Functional characterization of fifteen hundred transcripts from Ziarat juniper (Juniperus excelsa M.Bieb)

Humaira Abdul Wahid*, Muhammad Younas Khan Barozai, Muhammad Din

Abstract

Background: Ziarat juniper (Juniperus excelsa M.Bieb) is an evergreen and dominant species of Balochistan juniper forests. This forest is providing many benefits to regional ecosystems and surrounding populations. No functional genomics study is reported for this important juniper plant. This research is aimed to characterize the Ziarat juniper functional genome based on the analyses of 1500 transcripts.

Methods: Total RNA from shoot of Juniperus excelsa was extracted and subjected for transcriptome sequencing using Illumina HiSeq 2000 with the service from Macrogen, Inc., South Korea. The Illumina sequenced data was subjected to bioinformatics analysis. Quality assessment and data filtration was performed for the removal of low-quality reads, ambiguous reads and adaptor sequences. The high-quality clean reads data was deposited in the Sequence Read Archive (SRA) at NCBI, and used for downstream processes. Fifteen hundred transcripts were randomly chosen and used for functional characterization.

Results: As a result of homology search 80.3% transcripts showed significant similarities and were placed in significant similarities category, 19.3% transcripts showed low similarities and assigned to the “unclassified” category while 0.4% transcripts are defined as no hits. The functional characterization results showed that most (18%) of the transcripts are involved in metabolism, followed by 11.7% in transcription and 11.5% as structural protein. 8.8% transcripts are engaged in stress response, whereas the transcripts involved in growth and development constituted 6.7%. Transcripts involved in signal transduction represented 5.6%, while 3.5% facilitating transport and 34.1% are involved in hypothetical functions.

Conclusion: The functional annotation data produced in this study will be very useful for future functional genome analysis of Juniperus excelsa.
Introduction

Juniper belongs to genus *Juniperus*, which consists of about 70 species occurring throughout the northern hemisphere of the world [1,2]. In Balochistan juniper forests spread between latitude 30°9’ and 30°37’ N and longitude 67°11´ and 68°3´ E. in dry areas about 1200-3000m above sea level [3]. These juniper forests are pure tracks of *Juniperus excelsa* with one of the big block in Ziarat District. *Juniperus excelsa* is a large shrub or medium sized tree, spreads mainly in north-eastern Greece, southern Bulgaria, Turkey, Middle-East countries Syria and Lebanon, Saudi Arabia Iran, Pakistan, Oman and Caucasus Mountains at an altitude of 2000-4000m [4-6]. These plants are slow growing and can live up to 2000 years [7]. *Juniperus excelsa* forest of Ziarat have significance due to their old age, slow growth rate and provision of valuable services. In spite of a lot of significance presently no functional genomics study is reported for this plant.

In recent years, Next Generation Sequencing (NGS) has appeared as a high-throughput sequencing technology [8]. Various platforms use NGS, such as the 454 Roche Genome Sequencer, the SOLiDs ABI System, the Illumina Genome Analyzer, and showed to be efficient, powerful and cost-effective tools for advanced genomic research [9]. In the present study, Illumina sequencing HiSeq-2000 platform was used for transcriptome analysis of *Juniperus excelsa*. Fifteen hundred transcripts of *Juniperus excelsa* were randomly selected from Illumina sequenced data deposited in the Sequence Read Archive (SRA) at NCBI. The selected transcripts were used for functional characterization using BLASTn algorithm. To our knowledge, this is the first report on the characterization of the transcriptome of *Juniperus excelsa* and this new data-set will be a useful source for future genomic aspects such as to elucidate candidate genes for biotic and abiotic stress resistance, significant medicinal secondary metabolites and comparative genomics of *Juniperus excelsa*.

Methods

**Plant Material**

Samples (shoots) of *Juniperus excelsa* were collected from Sasnamana Valley, Ziarat, Balochistan, Pakistan. The collected samples were stored in air sealed polythene bags at -20°C for further analysis.

**RNA Extraction**

Total RNA extraction was performed by optimized RNA extraction method [10]. Briefly, 100 mg sample was ground in 1ml of preheated (65°C) total RNA extraction buffer 3% (w/v) CTAB, 3% (w/v) PVP, 100 mM Tris HCl (pH 8.0), 25 mM EDTA (pH 8.0), 2M NaCl, 2% SDS (w/v) and 2% 2-mercaptoethanol. The mixture was vigorously mixed and incubated at room temperature for 10 min with 2-3 times vortexing and centrifugation for 15 min at 15,500g at 4°C. Each sample was extracted twice by using an equal volume of chloroform:isoamyl alcohol (24:1), and phenol/chloroform/isoamyl alcohol (PCI) (25:24:1, v/v/v) at 15,500g for 20 min at 4°C. The resulting supernatant was carefully transferred into a new tube, mixed with 20ul of 5M sodium acetate and twice volume of 100% ethanol and was incubated overnight at -20°C for precipitation. After overnight precipitation RNA was collected by centrifugation at 15,500g at 4°C for 20 min. The pellet was washed with 70% ice cold ethanol. After air drying the RNA was dissolved in 80ul DEPC treated water and stored at -20°C until used. To check the integrity of total RNA samples were run on 1% agarose gel stained with ethidium bromide (EtBr), and visualized under UV light. Genova Nano spectrophotometer at an absorbance ratio of A260/230 and A260/280 nm was used to assess the quality and concentration of extracted total RNA. Quality analysis was performed through a Bioanalyzer (Agilent) at Macrogen, Inc., South Korea before deep sequencing.

**Sequencing and Bioinformatics Analysis**

Sequencing was performed through Illumina HiSeq-2000 platform with the service from Macrogen, Inc., South Korea according to the manufacturers’ protocols. The sequenced data was subjected to bioinformatics analysis, such as; data quality analysis by galaxy-FastQC read quality, removal of adaptor sequences by galaxy trimmomatic tool, homology search through BLAST algorithm and functional characterization by applying Gene-Ontology approaches. Sequenced reads were quality assessed with the quality assessment software FastQC (https://usegalaxy.org/). After quality assessment and data filtration reads with 94.67% at Quality score 20 were used for further analysis. The high quality reads produced have been deposited in the Sequence Read Archive (SRA) at NCBI.

How to Cite

(ncbi.nlm.nih.gov/sra) under accession number SRP082133.

Functional Annotation
Functional annotation was performed through BLASTn against the NCBI nonredundant (NR) DNA database (ftp://ftp.ncbi.nlm.nih.gov/BLAST/db/). The transcripts with query coverage more than 50% were regarded as significant similarity. The transcripts with below 50% query coverage were deemed to have no significant similarity and were assigned to the “unclassified” category and transcripts with no similarity were defined as “no hits” [11-13]. The transcripts were further categorized on the basis of available literature into different functional categories that was metabolism, stress response, signal transduction, structural protein, transcription, transport facilitation, growth and development and hypothetical function [14,15].

Results
Total RNA is the starting material to perform all downstream analysis, so it should be in good quality and quantity. A modified method was applied for the extraction of high-quality total RNA from juniper shoots. Total RNA extraction showed 200.95µg/ml yield per sample, with absorbance ratio 2.13 at optical density 260/230 and 1.94 at optical density 260/280. The total RNA integrity was assessed by two distinct 28S and 18S rRNA bands (Figure 1).

The functional annotation showed that out of 1500 randomly selected transcripts 1205 transcripts (80.3%) were observed with significant similarities to the GenBank non-redundant database and placed in significant similarities category, 289 transcripts (19.3%) showed low similarities and assigned to the “unclassified” category. The remaining 6 (0.4%) transcripts with no similarity are defined as no hits (Figure 2).

![Figure 1: Total RNA extracted from Juniper tree (Juniperus excelsa) and run on 1% agarose gel, stained with ethidium bromide, showing two intact rRNA bands.](image1)

![Figure 2: The pie chart showing transcripts categorization according to homology.](image2)

Profile of the putative functions presented that 18% (217) transcripts were engaged in metabolism, 11.7% (141) are playing role as transcription factors and 11.5% (139) were structural protein. 8.8% (106) transcripts were involved in stress responses, whereas 6.7% (81) were engaged in growth and development. Transcripts involved in signal transduction and transport represented 5.6% (67) and 3.6% (43) respectively. The 34.1% (411) transcripts were found with hypothetical protein nature (Figure 3).

![Figure 3: The pie chart showing transcripts categorization in different groups according to their putative function.](image3)
Majority (18%) of the newly profiled transcripts were involved in the metabolism, such as ribulose bisphosphate carboxylase, cinnamoyl CoA reductase (CCR2, CCR7), starch synthase, hydroquinone glucosyltransferase, carbonic anhydrase, glyceraldehyde transferase, zeta-carotene desaturase, fructose-bisphosphate aldolase. The 11.7% analyzed transcripts were engaged in transcription regulation including DEAD-box helicase 54, GTP-binding elongation factor, elongation factor 1-delta-like, MYB transcription factors (3R-MYB), zinc finger transcription factor and F-box/FBD/LRR-repeat protein. Transcripts engaged in transport activity were nitrate transporter, ATP-binding cassette (ABC) transporter, iron-inhibited ABC transporter, aromatic amino acid transporter and adenine nucleotide transporter.

The newly analyzed transcripts were also observed with roles in the process of cell signaling pathways like Rho small GTPase-activating proteins, receptor-like kinase, leucine-rich repeat receptor protein kinase (LRRK1) and cyclin-dependent kinase. Transcripts related with stress responses were heat shock proteins (HSP17, HSP70, HSP83, HSP88), galactinol synthase 1, endoplasm in homolog, Tobacco Mosaic Virus (TMV) resistance protein N-like.

**Discussion**

In the current research for functional characterization of Ziarat juniper transcripts, 80% transcripts were placed in significant similarities category. Similar homology categories were assigned to various plant by many researchers [16,17]. The rest 20% transcripts with no significant homologues in the NCBI-GenBank, representing the novel transcripts from Ziarat juniper that may have significant role in the maintenance of plant and human beneficial secondary metabolites.

Functional annotation of transcripts showed that maximum number of transcripts were involved in metabolism. Similar results were observed by Fei et al. [16] in leaves of mangrove plant. Among transcripts involved in metabolism, Fructose-bisphosphate aldolase (FBA) a key metabolic enzyme in glycolytic pathway of plants is also reported by Lu et al. [18] in Arabidopsis genome by using BLAST algorithms. Cinnamoyl CoA reductase (CCR) that catalyze the first specific step in lignin biosynthesis has been reported in the model plant Arabidopsis, Oryza and Populus [19-21]. CCR proteins have also been identified from and several economic plants such as Triticum aestivum and Zea mays [22,23]. Starch synthase an important enzymes for Starch biosynthesis in higher plants was also reported in the rice genome by Dian et al. [24] and Jiang et al. [25]. These identified transcripts with potential roles in metabolism would be helpful in understanding of metabolic pathways of juniper.

Numerous transcripts were found to be involved in transcription regulation like Zinc finger proteins and MYB transcription factors. The Zinc finger proteins (ZFPs) as significant transcription factors are widely reported in other plants [26,27]. Similarly, the MYB transcription factor is important for the regulation of several developmental and physiological processes in plants [28]. These functionally diverse transcription factors have been identified in Arabidopsis [29] and soybean [30]. Soler et al. [31] characterize the MYB family in Eucalyptus grandis genome, while Li et al. [32] identified and characterized MYB Transcription Factors in Pyrus bretschneideri. These transcripts represent a good information that could be utilized to improve the juniper forest based on transcription factors.

ATP-binding cassette (ABC) proteins that facilitate MgATP-energized transmembrane transport and regulate other transporters were reported by Rea [33] in Arabidopsis thaliana and Oryza sativa. Aromatic amino acids (AAAs) that function as precursors of several natural products and involved in synthesis of proteins, are also identified by Chen et al. [34] in Arabidopsis. For the growth and development of plant histone acetylation and deacetylation are very important [35]. Like our findings, involvement of transcripts with histone deacetylase activity is also reported in other plants by many researchers [36,37]. Findings regarding Glycine-rich cell wall structural proteins related transcripts were similar with the prior findings in bean and rice [38,39]. These findings would be of interest in future studies to determine the roles of juniper transcripts in growth and development, transport and as structural proteins.

Transcripts related to signal transduction such as Rho small GTPase-activating protein, have emerged as signal integrator and coordinators of a wide range of signaling pathways [40]. GTPases were discovered from a number of plant species such as barley and peach [41,42]. Genome sequencing studies showed that plant genomes encode a great number of receptor-like kinase (RLKs). For example, about 610 RLK coding sequences were
reported in \textit{Arabidopsis} genome \cite{43}. Leucine-rich repeat receptor protein kinase (LRR-RKs) that controls a wide variety of developmental and defense-related processes was also reported in rice, poplar and soybean \cite{44-46}. These findings may contribute in understanding of the mechanisms behind cell signaling pathways.

Heat shock proteins (HSPs) or chaperones play a crucial role in protecting plants against stress and assist in protein refolding. In our finding, many heat shock proteins are observed. According to Santhanagopalan et al. \cite{47} HSPs function, not only during stress, but also during specific developmental stages in plants. In plants, many HSP proteins have been identified in different species \cite{48,49}. Similarly, galactinol synthase catalyzes formation of galactinol associated with the responses to environmental stresses have been reported from \textit{Medicago falcata} \cite{50}. These findings regarding transcripts involved in cell signaling and stress responses would be significant candidates to understand the defense mechanism of Ziarat Juniper.

The profiling presented in this study provides the first overview of transcripts from the shoots of Ziarat juniper. This study analyzed 1500 transcripts with potential roles in metabolism, transcription, signal transduction, stress response, structural protein, growth and development and transport pathways of juniper plant. It is anticipated that the information obtained from this study such as; HSPs, ZFPs and LRR-RKs would be significant resources for future genomic studies and helpful to understand various mechanism under biotic and abiotic stresses in this plant.

\textbf{Acknowledgment}

This paper is a part of the research project (HEC-NRPU Project 20-1867/R&D/11) financed by the higher education commission (HEC) of Pakistan, Islamabad. The authors are highly thankful and acknowledge this financial support of the higher education commission (HEC) of Pakistan, Islamabad.

\textbf{References}


This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License. To read the copy of this license please visit: https://creativecommons.org/licenses/by-nc/4.0/