

## Nuclear and Plastid DNA Sequence-based Molecular Phylogeography of *Salvadora oleoides* (Salvadoraceae) in Punjab, India Reveals Allopatric Speciation in Anthropogenic Islands Due to Agricultural Expansion

Felix Bast\* and Navreet Kaur

Centre for Plant Sciences, Central University of Punjab, Bathinda, Punjab, 151001, India

\*Corresponding author: Bast F, Centre for Plant Sciences, Central University of Punjab, Bathinda, Punjab, 151001, India, Tel: +91 98721 52694; Fax: 713-500-5495; E-mail: felix.bast@gmail.com

Receiving date: May 25, 2017; Acceptance date: June 15, 2017; Publication date: June 26, 2017

Copyright: © 2017 Bast F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

*Salvadora oleoides* is a tropical tree species belonging to the little-known family Salvadoraceae and distributed in the arid regions of Africa and Asia. Aims of our study were to trace the microevolutionary legacy of this tree species with the help of sequence-based multi-local phylogeography and to find the comparative placement of family Salvadoraceae within angiosperm clade malvids. A total 20 geographical isolates were collected from different regions of North India, covering a major part of its species range within the Indian Subcontinent. Sequence data from nuclear-encoded Internal Transcribed Spacer region (ITS1-5.8S-ITS2) and plastid-encoded *trnL-F* spacer region, were generated for this species for the first time in the world. ITS-based Bayesian phylogeographic analysis revealed the existence of four clades while *trnL-F* spacer based Bayesian analysis revealed one clade for this species distributed in the Indian subcontinent. Between these two loci, ITS revealed more distinct phylogeographic clades, indicating the phylogeographic utility of this locus for the systematics of Salvadoraceae. Interestingly, the ITS phylogeny indicated the existence of allopatric speciation in this tree species. Factors such as habitat destruction through agricultural expansion of plains might have forced the remaining population of this threatened tree species to isolated patches of thorn forest biotas, the 'anthropogenic islands', analogous to an archipelago. Phylogenetic analyzes based on *trnL-F* spacer suggested a synonymy of this species with *Salvadora angustifolia*. Maximum Likelihood gene tree based on ITS sequence data revealed that Salvadoraceae belongs to Sapindales rather than Brassicales. However, in the gene tree based on *trnL-F* spacer sequence, this family clustered within Brassicales. An evolutionary congruence of *S. oleoides* isolates across its range in North India is revealed in this study. Given the conflicting results on the relative placement of Salvadoraceae in Brassicales and Sapindales, the need for further phylogenetic analyses of malvids using supermatrix approach is highlighted.

**Keywords:** Brassicales; Conservation genetics; Molecular systematics; Phylogenetics; Rosids; Sapindales

### Introduction

*Salvadora oleoides* Decne is a tropical tree-shrub species belong to the little-known family Salvadoraceae Lindl. This tree species has opposite individual undivided leaves, imbricated corolla lobes, tetramerous flowers with four stamens, and basally-fused connate petals. Although the natural range of this species extends from sub-Saharan Africa up to the island of Java, Indonesia, this species is under the stress of habitat destruction due to widespread deforestation and over-exploitation. Natural habitats of this tree, tropical thorn forests, are endangered by a variety of human activities, especially expansion of agricultural fields. While neither the official IUCN Red List, nor the recent estimate of threatened tree species worldwide [1] ascertain the conservation status of *S. oleoides*, there are some indications that this species is threatened [2]. In India, the tree is found in the isolated thorn forests of the arid regions of North India, especially in and around the Thar Desert.

As per the latest APG III system, the family Salvadoraceae is recognized in cabbage order Brassicales [3]. This family consists of poorly understood genera *Salvadora*, *Azima*, and *Dobera*. Similar to the other members of Brassicales, members of Salvadoraceae produces

a pungent phytochemical, mustard oil glucosinolate. Production of this chemical is thought to be a synapomorphic character of the entire crown group of Brassicales [4]. Since time immemorial plants of genus *Salvadora* had been used in traditional Ayurvedic medicine for a wide variety of purported health benefits; including hypoglycemic and hypolipidemic [5], antimicrobial [6] and as an antibacterial toothbrush. A recent analysis of the feces of a critically endangered Lemur species endemic to Madagascar, golden-crowned Sifaka (*Propithecus tattersale*), revealed that a major portion of its diet was encompassed of the leaves of *Salvadora angustifolia* [7], exposing a small facet of its hidden ecological importance.

Despite its ecological, conservational and commercial importance, systematics and taxonomy of this tropical tree species have neither been analyzed using molecular systematics, nor its evolutionary legacy been traced using sequence-based phylogenetic approach. For example, the phylogenetic position of *S. oleoides* within higher taxonomic hierarchies, such as Brassicales, malvids or rosids, had never been assessed to date. Prior to this study, no nucleotide sequence data had been available for this species in the GenBank database. Till date, there are only two reports available that analyzed relative position of genus *Salvadora* within malvids. The first study was based on chloroplast encoded *rbcL* gene and nuclear-encoded 18S sequences, and the study revealed phylogenetic affinity of this genus with Brassicales/Capparales [4]. The second study was based on *rbcL* gene

and that too corroborated phylogenetic affinity of genus *Salvadora* within Brassicales [8]. Genetic heterogeneity of *S. oleoides* from India had previously been assessed by RAPD method and concluded a very high genetic heterogeneity, with percent polymorphism at 90.09% [9]. Isozyme electrophoresis-based analysis also revealed a high genetic variation (HT=0.249) for this species distributed in North India [10].

Primary objective of this study was to assess the number, phylogenetic relationships and geographic distributions of evolutionary lineages of *S. oleoides* in North India- a first such phylogeographic assessment of a member of family Salvadoraceae throughout the world. We were especially motivated by the pithy dictum famously declared by Croizat [11], that the earth and life are co-evolving. Given that the natural range of this species in Indian Subcontinent lies along the North-Western fringe of Indian tectonic plate, we were interested whether biogeographic processes akin to the Gondwanaland and Indian plate contributed in the distribution of this species. For example, species-range expansion through geodispersion phenomena during immediate aftermath of the collision of Indian plate to the Eurasian plate in Eocene epoch of Cenozoic around 50 million years ago [12], or vicariance events happened subsequently after the formation of the Himalayas, might have shaped the current patterns of the distribution of genetic diversity and population differentiation within species. Brassicales are believed to have originated either in Turkey or Irano-Turanian region and diversified 85 million years ago, prior to the collision of the Indian plate to the Eurasian Plate [13]. Given that conservation status of *S. oleoides* is much contested amidst recommendations to classify it as a threatened species [14], a high-resolution assessment of genetic variation using DNA barcoding will be of paramount importance in its conservation genetics scenario as well. This is because extent and distribution of intraspecific variation is a reliable proxy for determining evolutionary potential and long-term survival of tree species [15]. Our second objective had been to resolve the phylogenetic placement of this

species and family Salvadoraceae within malvids. In many schemes of phylogenetic systematics Salvadoraceae had been considered as part of orders Brassicales/Capparales [3], Sapindales [16], Salvadorales [17], Oleales [18], Celastrales [19] and even as an “outsider”, *incerta sedis* [20]. As nucleo-ribosomal Internal Transcribed Spacers 1 and 2 are the most widely covered DNA barcode for angiosperms at the NCBI GenBank [21] and as sequence information at this locus is not yet available for Salvadoraceae, we were interested to generate ITS sequence data and test various phylogenetic hypotheses on the placement of this family within malvids. In addition to ITS, we also employed plastid-encoded *trnL-F* spacer region, as no DNA sequence data at this locus is available for *S. oleoides* at GenBank. *trnL-F* spacer region is also routinely employed for phylogenetic and phylogeographic studies of angiosperms.

## Materials and Methods

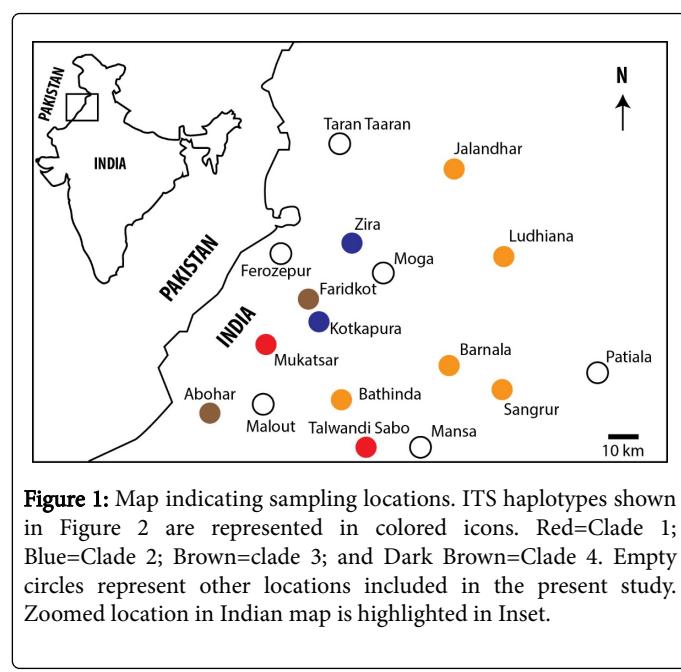
### Taxon sampling

Young leaves were collected from 20 geographic isolates of *Salvadora oleoides* covering the greater part of its range in the Indian subcontinent (Table 1 and Figure 1). Leaf samples from three individual trees from each location were separately analyzed, including PCR and sequencing at two loci, for intra-population heterogeneity. No special permission was required to collect these plant materials, as none of the sampling locations were part of the areas designated as protected by Government of Punjab. Collected specimens were transported to the laboratory in zip-lock polythene bags. Pressed vouchers were prepared and deposited in the Central National Herbarium, Botanical Survey of India, and Calcutta (*Index Herbariorum* Code: CAL). Samples for molecular analysis were stored at -80°C till further analysis.

Voucher No.	Location	Collected by	Collection date	Latitude (N)	Longitude (E)	ITS1-5.8S-ITS2 sequence generated (GenBank Accession)	trnL-F spacer sequence generated (GenBank Accession)
CUP VOUCHER-SO-2014-1	Faridkot	Sukhwant Singh and Navreet Kaur	27-02-2014	30°38'05.62"N	74°47'28.10"E	KP795947	KP795970
CUP VOUCHER-SO-2014-2	Sangrur	Sukhwant Singh and Navreet Kaur	02-03-2014	30°15'22.61"N	75°44'42.88"E	KP795949	KP795958
CUP VOUCHER-SO-2014-3	Mansa	Sukhwant Singh and Navreet Kaur	07-03-2014	30°06'18.20"N	75°10'39.55"E	-	KP795969
CUP VOUCHER-SO-2014-4	Talwandi Sabo	Sukhwant Singh and Navreet Kaur	07-03-2014	29°59'48.65"N	75°09'32.08"E	KP795946	KP795965
CUP VOUCHER-SO-2014-5	Barnala	Sukhwant Singh and Navreet Kaur	08-03-2014	30°19'35.18"N	75°30'10.70"E	KP795950	KP795968
CUP VOUCHER-SO-2014-6	Patiala	Sukhwant Singh and Navreet Kaur	08-03-2014	30°16'24.70"N	75°58'35.37"E	-	KP795964
CUP VOUCHER-SO-2014-7	Taran Taran	Gurbeer Singh	15-03-2014	31°10'46.35"N	74°54'12.08"E	-	KP795962
CUP VOUCHER-SO-2014-8	Malaut	Jasdeep Singh	18-03-2014	30°11'36.31"N	74°24'33.11"E	-	KP795961
CUP VOUCHER-SO-2014-9	Moga	Sukhwant Singh	02-04-2014	30°43'10.13"N	75°06'02.04"E	-	KP795960
CUP VOUCHER-SO-2014-10	Hoshiarpur	Gurpreet Singh	03-04-2014	31°38'08.66"N	75°50'14.66"E	-	-

CUP VOUCHER-SO-2014-11	Ferozpur	Sukhwant Singh and Navreet Kaur	04-04-2014	30°53'38.79"N	74°56'23.46"E	-	KP795967
CUP VOUCHER-SO-2014-12	Amritsar	Sukhwant Singh	07-04-2014	31°38'09.89"N	74°49'27.79"E	-	-
CUP VOUCHER-SO-2014-13	Zira	Sukhwant Singh and Navreet Kaur	09-04-2014	30°52'43.38"N	74°59'16.70"E	KP795951	KP795963
CUP VOUCHER-SO-2014-14	Kotkapura	Sukhwant Singh and Navreet Kaur	09-05-2014	30°38'20.34"N	74°49'46.35"E	KP795955	-
CUP VOUCHER-SO-2014-15	Bathinda	Navreet Kaur	21-05-2014	30°11'45.09"N	74°56'42.99"E	KP795948	KP795957
CUP VOUCHER-SO-2014-16	Chandigarh	Sukhwant Singh and Navreet Kaur	22-05-2014	30°46'37.19"N	76°45'34.70"E	-	-
CUP VOUCHER-SO-2014-17	Jalandhar	Sukhwant Singh and Sukhdeep Kaur	03-06-2014	31°07'91.78"N	75°33'57.29"E	KP795952	-
CUP VOUCHER-SO-2014-18	Ludhiana	Sukhwant Singh	06-06-2014	30°67'49.13"N	75°45'74.08"E	KP795956	-
CUP VOUCHER-SO-2014-19	Abohar	Harjoban Singh	12-06-2014	30°04'09.80"N	74°19'21.23"E	KP795953	KP795966
CUP VOUCHER-SO-2014-20	Muktsar	Harpreet Singh	13-06-2014	30°40'89.56"N	74°52'20.13"E	KP795954	KP795959

**Table 1:** Details of geographical isolates analyzed in this study. A dash (-) indicate no data available.



**Figure 1:** Map indicating sampling locations. ITS haplotypes shown in Figure 2 are represented in colored icons. Red=Clade 1; Blue=Clade 2; Brown=clade 3; and Dark Brown=Clade 4. Empty circles represent other locations included in the present study. Zoomed location in Indian map is highlighted in Inset.

ATTTGAACCTGGTGACACGAG 3') [24], 2  $\mu$ l of 10  $\mu$ M dNTP mixture containing dATP, dTTP, dCTP and dGTP (Genaxy Scientific, India), 0.2  $\mu$ l of 5 u/500  $\mu$ l Taq DNA Polymerase (Thermo Scientific, India), 4  $\mu$ l of template DNA and 2  $\mu$ l sterile water. PCR amplification were carried out in a programmable thermal cycler (Bio-Rad, India) and reaction profile included an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 52°C for 2 min and 72°C for 2 min and a final extension of 72°C for 10 min. Amplified products were electrophoresed on 1.5% agarose gel for 25 min at 250 V and visualized with ethidium bromide in order to determine approximate length and purity.

#### Purification of PCR product and DNA sequencing

Positive reactions were purified using Exo SAP-IT PCR clean-up kit following manufacturer's instructions (USB Corporation, Cleveland, OH, USA). The purified PCR products were subjected to chain termination reaction with the following composition of reaction mixture-buffer 2  $\mu$ l, sequence-ready reaction mixture 1  $\mu$ l, primer 2  $\mu$ l, PCR product 2  $\mu$ l and autoclaved distilled water 3  $\mu$ l. The PCR conditions for chain termination reaction were as follows -95°C for 5 min, followed by 50 cycles of denaturation at 95°C for 45 s, annealing at 52°C for 30 s and extension at 60°C for 4 min. For cleaning of sequencing reaction, Big-Dye X terminator kit (Applied Biosystems, Foster City, CA, USA) was used, following manufacturer's protocol. Purified PCR products were subjected to bidirectional Sanger sequencing using a dideoxy chain termination protocol with ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit v3.1 (Applied Biosystems, Foster City, CA, USA) and a programmable thermal cycler (Veriti, ABI, USA), as per [25]. DNA sequences were assembled using the computer program CodonCodeAligner (CodonCode Corporation, USA). Out of 20 isolates, ITS1-5.8S-ITS2 locus from only 11 isolates got amplified and its sequence length ranged between 605-658 bp. *TrnL-F* spacer locus from 14 isolates got amplified and its sequence length ranged between 188-478 bp. Out of 75 generated sequences, only 25 were unique, as three individual trees at each of the sampling populations (geographical locations) had

#### DNA extraction and PCR amplification

Total genomic DNA was extracted from the specimens using Hi PurATM Plant Genomic Extraction Kit (Hi Media Laboratories Pvt. Ltd., Mumbai) following manufacturer's protocol. The quality of DNA was checked on 0.8% agarose gel, and the quantity of DNA was determined with a spectrophotometer (NanoDrop2000, Thermo Scientific). Extracted DNA was stored at -20°C. For PCR amplification, 20  $\mu$ l reaction mixture was prepared containing 2  $\mu$ l of 10X reaction buffer with 1.8  $\mu$ l 3 mM magnesium chloride (Applied Biosystems, Foster City, CA, USA), 4  $\mu$ l each of 10  $\mu$ M primers: ITS1 (5' TCCGTAGGTGAACCTGG 3') [22], ITS4 (5' TCCTCCGCTTATTGATATGC 3') [22], *trnL-F* (5' CGAAATCGGTAGACGCTACG 3') [23], *trnL-R* (5'

exactly same sequences for both of the genomic loci analyzed. Unique 25 sequences were deposited in Genbank (Table 1).

## Multiple Alignment and Phylogenetic Analyses

Step-by-step protocol followed for contig assembly, multiple sequence alignment, ML ModelTest, and phylogenetic analyses are available at Nature Protocols [26]. In summary, sequences were assembled using CodonCode Aligner, BLASTN sequence homology search, Multiple Sequence Alignments, consensus sequence generation, and Bayesian Inference (BI) phylogenetic inference (using MrBayes add-in) were conducted in Geneious Pro (Biomatters, NZ), and Model Selection, pairwise distance estimation, and Maximum Likelihood phylogenetic inference were conducted in MEGA. For intraspecific phylogeographic assessments, an accession of *Khaya nyasica* (Sapindales) was used as an outgroup for ITS dataset and an accession of *Borthwickia trifoliata* (Brassicaceae) for *trnL-F* spacer dataset. For phylogenetic assessment of malvids, we used accessions of *Euphorbia* spp. as outgroups, as in Rodman et al. [4]. The rationale for choosing these outgroups was to have taxa sufficiently closely related to the in-group for resolving phylogenetic structures to the maximum extent. For example, as Malvid gene tree based on ITS revealed a phylogenetic affinity of *Salvadora* to Sapindales, we used another member of Sapindales rather than Brassicales as an outgroup for intraspecific phylogeographic assessment of *Salvadora oleoides*. We have concurrently tested a number of different outgroup taxa, including as far as *Ficus* (Moraceae), but the topology of phylogram remained similar indicating the robustness of the method.

## Results

### Intra-population heterogeneity

Neither within-individual polymorphisms (dual-peaks at electropherograms/heterozygosity due to imperfect homogenization of multiple rRNA genes, paralogs, pseudo-genes, etc.) nor within-population polymorphisms were observed at either of the two loci analyzed. Three sequences at each of the sampling populations had no nucleotide differences either at ITS1-5.8S-ITS2, or at *trnL-F* spacer region.

### BLASTN

BLASTN Homology search with NCBI nr nucleotide database (word size=10, Max E-value =1e-1 and gap cost-open extend=5.2) revealed interesting results. ITS sequences, both individually, as well as the consensus sequences generated for 11 of our ingroup, isolates, retrieved *Arfeuillea arborescens* (EU720461) as the closest hit with an E-value of 3.24e-93 and 79% identity. Particulars of top 10 hits are summarized in Table 2, and all 10 of them were from order Sapindales. This came as a surprise, as the current consensus among molecular phylogeneticists working on angiosperm systematics is that the family Salvadoraceae belongs to Brassicales. Top 100 results included orders as diverse as Fagales (*Nothofagus*) and Malpighiales (*Quiina*, *Lacunaria*) however not a single accession belonging to Brassicales. It was not because of the non-existence of ITS sequences of Brassicales in GenBank; as of this writing there are 5990 sequences of ITS locus belong to the members of Brassicales exist in the database. *TrnL-F* spacer sequences, both individually, as well as the consensus sequences generated for 14 of our ingroup isolates, retrieved *Salvadora angustifolia* (KC479309) as the closest hit with an E-value of 0 and

99.4% identity, well within generally accepted conspecific threshold (98.7%) [27,28]. Particulars of top 10 hits were summarized in Table 3, and all 10 of them were from order Brassicales. Comparable results were obtained in discontinuous Mega BLAST as well (results not presented); with all top ten hits of ITS belonging to Sapindales, and that of *trnL-F* spacer belong to Brassicales.

S. No.	Species	GenBank Accession	Percent Identity
1	<i>Arfeuillea arborescens</i>	EU720461	79
2	<i>Acer laurinum</i>	AM113543	80
3	<i>Acer laurinum</i>	AM113542	79
4	<i>Protium serratum</i>	KJ503508	80
5	<i>Protium craniopyrenum</i>	KJ503531	80
6	<i>Protium ravenii</i>	KJ503448	80
7	<i>Atalaya alata</i>	EU720425	79
8	<i>Protium glabrescens</i>	KJ503444	79
9	<i>Protium urophyllum</i>	AY375516	80
10	<i>Acer laurinum</i>	DQ366114	79

**Table 2:** Particulars of top 10 BLASTN hits of consensus sequence of ITS region from *Salvadora oleoides* isolates from Punjab, India. All ten were from order Sapindales.

S. No.	Species	GenBank Accession	Percent Identity
1	<i>Salvadora angustifolia</i>	KC479309	99.4
2	<i>Borthwickia trifoliata</i>	JQ733128	87.7
3	<i>Borthwickia trifoliata</i>	JQ733129	87.7
4	<i>Borthwickia trifoliata</i>	JQ733127	87.7
5	<i>Tovaria pendula</i>	FJ212280	83.8
6	<i>Tovaria pendula</i>	AY122465	83.4
7	<i>Pentadiplandra brazzeana</i>	AY122463	83.0
8	<i>Bretschneidera sinensis</i>	JF448517	83.0
9	<i>Moringa ovalifolia</i>	JX091844	82.7
10	<i>Moringa riva</i>	JX091846	82.7

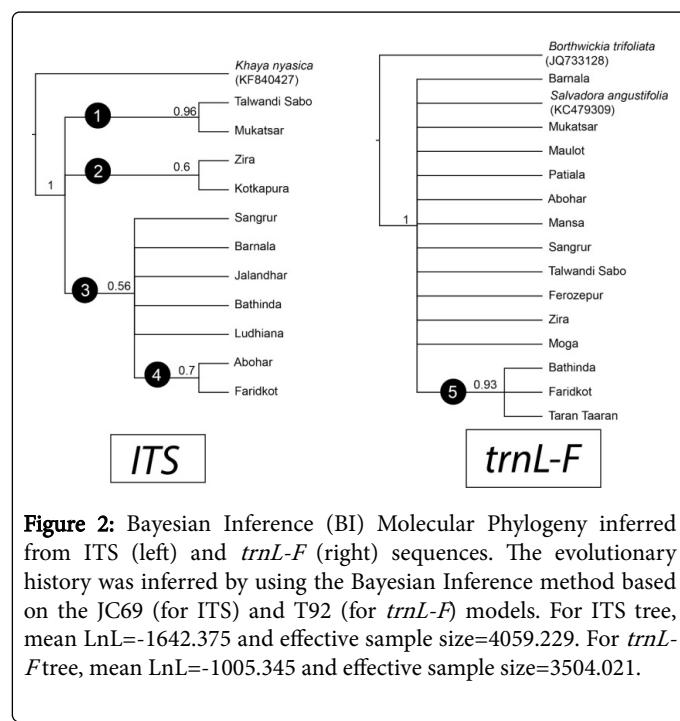
**Table 3:** Particulars of top 10 BLASTN hits of consensus sequence of *trnL-F* spacer region from *Salvadora oleoides* isolates from Punjab, India. All ten were from order Brassicales.

### Phylogeography

Results of phylogeographic assessment using ITS and *trnL-F* spacer regions are presented as Figure 2. In summary, ITS phylogram resolved intraspecific clades better than *trnL-F* spacer phylogram. Intraspecific genetic diversity (within-group mean distance) for ITS dataset (0.020) was found to be higher than that of *trnL-F* spacer dataset (0.003). Corrected (T92 model) pairwise distances of ITS dataset ranged

between 0, between isolates Jalandhar and Ludhiana, and 0.063 between isolates Zira and Talwandi Sabo. Corrected (JC69 model) pairwise distances of *trnL-F* dataset ranged between 0, between isolates Moga and Zira, and 0.008 between isolates Barnala and Zira.

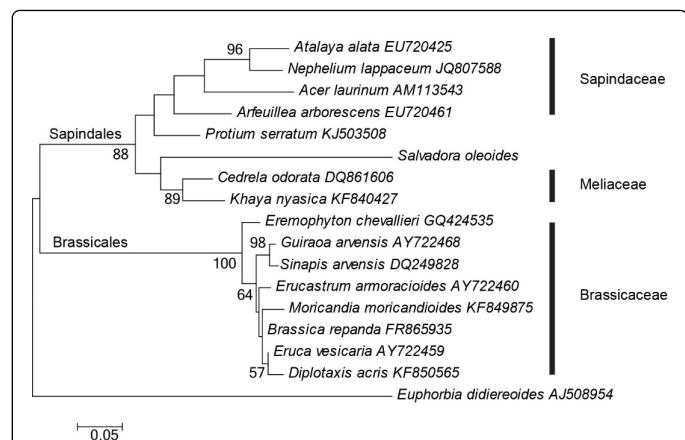
ITS phylogram resolved four clades. In general, geographically closer isolates clustered into distinct clades. The best-supported clade (Clade 1) consisted of isolates from Mukatsar and Talwandi Sabo, followed by a clade consisting of Abohar and Faridkot (Clade 4). The later clade (Clade 4) nested within a larger clade consisting of the majority of sampled isolates (Clade 3). Another, moderately supported, clade consisted of isolates from Kotkapura and Zira (Clade 2). Clades 1, 2 and 3 formed a trichotomy, with each clade being sister to the rest two clades. *trnL-F* phylogram resolved only one clade (Clade 5) that consisted of isolates from Bathinda, Faridkot, and Taran Taaran. All other isolates occupied a basal position in a giant polytomy, perhaps indicating low phylogenetic signal for sufficient resolution. Interestingly, an accession of *S. angustifolia* (KC479309) clustered within this polytomy. The *trnL-F* phylogram broach few possibilities; either a cryptic speciation inside the range of parent population (as *S. angustifolia* clustered within the clade), or synonymy of *S. oleoides* and *S. angustifolia* (as the group consisting of these two taxa was polyphyletic). Considering clades in ITS phylogram as haplotype variants, respective geographical locations of these haplotypes are labeled in Figure 1.



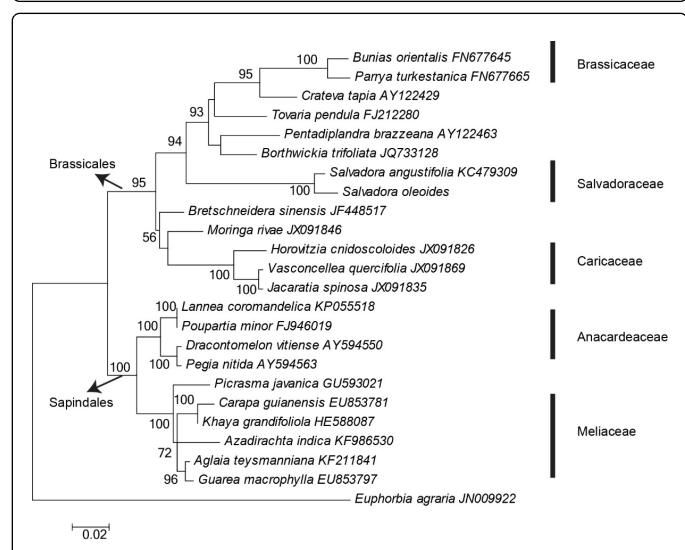
**Figure 2:** Bayesian Inference (BI) Molecular Phylogeny inferred from ITS (left) and *trnL-F* (right) sequences. The evolutionary history was inferred by using the Bayesian Inference method based on the JC69 (for ITS) and T92 (for *trnL-F*) models. For ITS tree, mean LnL=-1642.375 and effective sample size=4059.229. For *trnL-F* tree, mean LnL=-1005.345 and effective sample size=3504.021.

### Phylogenetic position of Salvadorales within Malvids

Results of our ITS sequence-based phylogenetic assessment of the comparative position of family Salvadoraceae within malvids corroborated our earlier observation of the BLASTN affinity of this family with Sapindales (Figure 3). *Salvadora* clustered within the strongly supported monophyletic clade of Sapindales and occupied a position phylogenetically closer to family meliaceae. On the other hand, our *trnL-F* spacer-based phylogram clustered Salvadoraceae within Brassicales with strong bootstrap support (Figure 4).



**Figure 3:** Maximum likelihood (ML) phylogram based on ITS1-5.8S-ITS2 sequences using K2P+G model of molecular evolution in MEGA phylogenetic framework. Numbers near nodes represent ML bootstrap proportion exceeding 50. LnL=-1860.2630. Bootstrap proportions (500 replicates) exceeding 50 were shown near nodes. Scale bar is on the unit of average nucleotide substitutions per site. The analysis involved 17 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 335 positions in the final dataset.



**Figure 4:** Maximum likelihood (ML) phylogram based on *trnL-F* sequences using T92+G model of molecular evolution in MEGA phylogenetic framework. Numbers near nodes represent ML bootstrap proportion exceeding 50. LnL=-2302.8388. Bootstrap proportions (500 replicates) exceeding 50 were shown near nodes. Scale bar is on the unit of average nucleotide substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 445 positions in the final dataset.

## Discussion

This study generated DNA barcode data for *Salvadora oleoides* for the first time in the world. Comparing with *trnL-F*, ITS locus exhibited more within-group distance, as well as resolved finer phylogeographic structures. It can be concluded that ITS is a useful locus for within species phylogeographic assessment in this species. Most probably this is due to differential evolutionary rates for these two loci; ITS- being an intron within the nucleoribosomal region- is more prone to mutations, while *trnL-F* being an intergenic spacer is more evolutionarily conserved. On the other hand, in phylogenetic analysis within malvids using ITS, a longer branch-length observed for the consensus sequence of our isolates suggest that it is a rapidly evolving Operational Taxonomic Unit with elevated substitution rates. Indeed, long stretch of insertions (22 nucleotides long) are observed for our consensus sequence near the end of ITS2 region, between sites 636 and 658, in the ITS alignment. As the method we adopted to infer the phylogeny was Maximum Likelihood-which had been proven to be less prone to Long Branch Artifacts, it is likely that this disproportionately Long Branch length was not due to any artifacts. One likely explanation is allopatric speciation [29]. Within-group mean distance observed for our isolates in ITS dataset (0.02) is indeed the highest ever recorded among angiosperms. Given that the isolates are sampled from area spanning ca 23000 km<sup>2</sup> with no apparent physical barriers between them, barriers in the form of anthropogenic islands due to habitat destruction could very well construe a vicariance mediator and likelihood for such a scenario of allopatric speciation is quite high. A word of caution in this conclusion is deemed necessary. While the current patterns of geographic differentiation may ultimately lead to long-term divergence and speciation, these might also be latest in several cycles of ephemeral patterns of genetic differentiation that this long-lived species have experienced.

Our phylogeographic assessment revealed the existence of five intraspecific clades of *S. oleoides*. Although the correlation between evolutionary distances and genetic distances were rather weakly negative (i.e., geographically close isolated had larger genetic distances between them,  $R^2=0.0211$ ), geographically closely located isolates were part of distinct clades in our phylogenograms. This observation in turn suggests an evolutionary congruence across geographic space. As DNA barcode data for this species is non-existent in public sequence repositories prior to this study, we were unable to perform finer phylogeographic assessments, including its center of origin and dispersal routes. It can be nevertheless concluded that factors such as habitat destruction through agricultural expansion of plains might have forced the remaining population of this tree to isolated patches of thorn forest biotas analogous to an archipelago. Our analyses were insufficient to conclude whether vicariance phenomena or range expansion through geodispersal shaped the current patterns of its distribution. Separation of once contiguous area of species distribution, as observed in this study, does not warrant these endemic spots are related geologically. We can only infer that these biotas were once part of a continuous range and most probable explanation for the apparent vicariance is the anthropogenic agricultural expansion.

An important aspect of our *trnL-F* phylogram is that it clustered *S. angustifolia* within our geographical isolates of *S. oleoides*. In addition, the percent identity between the consensus sequence of *S. oleoides* and *S. angustifolia*, 99.4%, was well within the intraspecific range commonly accepted for angiosperms. It can be inferred that these two species are conspecific. According to the principle of priority [30] the

binomen *S. oleoides* (Decne, 1844) should be preferred over *S. angustifolia* (Turril, 1918).

The most surprising finding is the affinity of salvadoraceae in Sapindales as revealed by our phylogram based on ITS locus. In our knowledge, this is the first study that revealed such an affinity using molecular phylogeny. This finding corroborates an earlier morphology-based systematic scheme that grouped Salvadoraceae under Sapindales. However, as per our *TrnL-F* phylogram, Salvadoraceae is affiliated in Brassicales, as currently considered by the majority of angiosperm systematicists. The reason for this contrasting observation remains elusive; this might be related to the differential manner in which these two loci evolve along plant lineages. ITS, being a nuclear intron, are prone to recombination and follows biparental inheritance. *trnL-F*, on the other hand, being an intergenic spacer in plastid genome, are not prone to recombination and follows the typical matrilineal pattern of inheritance. However, phylogeny based on 18S locus, another nuclear region contiguous to the ITS intron, revealed affiliation of Salvadoraceae in Brassicales [4]. Our conflicting finding warrants need for further scrutiny of angiosperm phylogeny based on multi-local concatenated approach with integrated morphometric dataset, for having the total evidence (*supermatrix* [31]) at our disposal. In earlier molecular systematic analyzes as well as comparative analyzes based on floral anatomy [32] and seed anatomy [33], Salvadoraceae had been shown to be evolutionarily closely related with Bataceae and Koberliniaceae. Although we were really interested in this affinity, as well as relative phylogenetic positions of families Bataceae and Koberliniaceae within malvids, neither ITS nor *trnL-F* sequences from any of its members are currently available at GenBank. A few other observations in our phylogenetic analysis of malvids are deemed noteworthy. Our ITS phylogram revealed non-monophyly of family sapindaceae, with *Acer*, a member of aceraceae, clustered within the clade consisting otherwise of the members of sapindaceae. As per our *trnL-F* phylogram, affiliation of *Picrasma*, currently considered to be a member of family Picramniaceae, within the family Meliaceae was obvious, as meliaceae+*Picrasma* had a strongly supported (BP=100) clade. Similarly, the genus *Crateva*, currently considered to be part of family cratevaceae, showed strong affinity (BP=95) to the family Brassicaceae.

This is the first DNA-sequence based phylogeographic assessment of *Salvadora* in the world. Sequence based phylogeographic assessment of tropical tree species is a relative rarity and in India, such investigations are almost non-existent. It is hopeful that the natural history of *S. oleoides* in Indian Subcontinent revealed in the present study would be beneficial for further, larger-scale phylogeographic assessment of this species for testing various hypotheses on geological, climatological and anthropogenic factors that might have shaped its current distribution. In addition, it is hopeful that our revelation of the affinity of Salvadoraceae to Sapindales would foster further, more detailed, "total evidence" assessment of malvids by molecular systematicists working on this field.

## Acknowledgements

We thank Dr. Kenneth J. Sytsma, University of Wisconsin, Madison for helpful insights. We are thankful to Dr. Pankaj Bhardwaj, Ms. Pooja Rani and Ms. Sheetal Sharma for support with experimental works. We are also thankful for the Vice-chancellor, the Central University of Punjab for his support with respect to the execution of this research. The study was supported by grant-in-aid "Research Seed Money" scheme of the Central University of Punjab.

## Author's Contributions

FB conceived the topic, interpreted the data and drafted the manuscript. NK collected the samples and performed the experiments.

## Funding

The study was funded by a grant from Research Seed Money programme of Central University of Punjab, India

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

1. Oldfield S, Lusty C, MacKinven A (1998) The world list of threatened trees. World Conservation Press, pp: 540-573.
2. Khan M, Shinwari ZK, Musharaf S (2011) Conservation and ecological characteristic of Trees in Tehsil Karak Pakistan. *JBES* 1: 155-164
3. Bremer B, Bremer K, Chase M, Fay M, Reveal J, et al. (2009) an update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161: 105-121.
4. Rodman J, Soltis P, Soltis D, Sytsma K, Karol K (1998) Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *Am J Bot* 85: 997-997.
5. Yadav J, Saini S, Kalia A, Dangi A (2008) Hypoglycemic and hypolipidemic activity of ethanolic extract of *Salvadora oleoides* in normal and alloxan-induced diabetic rats. *Indian J Pharmacol* 40: 23-27.
6. Kumar S, Dhankhar S, Arya VP, Yadav S, Yadav J (2012) Antimicrobial activity of *Salvadora oleoides* Decne. against some microorganisms. *J Med Plants Res* 6: 2754-2760.
7. Quémére E, Hibert F, Miquel C, Lhuillier E, Rasolondraibe E, et al. (2013) A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PLoS ONE* 8: 1-11.
8. Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, et al. (2000) Phylogeny of the eudicots: a nearly complete familial analysis based on rbcL gene sequences. *Kew Bull* 55: 257-309.
9. Yadav J, Kumar S, Yadav M, Kadyan S, Yadav S (2014) Assessment of genetic diversity using RAPD marker among different accessions of *Salvadora oleoides* of North-West India. *Biores Bull* 1: 1-6.
10. Saini S, Yadav J (2013) Genetic variation in natural populations of *Salvadora oleoides*: An important medicinal plant that needs conservation. *Asian J Plant Sci Res* 3: 20-27.
11. Croizat L (1964) Space, Time, and Form, the Biological Synthesis. Published by the author, Caracas
12. Patriat P, Achache J (1984) India-Eurasia collision chronology has implications for crustal shortening and driving mechanism of plates. *Nature* 311: 615-621.
13. Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science* 16: 108-116.
14. Khan M, Hussain F, Musharaf S (2013) Conservation status of trees in Tehsil Takht-e-Nasrati, Karak Pakistan. *Afr J Plant Sci* 7: 201-207.
15. Holsinger KE, Gottlieb L (1991) Conservation of rare and endangered plants: principles and prospects. *Genetics and conservation of rare plants*. Oxford University Press, NY, USA pp: 195-208.
16. Engler A, Diels L (1936) A. Engler's *Syllabus der Pflanzenfamilien*.
17. Dahlgren G (1989) an updated angiosperm classification. *Bot J Linn Soc* 100: 197-203.
18. Goldberg A (1986) Classification, evolution, and phylogeny of the families of *Dicotyledons* *Smithsonian* contributions to botany. *Smithsonian Institution Press*, Washington : Smithsonian Institution Press.
19. Takhtajan AL (1997) Diversity and classification of flowering plants. *Columbia University Press*.
20. Thorne RF (1992) An updated phylogenetic classification of the flowering plants. *Aliso* 13: 365-389.
21. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102: 8369-8374
22. White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, vol 18. PCR protocols: a guide to methods and applications. Academic Press, Waltham, Massachusetts pp: 315-322.
23. Poeaim A, Poeaim S, Soytong K, Krajangvuthi T (2012) Genetic diversity of *Ficus carica* L. based on non-coding regions of chloroplast DNA. Paper Presented at The 8th International Symposium on Biocontrol and Biotechnology.
24. Baraket G, Olfa S, Khaled C, Messaoud M, Mohamed M, et al. (2008) Chloroplast DNA analysis in Tunisian fig cultivars (*Ficus carica* L.): Sequence variations of the trnL-trnF intergenic spacer. *Biochem Syst Ecol* 36: 828-835.
25. Bast F, Rani P, Meenal D (2014) Chloroplast DNA phylogeography of Holy Basil (*Ocimum tenuiflorum*) in Indian subcontinent. *Sci World J* 2014.
26. Bast F (2013) Sequence similarity search, multiple sequence alignment, model selection, distance matrix and phylogeny reconstruction. *Nature Protocol Exchange*.
27. Schloss PD, Handelsman J (2004) Status of the microbial census. *Microbiol Mol Biol Rev* 68: 686-691.
28. Hoskin CJ, Higgle M, McDonald KR, Moritz C (2005) Reinforcement drives rapid allopatric speciation. *Nature* 437: 1353-1356.
29. McNeill J (2012) International Code of Nomenclature for algae, fungi and plants (Melbourne Code). Koeltz Scientific Books.
30. Sanderson MJ, Purvis A, Henze C (1998) Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol Evol* 13: 105-109.
31. De Craene LR, Wanntorp L (2009) Floral development and anatomy of Salvadoraceae. *Ann Bot* 104: 913-923.
32. Tobe H, Raven PH (2012) Seed morphology and anatomy in Salvadoraceae (Brassicaceae): Systematic and evolutionary implications. *APG* 63: 1-9.
33. Khan A (1996) Appraisal of ethno-ecological incentives to promote conservation of *Salvadora oleoides* Decne: The case for creating a resource area. *Biol Conserv* 75: 187-190.

This article was originally published in a special issue, entitled: "[Trends of Evolutionary Biology & Molecular Phylogenetics](#)", Edited by Xiong Momiao