Identification of *Nigella sativa* Seed and Its Adulterants Using DNA Barcode Marker

Sudhir S. P.[1, *], Alagappan Kumarappan[2], Lalit K. Vyas[3], Divya Shrivastava[1], Padma Deshmukh[4], H. N. Verma[1]

[1]Department of Life Science, Jaipur National University, Jaipur, India

[2]Department of Microbiology, Marine Biology, and Virology, University of Modern Sciences, Dubai, UAE

[3]Vidyabahrti Mahavidyalaya, Department of Cosmetic Technology, Amravati, India

[4]Department of Microbiology, Smt. C.H.M. College of Arts, Commerce and Science, Ulhasnagar, Mumbai, India

Email address:
spsjaipurnationaluniversity@gmail.com (Sudhir S. P.)
*Corresponding author

Abstract: Adulteration, misidentification, and substitution are the biggest challenges in maintaining safety and therapeutic efficacy of medicinal herbs. *Nigella sativa* seed, which is well known medicinal herb susceptible to adulteration or substitution due to its great therapeutic value. Adulteration and substitution by morphologically similar seeds are the primary concern in commercially available *Nigella sativa* seed. In this study, we have used DNA barcode marker to find out adulteration, misidentification, and substitution of *Nigella sativa* seed sold in various markets. We collected 10 samples, which were labelled as Black seed/*Nigella sativa* seed from open markets in India (1 No.), Pakistan (1 No.), Saudi Arabia(1 No.), Egypt (2 No.), Turkey (1 No.), Syria (1 No.), Tunisia(2 No.) and Oman (1 No.). All samples collected from different geographies were studied morphologically. Although few samples were quickly identified as *Nigella sativa* seeds, few were tough to detect and differentiate accurately. This is where DNA barcode marker proved to be useful. Plant DNA were obtained from seed coat cells of samples, was amplified by PCR with forward and reverse rbcl and matK primers as recommended by CBOL (The Consortium for the Barcode of Life). PCR amplification of plastid genome with matK was not very successful, while PCR amplification with rbcl primers was quite successful. We used rbcl sequences for alignment and further analysis. PCR products obtained were subjected to electrophoresis on 1.5% agarose plate. PCR products were sent to Macrogen (Seoul, South Korea) for DNA sequencing. DNA reads obtained with rbcl sequences were aligned and analyzed for nucleotide composition, conserved sites, variable sites, singleton sites and parsimony-informative sites, genetic distance and phylogenetic tree using MEGA 7. The phylogenetic tree was constructed using UPGMA method. NCBI Blast along with phylogenic tree and nucleotide characteristic were used to identify *Nigella sativa* seeds from different geographies and discriminate two adulterants as *Allium cepa* seed and *Clitoria guianensis* seed. Both of these adulterants are different regarding their active medicinal contents and therapeutic utility from *Nigella sativa* seed. This study proved the utility of DNA marker, especially rbcl loci in accurately identifying medicinal herb and its adulterants.

Keywords: *Nigella sativa*, Kalongi, DNA Barcoding, Molecular Markers, rbcl, Matk, Adulteration, Misidentification

1. Introduction

Worldwide trade of medicinal herb is about $ 60 billion dollar business annually. There are about 1000 companies from different countries involved in the trading of medicinal herbs, Business of medicinal herbs is growing at the rate of 15 to 20% per year [1]. This growth in the trade of herbal medicine is due to significant demand for natural, safe and reliable therapeutic agents. Patients want a more safe, secure and natural way of treatment of diseases.

Nigella sativa seed has been used by humankind for centuries as herbal medicine and spice. It is commonly known as Black Seed, Fennel Flower Black cumin, Love-in-a-mist., nutmeg flower, Roman coriander, a Barakah Shooneez, Habba Sauda, Habb al-barka, Krishana – Jiraka, Upakunchika and Kalonji. It is an annual flowering plant belong to buttercup (Ranunculaceae) family. *Nigella sativa* is a native to South and Southwest Asia but also domesticated in Europe.

Prophet Mohammed has described Black seed as the seed of blessings, which can cure all the ailments except death [2]. This belief has triggered a lot of research regarding establishing its therapeutic utility. Although Prophet Mohammed did not mention any particular seed, looking at significant therapeutic utility, *Nigella sativa* seed had been considered as cited seed. *Nigella sativa* seed has been reported to have antimicrobial, antioxidant, anti-aging, hair growth promoter, sun protection, anti-cancer activity, cardiovascular activity, anti-inflammatory activity, immunomodulatory activity, antioxytocic activity and wound healing [3, 4]. Because of its high medicinal value, it is used in its raw form as well as in other forms like seed oil, seed paste and different extracts. This great use in commercial formulations, keep this commodity high in demand and hence it is proven for commercial adulteration and substitution [5]. In few geographies due its nomenclature, it can be misunderstood with morphologically similar but biologically different seeds [6]. Hence it is necessary to identify *Nigella sativa* accurately before its use.

Identification of Herbal medicinal product can be made by following methods [7],

1) Microscopic and macroscopic analysis, where a botanical expert can identify the herb or compare with standard specimen samples.

2) Identification using phytochemical profiling, where one need to perform series of physical and chemical tests.

Both of the above methods had a limitation regarding requirements of the expert taxonomist, long analysis time and proven to error if the samples are in crushed or slurry form, aged, exposed to various environmental factors which would impact anatomical features, marker chemical composition of medicinal herb [8, 9]. Hence it is necessary to look for a new innovative approach for identification of medicinal herbs. DNA barcoding is one of the recent technology, which can help in accurate identification of plant and animal species [10, 11]. Species level identification can be made from the small fragment of Plastid or Genomic DNA. This DNA fragments(loci) are highly specific to particular species. These loci are ITS2, ITS, trnH-psbA, rbcl, and matK. DNA Barcode Identification can help the industry to overcome the problem of adulteration and substitution. DNA marker can surely be used as a method of accurate Identification of herbal medicinal ingredients and their adulterants or substitutions. In recent past, DNA barcodes have been used as an authentication tool for *Crocus sativus* [12], *Tulipa edulis* [13], *Cinnamomum* species [14] and *Ricinus communis* [15], which encouraged us to take the study of identification of *Nigella sativa* seeds and its adulterants using DNA barcode markers like rbcL and matK genes. MatK gene and rbcl genes are considered as standard loci generating quality sequence providing a high level of species discrimination by CBOL [16].

We collected *Nigella sativa* samples from various markets in India (1 No.), Pakistan (1 No.), Saudi Arabia (1 No.), Egypt (2 No.), Turkey (1 No.), Syria (1 No.), Tunisia (2 No.) and Oman (1 No.).
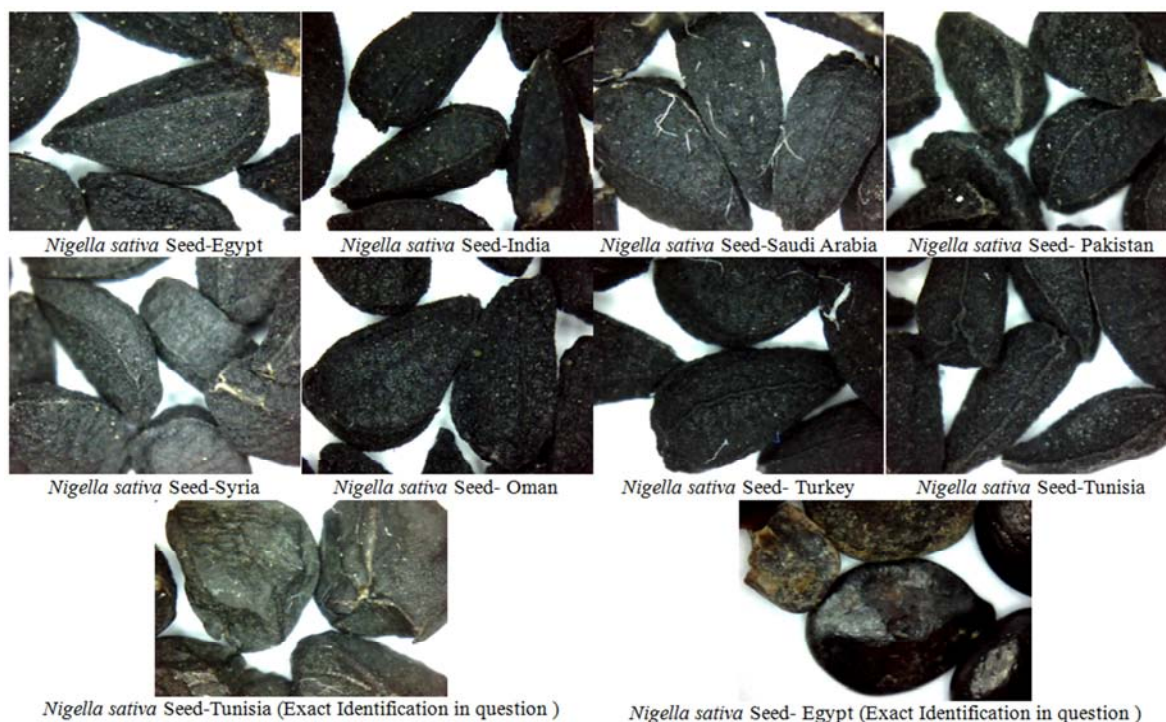

Figure 1. Photographs of seed samples collected from different markets.

This study is to report, the utility of rbcl and matK DNA barcode marker to identify substitution and adulteration in the *Nigella sativa* seed of various geographies.

## 2. Material and Methods

2.1. Samples

Samples of Nigella sativa seeds were collected from various geographies like India (1 No.), Pakistan (1 No.), Saudi Arabia (1 No.), Egypt (2 No.), Turkey (1 No.), Syria (1 No.), Tunisia (2 No.) and Oman (1 No.). Voucher specimens were deposited at Institute Herbarium.

2.2. DNA Extraction and PCR Amplification

Plant DNA was separated from Seed coat cells using plant/fungi DNA isolation kit from Norgen Biotek, Canada (DNA Isolation Kit Product # 26200) following manufacturer's protocol. Purified DNA was preserved at -20°C till further use. Further, extracted DNA was examined using 0.8% agarose gel electrophoresis stained with ethidium bromide.
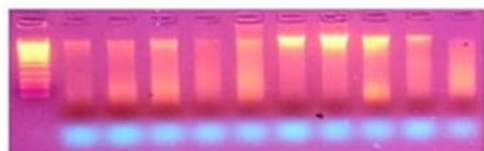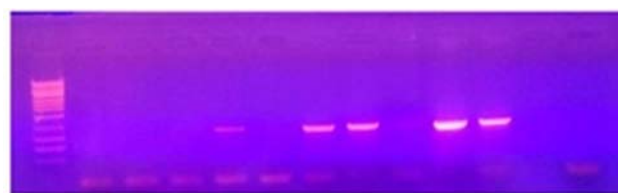

Well 1 – DNA ladder (10, 8, 6, 5, 4, 3, 2.5, 2, 1.5, 1, 0.7, 0.5, 0.3 Kb)
Well 2- 11 = DNA of seed samples (S1-10)

**Figure 2.** *Quality Check for extracted DNA from Seed samples on 0.8% Agarose Gel.*

2.3. DNA Amplification and Sequencing

The target DNA regions, namely rbcL and matK were amplified with respective universal DNA barcoding primers as prescribed by CBOL Plant working group, 2009 [16]. Universal primers for rbcl gene, rbcLa-F: ATGTCACCACAAACAGAGACTAAAGC and rbcLa-R: GTAAAATCAAGTCCACCRCG; for matK gene, matK-KIM1R: ACCCAGTCCATCTGGAAATCTTGGTTC and matK-KIM3F: CGTACAGTACTTTTGTGTTTACGAG were used. PCR was performed using a reaction mixture of a total volume of 50 µl for either of the genes: 25 µl of Taq PCR Master Mix (Norgen Biotek, Canada), 22 µl distilled water, 1 µl forward primer (10µM), 1 µl reverse primer (10 µM) and 1µl of the DNA template (50-80 ng/ µl). The PCR conditions maintained were as follows, one cycle (94°C for 3 min), 35 cycles (94°C for 1 min, 55°C for 1 min, 72°C for 1 min) and one cycle 72°C for 7 min. Amplified PCR products of rbcL and matK primers, each of 5 µl were checked on 1.5% agarose gel electrophoresis for the respective bands and sent to Macrogen (Seoul, South Korea) for DNA sequencing.

Well 1 – DNA ladder (10, 8, 6, 5, 4, 3, 2.5, 2, 1.5, 1, 0.7, 0.5, 0.3 Kb)
Well 2- 11 = mat K gene PCR product of samples (S1-10)
2, 3,4 6 wells shown light bands

Figure 3. *PCR amplification with matK primer.*
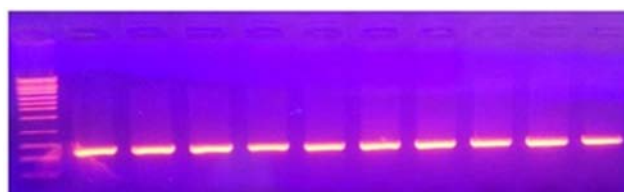
Well 1 – DNA ladder (10, 8, 6, 5, 4, 3, 2.5, 2, 1.5, 1, 0.7, 0.5, 0.3 Kb)
Well 2- 11 = rbcL gene PCR product of samples (S1-10)

Figure 4. *PCR amplification with rbcl primer.*

From above figures, it is clear that PCR amplification with rbcl primer was observed to be good in quality, while PCR amplification with matK primer not of high quality. Nucleotide bands of DNA with matK primer were not well separated. DNA sequencing was done by sending PCR products to specialized research laboratory Macrogen (Seoul, South Korea). As high-quality reads were obtained with the single direction, markers were sequenced in the single direction only.

3. Data Analysis

All data analysis were performed using MEGA 7 (Molecular Evolutionary Genetics Analysis Version: 7.0.18), NCBI (National Center for Biotechnology Information)website using 'blastn' application (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Microsoft Excel 2010.

Obtained sequences were aligned by MUSCLE [17], which generates multiple alignments of amino acid and nucleotide sequences. MUSCLE program is much better regarding speed and accuracy when compared with T-Coffee, MAFFT, and CLUSTALW in all tests. Aligned sequences by MUSCLE were used to locate conserved, variable, singleton, parsimony informative site and compared with other obtained sequences of other *Nigella sativa* seed and its adulterant samples using MEGA 7 [18]. Primary sequence analysis of nucleotide composition, conserved sites, variable sites, singleton sites, parsimony informative sites and phylogenetic tree provided adequate information to discriminate *Nigella sativa* seeds from adulterants. Further all aligned sequences were submitted to NCBI (National Center for Biotechnology Information) website and identified using blastn application (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Sequences of *Nigella sativa* (accession number - KU499880.1), *Nigella arvensis* (accession number - KM360895.1) and *Nigella damascene* (accession number - FJ626586.1) were included in nucleotide composition, genetic distance estimation. Distance estimation was carried out using Kimura 2-parameter model [19].

From rbcl sequence data, we have constructed the phylogenetic tree using UPGMA [20] method. Bootstrap support scoring was done for individual clade by running standard 500 bootstrap replicates of the data. The evolutionary distances were computed using the Kimura 2-parameter method [19] and in the units of the number of base substitutions per site.

The evolutionary history was inferred using the UPGMA method [21]. The optimal tree with the sum of branch length = 0.15209382 was drawn. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) were shown next to the branches [22]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [23] and in the units of the number of base substitutions per site.

4. Results

4.1. Sequence Characteristic

The Length of the rbcl sequence for all 13 samples were 533 bp. Average content of nucleotide were T(U) - 28.4%, C - 21.4%, A - 26.5% and G - 23.7%.

Table 1. Composition of nucleotide of samples and reported accession on NCBI.

| Sample ID	T(U)	C	A	G	Total
NS01 *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
NS02 *Nigella sativa*	28.1	21.6	26.5	23.8	533.0
NS03 *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
NS04 *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
NS05 *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
NS07 *Nigella sativa*	28.1	21.4	26.5	24.0	533.0
NS09 NS *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
NS10 NS *Nigella sativa*	28.3	21.4	26.5	23.8	533.0
Nigella sativa KU499880.1	28.3	21.4	26.5	23.8	533.0
Nigella arvensis KM360895.1	28.3	21.4	26.5	23.8	533.0
Nigella damascena FJ626586.1	28.1	21.4	26.1	24.4	533.0
NS06 NS Adulterant 1	28.3	21.8	28.0	22.0	533.0
NS08 NS Adulterant 2	29.6	21.6	25.9	22.9	533.0
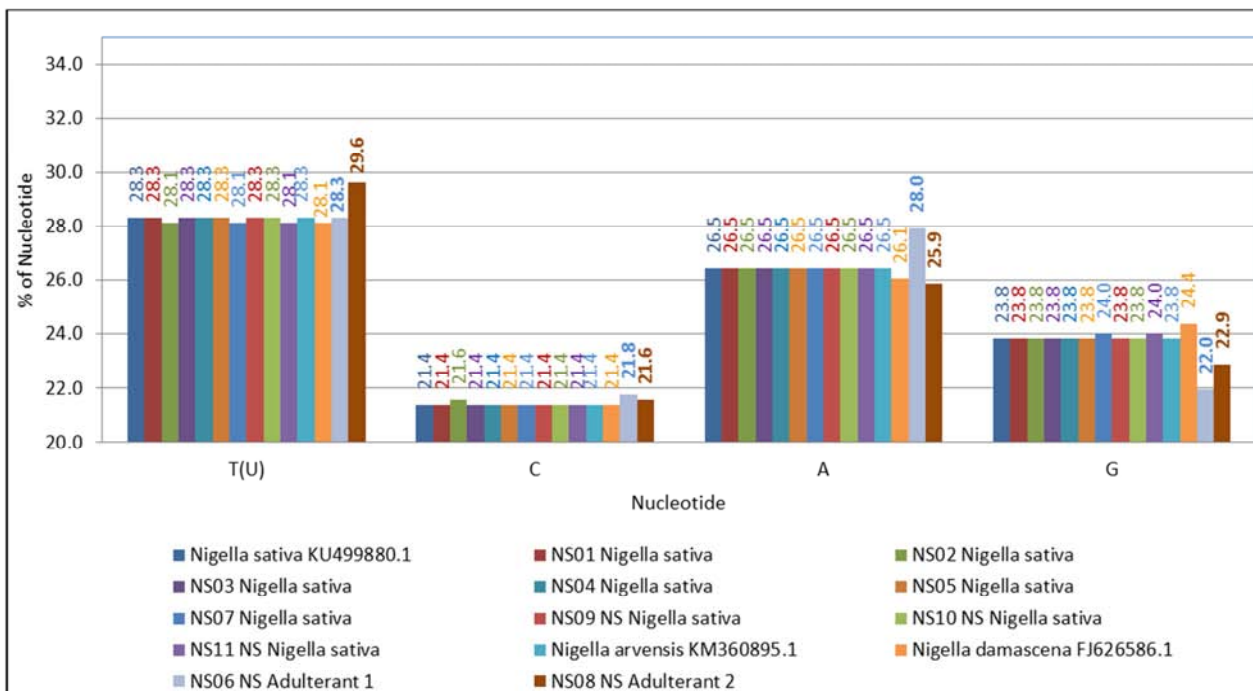
Figure 5. Variation in nucleotide contents of samples and reported accessions on NCBI.

From Table 1 and Figure 5, it is quite clear that nucleotide composition of two samples (NS06 and NS08) are entirely different from all 11 samples majorly in terms of A and G nucleotide, which indicate that these two samples (NS06 and NS08) have different genetic makeup and hence could belong to different plant genus or species. Sample NS01, NS02, NS03, NS04, NS05, NS07, NS09, NS10 and Accession KU499880.1, KM360895.1, FJ626586.1 shows very similar nucleotide composition, which indicates all these samples are belong to single plant genus or species.

Table 2. Details of nucleotide pair frequencies (Directional) observed in entire 13 nucleotide sequences.

	ii	si	sv	R	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG	Total	Domain Info
Avg	517	9	6	1.42	147	3	1	1	3	110	1	1	1	1	138	2	1	1	1	123	533	Data
1st	176	0	1	0.32	37	0	0	0	0	33	0	0	0	0	42	0	0	0	0	64	178	1st Pos Data
2nd	175	1	1	1.00	42	0	0	0	1	45	0	0	0	0	54	0	0	0	0	34	178	2nd Pos Data
3rd	166	7	4	2.00	68	3	0	1	2	31	0	1	0	1	43	2	1	0	1	25	177	3rd Pos Data

ii = Identical Pairs,si = Transitional Pairs,sv = TransversionalPairs,R = si/sv

In the entire group, the pair nucleotide frequencies provide the proper indication about diversity in the genetic makeup of various samples. In NS group, the nucleotide pair frequencies provide the precise evidence about highest % identical sites and lowest rate of Transversional Pairs in the group.

4.2. Analysis of Nucleotide Sequence

4.2.1. Conserved and Variable Sites

Out of 533 sites, 458 sites found to be conserved sites, while 75 sites found to be variable.

Table 3. Details of variable sites.

Site Number	106	118	111	124	120	133	135	148	151	150	169	175	194	202	217	211	228	232	230	246	241	242
Nigella sativa KU499880.1	C	G	T	T	T	C	A	T	G	C	C	A	C	C	C	C	C	A	A	A	T	G
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	.	.	.	G	C	.	T	.	C	T	T	.	T	A	T	G	T	G	.	.	G	C
NS08 NS Adulterant 2	T	T	C	G	.	T	T	.	A	T	.	G	T	A	.	.	.	.	G	T	G	C

Contd..

Site Number	252	255	258	276	309	324	330	333	339	349	351	358	361	366	369	370	371	379	380
Nigella sativa KU499880.1	C	C	T	A	G	G	A	G	C	C	A	C	A	C	T	G	T	G	T
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	T	T	.	.	.	T	.	C	A	T	.	.	C	T	C	C	C	T	C
NS08 NS Adulterant 2	T	T	C	.	T	.	G	C	.	.	G	T	C	.	.	C	C	T	C

Contd..

Site Number	381	387	396	405	420	441	444	474	483	489	492	495	498	501	510	516	525	527
Nigella sativa KU499880.1	T	C	C	T	G	T	T	A	A	T	T	G	C	T	G	T	C	G
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	C	T	.	C	A	.	A	.	.	C	A	A	.	C	A	C	A	.
NS08 NS Adulterant 2	.	T	T	C	.	C	C	T	G	.	.	.	T	C	A	.	.	.

(.)Identical site

Sample NS01, NS02, NS03, NS04, NS05, NS07, NS09, NS10 and Accession KU499880.1, KM360895.1 showed high level of conserved sites, except at site number 405 where NS07 samples got G in place of T. In case of sample NS02, G is replaced by T at site 405 and G is replaced by C at 527 site. This qualifies NS07 and NS02 as varieties under *Nigella sativa*. Accession FJ626586.1 has got five variable sites at 33 (G is replaced with C), 74 (C is replaced with G), 81 (A is replaced with G), 123 (T is replaced with G) and 276 (A is replaced with G). This accession is reported as a separate species as *Nigella damascene*. Sample NS06 and NS08 have 52 and 45 variable sites respectively, which shows wide genetic variation and hence can be considered as different plant genus or species from other NS01, NS02, NS03, NS04, NS05, NS07, NS09, NS10 and Accession KU499880.1, KM360895.1, FJ626586.1. This proves the capability of rbcl gene to discriminate plant genus and species.

4.2.2. Singleton Sites

A singleton site contains at least two types of nucleotides with, at most, one occurring multiple times. MEGA identifies a site as a singleton site if at least three sequences contain unambiguous nucleotides or amino acids.

Table 4. Details of Singleton sites.

Site Number										1	1	1	1	1	1	1	1	2	2	2	2	2	2	2
	3	3	5	7	8	8	9	9	0	1	2	2	3	4	5	6	0	1	1	2	3	3	4		
	6	8	0	3	7	4	1	4	3	8	1	0	3	8	1	9	5	7	1	8	2	0	6	7	
Nigella sativa KU499880.1	C	A	G	T	G	C	A	T	G	C	G	T	T	C	T	G	C	A	C	C	C	A	A	A	G
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	.	C	.	.	G	G	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	.	G	T	C	A	.	.	C	C	G	.	.	C	.	.	C	T	.	T	G	T	G	.	.	A
NS08 NS Adulterant 2	T	.	A	.	.	.	.	.	.	.	T	C	.	T	.	A	.	G	.	.	.	.	G	T	.

Contd…

Site Number	258	276	309	324	330	339	349	351	358	366	369	381	396	420	441	444	473	489	492	495	498	516	525	527
Nigella sativa KU499880.1	T	A	G	G	A	C	C	A	C	C	T	T	C	G	T	T	A	A	T	T	G	C	T	C
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	2	2	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5
Site Number	5	7	0	2	3	3	4	5	5	6	6	8	9	2	4	4	7	8	8	9	9	1	2	2
	8	6	9	4	0	9	9	1	8	6	9	1	6	0	1	4	4	3	9	2	5	6	5	7
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	.	.	.	T	.	A	T	.	.	T	C	C	.	A	.	A	.	.	C	A	A	.	C	A
NS08 NS Adulterant 2	C	.	T	.	G	.	.	G	T	.	.	.	T	.	C	C	T	G	.	.	.	T	.	.

(.)Identical site

When we studied singleton sites in 13 nucleotide sequences, it is observed that sample NS06 and NS08 showed 28 and 20 singleton sites respectively, which qualifies NS06 and NS08 samples as different plant genus or species from group consists of NS01, NS02, NS03, NS04, NS05, NS07, NS09, NS10 and Accession KU499880.1, KM360895.1.

4.2.3. Parsimony-Informative Site

A site is parsimony-informative if it contains at least two types of nucleotides (or amino acids), and at least two of them occur with a minimum frequency of two.

Table 5. *Details of Parsimony-informative sites.*

							1	1	1	1	1	2	2	2	2	3	3	3	3	3	3	3	4	5	5
Sites Number	1	3	3	4	5	6	1	3	5	7	9	4	4	5	5	3	6	7	7	7	8	8	0	0	1
	4	5	6	8	3	0	4	5	0	4	2	1	2	2	5	3	1	0	1	9	0	7	5	1	0
Nigella sativa KU499880.1	A	T	G	G	C	G	T	A	C	C	T	G	C	C	G	A	G	T	G	T	C	T	T	T	G
NS01 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS02 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
NS03 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS04 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS05 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS07 Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
NS09 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS10 NS Nigella sativa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella arvensis KM360895.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nigella damascena FJ626586.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NS06 NS Adulterant 1	C	A	A	T	T	A	G	T	T	T	A	G	C	T	T	C	C	C	C	T	C	T	C	C	A
NS08 NS Adulterant 2	C	A	A	T	T	A	G	T	T	T	A	G	C	T	T	C	C	C	C	T	C	T	C	C	A

Parsimony-informative sites indicate NS01, NS02, NS03, NS04, NS05, NS07, NS09, NS10 and Accession KU499880.1, KM360895.1 are quite similar in genetic makeup while, NS06 and NS08 are quite different.

4.3. Estimation of Genetic Distance Between Sequences

Estimation of genetic distance between sequences is done basis number of base substitutions per site from between sequences. Analyses were conducted using the Kimura 2-parameter model [23].

Table 6. *Intra and Inter-specific Genetic Distances and Species identification (Pairwise genetic distances).*

	Nigella_sativa_KU499880.1	NS01_Nigella_sativa	NS02_Nigella_sativa	NS03_Nigella_sativa	NS04_Nigella_sativa	NS05_Nigella_sativa	NS07_Nigella_sativa	NS09_NS_Nigella_sativa	NS10_NS_Nigella_sativa	Nigella_arvensis_KM360895.1	Nigella_damascena_FJ626586.1	NS06_NS_Adulterant_1	NS08_NS_Adulterant_2
Nigella_sativa_KU499880.1													
NS01_Nigella_sativa	0.00000												
NS02_Nigella_sativa	0.00376	0.00376											
NS03_Nigella_sativa	0.00000	0.00000	0.00376										
NS04_Nigella_sativa	0.00000	0.00000	0.00376	0.00000									
NS05_Nigella_sativa	0.00000	0.00000	0.00376	0.00000	0.00000								
NS07_Nigella_sativa	0.00188	0.00188	0.00188	0.00188	0.00188	0.00188							
NS09_NS_Nigella_sativa	0.00000	0.00000	0.00376	0.00000	0.00000	0.00000	0.00188						

	Nigella_sativa_KU499880.1	NS01_Nigella_sativa	NS02_Nigella_sativa	NS03_Nigella_sativa	NS04_Nigella_sativa	NS05_Nigella_sativa	NS07_Nigella_sativa	NS09_NS_Nigella_sativa	NS10_NS_Nigella_sativa	Nigella_arvensis_KM360895.1	Nigella_damascena_FJ626586.1	NS06_NS_Adulterant_1	NS08_NS_Adulterant_2
NS10_NS_Nigella_sativa	0.00000	0.00000	0.00376	0.00000	0.00000	0.00000	0.00188	0.00000					
Nigella_arvensis_KM360895.1	0.00000	0.00000	0.00376	0.00000	0.00000	0.00000	0.00188	0.00000	0.00000				
Nigella_damascena_FJ626586.1	0.00944	0.00944	0.01325	0.00944	0.00944	0.00944	0.01134	0.00944	0.00944	0.00944			
NS06_NS_Adulterant_1	0.10729	0.10729	0.10934	0.10729	0.10729	0.10729	0.10720	0.10729	0.10729	0.10729	0.11617		
NS08_NS_Adulterant_2	0.09021	0.09021	0.09222	0.09021	0.09021	0.09021	0.09013	0.09021	0.09021	0.09021	0.09870	0.09041	

From the above chart, it is clear that intra-specific genetic distance is from 0.0000 to 0.01325, which is tiny while inter-specific genetic distance among all sequences were almost 0.1161, which is quite high. This genetic distance further helped in identification of adulterants as sequence from NS06 samples had shown genetic distances of 0.10720 to 0.11617, while NS08 sample showed genetic distances of 0.09013 to 0.09870, which are quite high as compared to inter –specific genetic distance of a maximum of 0.01323.

4.4. Phylogenic Tree with UPGMA Method

Typically the evolutionary history is inferred using the UPGMA method [20]. On this study, we wanted to understand whether adulterants show different cluster in the phylogenic tree. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed Branches corresponding to partitions reproduced in less than 50% bootstrap replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Kimura 2-parameter method [19] and are in the units of the number of base substitutions per site.
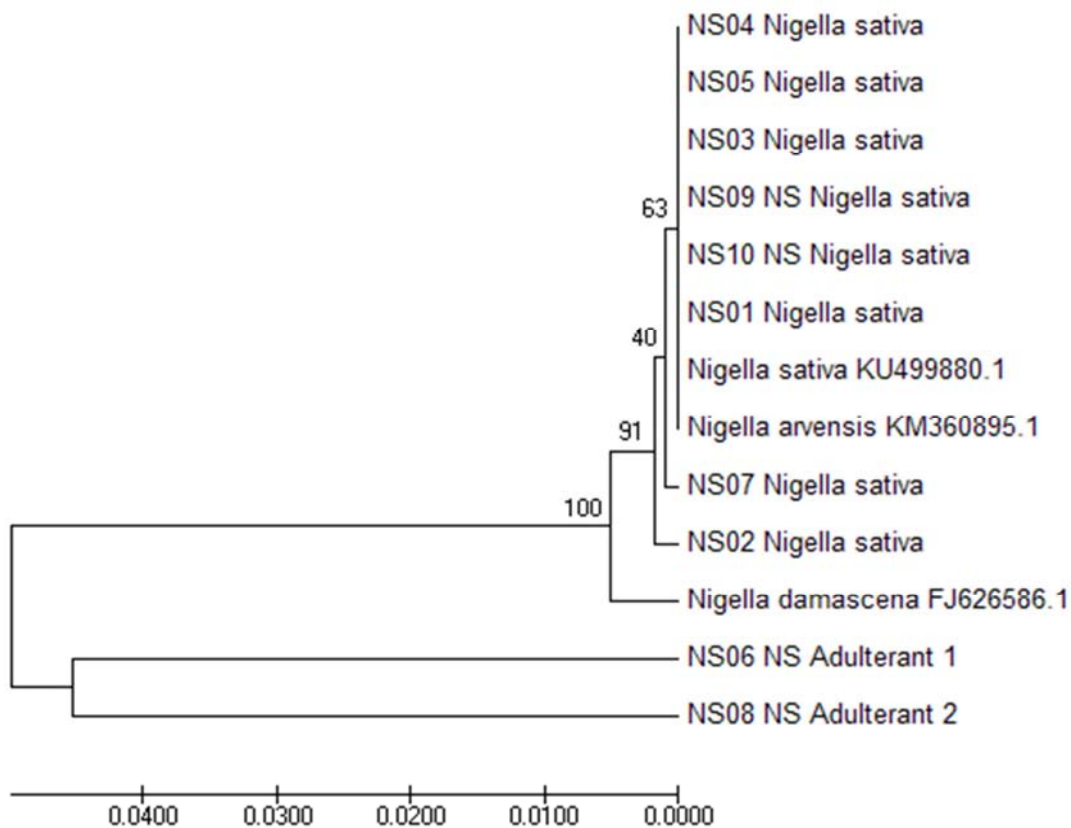
Figure 6. *Phylogenic tree of Nigella sativa (Original Tree).*
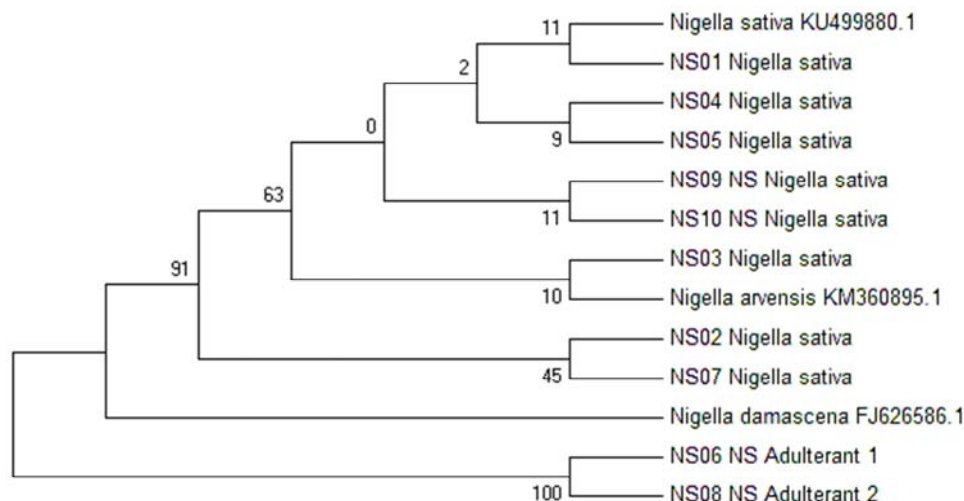
Figure 7. Phylogenic tree of Nigella sativa (The bootstrap consensus tree).

In figure 6 and 7, it can be clearly seen that sample NS06 and NS08 formed a separate clade since they are genetically different and evolved from different ancestor. This observation further confirms species discrimination power of rbcl sequences.

Further to above sequence analysis, aligned rbcl sequences of seed samples were blasted individually on NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) for identification of samples. Following are the details of identification of rbcl sequences from NCBI blastn tool.

Table 7. Matching and Identification of nucleotide sequences using NCBI blastn.

Sample ID	Matches with Accession	Description	Max score	Total score	Query cover	E value	Ident
NS01	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1027	1027	0.9	0	1
NS02	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1027	1027	0.91	0	0.99
NS03	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1025	1025	0.92	0	1
NS04	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1029	1029	0.9	0	0.99
NS05	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1033	1033	0.92	0	0.99
NS07	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1021	1021	0.9	0	0.99
NS09	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1027	1027	0.9	0	1
NS10	KU499880.1	*Nigella sativa* voucher A1 rbcL gene, partial cds; chloroplast	1042	1042	0.92	0	0.99
NS06_AD1	AB292286.1	Allium cepa chloroplast DNA, rbcL and ORF106, partial sequence	1037	1037	0.91	0	0.99
NS08_AD2	JQ591652.1	Clitoria guianensis voucher BioBot00900 rbcL gene, partial cds; chloroplast	1021	1021	0.9	0	0.99

Blastn tool identified rbcl sequences of two seed samples NS06 and NS08 as *Allium cepa* and *Clitoria guianensis.* These samples were further compared with standard morphological features of *Allium cepa* and *Clitoria guianensis.* Morphological features of samples found matching with physical samples used in study.

5. Discussion

Prophet Mohammed in Islamic literature has described black seed as the seed of blessings which has a property of curing any disease of humankind. *Nigella sativa* seed looking at its therapeutic utility can be considered as mentioned black seed.

Nigella sativa seed is one of the noble herbs which is extensively used as medicine and spices in the Middle East, South East Asia, and Europe It is one of the great spices used for culinary purposes. Recently many researchers proved the great therapeutic uses of *Nigella sativa* seeds and its extracts. Looking at huge benefits, its demand is growing and hence

can be easily susceptible to substitutions and adulteration for commercial benefits. Hence there is immediate need to have a quick, reliable and reproducible method of identification of *Nigella sativa* seeds and its adulterants. Traditional methods of identification of herbal medicine have few shortcomings regarding final output basis expertise of individual examiner, long time, difficulties in identification of sample in powdered/crushed/aged and slurry condition. To overcome these problems use of DNA barcoding has been suggested. Use of DNA barcode marker offers numbers of applications in the field plant taxonomy, identification, and authentication of herbal medicinal ingredients. Recently DNA barcoding was successfully used for identification of herbal medicinal ingredients few study to quote are Identification of *Achyranthis Bidentatae Radix* [24], DNA barcoding of 347 medicinal plants using rbcL marker [25], Identification of Physalis (Solanaceae) from its adulterants [26], Identification of Botanicals in Herbal Medicine and Dietary Supplements [27] and DNA-based identification of *Gentiana robusta* and related species [28].

In fact, *Chinese Pharmacopoeia 2010 edition adopted* allele-specific diagnostic PCR as a new method of identifying *Zaocys dhumnades* (Cantor). Recently US Food and Drug Administration has approved DNA barcoding as a method of seafood identification [29].

In this study we collected 10 samples of Black seed i.e. *Nigella sativa* seeds from open markets of India (1 No.), Pakistan(1 No.), Saudi Arabia(1 No.), Egypt(2 No.), Turkey (1 No.), Syria (1 No.), Tunisia(2 No.) and Oman(1 No.). In these countries, *Nigella sativa* seeds are regularly consumed as spices and herbal medicine. Morphological and microscopic study of seeds were done, where we could identify eight samples quickly, but few were tough to identify. This is where we decided to check the identity of seeds samples by DNA barcode marker.

There are lots of choices of molecular marker selections, like plastid DNA regions (atpF–atpH spacer, rbcL gene, matK gene, rpoB gene, rpoC1 gene, trnH–psbA spacer and psbK–psbI spacer) but on the basis of assessment conducted by CBOL Plant Working Group, rbcL and matK plant barcode were selected for assessments due to recoverability, sequence quality, and levels of species discrimination.

matK has been considered as most rapidly evolving plastid coding region having consistently high levels of discrimination power among angiosperm species, we got very poor amplification and low level of identification. This is in line with reported low routine success [30] and more patchy recovery [31, 32]. Performance of rbcl was entirely satisfactory regarding amplification, separation of and also provided good discriminating power.

DNA sequence reads obtained from with rbcl were of high quality. MEGA 7 software offered modules to check nucleotide composition, conserved sites, variable sites, singleton sites and parsimony-informative sites to discriminate *Nigella sativa* seeds from adulterants.

DNA sequences of rbcl were blasted on NCBI website. Blastn identified query DNA sequences as DNA sequences

of adulterant Number 1 as *Allium cepa* seed and adulterant Number 2 as *Clitoria guianensis* seed.

Further based on, Molecular identification we checked the morphology of adulterants, which has matched with the study samples.

Both of these adulterants are different from Nigella seeds regarding their active medicinal contents and therapeutic utility.

6. Conclusion

Based on the present research study, we can conclude that for quick, accurate identification of herbal medicine, like *Nigella sativa* seeds and its adulterants, DNA Barcode marker especially rbcl was found to be useful. It is recommended to include DNA barcode marker method of identification in the monograph of official books of herbal medicine standards. This would improve the quality of *Nigella sativa* in various markets and avoid the impact of consumption of adulterated *Nigella sativa* seeds.

References

[1] Showkat Hussain Ganie, Priti Upadhyay, Sandip Das, Maheshwer Prasad Sharma, Authentication of medicinal plants by DNA markers, Plant Gene, Volume 4, December 2015, Pages 83-99, ISSN 2352-4073, http://dx.doi.org/10.1016/j.plgene.2015.10.002.(http://www.sciencedirect.com/science/article/pii/S235240731500030X)

[2] Al-Bukhari. MI. Division (71) on medicine. In Sahi Al-Bukhari, the collection of authentic sayings of Prophet Mohammad (peace be upon him). 2nd ed. Hilal Yayinlari, Ankara, Turkey, 1976.

[3] Padmaa, M. and Paarakh (2010). Nigella sativa. Linn.A comprehensive review. Indian Journal of Natural Products and Resources 1(4): 409-429.

[4] Dermatological effects of Nigella sativa, Salih H. M. Aljabrea, Omar M. Alakloby a, Mohammad A. Randhaw, Journal of Dermatology & Dermatologic Surgery 19 (2015) 92–98.

[5] Handbook of herbs and spices, Volume 2,2004, Edited by,K. V. Peter, CRC Press Boca Raton Boston New York Washington, DC. Page 225-226.

[6] Rania Mubarak AwadObaid, Anti-Mycetoma, anti-oxidant and Phytochemical Screening of *Nigella sativa* seeds, A thesis for Master Degree in pharmacy (pharmacognosy),The National Ribat University Faculty of Graduate studies & Scientific Research, May 2015.

[7] Anjoo Kamboj (2012). Analytical Evaluation of Herbal Drugs, Drug Discovery Research in Pharmacognosy, Prof. OmboonVallisuta (Ed.), ISBN: 978-953-51-0213-7, InTech, Available from: http://www.intechopen.com/books/drug-discovery-research-in-pharmacognosy/analytical-evaluation-of-herbaldrugs.

[8] Xie PS, Leung AY. 2009. Understanding the traditional aspect of Chinese medicine in order to achieve meaningful quality control of Chinese materiamedica. Journal of Chromatography A. 1216: 1933- 1940.

[9] Yi L, Liang Y, Wu H, Yuan D. 2009. The analysis of Radix Angelicae Sinensis (Danggui). Journal of Chromatography A. 1216: 1991-2001.

[10] Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences. 2003;270 (1512): 313-321. doi: 10.1098/rspb.2002.2218.

[11] Kress WJ, Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS ONE 2(6): e508. doi: 10.1371/journal.pone.0000508

[12] Caterina Villa, Joana Costa, Liliana Meira, M. Beatriz P.P. Oliveira, Isabel Mafra, Exploiting DNA mini-barcodes as molecular markers to authenticate saffron (Crocus sativus L.), Food Control, Volume 65, July 2016, Pages 21-31, ISSN 0956-7135, http://dx.doi.org/10.1016/j.foodcont.2016.01.008.http://www.sciencedirect.com/science/article/pii/S0956713516300093)

[13] Hong-liang Ma, Zai-biao Zhu, Xiao-ming Zhang, Yuan-yuan Miao, Qiao- sheng Guo, Species identification of the medicinal plant Tulipa edulis (Liliaceae) by DNA barcode marker, Biochemical Systematics and Ecology, Volume 55, August 2014, Pages 362-368, ISSN 0305-1978, http://dx.doi.org/10.1016/j.bse.2014.03.038.http://www.sciencedirect.com/science/article/pii/S0305197814001240)

[14] Runglawan Sudmoon, Arunrat Chaveerach, Arisa Sanubol, Pansa Monkheang, Nantiya Kwanda, Sarocha Aungkapattamagul, TawatchaiTanee, KowitNoikotr, Chattong Chuachan and Napaporn Kaewdoungdee [e], Identifying Efficiency in Herbal medicine Cinnamomum Species (Lauraceae) Using Banding Patterns and Sequence Alignments of rpoB, rbcL and matK Regions, Chiang Mai J. Sci. 2014; 41 (5.1): 1094-1108 http://epg.science.cmu.ac.th/ejournal/Contributed Paper

[15] Mohamed Enan1, 2, Nael Fawzi1, 3, Mohammad Al-Deeb1, Khaled Amiri1, DNA Barcoding of Ricinuscommunis from Different Geographical Origin by Using Chloroplast matK and Internal Transcribed Spacers,American Journal of Plant Sciences, 2012, 3, 1304-1310 http://dx.doi.org/10.4236/ajps.2012.39157 Published Online September 2012 (http://www.SciRP.org/journal/ajps)

[16] COBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Nat AcadSci 106: 12794–12797.

[17] Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792-97.

[18] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance,and maximum parsimony methods. MolBiolEvol. 2011 Oct; 28(10): 2731-9. doi: 10.1093/molbev/msr121. Epub 2011 May 4. PubMed PMID: 21546353; PubMed Central PMCID: PMC3203626.

[19] Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16.

[20] Sneath P. H. A. and Sokal R.R. (1973).Numerical Taxonomy. Freeman, San Francisco.

[21] Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.Molecular Biology and Evolution 33: 1870-1874.

[22] Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791.

[23] Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

[24] Hou DY, Wang GP, Zhi LH, Xu 1 HW, Liang 1 HL, Yang 1 MM, Shi 1 GA. Molecular.

[25] Vassou SL, Nithaniyal S, Raju B, Parani M. Creation of reference DNA barcode library and authentication of medicinal plant raw drugs used in Ayurvedic medicine. BMC Complement Altern Med. 2016 Jul 18;16Suppl 1: 186. doi: 10.1186/s12906-016-1086-0. PubMed PMID: 27454470; PubMed Central PMCID: PMC4959393.

[26] Feng S, Jiang M, Shi Y, Jiao K, Shen C, Lu J, Ying Q, Wang H. Application of the Ribosomal DNA ITS2 Region of Physalis (Solanaceae): DNA Barcoding and Phylogenetic Study. Front Plant Sci. 2016 Jul 19;7: 1047. doi: 10.3389/fpls.2016.01047. eCollection 2016. PubMed PMID: 27486467; PubMed Central PMCID: PMC4949264.

[27] Parveen I, Gafner S, Techen N, Murch SJ, Khan IA. DNA Barcoding for the Identification of Botanicals in Herbal Medicine and Dietary Supplements: Strengths and Limitations. Planta Med. 2016 Jul 8. [Epub ahead of print] PubMed PMID: 27392246.

[28] U Xiong B, Zhao ZL, Ni LH, Gaawe D, Mi M. [DNA-based identification of Gentianarobusta and related species]. ZhongguoZhong Yao ZaZhi. 2015 Dec;40 (23): 4680-5.Chinese. PubMed PMID: 27141683.

[29] Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J. and Nguen, N. T. (2008) Potential Use of DNA Barcodes in Regulatory Science: Applications of the Regulatory Fish Encyclopedia. Journal of Food Protection, 71, 210-217.

[30] Fazekas AJ, et al. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802.

[31] Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE 2: e508.

[32] Renaud Lahaye, Michelle van der Bank, Diego Bogarin, Jorge Warner, Franco Pupulin, Guillaume Gigot, Olivier Maurin, Sylvie Duthoit, Timothy G. Barraclough, Vincent Savolainen (2008) DNA barcoding the floras of biodiversity hotspots. Proc Natl AcadSci USA 105: 2923–2928.