

Research Article

Phylogenetic Analysis by Implementation of Dna Barcoding Genes with Biogeographical Distribution of *Brassica Juncea*

Manisha Ray, Bijayini Panda, and Souvagyalaxmi Sahoo*

Tectona Biotech Resource Center, India

*Corresponding author

Souvagyalaxmi Sahoo, Tectona Biotech Resource Center, Bhubaneswar-751002, Odisha, India

Submitted: 21 December 2018

Accepted: 21 January 2019

Published: 23 January 2019

ISSN: 2576-1102

Copyright

© 2019 Sahoo et al.

OPEN ACCESS

Keywords

- *Brassica juncea*
- DNA Barcodes
- Phylogeny
- Biogeographical study

Abstract

Brassica juncea is one of the economically important plants, well known because of its medicinal and nutritive values. DNA barcoding technology has been used successfully in past years for species identification and phylogenetic construction. Thus, the present work was planned to study the divergence amongst 15 species of *Brassica juncea* through construction of phylogenetic tree using computational approach. Here, DNA barcoding genes such as ITS, rbcL and trnH-psbA of *Brassica juncea* were selected for tree construction using bioinformatics tools and techniques. Further, biogeographical study of *Brassica juncea* were performed for fourteen countries and continents using RASP, which resulted the world wide distribution as well as advocated for the origin of few subspecies of *Brassica juncea* from china and Asia. Thus, this work would throw lights on further genetic evaluation study of *Brassica juncea*.

ABBREVIATIONS

NCBI: National Centre for Biotechnology Information; MEGA7: Molecular Evolutionary Genetics Analysis Version 7; DNASP: DNA Sequence Polymorphism; PopART: Population Analysis with Reticulate Trees; PAUP: Phylogenetic Analysis Using PAUP; RASP: Reconstruct Ancestral State in Phylogenies; ITS: Internal Transcribed Spacer

INTRODUCTION

The evolution of molecular biology gave rise to a modern-day method primarily based on nucleotide series diversities amongst species known as DNA barcoding. Short standardized phase of the genome serving as a sample “barcode” has been proposed as a technique that gives to expedite accurate species identification and DNA barcodes, are increasingly more recognized for their capability to validate medicinal plant life [1]. It has been developing rapidly in recent years, and become a useful tool for biodiversity investigation and monitoring, and molecular phylogeny and evolution [2].

Brassica juncea (*B. juncea*) is an erect, often unbranched annual to biennial plant belongs to the family Brassicaceae having diverse medicinal uses. It is the build of several distinct forms which are grown for meals, oil and so forth and has numerous common names, along with Indian mustard, leaf mustard, brown mustard, Chinese mustard, and oriental mustard [3]. It has been pronounced that mustard leaves are beneficial in curing various dangerous diseases like diabetes, most cancers; in particular colon, lung, gastric and breast cancer and so forth [4,5]. Also the

leaves of *B. Juncea* are implemented to provide drug remedies which act as stimulants, diuretics and expectorants in addition to mustard leaf extract have own awesome organic activities at the side of antioxidant, anti inflammatory, renal ischemia-defensive and anti depressant and so forth.

The plant species and herbal pills are recognized via the usage of a selection of techniques, including analyses of morphologic traits, chemical compositions, and molecular genetic variations [6], and also the DNA barcodes are increasingly recognized for their ability to authenticate medicinal plants [1]. Thus this study includes three DNA barcodes regions i.e. internal transcribed spacer (ITS), ribulose biphosphate carboxylase (rbcL) and photosystem II protein D1 precursor (trnH-psbA) present in the nucleotide sequences of *B. juncea*, to molecular authenticate the 15 different sub species or species complexes (*B. juncea*, *B. juncea* subsp. *integrifolia*, *B. juncea* var. *crispifolia*, *B. juncea* var. *rugosa*, *B. juncea* var. *strumata*, *B. juncea* var. *integrefolia*, *B. juncea* var. *napiformis*, *B. juncea* var. *crassicaulis*, *B. juncea* var. *gemmaifera*, *B. juncea* var. *gracilis*, *B. juncea* var. *juncea*, *B. juncea* var. *multiceps*, *B. juncea* var. *multisecta*, *B. juncea* var. *tumida*, *B. juncea* var. *utilis*) of *B. juncea* by analyzing the genetic variation or haplotype diversity and evolutionary variation or phylogenetic study through computational approach.

The reciting of spatial patterns of biodiversity is a key detail to apprehend the ancient diversification of biota, and gain a better perceptive of ecological factors that imprint spatial patterns of biodiversity [7]. As a way to obtain biogeographical statistics at nearby and local scales which can be vital to recognize their

evolutionary and biogeographical histories, the biogeographical analysis for all the studied 15 species complexes of *B. juncea* was finished through the use of Bioinformatics tools and strategies to found its worldwide distribution, which is probably subjected to distinguish the plant species and its medicinal effects primarily based on its locality or neighborhood sicknesses.

MATERIALS AND METHOD

Retrieval of sequence

The nucleotide sequences of *B. juncea* were collected from nucleotide database of National Centre for Biotechnology Information (NCBI) web server (<https://www.ncbi.nlm.nih.gov/nucleotide/>). Further, entire resulted nucleotide sequences were filtered on the basis of important DNA barcoding markers such as Internal Transcribed Spacer I and II (ITS I and II), trnH-psbA and rbcL of *B. juncea* (Supplementary Table) Sequence alignment

Multiple sequence alignment: The collected nucleotide sequences of *B. juncea* were undergone for multiple sequence alignment using MUSCLE algorithm of Molecular Evolutionary Genetics Analysis (MEGA7) tool (<https://www.megasoftware.net/>).

Genetic diversity

All the aligned nucleotide sequences of the three marker based genes ITS, psbA and rbcL of *B. juncea* were subjected to Genetic

diversity analysis, which was done through DNA Sequence Polymorphism Software (DnaSP 6) (<http://www.ub.edu/dnasp/>), for the analysis of the uniqueness of each haplotype in the given sequence dataset. The resulted unique haplotypes were represented by a network using population genetics software i.e. Population Analysis with Reticulate Trees (PopART 1.7) (<http://popart.otago.ac.nz/index.shtml>).

Phylogenetic variance study

The Phylogenetic variance study or analysis was done among the nucleotide sequences of selected marker genes of *B. juncea* through Phylogenetic Analysis using Parsimony 4.0 (PAUP4.0) software (<https://paup.phylosolutions.com/>) to reconstructing and analyzing phylogenetic relationships or variations between molecular sequences and this generated tree was visualized through FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>).

Biogeographical analysis

The Biogeographical variation study was done through Reconstruct Ancestral State in Phylogenies (RASP 4.0β) software (<http://mnh.scu.edu.cn/soft/blog/RASP>) to analyze and reconstruct the historical biogeography of subspecies on the basis of worldwide distribution of *B. juncea*.

The above methodology was represented through a flow diagram (Figure 1).

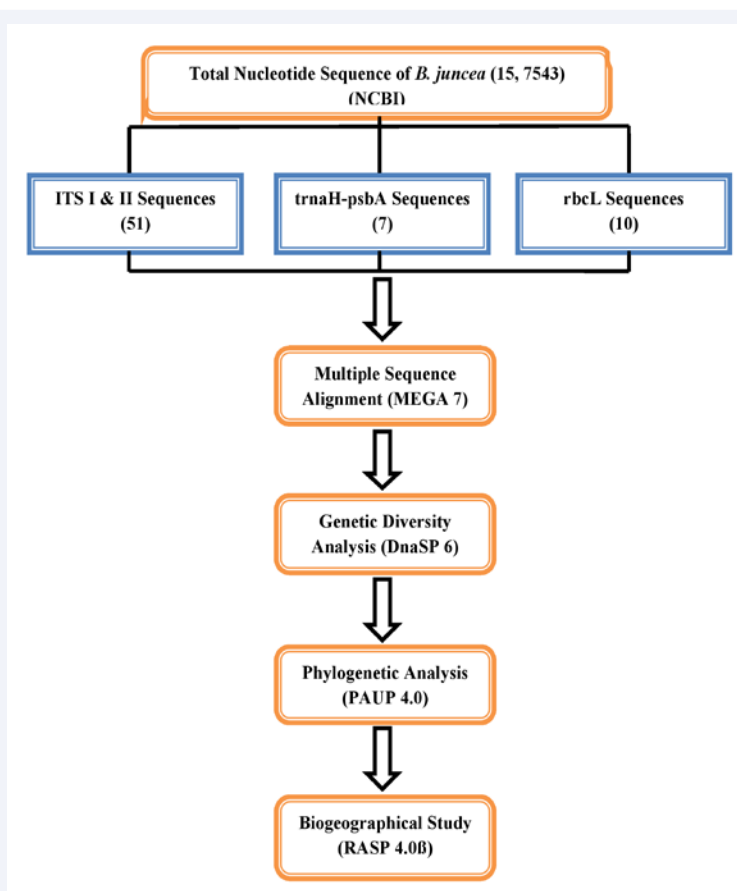


Figure 1 Schematic Diagrammatic Representation of Methodology.

RESULTS

Sequence retrieval and alignment

There have been overall 15, 7543 numbers of nucleotide sequences of *B. juncea* were present inside the nucleotide database of NCBI. As, presence of species-specific substitutions and DNA barcoding regions i.e. genes such as ITS, trnH-psbA and rbcL are important to authenticate any individual species therefore, only 68 nucleotide sequences were selected from the entire data set. Thus, In general 51 numbers of ITS I and II, 7 numbers of trnH-psbA and 10 numbers of rbcL markers related nucleotide sequences had been retrieved from nucleotide database of NCBI, and subjected to perform multiple sequence alignment using MUSCLE algorithm and UPGMA clustering method in MEGA 7, which is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, and testing evolutionary hypotheses [8].

Genetic diversity analysis

Genetic diversity in the population becomes envisioned by computing haplotype diversity and nucleotide variety. Here, the haplotype range as well as the nucleotide variety became analyzed through DnaSP (Table 1), which is designed for performing exhaustive population genetic analyses on multiple sequence alignments. It is likewise useful for comprehensive evaluation of DNA collection data version, the usage of a friendly Graphic User Interface [9]. So the genetic variability of each marker was described by the number of haplotype of the given sample, haplotype diversity, number of nucleotide sites, nucleotide diversity (Pi) value i.e. the number of nucleotide differences per site between the sequences, the DNA polymorphism data like GC content in the complete genomic region, number of polymorphic or segregating sites, total number of mutation, Tajima' D value which is based on the differences between the number of segregation sites and the average number of nucleotide differences etc for all those three marker genes i.e. ITS I and II, trnH-psbA and rbcL based sequences using DnaSP. There were total 16, 4 and 2 numbers of haplotypes were generated for ITS I and II, trnH-psbA and rbcL respectively, which was graphically represented by minimum spanning network view in PopART software (Figure 2) as it is used for inferring and visualizing haplotype networks among populations [10].

Table 1: Generated haplotype data through DNAsp.			
DNA Barcode Regions	ITS I & II	psbA	rbcL
No. of nucleotide sequence	51	7	10
No. of nucleotide sites	5832	328	631
No. of haplotypes(h)	16	4	2
Haplotype diversity(Hd)	0.7090	0.7143	0.2000
No. of polymorphic and segregating sites	20	80	1
GC content ,G-C	0.506	0.506	0.438
Tajima's D	-0.36021	-1.76477	-1.11173
Total no. of mutation	22	82	1
Nucleotide diversity	0.04619	0.13487	0.00038

Phylogenetic variance study and biogeographical analysis

Phylogenetic analyses are significant to many studies areas in biology and usually involve the identification of homologous sequences, their more than one alignment, the phylogenetic reconstruction and the graphical illustration of the inferred tree. So right here the phylogenetic evaluation among the aligned ITS I and II, trnH-psbA and rbcL contained nucleotide sequences of *B. juncea* turned into finished through PAUP software [11] and predicts the phylogenetic tree by using 1000 as bootstrap replicates, with using the Maximum Parsimony (MJ) tree format, as PAUP can used for best maximum parsimony method by taking *Arabidopsis thaliana* as an out group to show the variation between the plant species of family Brassicaceae and the resulted phylo tree with branch labels, tip labels and node labels was visualized through Fig Tree, as it is a graphical viewer of phylogenetic tree and present a schematic diagram based on used marker genes ITS I & II (Figure 3A), rbcL (Figure 3B), psbA (Figure 3C) respectively [12].

Although for figuring out and delimiting biogeographical areas, where the proof of the existence of the species complexes of *B. juncea* are present, had been accumulated which have been analyzed and the historical distribution location became reconstructed via the use of RASP software [13] containing S-Diva software analysis because S-Diva helps to statistically evaluate the alternative ancestral ranges at each node in a tree [14]. Thus to carry out this analysis the huge evolutionary distributions of *B. juncea* had been analyzed and ends with fourteen regions of taxonomic varieties, based totally on the presence of 11 species complexes of *B. juncea*'s out of 15 species complexes due to the absence of marker genes in rest of the varieties in addition to unavailability of biogeographical statistics data as shown in (Table 2) and represented in different colors based on the locations (Figure 4). These areas are: A (China {East Asia}), B (Serbia {Southeast Europe}), C (India {Southern Asia}), D (Central Asia), E (Europe), F (Japan {East Asia}), J (Eastern China), H (Middle China), I (Northern China), G (Western China), K (Mogolia {East Asia}), L (Northern Africa), M (Canada {North America}), and N (Southwest Asia) [15]. Thus this analysis concluded that most of the sub species of *B. juncea* originated or distributed in Asia and China.

DISCUSSION

The identification of vital medicinal plant species is a difficult undertaking and it desires expert taxonomists, thus as *B. juncea* have 15 numbers of species complexes, so this study follow the DNA barcoding principle by choosing ITS, trnH-psbA and rbcL regions contained nucleotide sequence, due to the fact that the reason of DNA barcoding is nucleotide series primarily based identification of multiple plant species with accuracy and is one of the extensively widespread technology [16]. Additionally sequence comparison of the ITS vicinity is widely utilized in taxonomy and molecular phylogeny as it is easy to increase even from small quantities of DNA and has a excessive degree of version even among closely related species [17], trnH-psbA possess attributes that are highly desirable in a plant DNA barcoding system, such as universality, sequence quality and

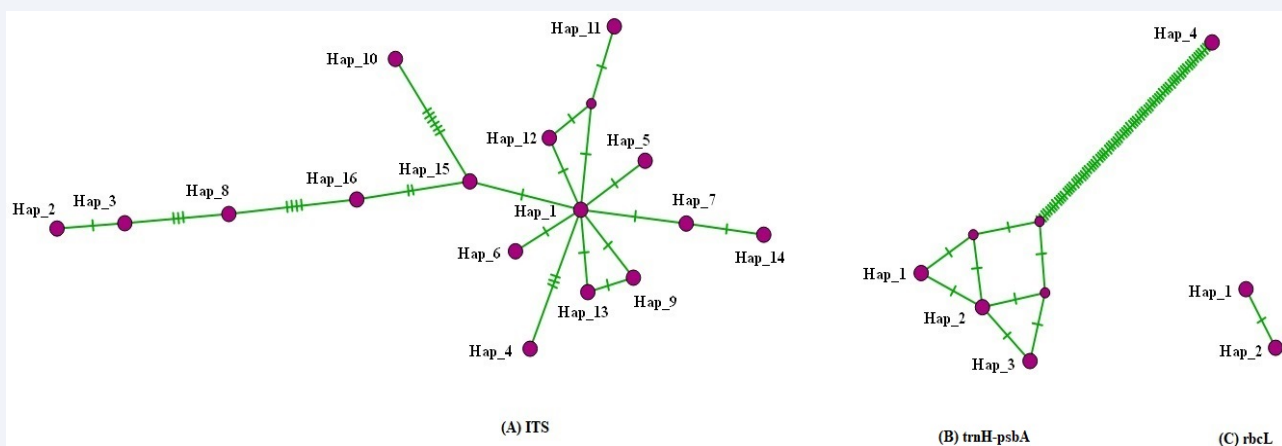


Figure 2 Generated Haplotype Networks of (A) ITS, (B), trnA-psbA, (C) rbcL.

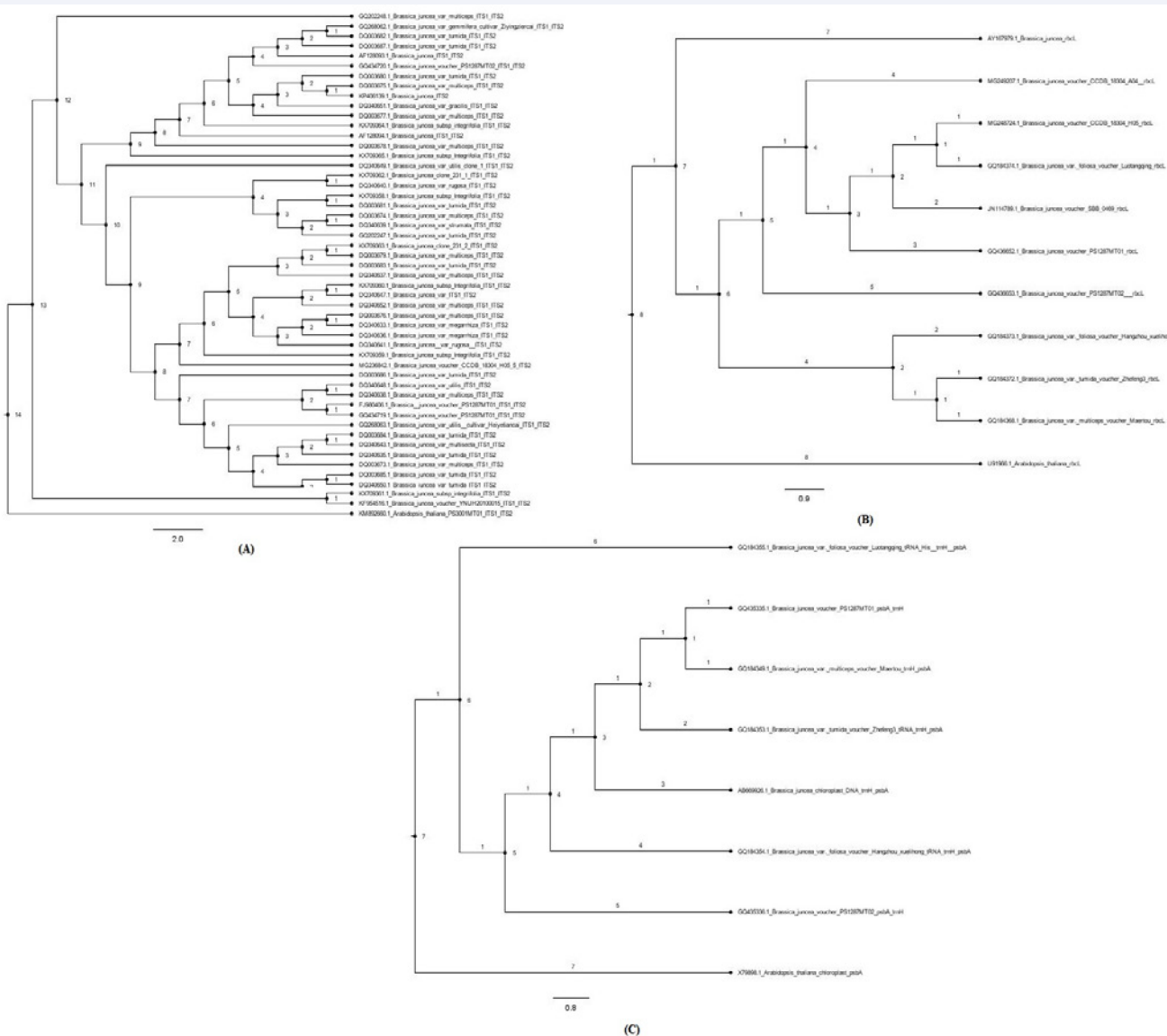
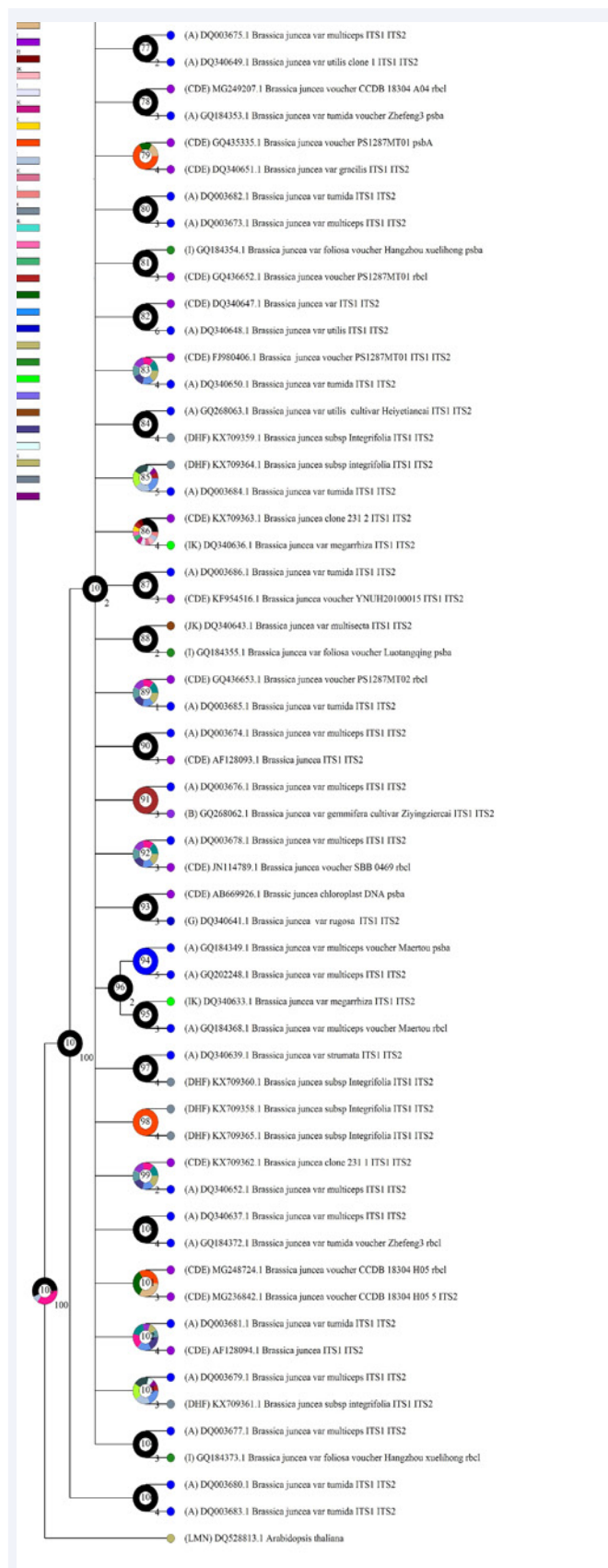
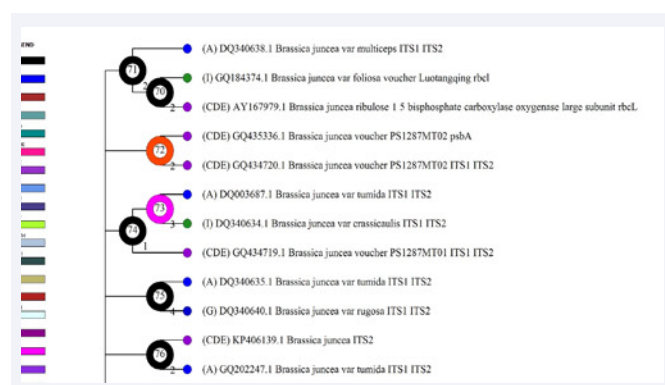


Figure 3 Phylogenetic Tree construction for *B.juncea* species complexes using ITS I&II (A), rbcL (B) and trnH-psbA (C) gene markers.

Table 2: The Biogeographical distribution with the indicated alphabets for each area of all studied species of *B. juncea*.

Sl. No.	Name of the Species	Geographical Location	Indicated Alphabets
1	<i>B. juncea</i>	India, Central Asia, Europe	C, D, E
2	<i>B. juncea</i> var. <i>napiformis</i>	Mongolia	K
3	<i>B. juncea</i> var. <i>foliosa</i>	Middle and Northern China	H, I
4	<i>B. juncea</i> var. <i>integrifolia</i>	Middle China, Japan	H, F
5	<i>B. juncea</i> var. <i>rugosa</i>	Eastern China	J
6	<i>B. juncea</i> var. <i>gemmifera</i>	Serbia	B
7	<i>B. juncea</i> var. <i>tumida</i>	China	A
8	<i>B. juncea</i> var. <i>utilis</i>	China	A
9	<i>B. juncea</i> var. <i>multiceps</i>	China	A
10	<i>B. juncea</i> var. <i>multisecta</i>	Western China, Mongolia	G, K
11	<i>B. juncea</i> var. <i>strumata</i>	China	A
12	<i>Arabidopsis thaliana</i>	Northern Africa, Canada, Southwest Asia	L, M, N

coverage, and discrimination ability [18] and *rbcl* is one of the first plant genes to be sequenced and is still among the most frequently sequenced segments of plant DNA, also located on cp genome as a single copy gene and has an enormous phylogenetic utility and has been used widely in systematic studies of land plants [19], for this reason those barcode areas authenticate, explore and solved the various evolutionary, phylogeny and taxonomic issues of selected species and its associated species complexes, so that it may play crucial role in controlling and treating of many illnesses. Though the genetic range in the populace was estimated via computing haplotype range and nucleotide range among the aligned nucleotide sequences of *B. juncea* in which haplotype range or gene diversity, represents the probability that two randomly sampled alleles are different and in other hand nucleotide diversity is defined as the average number of nucleotide differences in pair wise comparison among DNA sequences and resulted more gene diversification with 16 haplotypes of ITS I and II region based sequences. Then the phylogenetic study had a look at considered sequences, achieved through PAUP software discovered that species from the identical

**Figure 4** Biogeographical distribution analysis of selected species complexes in *B. juncea*.

circle of relatives or species complexes of a species are clustered collectively when nucleotide sequences are assessed with the aid of phylogenetic analysis and also a few diverged sequences were also located. The phylogenetic examine is crucial due to the fact the proper identification and assessment of biologically essential plant families play a key role to understand evolutionary records of many crucial plant species. Even each species spans a certain geographical range, the ecological and evolutionary distinctiveness, that restraint the geographical variety or may be essential for growing new species [16], for that reason to explain the spatial styles of biodiversity and provides better information of ecological factors, the biogeographical assay of *B. juncea* was completed among fourteen areas i.e. the place of foundation, rate of distribution of species complexes of *B. juncea* throughout the world discovered that the distribution is most in all likelihood found in Asia and China as China has a protracted history and a wealthy way of life of herbal drug treatments [20], this offers a course to preserve the subjected plant species and use its medicinal values to design novel putative pills to remedy many lethal illnesses by means of watching its evolutionary as well as ecological statistical data.

CONCLUSION

This whole study concluded that the omnipresent plant *B. juncea*, will be helpful within the coping distinct sicknesses. So, because of its authentication for medicinal value as well as to analyze its taxonomic species variations and international distributions the haplotype, phylogenetic and biogeographical study turned into concluded that from the 3 loci, namely, *rbcl*, *ITS*, and *trnH-psbA*, *ITS* indicates the more variations between the studied sub species of *B. juncea* and the biogeography of it miles maximum efficaciously disbursed in Asia and China respectively, so this statistics is probably helpful in taxonomic and floristic contexts of plant as well as deliver some thrilling values for ecosystems in geographic space and via geological time.

REFERENCES

- Cabelin VLD, Alejandro GJD. Efficiency of *matK*, *rbcl*, *trnH-psbA*, and *trnL-F* (cpDNA) to Molecularly Authenticate Philippine Ethnomedicinal Apocynaceae Through DNA Barcoding. *Pharmacog Maga*. 2016; 12: 384-388.
- Kang Y, Deng Z, Zang R, Long W. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Sci Rep*. 2017; 7.
- Kumar V, Thakur A, Barothia N, Chatterjee S. Therapeutic potentials of *Brassica juncea*: an overview. *Tang humanitas medicine*. 2011; 1: 2.1-2.16.
- Kwak Y, Lee J, Ju J. Anti-cancer activities of *Brassica juncea* leaves in vitro. *EXCLI J*. 2016; 15: 699-710.
- Sharma A, Kumar V, Kanwar MK, Thukral AK, Bhardwaj R. Phytochemical profiling of the leaves of *Brassica juncea*. L. using GC-MS. *International Food Research Journal*. 2017; 24: 547-551.
- Kim W, Ji Y, Choi G, Kang Y, Yang S, Moon B. Molecular identification and phylogenetic analysis of important medicinal plant species in genus *Paeonia* based on *rDNA-ITS*, *matK*, and *rbcl* DNA barcode sequences. *Gen Mol Res*. 2016; 15: 15038472.
- Lenormand M, Papuga G, Argagnon O, Barros GD, Alleaume S, Luque S. Biogeographical network analysis of plant species distribution in the Mediterranean region. *Ecology and Evolution*. 2018.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33: 1870-1874.
- Rozas J. DNA Sequence Polymorphism Analysis Using DnaSP, Methods in Molecular Biology Bioinformatics for DNA Sequence Analysis. *Methods Mol Biol*. 2009; 537: 337-350.
- Leigh JW, Bryant D. Popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*. 2015; 6: 1110-1116.
- Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, et al. Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Taxa. *PLoS ONE*. 2015; 10: 0118199.
- Czech V, Huerta-Cepas J, Stamatakis A. A Critical Review on the Use of Support Values in Tree Viewers and Bioinformatics Toolkits. *Molr Biol Evol*. 2017; 34: 1535-1542.
- Badik KJ, Jahner JP, Wilson JS. A biogeographic perspective on the evolution of fire syndromes in pine trees (*Pinus*: Pinaceae). *Roy Soc Ope Sci*. 2018; 5: 172412.
- Yu Y, Harris AJ, He X. S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*. 2010; 56: 848-850.
- Dixon GR. Origins and Diversity of Brassica and its Relatives. 2007.
- Sinwari ZK, Jan SA, Challis AT, Khan A, Ali M, et al. Identification and phylogenetic analysis of selected medicinal plant species from pakistan: dna barcoding approach. *Pak J Bot*. 2018; 50: 553-560.
- Das S, Deb B. DNA barcoding of fungi using Ribosomal ITS Marker for genetic diversity analysis: A Review. *Int J Pur App Biosci*. 2015; 3: 160-167.
- Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, Roy S. The Internal Transcribed Spacer (ITS) Region and *trnH-psbA* are Suitable Candidate Loci for DNA Barcoding of Tropical Tree Species of India. *PLoS ONE*. 2013; 8: 57934.
- Patwardhan A, Roy S, Ray A. Molecular Markers in Phylogenetic Studies-A Review. *J Phylo Evol Biol*. 2015; 2: 2.
- Ma B, Hu T, Li P, Yuan Q, Lin Z, Tu Y, et al. Phylogeographic and phylogenetic analysis for *Tripterygium* species delimitation. *Ecol Evol*. 2017; 7: 8612-8623.

Cite this article

Ray M, Panda B, Sahoo S (2019) Phylogenetic Analysis by Implementation of Dna Barcoding Genes with Biogeographical Distribution of *Brassica Juncea*. *J Bioinform, Genomics, Proteomics* 4(1): 1037.