THE MOLECULAR IDENTIFICATION OF ZANTHOXYLUM ARMATUM DC OF PAKISTAN BASED ON DNA BARCODING

SHAKILA UMER*, NAYAB SAFDAR* AND KHUSHI MUHAMMAD*

Department of Genetics Hazara University, Mansehra, Khyber Pakhtunkhwa-Pakistan 21300

Keywords: ITS; rbcL; DNA barcoding; Zanthoxylum armatum; Molecular phylogeny.

Abstract

Zanthoxylum armatum DC., belonged to the family Ruteaee, is a medicinal plant used to cure many diseases. DNA barcoding was used as a tool for molecular identification of Zanthoxylum armatum DC. species from Balakot Pakistan. In the present study four DNA barcodes including matk, rbcl, ITS and trnH-psbA were used. The sequenced data were analyzed by using BLASTn at NCBI, FASTA and Mega 7.0 software. During PCR analysis, 3 DNA barcodes ITS, rbcl and trnH-psbA were successfully amplified and showed the 100% sequencing success. Furthermore, these barcode markers showed 99-100% sequence similarity with the reference sequences at the BLASTn. The further analysis revealed the sequence similarity of investigating marker with Zanthoxylum armatum (MH016484.1), Zanthoxylum nitidum (FN599471.1) and Zanthoxylum bungeanum (MF097123.1) respectively. The current finding provides the basis for sequenced data of Z. armatum to be used in future for molecular discrimination among the plant species from Pakistan and it is concluded that combination of diverse kind of barcoding markers could be helpful in proper identification of species at lower taxonomic level.

Introduction

Initially plants were used by the peoples for their nourishing requirements. With the passage of time, the natural flora has become an important source across various human communities for health improvement and remedies against several diseases. Many species are used by peoples in many different parts of the world such as Africa, Asia and South America (Mustafa et al., 2017). However, more than 50,000 flowering plants out of 4,22,000 purposes reported from world are used for medicinal purpose (Uniyal et al., 2006). Globally medicinal plants constitute a single larger functional group of plants (Khan et al., 2011). Medicinal plants as a variety of natural bioactive products provides a rich source of structural biodiversity that have played a fundamental role in the discovery of a drug (Hussain et al., 2010). The acceptance and demand of these plants are increasing progressively (Jamshidi-Kia et al., 2018). But recently due to illegal exploitation decreasing populations of medicinal plants in the wild have led to discussion among ecologists, scientists and conservationists (Negi et al., 2010).

Zanthoxylum armatum (DC) is the important member of Rutaceae and it is known as Dambrary, Dambara (Pashtu) and Tamur (Urdu) in Pakistan (Alam et al., 2017; Ibrar et al., 2017). It is a small xerophytic shrub or tree, with leaflet blades usually having thorns (Barkatullah et al., 2013). The plant can be recognized by its shrubby habit, prickled trunk and branches, dense foliage, with pungent aromatic taste and small subglobose, red fruit (Paul et al., 2018). In Southeast Asia it is a common plant (Alam et al., 2017). It is reported in Pakistan from Rawalpindi, Hazara, Malakand, Murree hills, Dir, Swat and Buner and grows at an altitude that starts from about 800m up to 1500m in shady or semi shady habitat (Barkatullah et al., 2014).

*Corresponding author, E-mail: khushisbs@yahoo.com/ *Both authors are contributed equally.
It is an aromatic medicinal plant and the parts of this plants like fruit, bark, stem, leaves, roots and seeds possess medicinal properties and used in preparation of indigenous medicines against various diseases like rheumatism, varicose veins, bronchitis, dyspepsia, diarrhea, toothache, asthma, indigestion and cholera (Singh et al., 2015).

For the plant interaction with outside environment as well as for the regulation of development, growth and reproduction of plants, some of the plant derived bioactive molecules act as signalling molecules (Peng et al., 2012; Dhami et al., 2018). Plants that contain bioactive compounds could be an alternative source to control insect agents. Many of them have no or little damaging effect on the non-target organisms and environment (Zhang et al., 2018). However, Zanthoxylum armatum contains the phytochemicals like lignans, saponins, coumarins, alkaloids, flavonoids, sterols and phenolic compounds (Brijwal et al., 2013; Mirza et al., 2019).

Based on DNA identification, the current systems has the potential to facilitate both the discovery of new ones and proper identification/authentication of known species (Braukmann et al., 2017)

DNA barcoding is a novel technology which uses a short agreed upon fragment of DNA to accurately identify species. This method is widely used in plant study to assist biodiversity, differentiation and the discovery of new species. The consortium barcode of life CBOL proposed chloroplast genes such as rbcL, matK as universal barcode while the plastid intergenic spacer trnH-psbA along with nuclear ribosomal DNA internal transcribed region ITS suggested as supplementary barcode regions for DNA barcoding of plants (Zhao et al., 2020; Whitehurst et al., 2020).

Two species of Zanthoxylum have same morphological features and it is difficult to identify these species traditionally. Sometimes, the cultivar ‘Qinghuajiao’ habitually mix with Z. schinifolium Siebold & Zucc., but it is Z. armatum (Feng et al., 2020). Mostly in agriculture, improper recognition of Zanthoxylum armatum often result in economic losses because they have similar characters and name. To solve this problem, it is urgently needed to apply DNA barcoding for the identification of Zanthoxylum armatum. Many studies have been done on evolving the suitable markers to differentiate 97 diverse species of Zanthoxylum across various countries. Several molecular markers were used including, amplified fragment length polymorphism (AFLP) for distinguishing different species of Zanthoxylum (Gupta and Mandi, 2013), amplified polymorphism (SRAP) markers and ISSR markers (Feng et al., 2015), internal transcribed spacer ITS (Kim et al., 2019), along with these markers the chloroplast genome markers also used for the correct identification of Zanthoxylum. (Kumar et al., 2020).

Recently, various efforts have been made to validate the occurrence of several plastid and ribosomal markers and authenticated DNA barcodes have been reported from plant species of Pakistan (Khan et al. 2019a,b). Therefore, the current study is the continuity of our previous efforts and here we are investigating that how combined markers (ribosomal and plastid) could be effective in the molecular systematics and phylogenetic reconstruction of medicinally important plant species from Pakistan. The aim of this study to screen suitable DNA barcode region for the molecular identification of Z. armatum.

Materials and Methods

Collection, identification and preservation of plant materials

Plant sample Z. armatum was collected from Balakot Khyber Pakhtunkhwa, Pakistan found at an altitude of 974 m (3196 ft). Along with altitude of 34.2051° and longitude of 7.35213° from Balakot Khyber Pakhtunkhwa, Pakistan. The plant specimen was dried preserved and identified with the help of plant taxonomist dried plant specimen pasted on the herbarium sheets. Specimen
with voucher no 6255 dated 12.8.17 was submitted to the Herbarium of Hazara University Mansehra. Remaining dried plant specimen was used for the molecular study.

**Extraction, Purification and Quantification of DNA**

Total genomic DNA was extracted from dried plant tissue by using CTAB method with some modification (Verma and Biswas, 2020). A fine powder of weighed sample was made by grinding with the help of mortar and pestle. For each 100 mg of grinded tissue 800 µl of CTAB Extraction Buffer was used that was pre-warmed at 65 degrees. The mixture was vortex thoroughly. The homogenate was then incubated for 2 hours at 65°C. After the incubation period 600 µl of PCI was added and homogenate was centrifuge for about 20 minutes at 13,000 rpm. The supernatant along with 500 µl of ice cold iso-propanol in a new tube was mix by gentle inversion and leave on ice for almost 30 minutes. The samples were centrifuge for another 20 minutes at 13,000 rpm and then iso-propanol was removed to leave pellet or brown viscous layer in bottom of Eppendorf tube.

500 µl of 70% alcohol was added and the samples were centrifuge for 10 min at 13,000 rpm. The tubes were inverted for drying to remove alcohol completely. Then 60-80 µl of ddH₂O were added to each sample. The concentration and quality of extracted DNA was checked on 1% agarose gel.

**Details of DNA barcoding markers used in this study**

In this study four candidate barcoding markers namely *rbcL*, matK, *trnH-psbA* and ITS were evaluated for the investigation of *Zanthoxylum armatum* primer detail shown in table 2. The DNA barcoding markers were amplified by standard Polymerase chain reaction.

**PCR Amplification and Sequencing**

The PCR reaction was carried out using universal barcoding primers and standard protocols. The volume of 25 µl of PCR reaction mixture was prepared in 200 µl PCR tube and eachtube contained approximately 1 µl of DNA template, 3 µl of 10× PCR buffer, 3 µl *MgCl₂*, 3 µl of each *dTTP*, *dCTP*, *dATP*, *dGTP*, 0.5 units of Taq Polymerase kit (Catalog no.K0171) and 2 µl of each forward primer and reverse primer. The amplification was performed in an Applied Biosystems 2720 Thermal Cycler. The initial step for 10 min at 94°C was followed by 35 cycles for 2 min at 94°C, 58°C for 1:30 min and 72°C for about 2 min and 1 cycle at 72°C of 10 min. The amplified products were electrophoresed on 1.5% TAE agarose gel and then the PCR products were sent for sequencing to the sequencing centre of the National History Museum, London, United Kingdom and sequencing was done in both the direction with the PCR primers.

**Sequencing and sequence analysis**

For the analysis of *Z. armatum*, we selected one single species while other species were downloaded from NCBI GenBank, for the clustering of the *Zanthoxylum armatum* with its most close species as well as diverse species.

After the sequencing the quality of the sequences were checked in the Geneious software the messy sequence from the start and last were deleted to make the quality high. Moreover, the sequencher was used for editing of both directions to make the consensus sequences for the further analysis. The consensus sequences were used for the confirmation of the correct and similar identification on the Basic Local Alignment Search Tool (BLAST). To align the sequences Maff aligner was used. BioEdit removed the roughly arranged data and MEGA 7.0 software was used for sequence analysis.

**Results and Discussion**

In this study, three chloroplast region and one nuclear DNA regions were selected as the candidate DNA barcodes. The three regions namely, *ITS, trnH-psbA* and *rbcL* were amplified easily while *matK* was unable to amplify. The PCR success rate of the regions *trnH-psbA, ITS,* and
rbcl was (100%) and the sequencing success rate of these three regions was also (100%). The aligned length of ITS region was 679 bp having the GC content of (63%). The length of the plastid region rbcl was 1188 bp having the GC content of (44.8%). The aligned sequence length of trnH-psbA was 506 bp with the lowest GC content of (31.4%) (Table 1).

Table 1. GC content, amplification success rate and sequencing success rate of DNA barcode regions in Zanthoxylum armatum.

<table>
<thead>
<tr>
<th>Plant name</th>
<th>DNA barcoding regions</th>
<th>GC content</th>
<th>Amplification success rate %</th>
<th>Sequencing success rate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zanthoxylum armatum</td>
<td>ITS</td>
<td>63%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>rbcl</td>
<td>44.8%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>trnH-psbA</td>
<td>31.4%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Sequence analysis

Phylogenetic analysis was performed to identify the relationship between individual, species and genus. To confirm the monophyly of the species, therefore we performed 3 methods i.e., Maximum likelihood (ML), Maximum parsimony (MP) and Neighbour joining (NJ) methods using the data of amplified regions and the phylogenetic tree was constructed with bootstrap replicate 1000 (Sheng et al., 2020). A total of 10 sequences of ITS and trnH-psbA and 9 sequences of rbcl were selected including one from the present study and remaining collected from the GenBank NCBI for phylogenetic reconstruction. During (ML), (NJ) and (MP) analysis of ITS region, the studied species Z. armatum DC. Showed close similarity with Zanthoxylum armatum (MH016484.1) with bootstrap value of 100 and tree length of 1460.2434, 0.18824824 and 112 respectively (Fig 1. a, b, c). The branches less than 50% are collapsed. After trimming the ITS sequence of Z. armatum had a length of 746 bp out of which 617 were conserved sites,105 were variable sites, 38 were parsimony informative sites and 66 were singleton sites.

The Phylogenetic tree constructed by using Maximum Likelihood Method with highest likelihood log -1808.8398 of rbcl sequences composed of sequence length of 1403bp having 1342 conserved sites, 34 variable sites, 7 parsimony sites and 27 singleton sites. Further phylogenetic analysis based on kimura 2 parameter revealed that the studied species Z. armatum occurred in the same clade with Zanthoxylum nitidum (FN599471.1) with 62 bootstrap support as poor resolution power and these species are not highly related to each other in many genetic characters (Fig. 2a). On the other hand, in Neighbour Joining tree, Z. armatum shared same clade with Zanthoxylum nitidum (FN599471.1) with 63 bootstrap value and having the tree length of 0.01854648 (Fig. 2b). And the phylogenetic tree based on Maximum Parsimony method having tree length of 22 found in same clade with Zanthoxylum nitidum (FN599471.1) with the lowest bootstrap value of 29 which showed a diversity between these two species (Fig. 2c). Three clades were observed in MP tree and first clade had 2 species along with Z. armatum while 2 species in each 2 and 3 clades were grouped. However, the remaining one species lied in outer group (Fig. 2c).

Based on Maximum Likelihood, Neighbour Joining and Maximum Parsimony tree analysis the length of trees of trnH-psbA were -965.2520,0.11746074 and 54 respectively. Maximum Likelihood treedisplayed 3 clades with 4 species in one and three species each in other two clades that showed close resemblance of the present studied species with Zanthoxylum bungeanum (MF097123.1) with bootstrap value of 63 (Fig. 3a). After trimming of the present studied species sequence it was observed that it had a sequence of 519bp with 464 conserved sites, 46 variable sites, 20 parsimony sites and 25 singleton sites. Neighbour Joining tree having the length of
0.11719453 with 3 clades in the tree. *Z. armatum DC.* Show close resemblance with *Zanthoxylum bungeanum* (MF097123.1) with bootstrap value of 66 (Fig. 3b). While in Maximum Parsimony tree analysis it showed close resemblance with *Zanthoxylum bungeanum* (MF097123.1) with bootstrap value of 53 (Fig. 3c).

Fig. 1a-c. ITS based trees of Maximum Likelihood (ML), Neighbour Joining (NJ) and Maximum Parsimony (MP) Methods for Phylogenetic analysis.
Fig. 2a-c. *RbcI* based trees of Maximum Likelihood Method (ML), Neighbour Joining Method (NJ) and Maximum Parsimony Method (MP) for Phylogenetic analysis.
Fig. 3a-c. trnh-psbA based trees of Maximum Likelihood (ML), Neighbour Joining (NJ) and Maximum Parsimony (MP) Methods for Phylogenetic analysis.

The present work is the first report on the effectiveness of four candidate barcode region *Z. armatum* endemic to the Himalayan region of Pakistan.
The accurate and authenticated way of identifying the species of genus *Zanthoxylum* is important to ensure the use of this medicinal plant for drug discovery and for other traditional uses. DNA barcoding is an advanced technique by using short, standardized gene regions designed to provide automatic, accurate and rapid species identifications (Noh et al., 2020). An ideal DNA barcode should have high universality power and taxonomic coverage as universality is one of the most important benchmarks for an appropriate DNA barcode, i.e., high PCR and sequencing success (Srivastava and Manjurath, 2020) in addition the effectiveness of DNA barcode in the accurate identification of species depend on the monophyletic character of the same group species (Le et al., 2020). Based on these properties four barcode markers (*ITS, rbcL, matk and trnH-psbA*) were tested on the *Zanthoxylum armatum* from the present study.

In previous study, the comparison of *matk* with *ITS, rbcL* and *trnH-psbA* indicates that it is not a suitable DNA barcode in identifying *Zanthoxylum armatum* (Zhao et al., 2018). For all land plant *rbcL* and *matK* were thought to be core barcodes (Feng et al., 2020). But some previous study showed *matk* as a problematic because of its low amplification and sequencing success rate (Gao et al., 2019; Gostel et al., 2020). In the current study we find the same problems with *Zanthoxylum armatum*, where it showed poor amplification and sequencing success rate. Thus, based on these above mentioned results it is suggested that *matk* region have no potential of DNA barcoding in *Zanthoxylum armatum*.

*ITS* has all the characteristics of an ideal DNA barcode so that’s why it could be used as a single barcode our result match with previous results in which authors did not report any difficulty in amplification and sequencing in gymnosperm, the *ITS* is consider an ideal DNA barcode as it had no amplification, sequencing, alignment and editing problems (Chen et al., 2020). In the current study *ITS* almost showed a good amplification and sequencing success. Based on BLAST search the (*ITS*) showed a species identification of 99.12% with *Zanthoxylum armatum* (MH016484.1) which is more like previous literature showed clearly split two morphological same species without any contact with other species. Some of the previous literatures suggest *ITS* an ideal DNA barcode region based on species identification (Pang et al., 2010; Demirel et al., 2016; Dhivyaa et al., 2020; Kurian et al., 2020). Some studies showed the drawbacks of *ITS* region related to different species. The *ITS* failed in sequencing with *Calligonum* species due to which it is not considered a suitable DNA barcode region (Nguyen, 2020).

Compared with *ITS* and *trnH-psbA* the DNA barcode region *rbcL* determine a slightly better identification rate of 99.83% with *Zanthoxylum wutaiense* (FN599472.1) but the resolution power was so low which is more lower than other tested barcode regions due to which *rbcL* was not included in the suitable barcode list. Hence, more studies needed to find the correct identification. In most terrestrial plants, the *rbcL* marker offers great universality in terms of constant PCR amplification, high-quality bidirectional sequencing and reliable alignment of nucleotide sequences (Carneiro de Melo Moura et al., 2019).

According to the criteria of high divergence of sequences and universal application among species the *trnH-psbA* spacer is considered the most favourite single locus for land plant barcode (Gogoi et al., 2020). Due to excellent reliability for authentication in Rutaceae family *trnH-psbA* along with *ITS*2 can be used as a complementary barcode for a wide range of plant taxa (Timpano et al., 2020). Here in the present study, *trnH-psbA* is supposed to be a good DNA barcode region as it showed a species identification of 99.55% with *Zanthoxylum bungeanum* (MF097123.1) which is slightly lower than *rbcL* but it is acceptable barcode region because of high resolution power. It is thought to be a useful DNA barcode region across a wide range of angiosperms (Amar, 2020).
In the current study, the four DNA barcode regions were applied on the plant sample collected from Balakot Pakistan. The analysis was done for phylogenetic reconstruction of the species genotypes. The sample shows the good species identification with two DNA barcode regions (ITS, and trnH-PSbA) rbcL showed the lowest resolution power in all the three applied tree method, while matk needs more further research as it failed to amplify in the current study. The data of the DNA sequences present in this study could be helpful in the future for further investigation. Current study concluded that the trnH-PSbA and ITS are the most suitable barcode regions for the DNA barcoding and phylogenetic study of the Zanthoxylum armatum DC.

Acknowledgement
The authors are thankful to Higher Education Commission, Islamabad, Pakistan for funding (Grant no. 5711).

References


THE MOLECULAR IDENTIFICATION OF ZANTHOXYLUM ARMATUM


(Manuscript received on 21 June, 2020; revised on 19 November, 2020)