genes

MDPI

Article

Full-Length Transcriptome Analysis of the Halophyte *Nitraria sibirica* Pall

Huilong Zhang [1,2,†], Zhen Liu [3,4,†], Aishuang Hu [1,5], Haiwen Wu [1,2], Jianfeng Zhu [1,2], Fengzhi Wang [3,4], Pingping Cao [3,4], Xiuyan Yang [1,2] and Huaxin Zhang [1,2,*]

1 Institute of Ecological Protection and Restoration, Chinese Academy of Forestry, Beijing 100091, China; hlzhang@caf.ac.cn (H.Z.); hash0207@163.com (A.H.); 13621143840@163.com (H.W.); jfzhu@caf.ac.cn (J.Z.); yangxiuyan@caf.ac.cn (X.Y.)
2 The Comprehensive Experimental Center of Chinese Academy of Forestry in Yellow River Delta, Dongying 257000, China
3 Hebei Key Laboratory of Crop Salt-Alkali Stress Tolerance Evaluation and Genetic Improvement, Cangzhou 061001, China; liuzhen84575151@163.com (Z.L.); cznky420@126.com (F.W.); cp0820@126.com (P.C.)
4 Academy of Agriculture and Forestry Sciences, Cangzhou 061001, China
5 Institute of Coastal Agriculture, Hebei Academy of Agriculture and Forestry Sciences, Tangshan 063299, China
* Correspondence: zhanghx@caf.ac.cn; Tel.: +86-10-6288-9058
† These authors contributed equally to this work.

Abstract: Background: *Nitraria sibirica* Pall. is one of the pioneer tree species in saline–alkali areas due to its extreme salt tolerance. However, the lack of information on its genome limits the further exploration of the molecular mechanisms in *N. sibirica* under salt stress. Methods: In this study, we used single-molecule real-time (SMRT) technology based on the PacBio Iso-Seq platform to obtain transcriptome data from *N. sibirica* under salt treatment for the first time, which is helpful for our in-depth analysis of the salt tolerance and molecular characteristics of *N. sibirica*. Results: Our results suggested that a total of 234,508 circular consensus sequences (CCSs) with a mean read length of 2121 bp were obtained from the 19.26 Gb raw data. Furthermore, based on transcript cluster analysis, 93,713 consensus isoforms were obtained, including 92,116 high-quality isoforms. After removing redundant sequences, 49,240 non-redundant transcripts were obtained from high-quality isoforms. A total of 37,261 SSRs, 1816 LncRNAs and 47,314 CDSs, of which 40,160 carried complete ORFs, were obtained. Based on our transcriptome data, we also analyzed the coding genes of H^+-PPase, and the results of both bioinformatics and functional analyses indicated that the gene prediction via full-length transcripts obtained by SMRT technology is reliable and effective. In summary, our research data obtained by SMRT technology provides more reliable and accurate information for the further analysis of the regulatory network and molecular mechanism of *N. sibirica* under salt stress.

Keywords: full-length transcriptome analysis; *Nitraria sibirica*; H^+-PPase; salt stress

check for updates

Citation: Zhang, H.; Liu, Z.; Hu, A.; Wu, H.; Zhu, J.; Wang, F.; Cao, P.; Yang, X.; Zhang, H. Full-Length Transcriptome Analysis of the Halophyte *Nitraria sibirica* Pall. *Genes* **2022**, *13*, 661. https://doi.org/10.3390/genes13040661

Academic Editors: Zhanshan (Sam) Ma and Stefano Lonardi

Received: 15 February 2022
Accepted: 7 April 2022
Published: 8 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

N. sibirica is a typical salt-diluting halophyte which can complete its life cycle in desert and saline–alkali soils. Therefore, *N. sibirica* is considered to be the pioneer tree for afforestation in saline–alkali areas, and it is widely used in the landscaping and bioremediation of saline–alkali land [1]. In addition, *N. sibirica* can also be used as an economic plant in saline–alkali areas for its medicinal value and edible fruit [2,3]. To date, some studies have tried to reveal the physiological basis of salt tolerance in *N. sibirica* under salt stress [1,4–6]. However, due to the lack of a comprehensive and robust reference genome, most of these studies focused on physiological analyses, such as the accumulation of organic osmolytes [6], distribution of Na^+ and K^+ in various tissues [1,4,5] and changes in antioxidant enzyme activities [7].

Transcriptome data obtained based on high-throughput sequencing technology can reflect the gene types and expression changes in cells, which contributes to an overall, big-data-level analysis which can better analyze gene expression and regulatory networks [8,9]. High-throughput RNA sequencing can be carried out without relying on the reference genome, so this technology can be used for the analysis of most species [8,10–12]. So far, there has been a large amount of research based on second-generation sequencing technology. Some studies used the transcriptome technique to reveal the molecular regulation mechanism of *N. sibirica* under salt stress [13]. The massive data obtained using the second-generation sequencing technology laid the foundation for uncovering the regulatory mechanism for salt tolerance in *N. sibirica*. However, because of the constraints of the second-generation sequencing platform on the read length of the sequence, the RNA needs to be broken into short fragments (<500 bp) [14,15] for sequencing, and finally the transcripts are acquired by short read assembly. The transcripts obtained in this way are likely to be unsuccessfully or incompletely assembled [14,16,17].

Single-molecule real-time (SMRT) sequencing, such as that using the PacBio sequencing platform (PacBio, Menlo Park, CA, USA), is a third-generation sequencing technology which can effectively overcome the disadvantages of second-generation sequencing [14,16,18]. SMRT sequencing technology allows the direct reading of cDNA and obtains high-quality, complete and accurate reads [16]. Based on this technique, the average reading length range is from 10 kb to 15 kb and the maximum length can reach 70 kb, which makes it easier to identify the isomers of shared exons and does not require assembly [14,16,19,20]. In addition, the third-generation sequencing technology can also identify and obtain new transcripts or genes and can be used to supplement genome annotation information, which helps to accurately analyze more information, such as that found in homologous genes and superfamily genes. To date, based on the advantages of the third-generation sequencing technology via the PacBio sequencing platform, this technology has been successfully applied to some species, such as humans [20], red clover (*Trifolium pratense*) [21], Purslane (*Portulaca oleracea*) [17], *Ananas comosus* var *bracteatus* [22] and pear (*Pyrus betulifolia* Bunge) [23]. However, full-length transcript data based on SMRT sequencing technology has not been reported in *N. sibirica*.

In our previous study, we found that the growth of *N. sibirica* was promoted under 200 mM NaCl concentration [1,4] but decreased under higher than 300 mM or without NaCl treatment. This is a response specific to salt stress for most halophytes, most of which grow best in low salt concentrations. In other words, halophytes are uniquely adapted to salt environments. To better explain this characteristic of *N. sibirica*, 200 mM NaCl treatment was used for sequencing in this study. In the present study, we obtained the full-length transcriptome data of *N. sibirica* through SMRT sequencing via the PacBio sequencing platform. Based on the high-quality and accurate transcriptome data, CDS (sequence coding for amino acids in protein) identification, simple sequence repeat (SSR) prediction, long-chain noncoding RNA (LncRNA) analysis and function annotation were performed. Meanwhile, we identified coding genes of H^+-PPase based on full-length transcriptome data, analyzed their subcellular localization information and verified their expression in yeast using the DUAL membrane system. This study provides full-length transcript data to support the molecular and gene-function studies of *N. sibirica*, and also lays the foundation for the further clarification of the regulatory mechanism of *N. sibirica* under salt stress.

2. Materials and Methods

2.1. Plant Material and Salt Treatment

The seeds of *N. sibirica* were collected from saline–alkaline beach of Keluke lake in Qaidam Basin of Qinghai province, China. The seeds were treated in ddH_2O for 24 h at 40–50 °C, then buried in wet sand and placed in the conservatory for germination. After germination, seedlings were transferred to a basin filled with a vermiculite:perlite mixture (3:1 w:w) and cultured in the greenhouse of the Chinese Academy of Forestry at 28/22 °C day/night and at 60–80% relative humidity. After 9 weeks of culture, plants with

similar growth were selected for 200 mM NaCl treatment. The roots, stems and leaves were harvested and frozen in liquid N_2 and stored at −80 °C for RNA extraction.

2.2. Nucleic Acid Extraction and SMRT Sequencing

An RNA extraction kit (Takara, Kusatsu, Japan) was used to obtain the total RNA, and all samples were mixed. The Qubit Fluorometer was used to assess the concentration of RNA under the processing of the Qubit RNA IQ Assay Kit. The integrity of the RNA was assessed based on data measured using the 2100 Agilent Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA). Library construction can only be performed when the mass of RNA meets a 280/260 ratio of 1.9 to 2.1, a 230/260 ratio of 2.0 to 2.4 and a RIN value greater than 7.0. The library for this experiment was made by mixing equal amounts of total RNA extracted from three different tissues.

SMRTbell library was prepared and sequenced following the process of the PacBio Sequel platform (Pacific Biosciences, CA, USA). The cDNA fragments were obtained via the PCR process with oligo dT as a primer and screened according to length (two sizes, <4000 bp and >4000 bp). Then, after a series of modifications, the SMRT adaptor was connected to complete the library construction. Finally, the PacBio Sequel SMRT sequencer from Pacific Biosciences was used to obtain the raw data.

2.3. Preprocessing of Raw Data

SMRTlink 6.0 software was used to analyze the raw data and filter out all polymerase reads of less than 50 bp with a quality less than 0.90 to obtain the clean reads. A circular consensus sequence (CCS) was obtained from subread BAM files with min FullPass = 3 and min Predicted Accuracy = 0.9, CCS.BAM files were output. Furthermore, the BAM files were segmented into both full-length (FL) and non-full-length non-chimeric (FLNC) based on pbclassify. The FL and FLNC transcripts were identified through traces of the poly(A)-tail signal and the 5′ and 3′ cDNA primers in error-corrected reads of insert (ROI). The consistency sequences were obtained by iterative error correction (ICE) approximation clustering, and the FL consistency sequences in ICE were polished using Quiver. We selected the consensus data with a post-correction accuracy of >0.99 for subsequent analysis. The completeness of the transcriptome was assessed via BUSCO/v3.0.2 with the embryophyta_odb9 database [24].

2.4. CDS, SSR, Transcription Factors (TFs), LncRNA Analysis

The coding sequences were predicted by TransDecoder (https://github.com/TransDecoder/TransDecoder/releases, accessed on 31 August 2020), and the SSR identification was performed based on the MISA database (https://webblast.ipk-gatersleben.de/misa/, accessed on 31 August 2020). Transcription factors were predicted using ITAK software. LncRNA candidate sequences were screened from the putative protein-coding RNAs in the transcripts according to bricks with lengths of more than 200 nt. The obtained transcripts were further confirmed according to the CPC/CNCI/CPAT/Pfam database for their ability to distinguish protein-coding genes from non-coding genes [25].

2.5. Functional Annotation

The non-redundant transcript function information was obtained through BLASTx (version 2.2.26) [26] with an e-value < 10^{-5} based on the following databases: NCBI non-redundant protein sequences database (NR), Protein family (Pfam), a manually annotated and reviewed protein sequence database (Swiss-Prot), Gene Ontology Consortium (GO), Cluster of Orthologous Groups of proteins (COG), Clusters of Orthologous Groups of proteins (KOG), evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) [25].

2.6. Alternative Splice Detected

To obtain alternative splicing (AS) events for *N. sibirica*, BLAST with high-intensity settings was performed directly using Iso-SeqTM data. The screening of products for candidate AS events was based on the following criteria. First, two high-scoring segment pairs (HSPs) must exist in the sequence and have the same direction. There is at least one sequence that should be contiguous or have a small "overlap" size (less than 5 bp) in the same alignment. In addition, the "AS Gap" appears in another sequence, and consecutive sequences should be almost completely aligned with distinct sequences. We defined a sequence with a gap that should be greater than 100 basis points and at least 100 basis points from the 3′/5′ end as an AS Gap.

2.7. Phylogenetic and Subcellular Localization Analysis

MEGA-X software (Temple, Philadelphia, PA, USA) was used to reconstruct the phylogenetic tree with the neighbor-joining method, and the bootstrap value was set to 1000. TMHMM (http://www.cbs.dtu.dk/services/TMHMM/, accessed on 31 August 2020) was used for the transmembrane domain analysis of the screened candidate genes.

2.8. Function Analysis of H^+-PPase

RNA was obtained from *N. sibirica* following the manufacturer's protocol for the E.Z.N.A. Total RNA Kit (Omega Bio-Tek, Doraville, GA, USA). The first-strand cDNA was obtained from 1 µg of total RNA via HiFiScript gDNA Removal RT MasterMix (CoWin Biosciences, Beijing, China). The complete coding sequencing was amplified with the specific primers shown in Supplementary Table S1.

Function assays were performed following the DUAL membrane starter kit user Manual (Dualsystems Biotech AG, Schlieren, Switzerland). The coding sequences of two H^+-PPase genes were obtained via PCR amplification, and the specific primers for the target genes are listed in Supplementary Table S1. The vectors were cotransformed into yeast strain NMY51 according to the standard process. The empty vectors were used as negative controls. Function was analyzed based on their ability to grow on Leu/Trp/His/Ade+X-α-Gal plat.

3. Results

3.1. The Sequencing Data Analysis of N. sibirica

The qualified library was sequenced using SMRT, and the redundant or error consensus sequences were removed from the original sequences. A total of 19.26 Gb raw data were identified. The raw reads were mainly distributed in the range from 1 to 5000 bp, and the mean length was 2080 bp (Figure 1).

In order to further reduce the error rate, we carried out circular consensus sequence (CCS) analysis to obtain accurate data. In the present study, we obtained a total of 234,508 CCSs based on the full-length transcriptome data, with an average length of 2121 bp. In addition, we also constructed a consensus-length distribution map to show the calculation results based on the consensus length and found that the consensus sequence was dominated by a length of 1–4000 bp (Figure 2). The number of full-length non-chimeric (FLNC) readings was 195,027, accounting for 83.16%, of which the number of high-quality consensus isomers was 92,116. In order to further remove the redundant sequences, we used the CD-HIT program to filter them and finally obtained 49,240 transcripts for subsequent analysis. The analysis of transcriptome completeness with BUSCO is shown in Figure 3. Using the 1440 BUSCO plant set (embryophyta_odb9 database), a total of 1120 (77%) genes from the transcriptome were complete, with 662 (46%) single-copy genes and 458 (34.8%) duplicated; 96 (6.7%) were fragmented and 224 (15.5%) were missing BUSCOs (Figure 3).

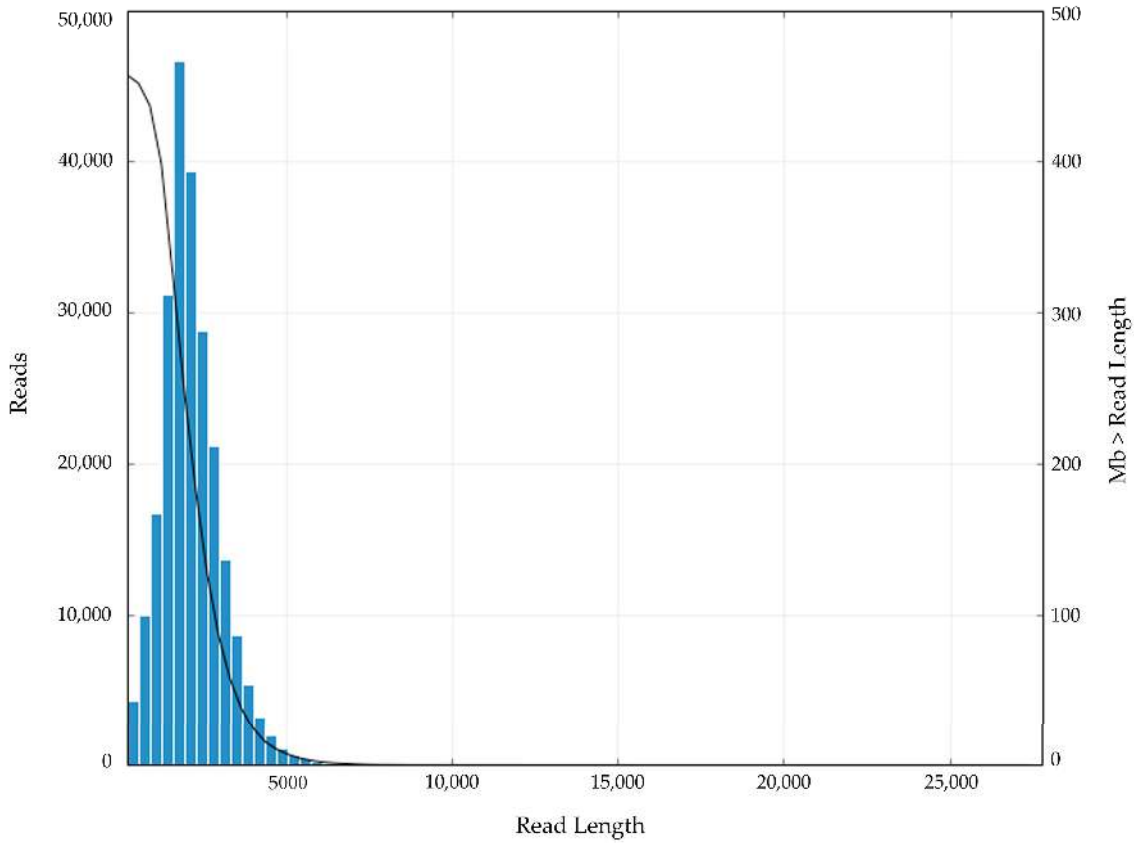

Figure 1. Distribution of the number and length of raw reads.

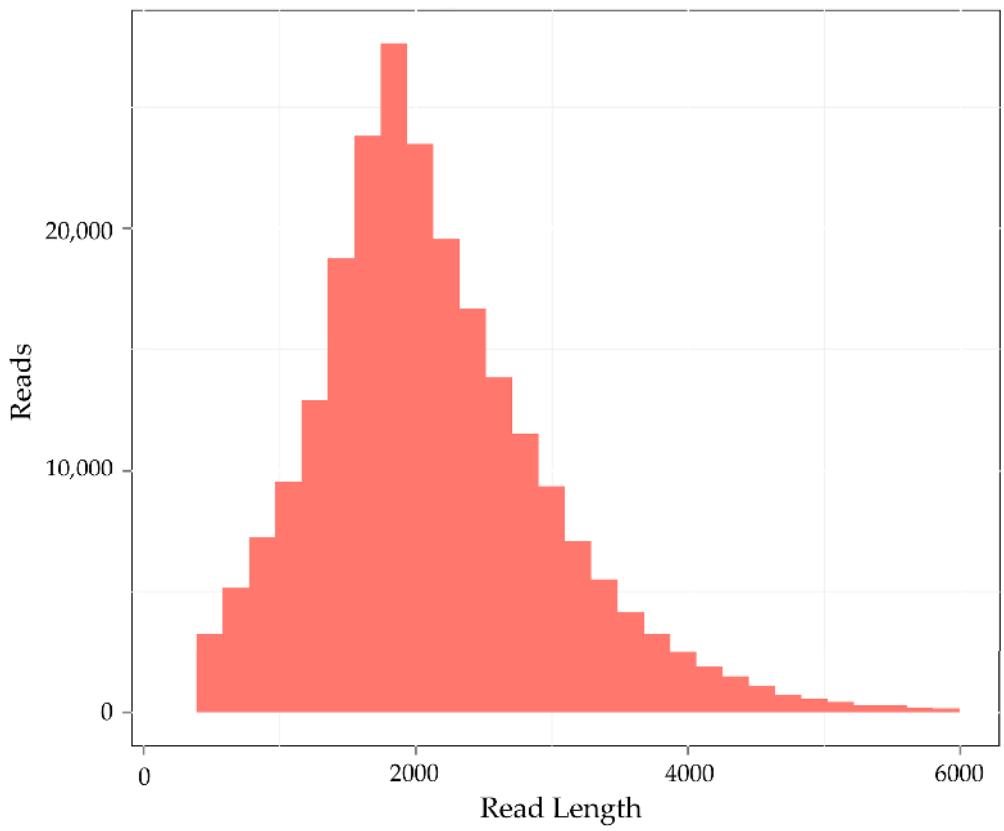

Figure 2. Distribution of the read number and length of CCS.

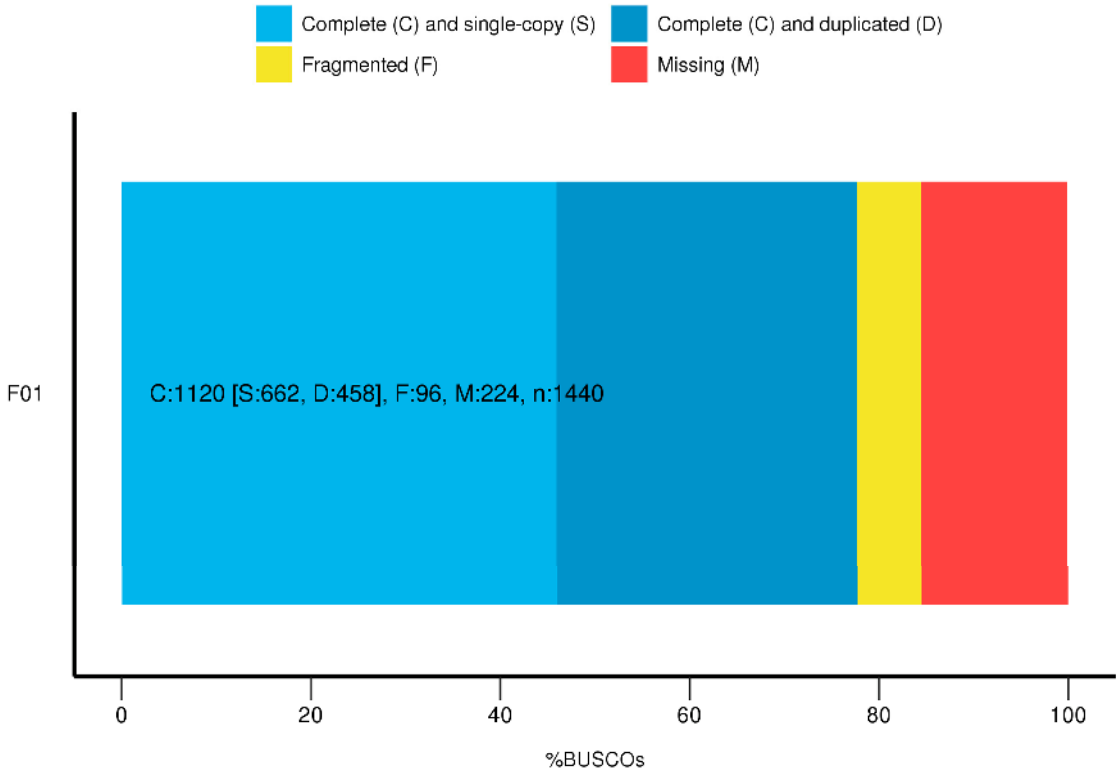

Figure 3. Result of the BUSCO assessment.

3.2. Prediction of Coding Sequences

We used TransDecoder sequencing data to screen the coding sequence. A total of 47,314 coding sequences were predicted (Supplementary Table S3), of which 40,160 coding sequences had initial and termination codons, and these sequences were defined as a complete open reading frame (ORF). The predicted length distribution of the protein encoded by the complete ORF region is shown in Figure 4, in which amino–acid sequence lengths $\leq$ 1000 aa accounted for 97.81% (39,287), followed by those from 1000 to 2000 aa, which accounted for 2.15% (867), and for those >2000 aa, there were only six.

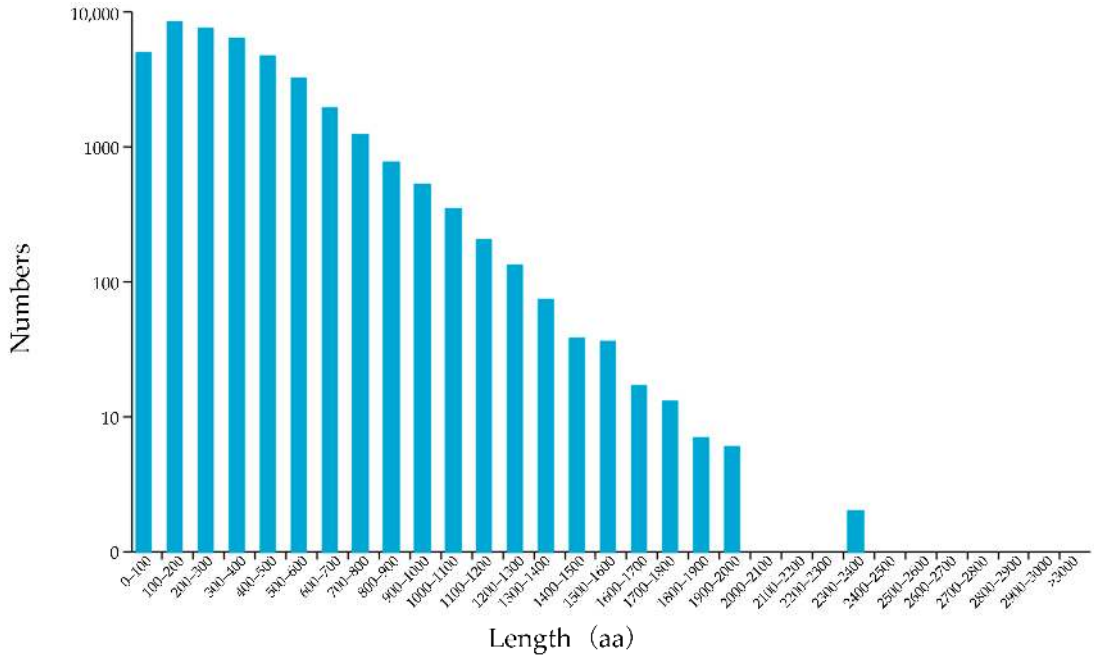

Figure 4. The predicted length distribution of the protein encoded by the complete ORF.

Transcription factors (TFs) play a critical role in the transcriptional regulation system and widely participate in the regulation of diverse, crucial cellular activities in plants. A total of 4596 transcription factors were confirmed based on the iTAK software, and RLK-Pelle_DLSV (219), bZIP (125), C3H (125) and MYB-related (119) were the top four TF families (Figure 5 and Supplementary Table S4). These TFs identified from the full-length transcriptome lay a foundation for exploring the molecular regulation mechanism of salt tolerance in *N. sibirica*.

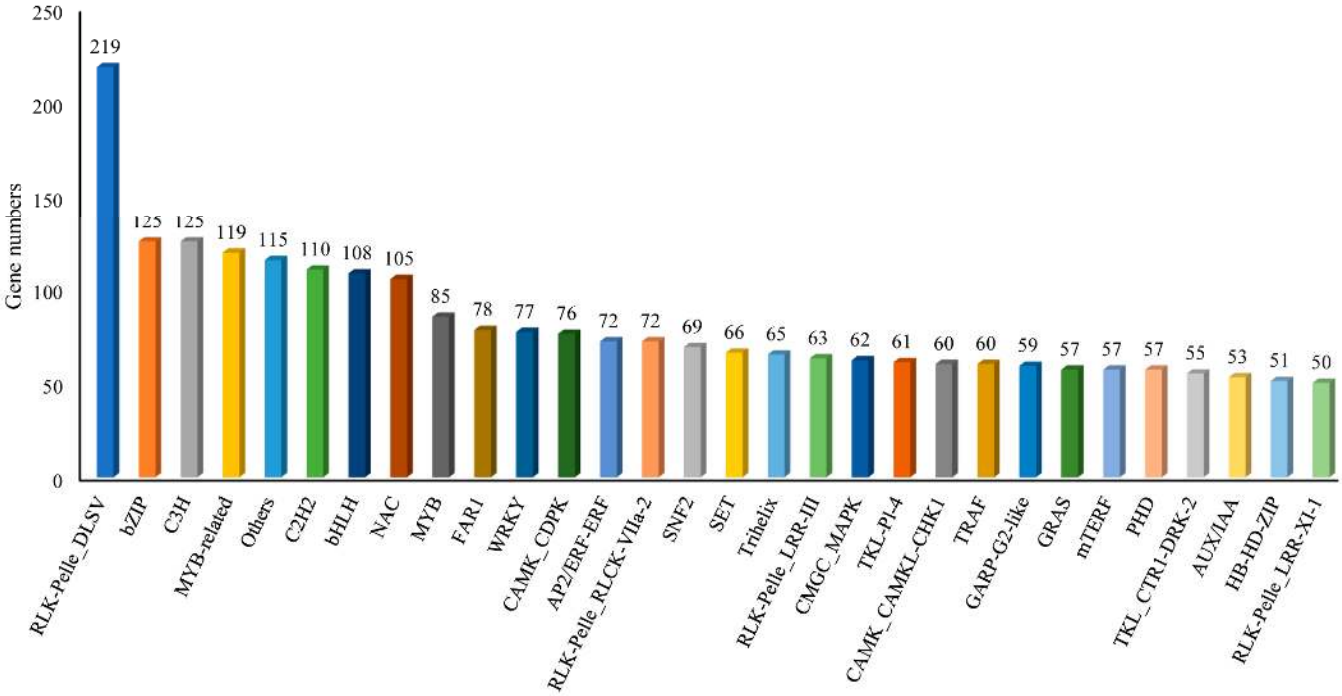

Figure 5. Numbers and families of the top 30 TFs in *N. sibirica* under salt stress.

3.3. SSR Discovery

The MIcroSAtellite identification tool (MISA) is software for identifying simple repetitive sequences. In our analysis, the MISA software was used for SSR detection from the full-length transcript data. In total, 37,261 SSR candidates were identified from our full-length transcriptome; the mononucleotide motif (21,885, 58.73%) was the main type of SSR, followed by the dinucleotide motif (9024, 24.21%), trinucleotide (5553, 14.90%), tetranucleotide (466, 1.25%), hexanucleotide motif (209, 0.56%) and pentanucleotide (124, 0.33%) (Table 1). In addition, the SSR density result also showed similar results (Figure 6 and Supplementary Table S5).

Table 1. SSRs obtained from transcripts.

Search Item	Numbers
Total number of sequences examined	48,442
Total size of examined sequences (bp)	109,925,036
Total number of identified SSRs	37,261
Number of SSR-containing sequences	21,573
Number of sequences containing more than 1 SSR	9024
Number of SSRs present in compound formation	5280
Mononucleotide	21,885
Dinucleotide	9024
Trinucleotide	5553
Tetranucleotide	466
Hexanucleotide	209
Pentanucleotide	124

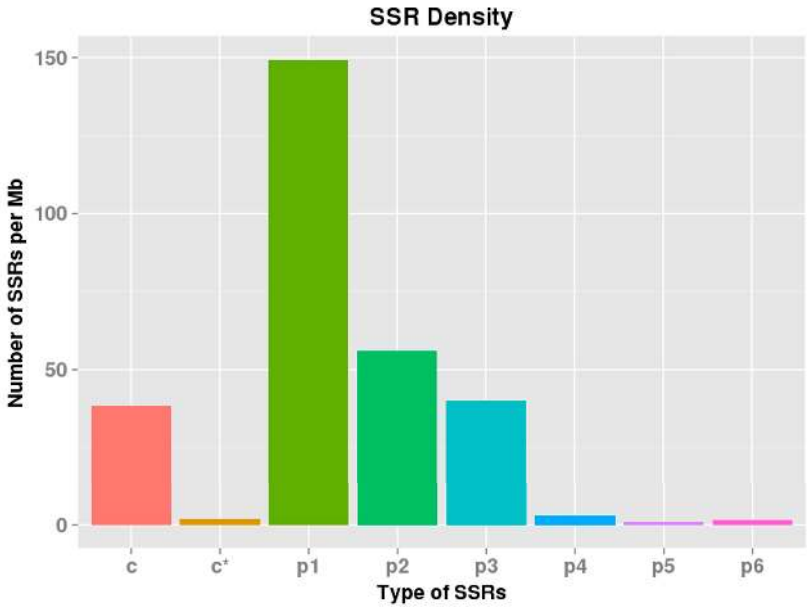

Figure 6. SSR density distribution.

3.4. LncRNA Prediction and Alternative Splicing Analysis

As lncRNA does not encode protein, by screening the coding potential of the transcript, we can judge whether it has the coding potential, filter out the transcript with the coding potential and then obtain the candidate lncRNA. In the present study, we used four tools to identify the lncRNA from the transcriptome. In addition, we constructed a Venn diagram to describe the lncRNAs screened via the four methods to identify common sequences (Figure 7). The numbers identified by CNCI, CPC, Pfam and CPAT were 3409, 3346, 9292 and 9072, respectively. A total of 1816 common sequences were identified in four tools at the same time (Figure 7). In addition, the target genes of lncRNAs were predicted (Supplementary Table S6). The information of alternatively spliced (AS) isoforms in *N. sibirica* is still unclear. In the present study, candidate AS events were generated from a full-length transcriptome via the all-vs-all BLAST with high identity settings. In our result, there were 2775 candidate AS events that met all the criteria and are considered AS events in *N. sibirica*.

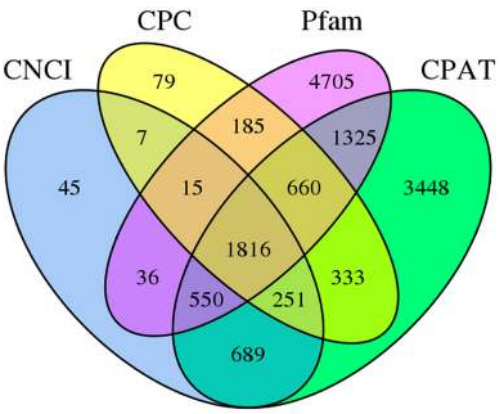

Figure 7. LncRNAs predicted based on CNCI, CPC, Pfam and CPAT.

3.5. Functional Annotation

The non-redundant transcript sequences were obtained based on BLAST software (version 2.2.26) and annotated via the COG, GO, KEGG, KOG, Pfam, SwissProt, eggnog and NR databases, and 46,969 transcripts were identified (Figure 8A). The comparison results based on the NR database suggested that 46,799 transcripts were successfully annotated, and sequence similarity with *Citrus sinensis* (14,903) was the highest (Figure 8B).

A total of 19,714 and 30,675 transcripts were annotated in the COG and KOG, respectively (Figure 8C,D). There were 20,554 transcripts that were mapped to 129 KEGG pathways and most pathways clustered significantly in the carbon metabolism, biosynthesis of amino acids and spliceosome (Figure 9).

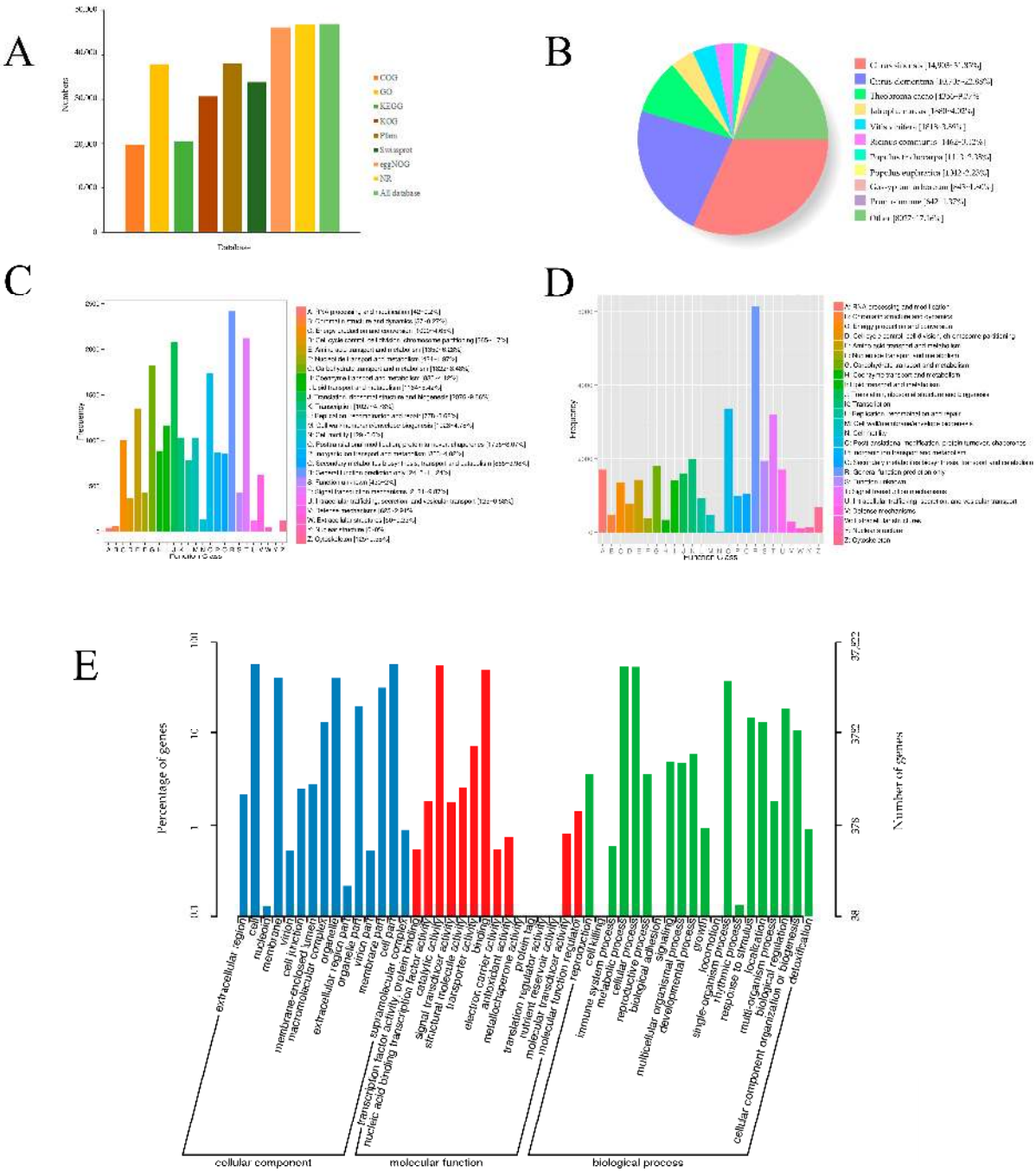

Figure 8. Functional annotation of the corrected transcripts: (**A**) Function annotation of *N. sibirica* transcripts in eight databases. NCBI non-redundant protein sequences database (NR); Protein family (Pfam); a manually annotated and reviewed protein sequence database (Swiss-Prot); Gene Ontology Consortium (GO); Cluster of Orthologous Groups of proteins (COG); Clusters of Orthologous Groups of proteins (KOG); evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG); and Kyoto Encyclopedia of Genes and Genomes (KEGG). (**B**) NR-annotated homologous species distribution. The best hits with an e-value = 10^{-6} for each query were grouped according to species. (**C**) COG annotation of transcript sequences. (**D**) KOG classification diagram of *N. sibirica* transcripts. (**E**) Classification of the all transcripts annotated by the GO.

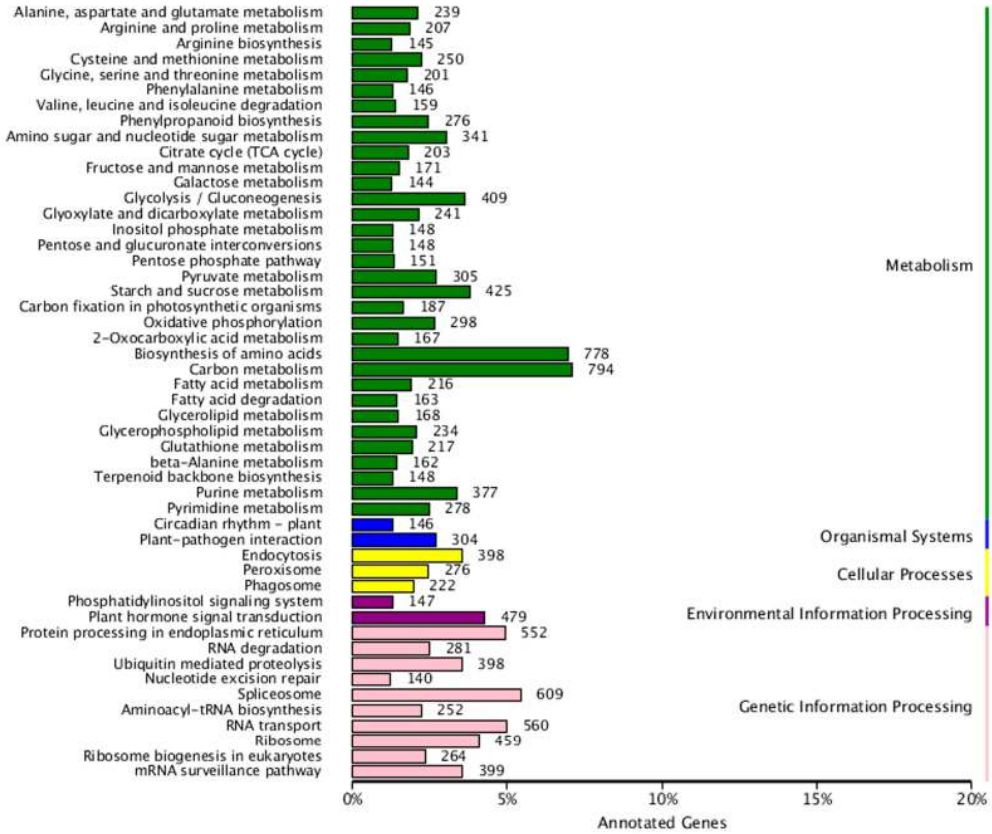

Figure 9. KEGG pathway classification diagram of *N. sibirica* transcripts.

Finally, numerous transcripts in the "cellular component" were mainly associated with the cell, cell part, membrane, organelle and membrane part. The catalytic, binding, transporter, structural molecular activities and nucleic acid binding transcription factor were mainly present in the category "molecular functions". "Biological processes" such as the metabolic process, cellular process, single-organism process, biological regulation and response to stimulus contain most of the genes (Figure 8E).

3.6. Characterization of H^+-PPase Genes Based on Full-Length Transcriptome

With the advantage of SMRT sequencing technology, the coding sequences of H^+-PPase were identified via the full-length transcriptome data. As a result, eight coding sequences of H^+-PPase were screened out and named with their coding number in the full-length transcriptome (Supplementary Table S2). Structural analysis showed that all the identified genes contained the Pf03030 domain. The phylogenetic tree constructed by the neighbor-joining method indicated that these genes belonged to the PPase family and were classified into two groups (Figure 10A). After multiple sequence alignment, the two genes (F01_transcript_4398 and F01_transcript_3275) with the longest coding sequences were selected for subsequent analysis. The subcellular location pattern of two candidate genes was predicted based on the TMHMM database. The result showed that F01_transcript_4398 and F01_transcript_3275 were more likely to be localized on the cell membrane (Figure 10B,C). In addition, the predicted TMH numbers of F01_Transcript_3275 and F01_Transcript_4398 were 13 and 14, respectively (Figure 10B,C).

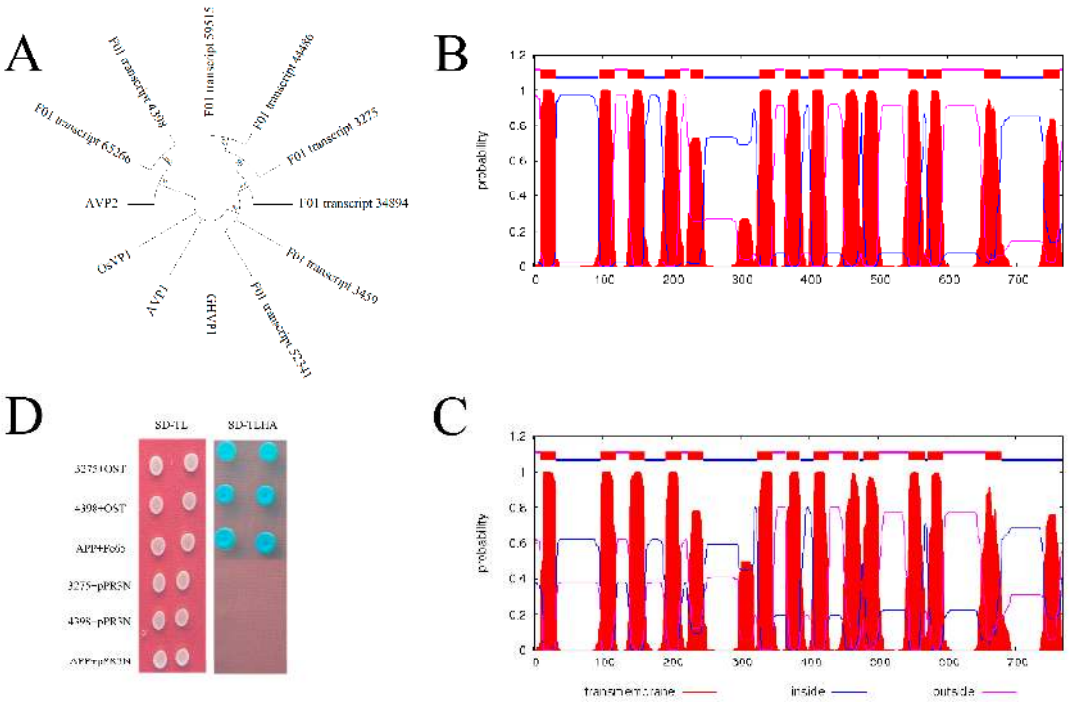

Figure 10. (**A**) Phylogenetic analysis of H^+-PPase genes. (**B**,**C**) H^+-PPase transmembrane helix-domain prediction by TMHMM. (**D**) Function analysis by DUAL membrane system, 3275: F01_Transcript_3275-pBT-3N; 4398: F01_Transcript_4398-pBT-3N; OST: positive prey control vector, pOst1-NubI; pPR3N: empty prey vector; Fe65: positive prey vector, pNubG-Fe65; APP: positive bait vector, pTSU2.

The DUAL membrane system vector, which is a yeast-based screening assay to identify and characterize membrane-associated proteins, was constructed and cotransformed into NMY51 (Figure 10D). The results showed that all cotransformed yeasts grew normally on the TD–LT medium, indicating that the cotransformation was successful. The growth phenotype on the TD–LTHA+α-β-gal plates showed that the F01_transcript_4398 and F01_transcript_3275 grew normally when cotransformed with the positive vector OST (Figure 10D), which indicated that both genes had transmembrane domains and could be used for follow-up experiments. The results of the control group show that our experimental system is reliable (Figure 10D).

4. Discussion

N. sibirica is a typical halophytic shrub with strong environmental adaptability, which is broadly found in desert, saline–alkali and coastal saline–alkali areas. Therefore, *N. sibirica* is a major landscaping species and native economic tree in these areas. To date, some studies on the salt tolerance and germplasm resources of *N. sibirica* have been carried out [1,2,4–6,13,24]. In our previous study, we not only completed the analysis of phenotype and physiology of *N. sibirica* under different salt-stress treatments [1,4,5,7] but also compared the transcriptome data under different salt-treatment concentrations based on NGS sequencing technology [13,27]. We found that the growth of *N. sibirica* was promoted under a 200 mM NaCl concentration but decreased under concentrations higher than 300 mM or without NaCl treatment, while glycophytic plant growth could be seriously damaged under the same conditions [28,29]. That is a specific characteristic of halophyte plants. However, the lack of whole genome information on *N. sibirica* seriously hinders the further development of related molecular research, which is not consistent with its important ecological and economic value. In order to obtain more information on this salt condition, we carried out the full-length transcriptome sequencing based on SMRT technology in this

project, aiming to provide more detailed and comprehensive data on this salt-condition characteristic of *N. sibirica* at the molecular level.

In the present work, we analyzed RNA samples from *N. sibirica* with salt-stress treatments and obtained full-length transcriptome data via the PacBio Iso-Seq platform, which provided more comprehensive genetic information. As a result, sufficient full-length transcriptome data of *N. sibirica* were obtained, including 19.26 Gbp of clean data and 234,508 CCSs, of which 195,027 were screened as FLNC transcripts. Furthermore, based on transcript cluster analysis, 93,713 consensus transcripts were obtained with an average length of 2080 bp, which is significantly better than the average length (555 bp) of transcripts obtained based on the NGS platform in the previous study [27]. This result indicated the advantage of long read-length based on the PacBio Iso-Seq platform, and has also been validated in other species, such as *Sogatella furcifera* (average length 2994 bp) [25], *A. comosus* var. *bracteatus* (average length 2404 bp) [22], and *P. betulifolia* Bunge (average length 1502 bp) [23]. Furthermore, 47,314 coding sequences were generated using TransDecoder sequencing data from the full-length transcriptome data. The number of coding sequences generated by second-generation sequencing was 78,927 [27], which is more than the number in this study. This may be due to the fact that the second-generation sequencing requires the assembly of short reads to obtain the complete sequence, but different assembly methods may not obtain the same results [30]. This result suggests that full-length transcriptome data obtained grounded on SMRT technology may better reflect the complete information about the mRNA in the organism [14,23,31].

LncRNAs are a kind of non-protein coding transcript of longer than 200 nt, which is of great significance in many biological pathways [32–34]. Using the non-protein-coding characteristic of lncRNAs, we obtained candidate lncRNAs by filtering out transcripts with coding potential and removing them. At last, 1816 lncRNAs were identified based on CNCI, CPC, Pfam and CPAT methods, and their target genes were also obtained. This number was more than the number in *A. comosus* var. *bracteatus* (329) [22] but less than *P. betulifolia* (7094) [23] and *Ginkgo biloba* L. (4363) [35]. In recent years, the functions of lncRNAs in salt tolerance, such as *Pistacia vera* [36], duckweed [37] and *C. sinensis* [38], were reported. Although crucial progress has been made in the study of lncRNA function in some plants, lncRNA function in *N. sibirica* still needs more studies in the future.

Tissue salt-tolerance is one of the most important strategies for plants to adapt to a saline environment, and one which widely exists in a variety of halophytes. *N. sibirica* is a typical halophyte of the dilute-salt type. Previous studies showed that *N. sibirica* mainly transports Na^+ into leaves and compartmentalizes it into vacuoles to decrease the toxicity caused by Na^+ under salt stress [1,4,5]. H^+-PPase plays a crucial role in the process of vacuolar Na^+ sequestration, but the molecular network in *N. sibirica* under salt stress remains unclear [1,39,40]. Based on the advantages of the full-length transcriptome, the database obtained by the full-length transcriptome can be used for the screening and identification of functional genes [16,19,31]. In the present work, we identified eight coding sequences of H^+-PPase from our transcriptome data. The results of subcellular localization showed that the two major H^+-PPase (F01_transcript_4398 and F01_transcript_3275) had multiple transmembrane domains (Figure 10B,C). Next, their functions were verified based on the DUAL membrane system (Figure 10D), and the results showed that they all had transmembrane regions [40], which is similar to the results of previous studies [40,41]. In addition, the two main H^+-PPase identified in this study were most closely related to AVP1 in *Arabidopsis*, and F01_transcript_4398 had one more transmembrane structural domain than F01_transcript_3275. This result may imply that the function of H^+-PPase is redundant in *N. sibirica*, which can help this species to better adapt to the salt habitat. However, further work is needed to reveal the functions of these two H^+-PPases.

5. Conclusions

In the present study, SMRT technology based on the PacBio Iso-Seq sequencing platform was applied to obtain complete and accurate transcriptome data of *N. sibirica*. A total

of 37,261 SSRs, 1816 LncRNAs and 47,314 CDSs, of which 40,160 carried complete ORFs, were obtained. These results were easily obtained without reference genome species. In summary, our results show that the SMRT technology based on the PacBio Iso-Seq sequencing platform is a suitable and effective method to obtain accurate and reliable full-length transcripts of *N. sibirica*. Based on our transcriptome data, we also analyzed the coding genes of H^+-PPase, and the results of both bioinformatics and functional analyses indicated that the gene prediction via full-length transcripts obtained by SMRT technology was complete and accurate. Our data provide a set of reference sequences containing a high-quality and valuable full-length transcript database, which provides useful and valuable support for the further analysis of the regulatory network and molecular mechanism of *N. sibirica* under salt stress.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13040661/s1. Table S1: Gene-specific primers used in this study; Table S2: Sequence information of all identified H^+-PPases; Table S3: The predicted information of CDS sequences; Table S4: The predicted information of TF annotation and sequence; Table S5: The predicted information of SSR; Table S6: The predicted information of lncRNA and their target genes.

Author Contributions: H.Z. (Huilong Zhang) and X.Y. conceived the idea, designed the study, and revised the manuscript. H.Z. (Huilong Zhang) and A.H. performed the experiments. H.Z. (Huilong Zhang) and J.Z. performed the bioinformatics analysis. H.Z. (Huilong Zhang) wrote the manuscript. H.W., P.C. and Z.L. helped to collect materials and observe the phenotype. F.W. and H.Z. (Huaxin Zhang) critically revised the manuscript. All the authors have reviewed and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundamental Research Funds of CAF (project no. CAFYBB2020MA007 and CAFYBB2021MB004), the National Natural Science Foundation of China (grant 31870663), the Hebei Key Laboratory of Crop Salt-Alkali Stress Tolerance Evaluation and Genetic Improvement (SZX2019010).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNA sequencing datasets generated in this study have been deposited in the Sequence Read Archive (SRA) in NCBI under the accession number PRJNA804704. Other data supporting our findings are available in the manuscript file or from the corresponding author upon request.

Acknowledgments: We acknowledge everyone who contributed to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tang, X.; Zhang, H.; Sergey, S.; Li, H.; Yang, X.; Zhang, H. Tissue Tolerance Mechanisms Conferring Salinity Tolerance in a Halophytic Perennial Species *Nitraria sibirica* Pall. *Tree Physiol.* **2020**, *7*, 1264–1277. [CrossRef] [PubMed]
2. Song, Q.B.; Zhao, H.D.; Fu, Z.L.; Chen, D.F.; Lu, Y. Antioxidant and Anticomplement Compounds Isolated from *Nitraria sibirica* Fruit by High-Speed Counter-Current Chromatography. *Pharmacogn. Mag.* **2018**, *14*, 541.
3. Su, W.; Zhang, W. Study on Extraction of Red Pigment from *Nitraria sibirica* Fruit and Its Stability. *J. Tianjin Agric. Coll.* **2002**, *4*, 5–11.
4. Tang, X.; Yang, X.; Li, H.; Zhang, H. Maintenance of K^+/Na^+ Balance in the Roots of *Nitraria sibirica* Pall. in Response to NaCl Stress. *Forests* **2018**, *9*, 601. [CrossRef]
5. Tang, X.; Li, H.; Yang, X.; Wu, H.; Liu, X.; Zhang, H. Effect of NaCl Stress on Root Growth and K^+/Na^+ Balance of *Nitraria sibirica* Pall. Seedlings. *J. Northwest A F Univ. Nat. Sci. Ed.* **2019**, *8*, 83–89.
6. Chen, G.; Wang, C.; Chen, J. Osmosis-Regulating Substance and Cation Contents in *Nitraria sibirica* Pall. Test-Tube Plantlets with NaCl Stress. *Acta Bot. Boreali-Occident. Sin.* **2009**, *6*, 1233–1239.
7. Rina, S.A.; Chen, G. Effect of Exogenous Spermidine on Antioxidant Enzyme System in Leaves of *Nitraria sibirica* Pall. Seedlings under Salt Stress. *Acta Bot. Boreali-Occident. Sin.* **2013**, *2*, 352–356.
8. Bashir, K.; Matsui, A.; Rasheed, S.; Seki, M. Recent Advances in the Characterization of Plant Transcriptomes in Response to Drought, Salinity, Heat, and Cold Stress. *F1000Research* **2019**, *8*, 658. [CrossRef]

9. Slatko, B.E.; Gardner, A.F.; Ausubel, F.M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **2018**, *122*, e59. [CrossRef]
10. Huang, J.; Zhao, X.; Chory, J. The Arabidopsis Transcriptome Responds Specifically and Dynamically to High Light Stress. *Cell Rep.* **2019**, *29*, 4186–4199. [CrossRef]
11. Nobori, T.; Velásquez, A.C.; Wu, J.; Kvitko, B.H.; Kremer, J.M.; Wang, Y.; He, S.Y.; Tsuda, K. Transcriptome Landscape of a Bacterial Pathogen under Plant Immunity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3055–3064. [CrossRef] [PubMed]
12. Petre, B.; Lorrain, C.; Stukenbrock, E.H.; Duplessis, S. Host-Specialized Transcriptome of Plant-Associated Organisms. *Curr. Opin. Plant Biol.* **2020**, *56*, 81–88. [CrossRef] [PubMed]
13. Li, H.; Tang, X.; Yang, X.; Zhang, H. Comprehensive Transcriptome and Metabolome Profiling Reveal Metabolic Mechanisms of *Nitraria sibirica* Pall. to Salt Stress. *Sci. Rep.* **2021**, *11*, 12878. [CrossRef] [PubMed]
14. Ardui, S.; Ameur, A.; Vermeesch, J.R.; Hestand, M.S. Single Molecule Real-Time (SMRT) Sequencing Comes of Age: Applications and Utilities for Medical Diagnostics. *Nucleic Acids Res.* **2018**, *46*, 2159–2168. [CrossRef] [PubMed]
15. Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-Read Human Genome Sequencing and Its Applications. *Nat. Rev. Genet.* **2020**, *21*, 597–614. [CrossRef]
16. Nakano, K.; Shiroma, A.; Shimoji, M.; Tamotsu, H.; Ashimine, N.; Ohki, S.; Shinzato, M.; Minami, M.; Nakanishi, T.; Teruya, K.; et al. Advantages of Genome Sequencing by Long-Read Sequencer Using SMRT Technology in Medical Area. *Hum. Cell* **2017**, *30*, 149–161. [CrossRef]
17. Du, H.; Zaman, S.; Hu, S.; Che, S. Single-Molecule Long-Read Sequencing of Purslane (*Portulaca oleracea*) and Differential Gene Expression Related with Biosynthesis of Unsaturated Fatty Acids. *Plants* **2021**, *10*, 655. [CrossRef]
18. Zhao, L.; Zhang, H.; Kohnen, M.V.; Prasad, K.V.S.K.; Gu, L.; Reddy, A.S.N. Analysis of Transcriptome and Epitranscriptome in Plants Using PacBio Iso-Seq and Nanopore-Based Direct RNA Sequencing. *Front. Genet.* **2019**, *10*, 253. [CrossRef]
19. Liu, D.; Chen, L.; Chen, C.; An, X.; Zhang, Y.; Wang, Y.; Li, Q. Full-Length Transcriptome Analysis of *Phytolacca americana* and Its Congener *P. icosandra* and Gene Expression Normalization in Three Phytolaccaceae Species. *BMC Plant Biol.* **2020**, *20*, 396. [CrossRef]
20. Sharon, D.; Tilgner, H.; Grubert, F.; Snyder, M. A Single-Molecule Long-Read Survey of the Human Transcriptome. *Nat. Biotechnol.* **2013**, *31*, 1009–1014. [CrossRef]
21. Chao, Y.; Yuan, J.; Li, S.; Jia, S.; Han, L.; Xu, L. Analysis of Transcripts and Splice Isoforms in Red Clover (*Trifolium pratense* L.) by Single-Molecule Long-Read Sequencing. *BMC Plant Biol.* **2018**, *18*, 300. [CrossRef] [PubMed]
22. Ma, J.; Xiang, Y.; Xiong, Y.; Lin, Z.; Xue, Y.; Mao, M.; Sun, L.; Zhou, Y.; Li, X.; Huang, Z. SMRT Sequencing Analysis Reveals the Full-Length Transcripts and Alternative Splicing Patterns in *Ananas comosus* Var. bracteatus. *PeerJ* **2019**, *7*, e7062. [CrossRef] [PubMed]
23. Yang, Q.; Wang, J.; Wang, Z.; Li, X.; Kan, J.; Lin, J. SMRT Sequencing of Full-Length Transcriptome of Birch-Leaf Pear (*Pyrus betulifolia* Bunge) under Drought Stress. *J. Genet.* **2021**, *100*, 29. [CrossRef]
24. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
25. Yang, H.; Xu, D.; Zhuo, Z.; Hu, J.; Lu, B. SMRT Sequencing of the Full-Length Transcriptome of the *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae). *PeerJ* **2020**, *8*, 9133. [CrossRef]
26. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]
27. Li, H.; Tang, X.; Zhu, J.; Yang, X.; Zhang, H. De Novo Transcriptome Characterization, Gene Expression Profiling and Ionic Responses of *Nitraria sibirica* Pall. under Salt Stress. *Forests* **2017**, *8*, 211. [CrossRef]
28. Yuan, F.; Guo, J.; Shabala, S.; Wang, B. Reproductive Physiology of Halophytes: Current Standing. *Front. Plant Sci.* **2018**, *9*, 1954. [CrossRef]
29. Rahman, M.M.; Mostofa, M.G.; Keya, S.S.; Siddiqui, M.N.; Ansary, M.M.U.; Das, A.K.; Rahman, M.A.; Tran, L.S.-P. Adaptive Mechanisms of Halophytes and Their Potential in Improving Salinity Tolerance in Plants. *Int. J. Mol. Sci.* **2021**, *22*, 10733. [CrossRef]
30. Madritsch, S.; Burg, A.; Sehr, E.M. Comparing de Novo Transcriptome Assembly Tools in Di- and Autotetraploid Non-Model Plant Species. *BMC Bioinform.* **2021**, *22*, 146. [CrossRef]
31. Zhu, L.; Lu, L.; Yang, L.; Hao, Z.; Chen, J.; Cheng, T. The Full-Length Transcriptome Sequencing and Identification of Na^+/H^+ Antiporter Genes in Halophyte *Nitraria tangutorum* Bobrov. *Genes* **2021**, *12*, 836. [CrossRef] [PubMed]
32. Meng, X.; Li, A.; Yu, B.; Li, S. Interplay between MiRNAs and LncRNAs: Mode of Action and Biological Roles in Plant Development and Stress Adaptation. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2567–2574. [CrossRef] [PubMed]
33. Yu, Y.; Zhang, Y.; Chen, X.; Chen, Y. Plant Noncoding RNAs: Hidden Players in Development and Stress Responses. *Annu. Rev. Cell Dev. Biol.* **2019**, *35*, 407–431. [CrossRef] [PubMed]
34. Zhang, X.; Wang, W.; Zhu, W.; Dong, J.; Cheng, Y.; Yin, Z.; Shen, F. Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *Int. J. Mol. Sci.* **2019**, *20*, 5573. [CrossRef] [PubMed]
35. Sun, S.; Li, Y.; Chu, L.; Kuang, X.; Song, J.; Sun, C. Full-Length Sequencing of Ginkgo Transcriptomes for an in-Depth understanding of Flavonoid and Terpenoid Trilactone Biosynthesis. *Gene* **2020**, *758*, 144961. [CrossRef] [PubMed]

36. Jannesar, M.; Seyedi, S.M.; Jazi, M.M.; Niknam, V.; Ebrahimzadeh, H.; Botanga, C. A Genome-Wide Identification, Characterization and Functional Analysis of Salt-Related Long Non-Coding RNAs in Non-Model Plant *Pistacia vera* L. Using Transcriptome High Throughput Sequencing. *Sci. Rep.* **2020**, *10*, 5585. [CrossRef] [PubMed]
37. Fu, L.; Ding, Z.; Tan, D.; Han, B.; Sun, X.; Zhang, J. Genome-Wide Discovery and Functional Prediction of Salt-Responsive LncRNAs in Duckweed. *BMC Genom.* **2020**, *21*, 212. [CrossRef] [PubMed]
38. Wan, S.; Zhang, Y.; Duan, M.; Huang, L.; Wang, W.; Xu, Q.; Yang, Y.; Yu, Y. Integrated Analysis of Long Non-Coding RNAs (LncRNAs) and MRNAs Reveals the Regulatory Role of LncRNAs Associated with Salt Resistance in *Camellia sinensis*. *Front. Plant Sci.* **2020**, *11*, 218. [CrossRef] [PubMed]
39. Kotula, L.; Garcia Caparros, P.; Zörb, C.; Colmer, T.D.; Flowers, T.J. Improving Crop Salt Tolerance Using Transgenic Approaches: An Update and Physiological Analysis. *Plant Cell Environ.* **2020**, *43*, 2932–2956. [CrossRef] [PubMed]
40. Segami, S.; Asaoka, M.; Kinoshita, S.; Fukuda, M.; Nakanishi, Y.; Maeshima, M. Biochemical, Structural and Physiological Characteristics of Vacuolar H^+-Pyrophosphatase. *Plant Cell Physiol.* **2018**, *59*, 1300–1308. [CrossRef] [PubMed]
41. Zhang, Y.; Feng, X.; Wang, L.; Su, Y.; Chu, Z.; Sun, Y. The Structure, Functional Evolution, and Evolutionary Trajectories of the H^+-PPase Gene Family in Plants. *BMC Genom.* **2020**, *21*, 195. [CrossRef] [PubMed]