Electronic Journal of Biotechnology 35 (2018) 48–56

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research article

A whole-transcriptome approach to evaluate reference genes for quantitative diurnal gene expression studies under natural field conditions in *Tamarix ramosissima* leaves

Xia Yan [a,b], Chaoju Qian [a], Xiaoyue Yin [a], Xingke Fan [a], Xueyong Zhao [a,c], Menghe Gu [d], Tao Wang [d], Xiao-Fei Ma [a,*]

[a] Key Laboratory of Stress Physiology and Ecology in Cold and Arid Regions of Gansu Province, Cold and Arid Regions Environmental and Engineering Research, Chinese Academy of Sciences, Lanzhou 730000, China
[b] Key Laboratory of Inland River Ecohydrology, Cold and Arid Regions Environmental and Engineering Research, Chinese Academy of Sciences, Lanzhou 730000, China
[c] Naiman Desertification Research Station, Cold and Arid Regions Environmental and Engineering Research, Chinese Academy of Sciences, Lanzhou, 730000, China
[d] Key Laboratory of Desert and Desertification, Cold and Arid Regions Environmental and Engineering Research, Chinese Academy of Sciences, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 13 September 2017
Accepted 7 August 2018
Available online 24 August 2018

Keywords:
Desert plants
Desert restoration
Drought adaptation
Fragile ecosystem
Functional analysis
Gene expression
qPCR
Reference gene
RNA-seq
Tamarix ramosissima
Water transport

ABSTRACT

Background: Tamarix ramosissima is a desert forest tree species that is widely distributed in the drought-stricken areas to sustain the fragile ecosystem. Owing to its wide usage in the desert restoration of Asia, it can be used as an ecophysiological model plant. To obtain reliable and accurate results, a set of reference genes should be screened before gene expression. However, up to date, systematical evaluation of reference genes has not been conducted in *T. ramosissima*.

Results: In this study, we used eigenvalues derived from principal component analysis to identify stable expressed genes from 72,035 unigenes from diurnal transcriptomes under natural field conditions. With combined criteria of read counts above 900 and CV of FPKM below 0.3, a total of 7385 unigenes could be qualified as candidate reference genes in *T. ramosissima*. By using three statistical algorithm packages, *geNorm*, *NormFinder*, and *BestKeeper*, the stabilities of these novel reference genes were further compared with a panel of traditional reference genes. The expression patterns of three aquaporins (AQPs) suggested that at least UBQ (high expression), EIF4A2 (low expression), and GAPDH (moderate expression) could be qualified as ideal reference genes in both RT-PCR and RNA-seq analysis of *T. ramosissima*.

Conclusions: This work will not only facilitate future studies on gene expression and functional analysis of genetic resources of desert plants but also improve our understanding of the molecular regulation of water transport in this plant, which could provide a new clue to further investigate the drought adaptation mechanism of desert plant species under harsh environments.

How to cite: Yan X, Qian C, Yin X, et al. A whole-transcriptome approach to evaluating reference genes for quantitative diurnal gene expression studies under natural field in *Tamarix ramosissima* leaves. Electron J Biotechnol 2018;35. https://doi.org/10.1016/j.ejbt.2018.08.004.

1. Introduction

Gene expression analysis is increasingly important in many fields of biological research. Understanding gene expression patterns is expected to provide insights into complex regulatory networks and helps to identify genes that are relevant to new biological processes. Plants fluctuate diurnally for approximately 24 h under diurnal signals such as light–dark or temperature cycles with diverse biological activities and physiological output [1,2]. Most transcripts were diurnally expressed following those signals in plants. In the dicot model *Arabidopsis thaliana*, this number amounts to approximately 80% [3,4,5,6]. In rice, poplar, maize, tomato, and soybean, similar results were observed [7,8,9,10]. In addition to rhythmic changes, complex and noise environmental signals also affect gene expression. Many studies have dissected these signals under artificial constant conditions to further investigate the contribution of environmental factors to the molecular mechanism in plants [2,11,12]. However, many results showed discrepancy with those obtained under natural field conditions, which is even more serious in transgenic crop plants [13,14,15,16]. Natural conditions are increasingly taken into

* Corresponding author.
E-mail address: maxiaofei@lzb.ac.cn (X.-F. Ma).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

https://doi.org/10.1016/j.ejbt.2018.08.004

consideration to meet the required arguments in modern agriculture and ecology [17,18].

qPCR is the best method for quantifying gene expression owing to its simplicity, accuracy, and low cost. To obtain reliable and accurate expression levels, one or a set of reliable reference genes with low variation in expression across diverse sample types is a prerequisite [19,20,21,22,23,24]. Usually, a handful of selected reference genes such as actin (*ACT*), elongation factor 1 alpha (*EF-1α*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and ubiquitin (*UBQ*) are commonly used without any validation. Unfortunately, these traditional reference genes can exhibit surprisingly high expression variance in some species or under different environmental conditions [19,25,26,27,28,29]. Consequently, bias of expression data occurred, or incorrect conclusions were drawn [30]. Many statistical methods for analyzing expression variability have been developed, including *geNorm*, *BestKeeper*, and *NormFinde*r [23,31,32], but these programs can only analyze the expression data from a handful of selected genes from a small *priori* set of genes. With the development of sequence technology, reference genes from a genome-wide background selected by high-throughput technologies have prevailed. For example, microarray has been used in *A. thaliana*, *Eucalyptus*, and soybean [26,33,34]; RNA-seq has been used in other plant species [35,36,37,38]. The genes selected by high-throughput technologies are generally better than commonly used reference genes after validation with experiments [39]. At present, with advantages such as fast output, inexpensive, and minimal variation across technical replicates, RNA-seq is more attractive than microarray and is more widely used to select reference genes [21,40,41,42,43].

A large number of studies on systematic validation of reference genes have been reported in classic plants such as *Arabidopsis* [26], rice [44], poplar [45], soybean [46], wheat [47], barley [48], tomato [49], grape [50], and potato [51]. Some stability evaluation of orthologous genes of these plants was also reported in nonmodel plant species [52]. These studies were limited to specific contexts and generally unsuitable for any context, including any considerable alteration in different biological samples [19,29]. Thus, selection of a series of new reference genes on nonmodel plants was widely developed, for example, in bamboo [53], peach [54], *Caragana intermedia* [55], eggplant [56], *Mimulus* [20], andromonoecious *Taihangia rupestris* [22], watermelon [40], lettuce [12], and *Reaumuria songarica* [57].

Tamarix ramosissima is a typical halophytic plant. It can sequester high concentrations of sodium and other salts in their above-ground tissue and can secrete these concentrated salts through specialized salt glands to the surface of leaves or shoots [58,59]. It is also a desert forest tree species widely distributed in the drought-stricken areas, with annual precipitation less than 200 mm in North China [60]. With the pursuit to understand the desertification of Asia, the mechanism by which *T. ramosissima* adapts to the harsh environment since millions of years has attracted increasing ecological and physiological interests in recent years. Some studies have investigated the relationship between gene expression and water balance in other species of *Tamarix* [61,62,63,64,65]. The results may lead to bias regarding the stability of the reference genes. In addition, studies have revealed that water from leaves has been regarded as an important subsidiary to mitigate the deleterious effects of soil water deficits [66,67,68]; aquaporins (AQPs) in leaves are responsible for absorbing water from moisture in air at night in *Tamarix* [65]. On the basis of the experiments, a diurnal transcriptome of *T. ramosissima* under natural field conditions has been profiled to reveal the molecular mechanism of foliar water uptake. There is an urgent need to develop a set of candidate reference genes to finely analyze gene expression in response to water availability [69]. In addition, there is no study to date on the systematic evaluation of reference genes in *T. ramosissima*.

In this study, we used eigenvalues derived by using the mathematical method and principal component analysis (PCA) to identify stable expressed genes from 72,035 unigenes from diurnal transcriptomes in *T. ramosissima* under natural field conditions. With combined criteria of read counts above 900 and FPKM (fragments per kilobase per million reads) coefficient of variation (CV) below 0.3, a total of 7385 unigenes were selected as candidate reference genes in *T. ramosissima*. Three statistical algorithm packages, *geNorm* [31], *NormFinder* [23], and *BestKeeper* [32], were used to compare stability between a set of novel reference genes from these 7385 unigenes and a panel of traditional reference. Further validation was quantified on the expression pattern of three AQPs against three representations of the reference genes *UBQ* (high expression), *EIF4A2* (low expression), and *GAPDH* (moderate expression) with RT-PCR and RNA-seq, respectively. This work will provide information for future studies about not only gene expression and functional analysis of the genetic resources in desert plants but also the molecular regulation of foliar water uptake in *T. ramosissima*, which could provide a new clue to us to understand the genetic mechanism of desert plants in their adaptation to drought environments.

2. Material and methods

2.1. Ethics statement

T. ramosissima is a desert tree species widely distributed in Jingtai County, Gansu Province, and other arid regions; it has not been included in any list of endangered or protected species. Before collecting the samples, oral permission was obtained from the local management of forestry after sending introduction letters from the CAREERI (Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences).

2.2. Plant material and cDNA preparation

RNA was extracted from at least 5 g of leaves harvested from one *T. ramosissima* plant from 6:00 AM to 24:00 PM on 25 June in 2013, and it served as samples for both RNA-seq and qPCR. RNA concentration and integrity were assessed using a NanoDrop Lite spectrophotometer (Thermo Fisher Scientific, DE). RNA with high quality isolated from leaves of at least three branches was mixed equally for replicate experiments.

Part of the RNA sample was used to construct the CDNA library by using the Illumina HiSeq platform. Expression levels in FPKM were determined for a total of 72,035 unigenes with a mean length of 831 bp and N50 length of 1494 bp (unpublished data).

Part of the RNA sample was used for RT-PCR analysis. For each sample, 1 μg of total RNA was reverse transcribed in a 20 μl reaction volume with oligo(dT) primers, using the RevertAid™ First Strand cDNA Synthesis Kit (Fermentas/Thermo Fisher Scientific). The cDNAs were diluted as 1:10 with nuclease-free water and stored at −20°C.

2.3. Reference gene and AQP gene selection

As low-expression genes would make poor qPCR reference genes owing to the difficulties in detecting and quantifying their expression, genes with expression levels lower than 5 FPKM in any of the eight transcriptomes were excluded from further stability analyses. Eigenvalues derived by using the mathematical method with PAST software were used to determine the comprehensive value, which indicated the common expression pattern along time series. Genes with slight or changed curve during the time series were selected to meet the requirements of a reference gene. To further support selection with eigenvalues, read counts and CV of FPKM were included. Genes with both a mean of read counts above 500 and a CV below 0.3 were considered as stably expressed [26,52]. According to nonredundant (Nr) annotation, homologs of the traditional reference genes used in previous diurnal studies [12,22,70] were also included.