annotated as JAZ

(K13464), with up-regulation in T1, T2 and T3. In contrast, most of ETH signaling genes exhibited higher expression in T3, and all were down-regulated in T2.

Verification of gene expression by qRT-PCR

To confirm the accuracy of the genes obtained by RNA-seq, twelve genes including six plant hormone signal transduction genes, three TFs and three genes belonging to other classes were randomly selected for quantitative real-time RT-PCR (qRT-PCR) analysis. Good reproducibility between the qRT-PCR and RNA-seq results were indicated by Pearson correlation analysis, and these results verified the accuracy and reliability of the RNA-seq data (Additional file 8: TableS4).

Discussion

Caucasian clover's rhizome is unique among legume species, endowing it with particular clonal reproduction characteristics and resistance to stress. In this study, we obtained high-quality transcript sequences for the caucasian clover rhizome by PacBio and Illumina sequencing, and the results will contribute to our understanding of rhizome growth and development and lay a molecular foundation for further study.

AS is a vital mechanism regulating gene expression and producing protein diversity[35]. The numbers of AS events identified for the first time in the caucasian clover rhizome was found to be lower than that in *Medicago sativa* (7,568)[36] but higher than that in *Trifolium pratense* (5,492)[9]. Our study on the special characteristics of the caucasian clover rhizome was hindered by the lack of a reference genome, and it is impossible to judge the type of AS.

We found that the most transcripts (1,396) were annotated as related to the amino acid biosynthetic pathway in KEGG (Fig.2c). Many amino acids (phenylalanine, tyrosine, and tryptophan) are not only important components of proteins but are also precursors of many secondary metabolites. These secondary metabolites are crucial for plant growth[37]. Similar to the rhizome of other plants (*Oryza longistaminata, Miscanthus lutarioriparius*)[15, 8], some basal metabolism has an important role in the rhizome of caucasian clover, for example, carbon metabolism (1,389) and starch and sucrose metabolism(1,189).

By analyzing specific expression in each tissue, we found that MAPK3 (F01_cb16574_c994/f1p0 /1592) is mainly specifically expressed gene in T2 (Table 2). The MAPK family has been studied in tobacco and may be involved in growth, development, response to plant hormone signals and environmental signals[38]. T5 is very different from root tissue (T1, T2 and T3) (Fig.3a), with the greatest number of DEGs; in this tissue, a new part is formed from this tissue. The highest specific expression in T5 was observed for betaine aldehyde dehydrogenase 1 (F01_cb7158_c94/f1p0/1000) (Table 2). BADHs are involved in glycine betaine synthesis and act as plant osmotic regulators, with important roles in abiotic stress[39]. Experiments have shown that BADH can increase the abiotic stress tolerance of sweet potato and carrot, such as salt stress, oxidative stress and low temperature stress, maintaining cell membrane

integrity[40, 41]. BADH was specifically expressed in a large amount in T5. This may be because T5 is relatively more fragile than other tissues, and BADHs may protect T5 for promote the growth of new plants. Thus, we speculate that defense and stress response play a vital role in the development of caucasian clover. This may be the reason for its ability to grow in extreme winter temperatures[4], or it may be a necessary condition for a large rhizome system.

TFs can offer knowledge into the gene-regulating networks controlling developmental programs and are recognized as major players in better understanding root tissue differentiation and root development in response to internal growth regulators as well as environmental signals[42, 43]. It has been reported that the genes involved in hormone metabolism, cellulose synthesis energy, metabolism substance synthesis and transportation stress as well as expansion-related protein genes and TFs such as HLH, TCP WRKY, bZIP, MYB, and NAC participate in the formation of lotus root rhizomes. In addition, the AP2/ERF TF family had the greatest number in our rhizome(Fig.4a), among which ethylene response factors, such as BBM/PLT4 and PLT1-3, have been described as master regulators of root meristem initiation and maintenance in *Arabidopsis thaliana*[44, 45]. In *Raphanus sativus* and alfalfa[46, 47], the abiotic stress response mechanism regulated by AP2/ERF has been carefully studied. The *Arabidopsis* NAC family member NAC1 transduces auxin signals downstream of TIR1 to promote lateral root development[48]. In the WRKY family, WRKY75 is reported to be involved in regulating nutrient starvation response and root development[49]. It is worth noting that we performed WGCNA clustering for TFs and found that 3 of 11 hub genes belong to the HB-KNOX family in MEgreen-GA3 (Fig. 4d). GA can suppress the effect of elevated KNOX gene expression, and there is a possibility of modifying KNOX gene expression to alter plant structure through local changes in GA levels[50]. In *Arabidopsis*, the GA20ox1 mRNA level is reduced in leaves overexpressing the KNOX proteins STM or BREVIPEDICELLUS[51]. Moreover, in model plants , such as *Arabidopsis*, maize and tobacco, KNOX gene expression is confined to the shoot meristem and stem[52]. However, in the underground rhizome of caucasian clover, KNOX genes were identified as hub genes, especially in the main tissue of the swollen taproot (T3). Whether it is consistent with KNOX regulation of *Arabidopsis* meristem stems and buds is worth further investigation[53]. Xi Cheng found that in pear plants coexpressing KNOX and PbKNOX1, these factors are involved in cell wall thickening and lignin biosynthesis, with inhibition of key structural genes involved in lignin synthesis[54].

There is growing evidence that hormones affect tillering growth and the formation of storage organs[55, 8]. Research reported 600mg·L^{-1} GA_3 can promote the growth and development of caucasian clover rhizome and increase the content of endogenous IAA ,ZT and GA_3[56] .T1 and T2 showed high GID1 expression, and GA may be involved in photoperiod induction and regulation of the formation of storage organs and rhizome elongation[25]; therefore, TI and T2 may be key organs for nutrient storage. In addition, JAZ accumulated in roots (T1, T2 and T3). It is believed that when pathogens invade and abiotic stress occurs, JAZ-MYC form an immune network, followed by JAZ protein degradation and MYC TF release[57]. T3 is closely linked to ETH and CTK-mediated pathways, which may be responsible for root swelling. It has been demonstrated that auxin activates root formation and that cytokinins mediate

root identity, early primordial disintegration and early loss of bud development initiation[58]. CTKs are conducive to rhizome enlargement but not to rhizome induction.

Genes related to the IAA anabolic pathway were downregulated in T3 and T4. It may be that IAA does not participate in root enlargement or induce bud production but that it is closely related to lateral root development. T1, T2 and T3 might mainly function as storage organs, providing energy for plant growth. In addition, T4 tissue may be relatively fragile and require the SA pathway to mediate immunity to prevent pathogen infection and to grow new plants. BR signaling is mainly involved in plant growth and plant morphology development, and related genes were upregulated in T5[59]. T4 and T5 are mainly associated with resistance to stress and secondary metabolic pathways. Of course, hormones are not the only factors that regulate the development of apical meristems and lateral organs; they often cooperate with TFs to balance the maintenance of meristems and organogenesis.

Conclusion

In summary, we provide a full-length transcriptome of the caucasian clover rhizome based on PacBio sequencing and Illumina sequencing, revealing gene expression and annotation for different tissues. We highlighted the role of hormones and TFs in the rhizome of caucasian clover, investigating the expression of hormone-pathway related genes in different tissue of caucasian clover and identified 11 candidate genes in TF- and GA-related modules by WGCNA. In this study, a set of genes related to rhizome development was identified, laying the foundation for further functional genomics research on rhizome development.

Methods

Plant materials and RNA preparation

The taproot (T1), horizontal rhizome (T2), swelling of taproot (T3), rhizome bud (T4) and rhizome bud tip (T5) of 3-year-old caucasian clover (Fig.6) were collected from a test field at Northeast Agricultural University (E 126º14';N 45º05') in August 2018), with three replicates from five individual plants for each tissue has, showing good correlation (R^2>0.8; Additional file 9: Figure S5). Original sources of plant materials were introduced from Inner Mongolia Grass Variety Engineering Technology Research Center of Inner Mongolia Agricultural University. The Inner Mongolia Grass Variety Engineering Technology Research Center of Inner Mongolia Agricultural University undertook the formal identification of the samples, provided details of specimen deposited and allowed to collect. Caucasian clover's IPNI Life Sciences Identifier (LSID) is urn:lsid:ipni.org:names:522843-1. Plants were removed from the soil bed, and the roots were washed gently with running water, frozen in liquid nitrogen and immediately stored at -80℃. Total RNA was extracted using Trizol. RNA degradation and contamination were monitored by 1.2% agarose gel electrophoresis. The quantity and integrity of the extracted total RNA were determined using a NanoDrop and an Agilent 2100 bioanalyzer. For each RNA sample, 1μg was pooled and

sequenced by PacBio single-molecule long-read sequencing (PacBio Sequel, Menlo Park, USA) and Illumina sequencing (Illumina NovaSeq6000, California, U.S.A) in parallel.

PacBio cDNA library preparation and sequencing

Full-length cDNA was synthesized using SMARTerTM PCR cDNA Synthesis Kit and then subjected to full-length cDNA PCR amplification and repair of cDNA ends. The concentration and quality of the cDNA library were determined using a qubit 2.0 fluorometer and an Agilent 2100 bioanalyzer. The 1-6-k library was sequenced via PacBio Sequel.

Illumina cDNA library construction and sequencing

First, 15 samples of eukaryotic mRNA were enriched with magnetic beads with oligo(dT) and randomly broken into small fragments in fragmentation buffer. First-strand cDNA was synthesized using six-base random hexamers with a small fragment of mRNA as a template. The second cDNA strand was synthesized by adding buffer, dNTPs, RNase H and DNA polymerase I, and the cDNA was purified by AMPure XP beads. The purified double-stranded cDNA was subjected to end repair, a tail was added, and the sequencing linker was ligated; the fragment size was then selected using AMPure XP beads. The final cDNA library was assessed by PCR, and the quality of the cDNA library was determined using an Agilent 2100 Bioanalyzer (Santa Clara, CA). The libraries were sequenced from both 5’ and 3’ ends using Illumina NovaSeq.

PacBio sequencing data analysis

Raw reads were processed into error-corrected reads of insert (ROIs) using the Iso-seq pipeline with minFullPass=0 and minPredictedAccuracy=0.80. Next, full-length, non-chimeric (FLNC) transcripts were determined by searching for the polyA tail signal and the 5’ and 3’ cDNA primers in ROIs. ICE was used to obtain consensus isoforms and FL consensus sequences from ICE data, which were further processed using Quiver. High-quality FL transcripts were classified with the criterion postcorrection accuracy above 99%. Iso-Seq high-quality FL transcripts were obtained, and redundancy was removed using cd-hit (identity > 0.99)[60].

Illumina sequencing data analysis

Raw data (raw reads) in FASTQ format were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapters, reads containing poly-N and low-quality reads from the raw data. At the same time, the Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. All downstream analyses were based on these clean data with high quality. These clean reads were then mapped to the PacBio reference genome sequence. Clean data were normalized by converting the fragment counts to fragments per kilobase of transcript per million mapped reads (FPKM). Differential expression analysis was carried out using DESeq (v 1.10.1); a fold change ≥ 4 and FDR < 0.01 based on DESeq was considered differential expression[61].

Detection of SSR, ORFs, AS and lncRNA

Simple sequence repeats (SSRs) in the transcriptome were identified using MISA (http://pgrc.ipk-gatersleben.de/misa/), with only transcripts ≥ 500 bp being detected.

TransDecoder software (https://github.com/TransDecoder/TransDecoder/releases) was employed to identify reliable potential coding sequences (CDSs) from transcript sequences, as based on open reading frame (ORF) length, log-likelihood score, and amino acid sequence comparison in the Pfam database.

We used Iso-SeqTM data directly to run all-vs-all BLAST with high-identity settings[62]. BLAST alignments that met all criteria were considered products of candidate alternative splicing (AS) events[63], with two HSPs (high segmentation pairs)≥1000 bp in the alignment. Two HSPs have the same forward/reverse orientation, and one sequence should be contiguous in the same alignment or have a small overlap of less than 5 bp. The other should be different to show "AS Gap", and the contiguous sequence should align completely with the different sequences. The AS gap should be greater than 100 bp and at least 100 bp from the 3'/5' end.

Four computational approaches include CPC [64] /CNCI/ CPAT [69]/ Pfam[65] were combined to sort nonprotein-coding RNA candidates from putative protein-coding RNAs among the transcripts. Putative protein-coding RNAs were filtered using a minimum length and exon number threshold. Transcripts with lengths more than 200 nt and more than two exons were selected as long noncoding RNA (lncRNA) candidates and further screened using CPC/CNCI/CPAT/Pfam, which have the power to distinguish protein-coding genes from noncoding genes.

Functional annotation

Annotation information on the obtained nonredundant transcript sequences was based on BLAST in the following databases[66]: NR; Pfam [66]; KOG [67]; COG; swiss-Prot[68]; KEGG [69]; and GO [70].

Real-time RT- PCR

Quantitative real-time RT-PCR (qRT-PCR) was carried out in a 10-μl volume containing 0.5μl diluted cDNA, 0.2μl forward primer, 0.2μl reverse primer, and 1×SYBR Premix Ex Taq II (TaKaRa) with the following conditions: 95ºC for 180 sec, followed by 40 cycles of 95ºC for 15 sec, 59ºC for 15 sec and 72ºC for 15 sec. The $2^{-\triangle\triangle}$Ct method was used to calculate relative expression levels. All reactions were performed with three replicates. All primers used are shown in additional file 10: Table S5).

Abbreviations

OFRs: Open reading frames; CDSs: Coding sequences; SSRs: Simple sequence repeats; AS: Alternative splicing; LncRNAs: Long non-coding RNA; TF: Transcription factors; ROI: Reads of Insert; ICE: Iterative isoform-clustering; FL: full-length; FLNC: Full-length non chimera; CNCI: Coding-Non Coding-Index; CPC: Coding potential≥CPAT: Coding potential assessment tool calculator; FLNC: Full-length non chimera; COG:

Clusters of orthologous groups; NR: NCBI non-redundant protein sequences; GO: Gene Ontology; KOG: Eukaryotic ortholog groups. Pfam: protein family; KEGG: Kyoto encyclopedia of genes and genomes; DEG: Differentially expressed genes; IAA: Auxin; ABA: Abscisic acid; ETH: Ethylene; CTK: Cytokinin; GA: Gibberellic acid; BR: Brassinosteroid; JA: Jasmonic acid; SA: Salicylic acid; BADHs: Betaine aldehyde dehydrogenases; MAPK: Mitogen-activated protein kinase; qRT-PCR: Quantitative real-time RT-PCR.

Declarations

Ethics approval and consent participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Natural Science Foundation of China (31802120), Research and Demonstration of Large-scale Artificial Grassland Combined Plant and Circular Mode (2017YFD0502106) and Academic Backbone Fund Project of Northeast Agricultural University. The funders did not design the experiment, or draft and revise the manuscript.

Authors' contributions

XY designed the experiment and revised the manuscript. KY performed data processing and drafted the manuscript. YZ, YH and XC prepared the materials and performed the experiments. TH and JL assisted in manuscript preparation. GC conceived the study. All authors read and approved the final version of the manuscript.

Acknowledgements

We thank the National Natural Science Foundation of China (31802120) supporting this work.

Availability of data and materials

Raw reads of one combined PacBio library and one Illumina RNAseq library generated in this study are available from BioProject at NCBI (https://www.ncbi. nlm.nih.gov/bioproject/) under accession numbers PRJNA586585 and PRJNA588309, respectively.

References

1. Black AD, Lucas RJ. Caucasian clover was more productive than white clover in grass mixtures under drought conditions. In: New Zealand Grassland Association. 2000;62:183-8.
2. Davis PH. Flora of Turkey and the East Aegean islands. Quarterly Review of Biology. 1965;35(1):103.
3. Brummer EC, Moore KJ. Persistence of perennial cool-season grass and legume cultivars under continuous grazing by beef cattle. Agronomy Journal. 2000;92(3):466-71.
4. Sheaffer CC, Marten GC. Kura clover forage yield, forage quality, and stand dynamics. Canadian Journal of Plant Science. 1991;71(4):1169-72.
5. Smith NLTRR. Kura Clover (*Trifolium ambiguum* M.B.) Breeding, Culture, and Utilization. Advances in Agronomy. 1997;63(08):153-78.
6. Andrzejewska J, Contreras-Govea FE, Pastuszka A, Albrecht KA. Performance of Kura clover compared to that of perennial forage legumes traditionally cultivated in central Europe. Acta Agr Scand B-S P. 2016;66(6):516-22.
7. Speer GS, Allinson DW. Kura clover (*Trifolium ambiguum*): Legume for forage and soil conservation. Economic Botany. 1985;39(2):165-76.
8. Hu RB, Yu CJ, Wang XY, Jia CL, Pei SQ, He K, He G, Kong YZ, Zhou GK. De novo transcriptome analysis of *Miscanthus lutarioriparius* identifies candidate genes in rhizome development. Frontiers In Plant Science. 2017;8(482).
9. Chao YH, Yuan JB, Li SF, Jia SQ, Han LB, Xu LX. Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. Bmc Plant Biol. 2018;18:300-12
10. Jang CS, Kamps TL, Tang H, Bowers JE, Lemke C, Paterson AH. Evolutionary fate of rhizome-specific genes in a non-rhizomatous *Sorghum* genotype. Heredity. 2009;102(3):266-73.
11. Zhang T, Zhao XQ, Wang WS, Huang LY, Liu XY, Zong Y, Zhu LH, Yang DC, Fu BY, Li ZK. Deep transcriptome sequencing of rhizome and aerial-shoot in *Sorghum propinquum.* Plant Mol Biol. 2014;84(3):315-27.
12. Wang KH, Peng HZ, Lin EP, Jin QY, Hua X, Sheng Y, Bian HW, Ning H, Pan JW, Wang JH. Identification of genes related to the development of bamboo rhizome bud. J Exp Bot. 2010;61(2):551-61.
13. Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, Mcnally K, Khush GS, Paterson AH. Convergent evolution of perenniality in rice and *sorghum*. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(7):4050-4.
14. Hu F. Identification of rhizome-specific genes by genome-wide differential expression analysis in *Oryza longistaminata*. Bmc Plant Biol. 2011;11(1):18-32.
15. He RF, Salvato F, Park JJ, Kim MJ, Nelson W, Balbuena TS, Willer M, Crow JA, May GD, Soderlund CA, et al. A systems-wide comparison of red rice (*Oryza longistaminata*) tissues identifies rhizome specific genes and proteins that are targets for cultivated rice improvement. BMC Plant Biol. 2014;14 (1):46-67.

16. Balbuena TS, He RF, Salvato F, Gang DR, Thelen JJ. Large-scale proteome comparative analysis of developing rhizomes of the ancient vascular plant *Equisetum hyemale*. Frontiers In Plant Science. 2012;3(131).
17. Yang BW, Hahm YT. Transcriptome analysis using de novo RNA-seq to compare ginseng roots cultivated in different environments. Plant Growth Regul. 2018;84(1):149-57.
18. Ruifeng H, Min-Jeong K, William N, Balbuena TS, Ryan K, Robin K, Crow JA, May GD, Thelen JJ, Soderlund CA. Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, *Phragmites australis* (Poaceae), reveals genes involved in invasiveness and rhizome specificity. American Journal of Botany. 2012;99(2):232-47.
19. Yang M, Zhu L, Pan C, Xu L, Liu Y, Ke W, Yang P. Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical Lotus (*Nelumbo nucifera*). Sci Rep. 2015;5:13059.
20. Huang QQ, Huang X, Deng J, Liu HG, Liu YW, Yu K, Huang BS. Differential gene expression between leaf and rhizome in *Atractylodes lancea*: a comparative transcriptome analysis. Frontiers In Plant Science. 2016;7(348).
21. Song T, Liu ZB, Li JJ, Zhu QK, Tan R, Chen JS, Zhou JY, Liao H. Comparative transcriptome of rhizome and leaf in *Ligusticum Chuanxiong.* Plant Syst Evol. 2015;301(8):2073-85.
22. Koo HJ, McDowell ET, Ma XQ, Greer KA, Kapteyn J, Xie ZZ, Descour A, Kim H, Yu Y, Kudrna D, et al. Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. BMC Plant Biol. 2013;13(1):27-44.
23. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends in Genet. 2008;24(3):142-9.
24. Mason CE. Faster sequencers, larger datasets, new challenges. Genome Biol. 2012;13(3):314.
25. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genom Proteom Bioinf. 2015;13(5):278-89.
26. Li PH, Ponnala L, Gandotra N, Wang L, Si YQ, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al. The developmental dynamics of the maize leaf transcriptome. *N*at Genet. 2010;42(12):1060-7.
27. Wang T, Wang H, Cai D, Gao Y, Zhang H, Wang Y, Lin C, Ma L, Gu L. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). Plant J. 2017;91(4):684-99.
28. Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, Zhu YJ, Ji AJ, Zhang B, Hu SN, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. Plant Journal. 2015;82(6):951-61.
29. Chao Q, Gao ZF, Zhang D, Zhao BG, Dong FQ, Fu CX, Liu LJ, Wang BC. The developmental dynamics of the *Populus* stem transcriptome. Plant Biotechnol J. 2019;17(1):206-19.
30. Lin WH, W HR, Stephen S. Physiological responses of five species of *Trifolium* to drought stress. Chinese Journal of Applied and Environmental Biology. 2011;17(4):580-4.

31. Schmid B, Bazzaz FA. Clonal integration and population structure in perennials: effects of severing rhizome connections. Ecology. 1987;68(6):2016-22.
32. Schmid B, Bazzaz FA. Growth responses of rhizomatous plants to fertilizer application and interference. Oikos. 1992;65(1):13-24.
33. Humphrey LD, Pyke DA. Clonal foraging in perennial wheatgrasses: a strategy for exploiting patchy soil nutrients. Journal of Ecology. 1997;85(5):601-10.
34. Huber-Sannwald E, Pyke DA, Caldwell MM, Durham S. Effects of nutrient ptches and root systems on the clonal plasticity of a rhizomatous grass. Ecology. 1998;79(7):2267-80.
35. Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, Syed NH. Alternative splicing and protein diversity: plants versus animals. Front Plant Sci. 2019;10(10):14.
36. Chao YH, Yuan JB, Guo T, Xu LX, Mu ZY, Han LB. Analysis of transcripts and splice isoforms in *Medicago sativa* L. by single-molecule long-read sequencing. Plant Mol Biol. 2019;99(3):219-35.
37. Tzin V, Galili G. New Insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Mol Plant. 2010;3(6):956-72.
38. ZHANG S: MAPK cascades in plant defense signaling. *Trends Plant Sci.* 2001, 6.
39. Fitzgerald TL, Waters DLE, Henry RJ. Betaine aldehyde dehydrogenase in plants. Plant Biology. 2009;11(2):119-30.
40. Kumar S, Dhingra A, Daniell H. Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots, and leaves confers enhanced salt tolerance. Plant Physiol. 2004;136(1):2843-54.
41. Fan WJ, Zhang M, Zhang HX, Zhang P. Improved tolerance to vrious abiotic stresses in tansgenic sweet potato (*Ipomoea batatas*) expressing spinach betaine aldehyde dehydrogenase. PLoS One. 2012;7(5):14.
42. Jin JP, Zhang H, Kong L, Gao G, Luo JC. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014;42(D1):D1182-7.
43. Montiel G, Gantet P, Jay-Allemand C, Breton C. Transcription factor networks. Pathways to the knowledge of root development. Plant Physiol. 2004;136 (3):3478-85.
44. Licausi F, Ohme-Takagi M, Perata P. APETALA/Ethylene responsive factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol. 2013;199(3):639-49.
45. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. Bba-Gene Regul Mech. 2012; 1819 (2):86-96.
46. Karanja BK, Xu L, Wang Y, Tang M, M'Mbone Muleke E, Dong J, Liu L. Genome-wide characterization of the AP2/ERF gene family in radish (*Raphanus sativus* L.): Unveiling evolution and patterns in response to abiotic stresses. Gene. 2019;718:144048.
47. Jin XY, Yin XF, Ndayambaza B, Zhang ZS, Min XY, Lin XS, Wang YR, Liu WX. Genome-wide identification and expression profiling of the ERF gene family in *Medicago sativa* L. under various abiotic stresses. DNA Cell Biol. 2019;13.

48. Xie Q, Frugis G, Colgan D, Chua NH. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes & development. 2000;14(23):3024-36.
49. Devaiah BN, Karthikeyan AS, Raghothama KG. WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. Plant Physiol. 2007;143(4):1789-801.
50. Singh DP, Filardo FF, Storey R, Jermakow AM, Yamaguchi S, Swain SM. Overexpression of a gibberellin inactivation gene alters seed development, KNOX gene expression, and plant development in *Arabidopsis*. Physiol Plant. 2010;138(1):74-90.
51. Bolduc N, Hake S. The maize transcription factor KNOTTED1 directly regulates the gibberellin catabolism gene ga2ox1. Plant Cell. 2009;21(6):1647-58.
52. Hay A, Tsiantis M. KNOX genes: versatile regulators of plant development and diversity. Development. 2010;137(19):3153-65.
53. Xu L, Shen WH. Polycomb silencing of KNOX Genes confines shoot stem cell niches in *Arabidopsis.* Curr Biol. 2008;18(24):1966-71.
54. Cheng X, Li ML, Abdullah M, Li GH, Zhang JY, Manzoor MA, Wang H, Jin Q, Jiang TS, Cai YP,et al. In silico genome-wide analysis of the pear (*Pyrus bretschneideri*) KNOX family and the functional characterization of PbKNOX1, an arabidopsis BREVIPEDICELLUS orthologue gene, involved in cell wall and lignin biosynthesis. Front Genet. 2019;10(17).
55. Dello Ioio R, Linhares FS, Scacchi E, Casamitjana-Martinez E, Heidstra R, Costantino P, Sabatini S. Cytokinins determine *Arabidopsis* root-meristem size by controlling cell differentiation. Curr Biol. 2007;17(8):678-82.
56. Yi K, Zhao YH, Hu Y, Liu JX, He TT, Li X, Song P, Cui GW, Yin XJ. Effect of GA_3 and 6 -BA on rhizome segment growth and endogenous hormone content of caucasian clover. Acta Prataculturae Sinica. 2020;29(2):22-30.
57. Garrido-Bigotes A, Valenzuela-Riffo F, Figueroa CR. Evolutionary analysis of JAZ proteins in plants: an approach in search of the ancestral sequence. Int J Mol Sci. 2019;20(20).
58. Pernisova M, Grochova M, Konecny T, Plackova L, Harustiakova D, Kakimoto T, Heisler MG, Novak O, Hejatko J. Cytokinin signalling regulates organ identity via the AHK4 receptor in *Arabidopsis.* Development. 2018;145(14):11.
59. Nie SM, Huang SH, Wang SF, Mao YJ, Liu JW, Ma RL, Wang XF. Enhanced brassinosteroid signaling intensity via SlBRI1 overexpression negatively regulates drought resistance in a manner opposite of that via exogenous BR application in tomato. Plant Physiol Biochem. 2019;138:36-47.
60. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.
61. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):106-18.
62. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research.

1997;25(17):3389-402.
63. Liu XX, Mei WB, Soltis PS, Soltis DE, Barbazuk WB. Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. Mol Ecol Resour. 2017;17(6):1243-56.
64. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35:345-9.
65. Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Research. 2013;41(6):7.
66. Lamb J, Jarmolinska AI, Michel M, Menendez-Hurtado D, Sulkowska JI, Elofsson A. PconsFam: an interactive database of structure predictions of pfam families. J Mol Biol. 2019;431(13):2442-8.
67. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, et al, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 2004;5(2):28.
68. Renaux A, UniProt C. UniProt: the universal protein knowledgebase. Nucleic Acids Research. 2018;46(5):2699.
69. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32:D277-80.
70. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.

Figures

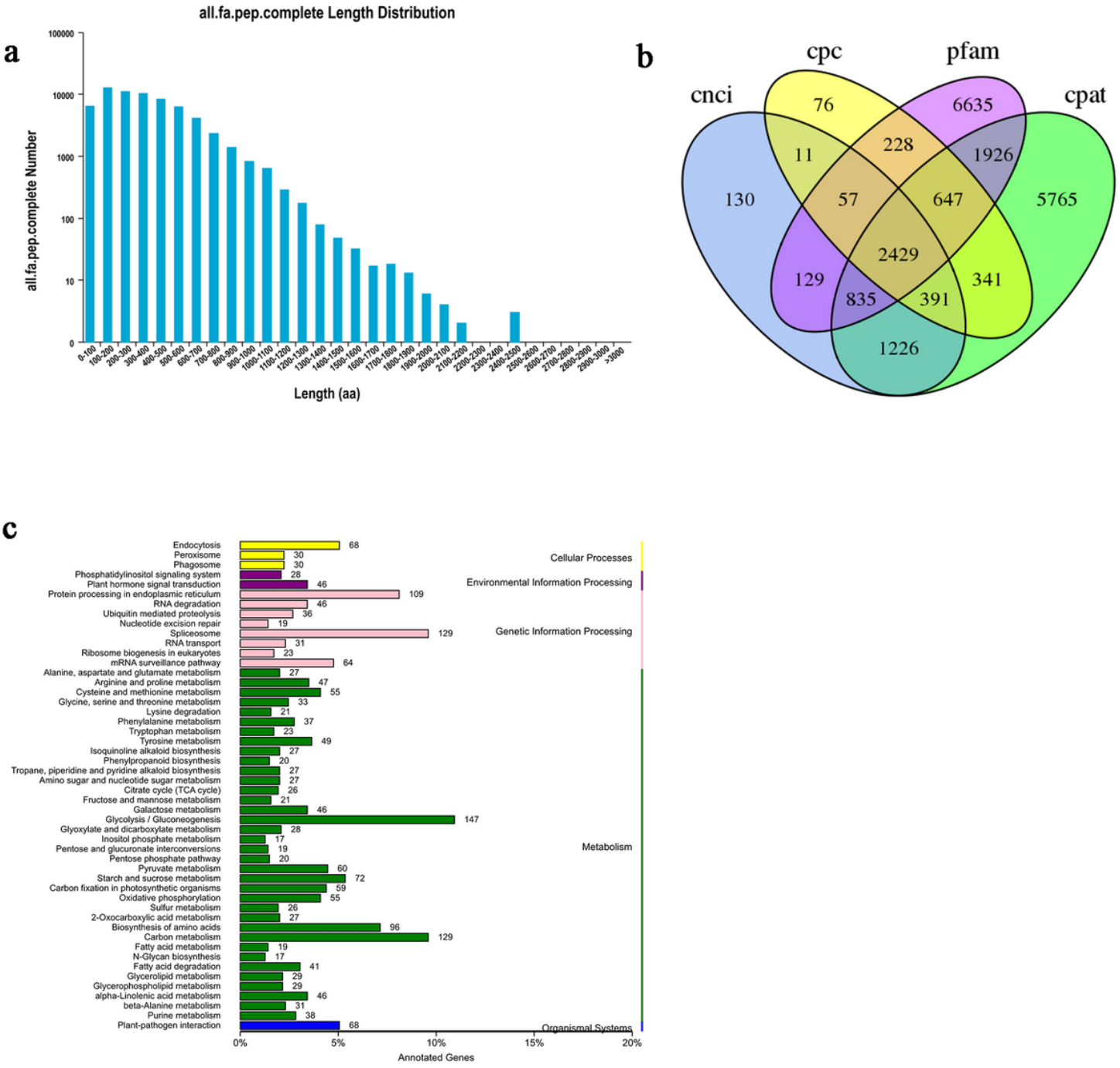

Figure 1

Prediction of CDS, lncRNAs and AS. a The distribution of CDS lengths with a complete open reading frame. b Venn diagram of the number lncRNAs predicted. c KEGG pathways of genes related to AS.

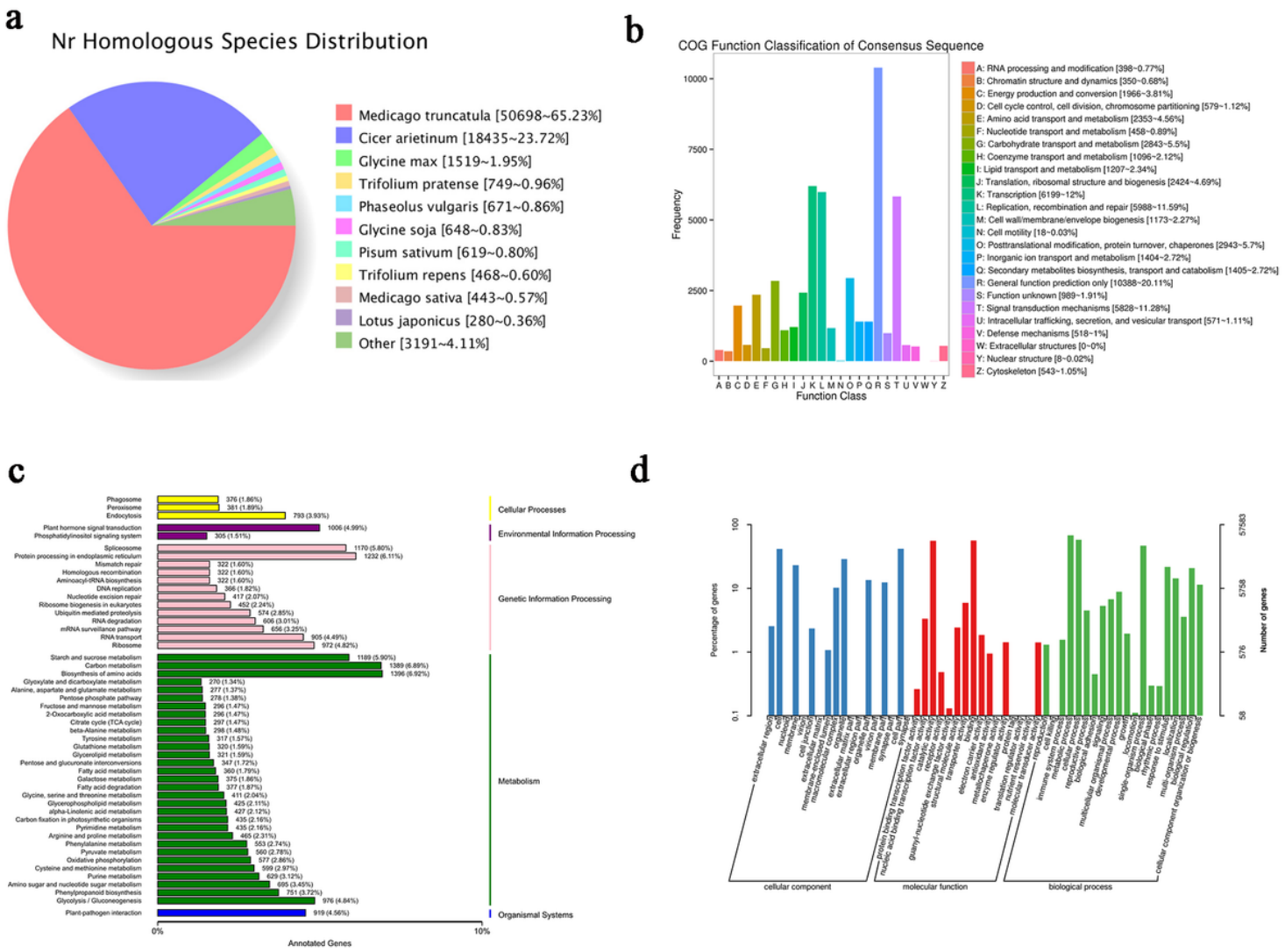

Figure 2

Transcripts annotated in four databases. a NR homologous species distribution diagram of transcripts. b COG function classification of transcripts. c KEGG pathway classification of transcripts. d Distribution of GO terms for all annotated transcripts.

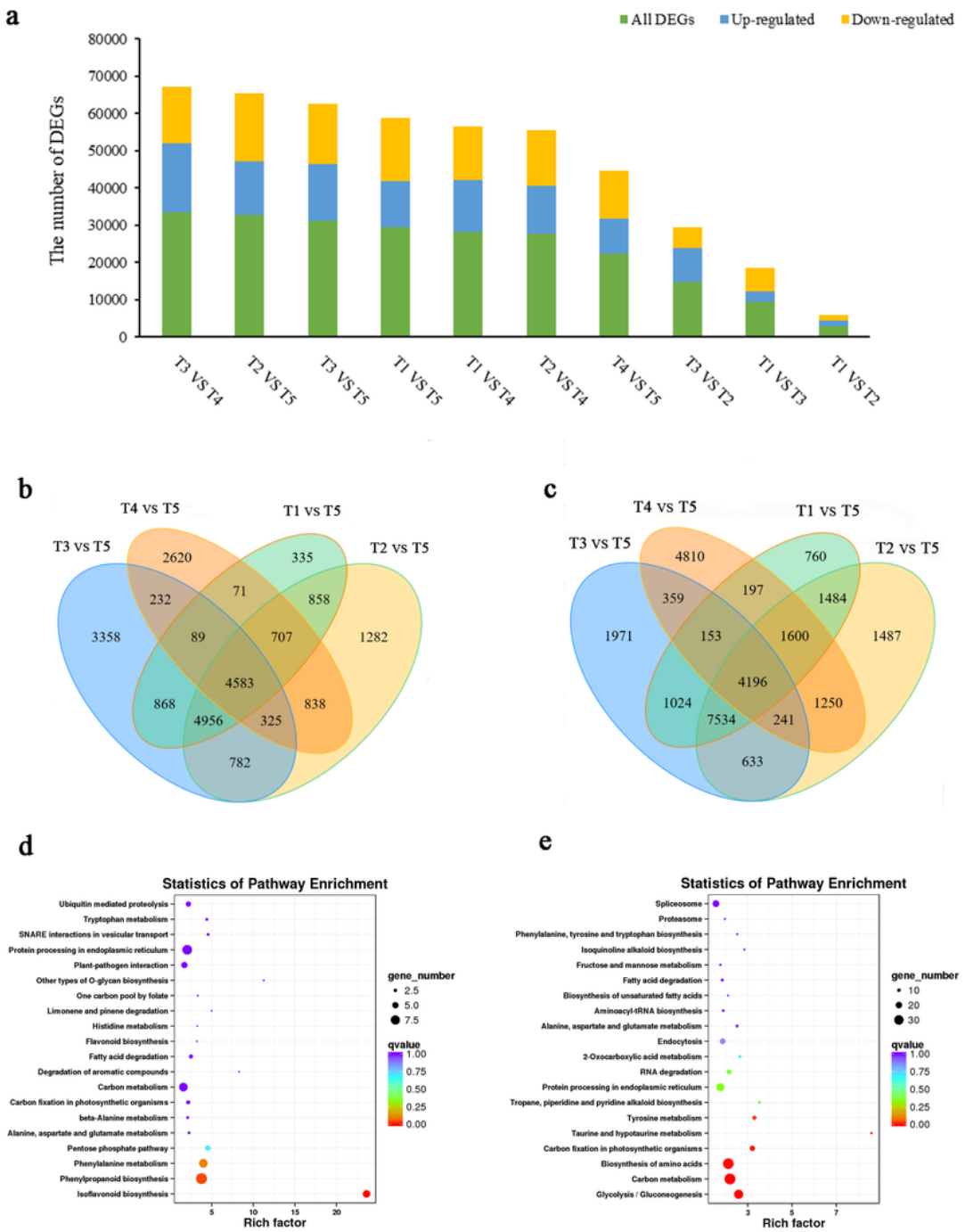

Figure 3

Differentially expressed genes statistics in the rhizome of caucasian clover. a The number of DEGs in different tissues. b Venn diagram showing upregulated genes in tissues compared to T5. c Venn diagram showing down-regulated genes in tissues compared T5. d Co-upregulated genes in KEGG enrichment for T5. e Co-downregulated genes in KEGG enrichment for T5.

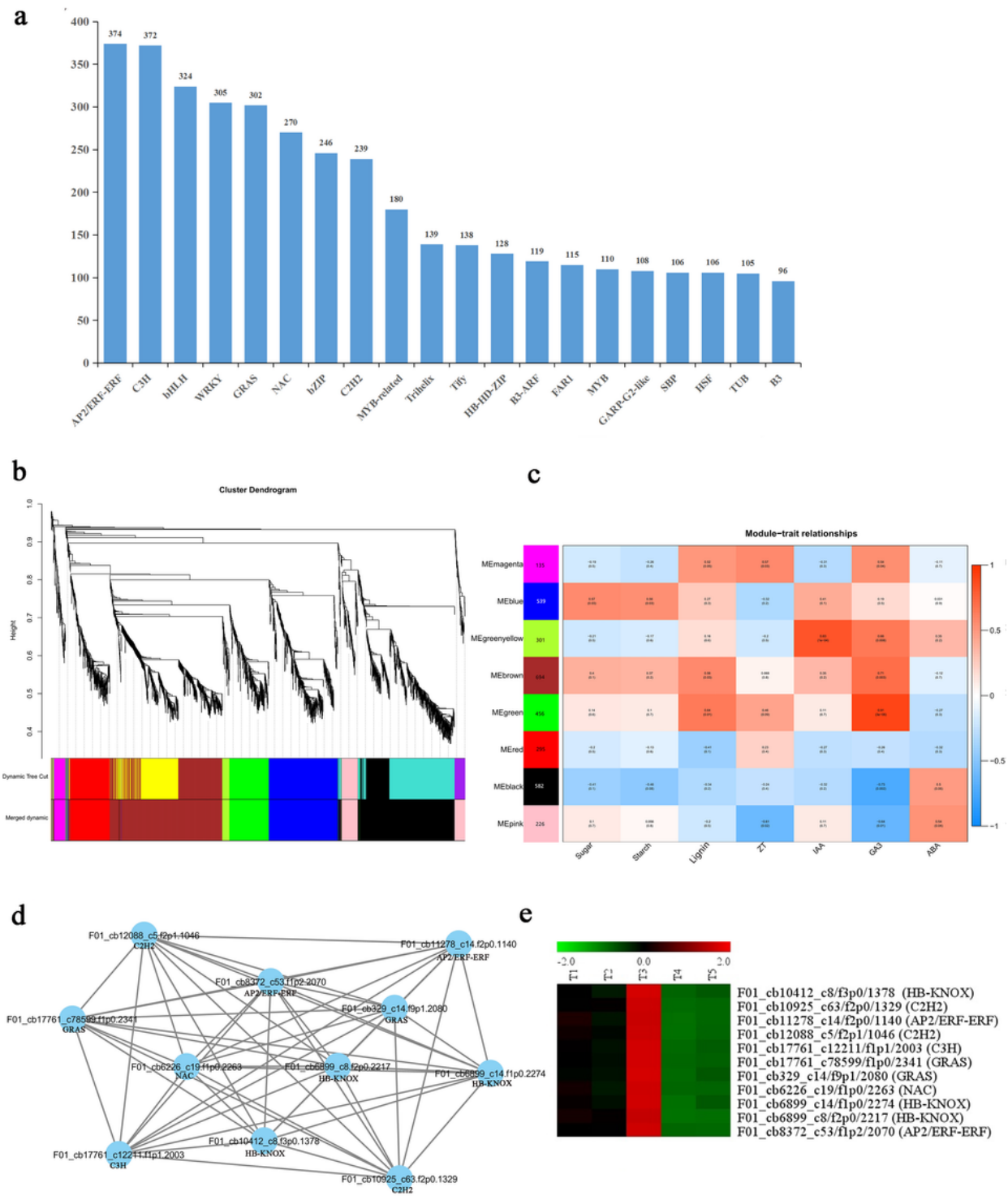

Figure 4

The results of TF WGCNA. a Number of top 20 TFs. b Hierarchical cluster tree showing co-expression modules identified by WGCNA. c Module-sample association relationships. d Correlation networks of hub genes in the green module. e Heatmap of hub genes in the green module.

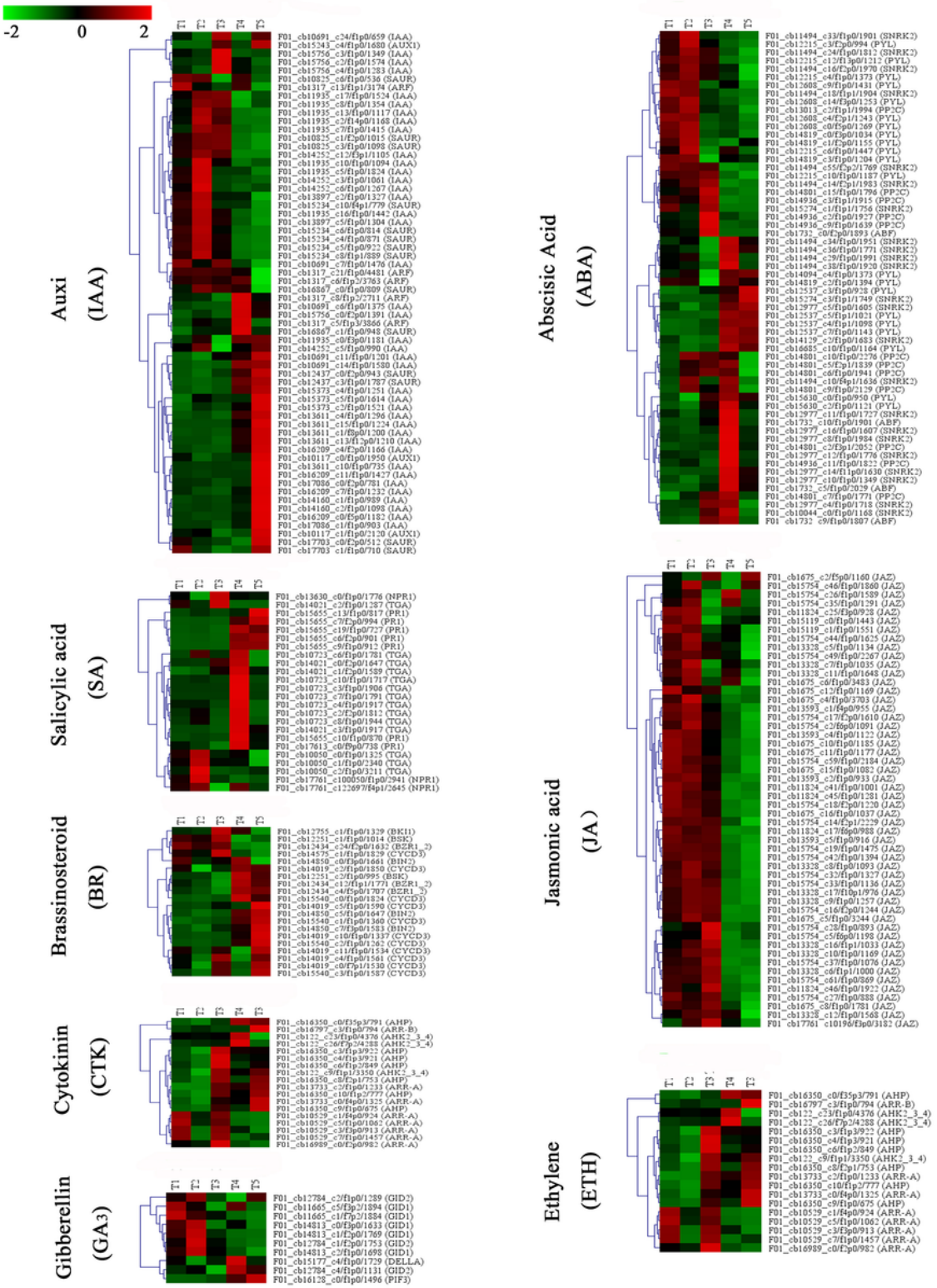

Figure 5

Heatmap of hormone signaling-related genes in the five tissues.

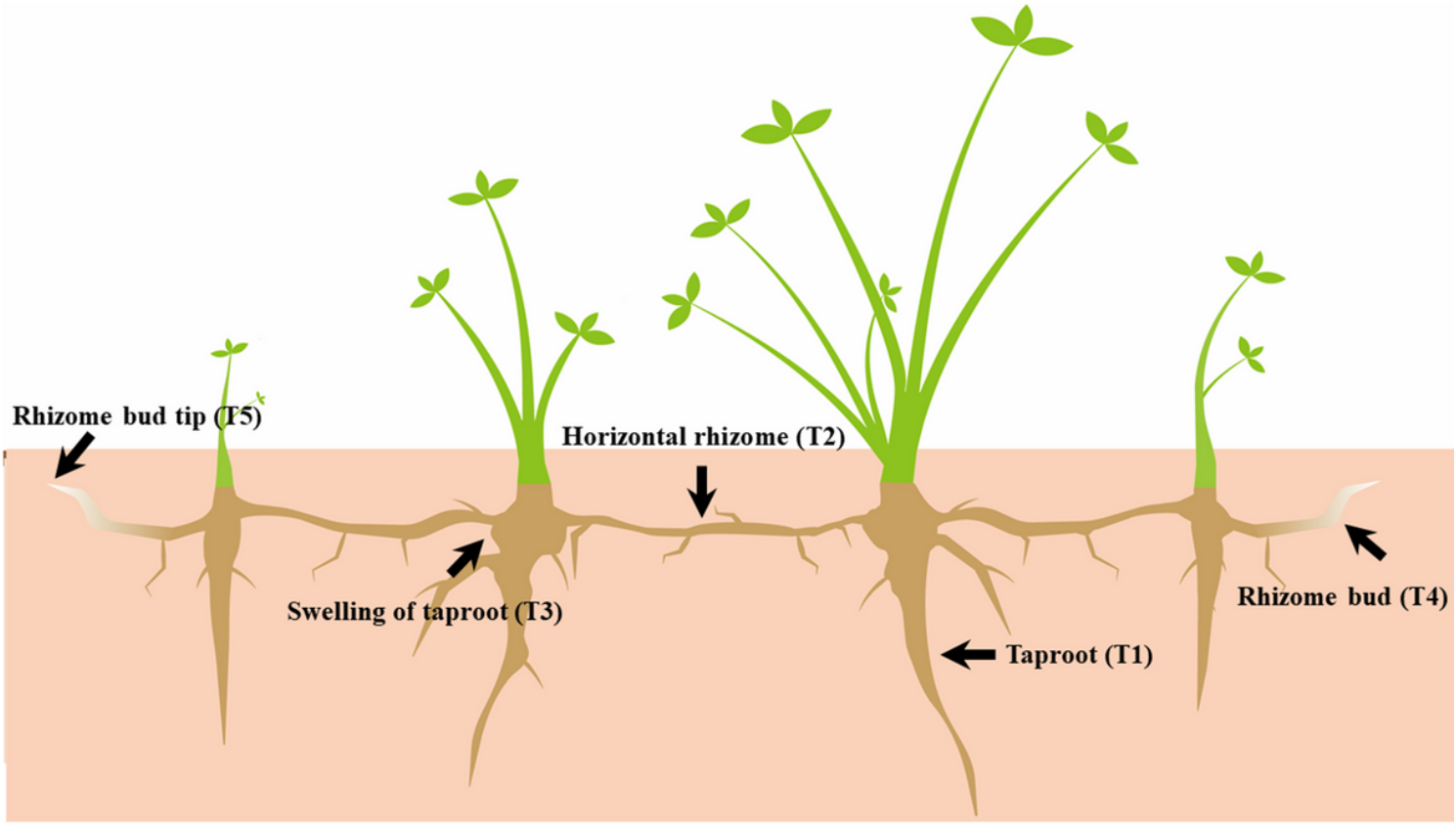

Figure 6

Schematic graph of tissues collected for PacBio sequencing and Illumina sequencing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2tableS1.xlsx
- Additionalfile10TableS5.xlsx
- Additionalfile6TableS3.xlsx
- Additionalfile8TableS4.xlsx
- Additionalfile9FigureS5.png
- Additionalfile7FifureS4.png
- Additionalfile1FigureS1..png
- Additionalfile3FigureS2.png
- Additionalfile4FigureS3.png
- Additionalfile5TableS2.xlsx