Current Science International Volume : 07 | Issue : 04| Oct.- Dec. 2018
ISSN: 2077-4435 Pages: 845-855

Authentic Identification and Phylogenetic Analyses of Nine Cultivated Flax Genotypes (*Linum usitatissimum* L.) Based on *rbcL* Gene Sequence

Mohamed Z.S. Ahmed[1], Injy M. Masoud[1] and Said Z.A. Zedan[2]

[1]*Genetic Resources Department, Desert Research Center (DRC), 1, Mathaf El-Matarya Street, El-Matarya B.O.P 11753 El-matarya, Cairo, Egypt.*
[2]*Fiber Crops Research Department, Field Crops Research Institute (FCRI), ARC, 9, El-Gamaa St. Giza, Egypt.*

Received: 30 Sept. 2018 / Accepted: 15 Dec. 2018/ Publication date: 30 Dec. 2018

ABSTRACT

DNA barcoding is a power technique for the quick identification and phylogenetic relationship analysis and has several applications in classification of plants into wild and genotypes cultivated. The present study compared among 9 Flax cultivars (*Linum usitatissimum* L.) included, 7 Egyptian flax cultivars, and 2 foreign flax cultivars with 54 species belong to subfamilies: *Linoideae* and *Hugonioideae* based on the single locus marker *rbcL* gene sequence data. The results showed a single partial nucleotide fragment of the *rbcL* gene (about ~ 600 bp) that was obtained from each sample. The size and reading sequence were determined by Macrogen analysis as follow: 550bp, 551bp, 566bp, 556bp, 536bp, 548bp, 498bp, 544bp and 531bp with Electra, Giza-10, Giza-11, Giza-12, Sakha-3, Sakha-5, Sakha-6, Strain-562, Blanka cultivars, respectively. Multiple alignments of *rbcL* partial gene nucleotide sequence were done using Clustal Omega program between consensus *rbcL* partial gene sequence (566bp) of the nine cultivated flax genotypes and 54 flax species showed that *rbcL* gene partial sequence are shared high identity and similarity more than 94.92 to 100%. The phylogenetic analyses among consensus *rbcL* gene sequence (566bp) of the 9 cultivated flax genotypes (*Linum usitatissimum* L.), and 54 species representing all eight genera of subfamily *Linoideae* included 5 species of *Hugonia* (Family *Linaceae* - subfamily *Hugonioideae*) as out-group were analyzed with data from the chloroplast *rbcL* sequences that are more closely related to *Linum*. They discriminated and confirmed the identification of flax cultivated in this study for the genus *Linum*, family *Linaceae* at the molecular level, using *rbcL* regions through *rbcL* reference sequences in NCBI database. The results using *rbcL* partial gene sequencing tools a single locus marker could be helpful for authentic identification, resolving authentication, and to monitor the adulteration of flax seed varieties market under national and international trade.

Keywords: DNA barcoding, Chloroplast *rbcL* gene, Molecular Phylogenetic, Flax (*Linum usitatissimum* L.), authentic identification, Family Linaceae - subfamily Linoideae

Introduction

Flax (*Linum usitatissimum* L.), is a self-pollinating annual crop species, it is grown commercially for several purposes as the source of linen fiber, linseed oil and medicinal value. Flax is a source of food, feed, fiber, and medicine for more than 8,000 years (van Zeist and Bakker-Heeres, 1975). Linseed oil provides health benefits mainly due to its high content in omega-3 alpha linolenic acid (55-57%) (Przybylski, 2001). The lignans present in linseeds have beneficial properties against breast, colon, prostate and thyroid cancer, and in lowering relative risk factors for heart disease (Westcott and Muir, 2003). Lignans appear to be anticarcinogenic compounds (Power *et al*., 2004; Thompson *et al*., 2004).

The genus *Linum*, a representative of the Linaceae family, comprises about 300 species (Hickey and King, 1988). The taxonomy of the genus is still unclear, although a subdivision to five sections (*Linum*, *Dasylinum*, *Linastrum*, *Syllinum* and *Cathartolinum*) is generally accepted (Ockendon and Walters, 1968). Gill, (1987) proposed the same sub sections based on chromosome number, floral morphology, and interspecific compatibility. Rogers, (1982) also divided section *Linopsis* (*Linastrum*) into five subsections. Only some species from the section *Linum* are

Corresponding Author: Mohamed Z.S. Ahmed, 1Genetic Resources Department, Desert Research Center (DRC), 1, Mathaf El-Matarya Street, El-Matarya B.O.P 11753 El-matarya, Cairo, Egypt. E-mail: zoelfakar99@yahoo.com

used for agronomical purposes. Where *Linum grandiflorum* Desf. and *L. perenne* L. serve as ornamentals, cultivated flax (*L. usitatissimum*) is commercially grown for its fiber and linseed oil and the *Linaceae* is a relatively small but nearly cosmopolitan family of flowering plants, distributed throughout the temperate and tropical regions of the world. Including cultivated flax (*Linum usitatissimum* L.), the source of linen fiber and linseed oil, the family contains approximately 260 species in two subfamilies: *Linoideae* and *Hugonioideae*. *Linoideae* is the larger subfamily, with an estimated 210 species in eight genera, including *Linum*, the type genus of the family. The distribution of *Linoideae* is essentially temperate modified from McDill *et al*., 2009. Several flaxes are cultivated as garden ornamentals, including the blue-flowered species blue flax (*L. narbonense*), Lewis' blue flax (*L. lewisii*), and perennial blue flax (*L. perenne*), the red-flowered scarlet flax (*L. grandiflorum*), and the yellow-flowered golden flax (*L. flavum*). All of the families of *Linales sensu* Cronquist, including Linaceae, are placed in the large order *Malpighiales* (APG II, 2003; APG III, 2009; APG IV. 2016).

The search for chloroplast sequences that are optimal for phylogenetic inference started in the 1980s. In early work, the *rbcL* gene, which encodes the large subunit of ribulose-1, 5- bisphosphate carboxylase/oxygenase (RuBisCO) was used (Taberlet *et al*., 1991). In the early 1990s, researchers focused on single copy, noncoding regions of the chloroplast genome, mainly introns and intergenic spacers in order to identify sequences with higher rates of evolution (Gielly and Taberlet, 1994). Olmstead *et al*., (1992) conducted 57 angiosperm *rbcL* sequences to test the monophyly of the *Asteridae* and to identify major lineages within the *Asteridae*. The data of DNA sequences from 475 and 499 species of seed plants are exclusively from the chloroplast gene *rbcL*, which codes for the large subunit of ribulose-l .5-bisphosphate carboxylase/ oxygenase (RuBisCO or RuBPCase) used in major taxonomic groups (Chase *et al*., 1993). Phylogenetic analysis of flowering plants based on a second plastid gene, *atpB*, analyzed separately and in combination with *rbcL* sequences for 357 taxa (Chase *et al*., 2002). Chloroplast gene *rbcL*, which is the most frequently sequenced plastid gene with a single copy gene, approximately 1430 base pairs in length, encodes the large subunit of ribulose-1, 5-bisphosphate carboxylase/oxygenase (RUBISCO). It is widely sequenced from numerous plant taxa which free from length mutations except at the far 30 end (Bousquet *et al*., 1992; Gaut *et al*., 1992; Morgan and Soltis, 1993). DNA barcoding, a power tool to identify species based on sequences from a short, standardized DNA region, opens up a unique avenue to identify species, promises a rapid and efficient method to explore the dimensions of biodiversity. The use of one or more short DNA sequences for biological identifications with the ultimate goal of quick and reliable species-level identifications across all forms of life, including animals, plants, and microorganisms (Hebert *et al*., 2003; Hebert and Gergory, 2005). DNA barcodes methods for identifying species by using short orthologous DNA sequences was proposed by Kress *et al*., (2005) and was initiated to facilitate biodiversity studies, identify juveniles, associate sexes, and enhance forensic analyses.

An ideal DNA barcoding region should be short enough to amplify from degraded DNA and analyzed via single-pass sequencing. One benefit of this region is the large amount of existing information; there are more than 10,000 *rbcL* sequences already in GenBank (Chase *et al*., 2005; Newmaster *et al*., 2006). Three regions of plastid DNA as a standard protocol for barcoding all land plants were used by Chase *et al*., (2007). A major goal of application of DNA barcoding for plant use universal primary to identify species in local floras and ecological communities. With the consensus of a 2-locus DNA barcode (*rbcL*+*mat*K) by the Consortium for the Barcode of Life (CBOL, 2009), barcoding efforts have begun to focus on building the barcode library for land plants. Recently for plants, (CBOL, 2009) recommend the 2-locus combination of *rbcL*+*mat*K from 7 candidate plastid DNA regions (*atpF-atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer, and *trnH-psbA* spacer) the plant barcode based on assessments of recoverability, sequence quality, and levels of species discrimination use universal primary gene regions. Soltis *et al*., (2000) used phylogenetic study of *rbcL*, *atpB* and nuclear ribosomal gene 18S and found that *Humiriaceae* (represented by *Humiria* Aubl. and *Vantanea* Aubl.), *Ixonanthaceae* (*Ochthocosmus* Benth. and *Ixonanthes* Jack) and *Erythroxylaceae* (*Erythroxylon* L. and *Nectaropetalum* Engl.) were dispersed throughout *Malpighiales*; none of these families was found to be sister to *Linaceae*, and *Linales sensu* Cronquist appeared to be non-monophyletic. The close relationship of *Linoideae* and *Hugonioideae* has been found consistently in phylogenetic studies that have included members of both subfamilies (e.g. Davis *et al*., 2005; Wurdack and Davis, 2009).

Analyzed sequences of *rbcL* from 48 species of *Linaceae*, suggested that *Linaceae* and subfamily *Linoideae* are monophyletic, but *Linum* is not. *Anisadenia*, *Reinwardtia*, and *Tirpitzia* were found to be the basal members of *Linoideae* by McDill *et al*., (2009). Phylogenetic analyses of 44 species representing all the eight genera of subf. *Linoideae* and 37 species of *Linum*, with data from the chloroplast (*ndhF, trnL-F, trnK 3′* intron) and the nuclear ITS, with *Hugonia* (*Linaceae* subf. *Hugonioideae*) as out-groups were studied. China Plant BOL Group, (2011) proposed combinations of the data sets and followed the standards and guidelines of the CBOL Plant Working Group to amplify and sequence four DNA-barcoding regions, i.e., plastid *rbcL, matK*, *trnH–psbA*, and nuclear ribosomal (nr) ITS. McDill and Simson, (2011) examined the phylogeny of *Linaceae*, with sampling from the 13 commonly recognized genera of the family and sequence data from the plastid genes *matK* and *rbcL*. Burgess *et al*., (2011) examined markers singly and in combination to evaluate discriminatory power of the *rbcL+matK* barcode among 436 species in 269 genera of land plants from Canada. Schori and Showalter, (2011) used DNA barcoding to identify several medicinal plants found in Pakistan and distinguished them from other similar species. Bafeel *et al*., (2012) showed that using *rbcL* gene sequences enabled identification of the majority of the samples (92%) to genus level and only 17% to species level. Many authors have proposed DNA barcoding as an integrated approach with classical taxonomy for species identification and authentication in the postgenomics era (Sahare and Srinivasu, 2012; Vohra and Khera, (2013)). Techen *et al*., (2014) discussed the genomic regions and molecular methods selected to provide barcodes, available databases and the potential future of barcoding using next generation sequencing. The database should provide DNA barcodes for data retrieval and similarity search. Several molecular methods have been applied to develop markers that aid with the authentication and identification of medicinal plant materials. DNA barcoding for all groups of living organisms is a good tool for germplasm documentation, conservation, and utilization in the world's biodiversity. For land plants the core DNA barcode markers are two sections of coding regions within the chloroplast, part of the genes, *rbcL* and *matK*, de Vere *et al.,* (2015). Mishra *et al*., (2016) developed multi-locus strategy used for herbal plant identification, *matK*, *rbcL*, *trnH-psbA*, ITS, *trnL-F*, genes, 5S-rRNA and 18S-rRNA as successful DNA barcodes for identification of herbal drug industry to offer an alternative and feasible taxonomic tool for rapid and robust species with the herbal market identification. Bell *et al*., (2017) succeeded in species-level identifications using *rbcL* database and ITS2 for 8 species and a family-level identification of the 9th species. DNA metabarcoding has broad-ranging applications in ecology, aerobiology, biosecurity, and forensics. Kress, (2017) developed 4 plant DNA barcode markers as a new biological tool for organismal biologists to increase their understanding of the natural world, *rbcL*, *mat*K, *trnH-psbA*, and ITS2, for used to address basic questions in systematics, ecology, evolutionary biology and assessing priority areas for environmental protection. Maloukh *et al*., (2017) suggested simple method through *rbcL* as promising barcode locus for identification of 51 species of desert plant taxa (legumes and grasses). The primer universality and discriminatory power of *rbcL* is 100%, while it is 35% for *matK* locus for these plant species.

The aim of this work is to discriminate and confirm the identification of seven Egyptian flax cultivars and another two foreign flax cultivars at the molecular level, using Chloroplast *rbcL* gene partial sequences region as DNA barcoding. This study would further be helpful for identification, authentication, and to monitor the adulteration of seed varieties under national and international trade.

Materials and Methods

I. Plant materials collection

Plant flax (*Linum usitatissimum* L.) as nine genotypes; seven Egyptian flax cultivars, i.e., Giza-10, Giza-11, Giza-12, Sakha-3, Sakha-5, Sakha-6, Strain-562 and another two foreign flax cultivars; Electra (Belgium) and Blanka (The Leibniz Institute of plant Genetics and Crop plant Research (IPK), Gatersleben, Germany) were collected as bulk and ground into fine powder using liquid nitrogen (LN). The frozen samples of all genotypes were kept frozen at -80°C till use. In this study, selected 54 species representing the 13 currently recognized genera of *Linaceae* for inclusion were compared with the 9 cultivars. Each genus is represented by multiple species when possible, and sampling of the larger genera *Linum* approximated to taxonomic diversity and geographical range. The used 54 flax germplasm genotypes belonging to the genus *Linaceae* was chosen according to McDill and Simson,

(2011). The 54 sequence data of closer accessions of genera of *Linaceae,* subfamily *Linoideae* and *Hugonioideae*, out-group were downloaded from NCBI database. The list of species, BLASTn hits against the GenBank database, similarity score and GenBank accession numbers in NCBI are presented in Table (1).

II. DNA extraction, primers design and PCR amplification

Genomic DNAs were extracted from frozen tissue of the individual nine flax samples following the procedure described by DNeasy Plant Mini Kit (Qiagen Inc., Cat.no.69104, USA), this was performed following the manufacturer's instruction. The concentrations and quality of the genomic DNA samples were estimated on spectrophotometer ND-2000 (Nanodrop, USA). Finally, all the genomic DNA samples were diluted to a final concentration of 40 ng/µl with TE buffer (10 mM Tris-HC1, pH 8.0; 1 m M EDTA) and stored at -20°C for further use. DNA fragments were amplified via standard polymerase chain reaction (PCR). The entire coding plastid *rbcLa* (first part of *rbcL* gene ~ 700bp or less) was amplified using the primer pairs *rbcLa*_Fwd (5′-ATG TCA CCA CAA ACA GAG ACT AAA GC-3′) and *rbcLa*_Rev (5'-GTA AAA TCA AGT CCA CCR CG-3') and sequence a 599 bp of the *rbcL* gene as previously described by Levin *et al*., (2003) and Olmstead *et al*., (1992).

The amplification reaction was carried out in 10 µl reaction volume containing 6.32 µl sterile water, (2 µl) 5X HF Buffer (with $MgcL_2$), (0.3 µl) 100% DMSO, (0.056 µl) 10mM dNTPs), (0.1 µl of) Fwd primer (0.10 µM), (0.1 µl of) Rev primer (0.10 µM), (0. 125 µl) Phusion High Fidelity (5U/ µl) and 1 µl (50ng) template DNA was performed in a Perkin Elmer thermal cycler 9700. PCR program for *rbcL* regions*,* the temperature profile for in the different cycles was as follows: an initial strand separation cycle at 98°C for 3 min followed by 35 cycles comprised of a denaturation step at 98°C for 30 seconds, an annealing step at 55°C for 30 seconds and an extension step at 72°C for 45 seconds. The final cycle was a polymerization cycle for 7 min at 72°C.

III. Gel electrophoresis

PCR amplifications samples were resolved by electrophoresis in a 1.5% agarose gel containing ethidium bromide (0.5ug/ml) in 1X TAE buffer at 95 volts. A 1Kbp DNA ladder was used as DNA standard size marker. PCR products were visualized on UV light and photographed using a Gel Documentation System (BIO-RAD 2000).

IV. QIAquick® gel extraction kit protocol for purification

QIAquick gel extraction kit (Qiagen, Cat. No. 28704 -28706) was used for PCR product purification or fast cleanup from the agarose gel up to 10 µg can bind to each QIAquick column. This was performed following the manufacturer's instruction.

V. Sequence editing, alignment and phylogenetic inference

Sequencing chromatograms *rbcL* regions obtained were analyzed by Macrogen, Seoul, south of Korea. In all the sequenced *rbcL* regions, Protein sequences were predicted from the obtaining Open Reading Frame (ORF) *rbcL* nucleotide sequence using the web:www.Expasy.org/translate. All nucleotide sequences of *rbcL* gene were searched in NCBI database, The National Center for Biotechnology Information GenBank Database, (http://www.ncbi.nlm.nih.gov), whiles the European Molecular Biology Laboratory (EMBL) web site, (https://www.ebi.ac.uk/services) were used for sequence analyses. Homology comparison and phylogenetic relationship were analyzed using the CLUSTALW multiple sequence alignment algorithm (Thompson *et al*., 1994). The obtained *rbcL* data sequences were aligned on line using the NCBI (https:// www.ncbi.nlm.nih.gov/) web site through Basic Local Alignment Search Tool (BLAST). Then multiple alignment and phylogeny tree of nucleotide were done using the Clustal Omega free web site: (https://www.ebi.ac.uk/Tools/msa/clustalo/) for both phylogenies (Sievers *et al*., 2011); consensus of the nine sequences between the 54 *rbcL* data nucleotide sequences for *linaceae* family from NCBI.

Results and Discussion

DNA extraction was successful for all the nine samples (the 7 Egyptian flax cultivars, and another two foreign flax cultivars with high DNA quality and good yield (i.e. 25–30 ng mL-1). The

extracted DNA was amplified through PCR using the universal primers *rbcL*-Fwd1 and *rbcL*-Rev599 through the single locus marker of *rbcL* partial gene sequence data. The single fragment was purified and sequenced. Specific single fragment was obtained for each sample, size and reading sequence were determined by Macrogen analysis of molecular sizes;: 550bp, 551bp, 566bp, 556bp, 536bp, 548bp, 498bp, 544bp and 531bp with Electra, Giza-10, Giza-11, Giza-12, Sakha-3, Sakha-5, Sakha-6, Strain-562, Blanka, respectively, Figure (1).

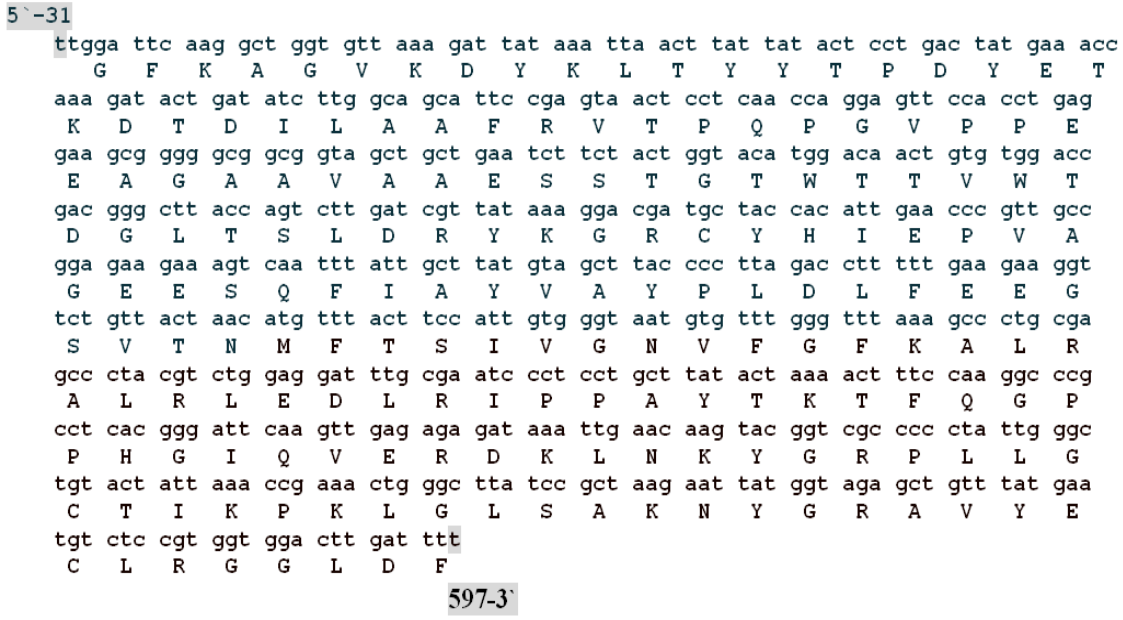

Fig. 1: Consensus sequence of the 9 *rbcL* partial gene ORF region in the nine flax cultivated genotypes from ATG position 31 to 597bp and deduced amino acid sequence. Nucleotide sequence and deduced sequence of amino acid residues of *rbcL* partial gene are indicated by a single letter code used by https://web.expasy.org/translate/.

Each nucleotide sequence was aligned individually at BLAST to confirm each species identity and to get other related sequences and multiple sequence alignments were done using Clustal Omega program online for the nine sequences for consensus length about 566bp as shown in Figure (2). Multiple alignments and phylogenetic relationship tree for the nine flax genotypes showed that consensus was closely related and shared high identity and similarity (100%) to 7 accessions no. MG247942.1 (*Linum bienne*), KM360855.1 (*Linum bienne*), JX664057.1 (*Linum usitatissimum*), JN892708.1 (*Linum perenne*), MG946893.1 (*Linum usitatissimum*), HM544063.1 (*Linum bienne*), KU645802.1 (*Linum usitatissimum*) and other 13 accessions less than 100% - 94.92% with E-value = Zero as they were grouped in one main cluster (Section 1 *Linum sect. Linum*) which illustrated the orthologous variants that occurred between the species (Figure 3).

The present investigation is basically designed to assess the molecular phylogenetic relationship analysis between consensus *rbcL* partial gene sequence of the nine genotypes and 54 accessions of flax germplasm selected from NCBI based on nucleotide sequence analysis of the *rbcL* region. First of all after subsequent cleaning of the *rbcL* sequence, we performed homology searches against the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/), using BLASTn, which showed 100% to 94.92% similarity with the previously sequenced of *rbcL* regions, providing a proof for the good trust worthiness of *rbcL* regions sequenced as mentioned in Table (1). Species identification based on *rbcL* chloroplast region agreed with Kress *et al*., (2005); Chase *et al*., (2005); Bell *et al*., (2017) and Maloukh *et al*., (2017).

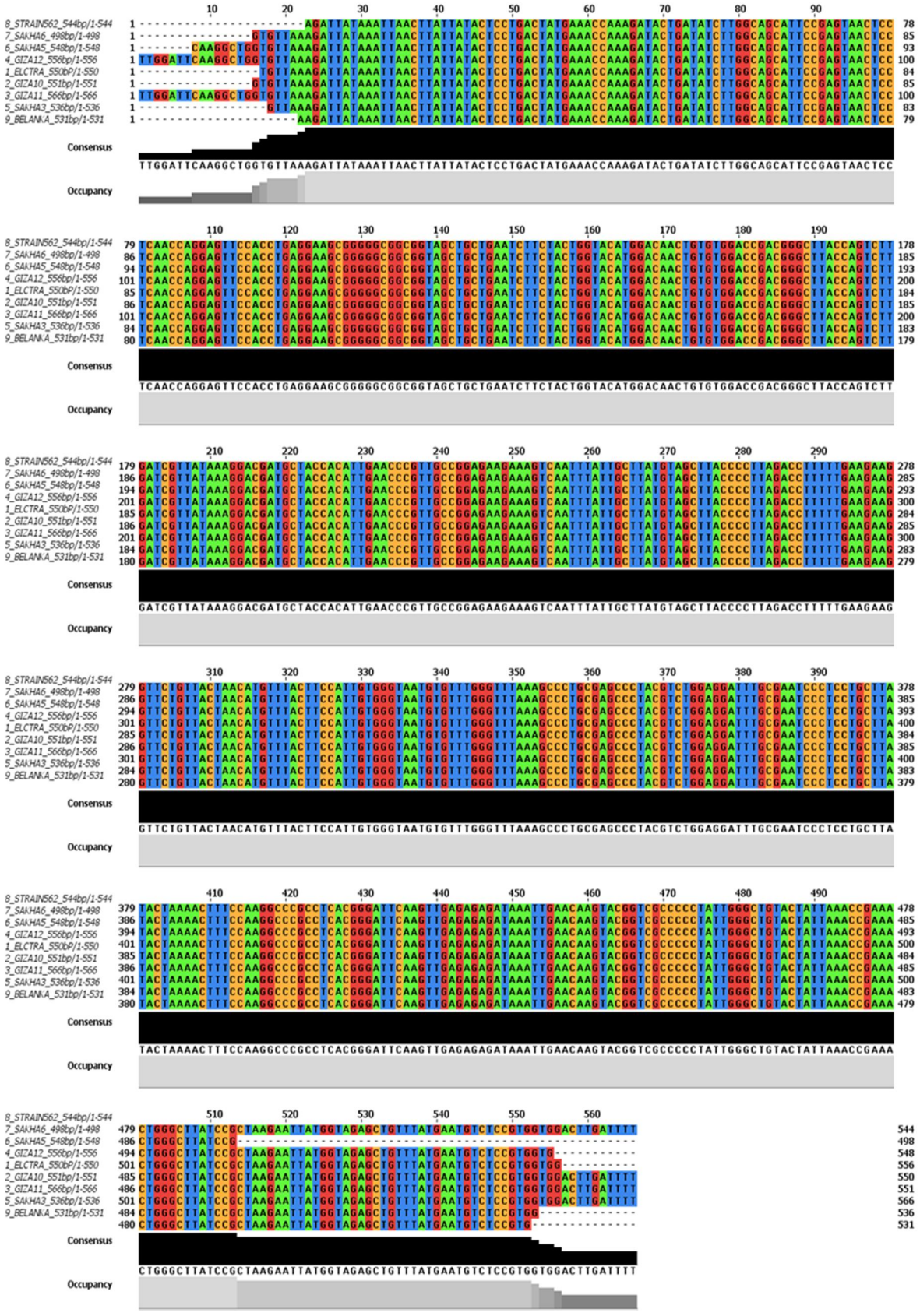

Fig. 2: Multiple sequence alignment of nucleotide sequence of *rbcL* partial gene among the nine flax cultivated genotypes (*Linum usitatissimum* L.) and consensus (566bp). Software packages, https://www.ebi.ac.uk/Tools/msa/clustalo/, were used to do multiple sequences alignment of nucleotide sequences.

Table 1: Accession names, number and similarity % of *linum* species and their related nine cultivated flax genotypes (*Linum usitatissimum* L.) *rbcL* partial gene nucleotide sequenced, BLASTn hits against GenBank database, similarity score and GenBank accession numbers with E-value = Zero. (Code 1 to 20 (Section 1 *Linum* sect. *Linum*), 21 to 25 (Section 2 *Linum* sect. *Dasylinum*), 26 to 30 (Section 3 *Linum* sect. *Linopsis*), 31 to 35 (Section 4 *Linum* sect. *Cathartolinum*), Code 36 to 40 (Section 5 *Linum* sect. *Syllium*), 41 to 42 (Genera *Cliococca*), 43 to 44 (Genera *Hesperolinon*), 45 to 47 (Genera *Radiola*), 48 to 49 (Genera *Sclerolinon*), and 50 to 54 (subfamily II. *Hugonioideae* (*Hugonia*) out-group).

Code	Scientific name (rbcL gene length bp)	GenBank Accession No.	Similarity %	Code	Scientific name (*rbcL* gene length bp)	GenBank Accession No.	Similarity %
---	Current study cons. of 9 sequence (566bp)		---	28-	*Linum tenuifolium* (853bp)	HE963540.1	97.34%
1-	*Linum bienne* (553bp)	MG247942.1	100.00%	29-	*L. suffruticosum* (1425bp)	FJ169591.1	97.35%
2-	*Linum bienne* (1408bp)	KM360855.1	100.00%	30-	*L. tenuifolium* (1428bp)	FJ169594.1	97.17%
3-	*Linum usitatissimum* (1425bp)	JX664057.1	100.00%	31-	*Linum catharticum* (552bp)	MG246795.1	97.26%
4-	*Linum perenne* (502bp)	JN892708.1	100.00%	32-	*L. catharticum* (1428bp)	FJ169570.1	97.17%
5-	*L. usitatissimum* (1419bp)	MG946893.1	100.00%	33-	*L. nodiflorum* (552bp)	MG247581.1	97.26%
6-	*Linum bienne* (1280bp)	HM544063.1	100.00%	34-	*Linum catharticum* (552bp)	MG248295.1	97.26%
7-	*L. usitatissimum* (1428bp)	KU645802.1	100.00%	35-	*L. catharticum* (1293bp)	HM544064.1	97.22%
8-	*L. decumbens* (1428bp)	FJ169573.1	98.76%	36-	*Linum flavum* (1293bp)	HM544065.1	97.41%
9-	*L. monogynum* (1274bp)	HM544069.1	99.07%	37-	*Linum flavum* (1428bp)	FJ169574.1	97.35%
10-	*L. usitatissimum* (1428bp)	FJ169596.1	99.12%	38-	*Linum arboreum* (1354bp)	AY380351.1	97.17%
11-	*Linum bienne* (1415bp)	FJ169568.1	99.12%	39-	*Linum arboretum* (1428bp)	FJ169565.1	97.17%
12-	*L. narbonense* (1428bp)	FJ169580.1	98.76%	40-	*Linum arboretum* (1293bp)	HM544062.1	97.59%
13-	*L. narbonense* (1293bp)	HM544070.1	98.70%	41-	*Cliococca selaginoides* (1415bp)	FJ169558.1	96.82%
14-	*L. grandiflorum* (1428bp)	FJ169575.1	98.76%	42-	*L. striatum* (552bp)	MG249630.1	96.72%
15-	*L. perenne* (553bp)	MF349746.1	98.18%	43-	*Hesperolinon drymarioides* (1415bp)	FJ169560.1	96.64%
16-	*Linum austriacum* (552bp)	MG246391.1	98.18%	44-	*Hesperolinon disjunctum* (1280bp)	HM544050.1	96.11%
17-	*Linum lewisii* (552bp)	MG247926.1	98.18%	45-	*Radiola linoides* (1415bp)	FJ169598.1	96.64%
18-	*Linum lewisii* (1427bp)	FJ169577.1	97.53%	46-	*Radiola linoides* (1408bp)	KM360952.1	96.64%
19-	*Linum perenne* (1428bp)	FJ169582.1	97.53%	47-	*Radiola linoides* (1293bp)	HM544074.1	96.67%
20-	*Linum alpinum* (1315bp)	KF602153.1	98.17%	48-	*Sclerolinon digynum*(1355bp)	FJ169601.1	94.92%
21-	*L. pubescens* (1341bp)	FJ169585.1	97.95%	49-	*Sclerolinon digynum*(1324bp)	AB233896.1	96.47%
22-	*Linum viscosum* (1228bp)	FJ169584.1	97.00%	50-	*Hugonia castaneifolia* (1225bp)	FJ169562.1	96.82%
23-	*Linum hirsutum* (1249bp)	HM544066.1	97.78%	51-	*Hugonia gabunensis* (1404bp)	FJ169563.1	97.35%
24-	*Linum hirsutum* (1428bp)	FJ169583.1	97.88%	52-	*Durandea pentagyna* (1402bp)	FJ169559.1	96.82%
25-	*L. hypericifolium* (839bp)	HM544067.1	97.78%	53-	*Roucheria calophylla* (1398bp)	FJ169600.1	97.35%
26-	*Linum strictum* (1429bp)	FJ169590.1	97.17%	54-	*Roucheria schomburgkii* (1395bp)	FJ169603.1	96.28%
27-	*Linum volkensii* (1209bp)	FJ169597.1	97.17%				

The phylogenetic analyses among consensus *rbcL* partial gene sequence (566bp) of the nine cultivated flax genotypes (*Linum usitatissimum* L.) and the selected 54 species representing all the eight genera of subfamily *Linoideae* ; (Code 2 to 21 (Section 1 *Linum* sect. *Linum*), 22 to 26 (Section 2 *Linum* sect. *Dasylinum*), 27 to 31(Section 3 *Linum* sect. *Linopsis*), 32 to 36 (Section 4 *Linum* sect. *Cathartolinum*), Code 37 to 41 (Section 5 *Linum* sect. *Syllium*), 42 to 43(Genera *Cliococca*), 44 to 45 (Genera *Hesperolinon*), 46 to 48 (Genera *Radiola*), 49 to 50 (Genera *Sclerolinon*), and 51 to 55 (subfamily II. *Hugonioideae* (Hugonia) out-group). They were analyzed with data from the chloroplast *rbcL* partial gene sequences that are more closely related to L*inum* (*Linum usitatissimum* L.) as shown in Figure 3.

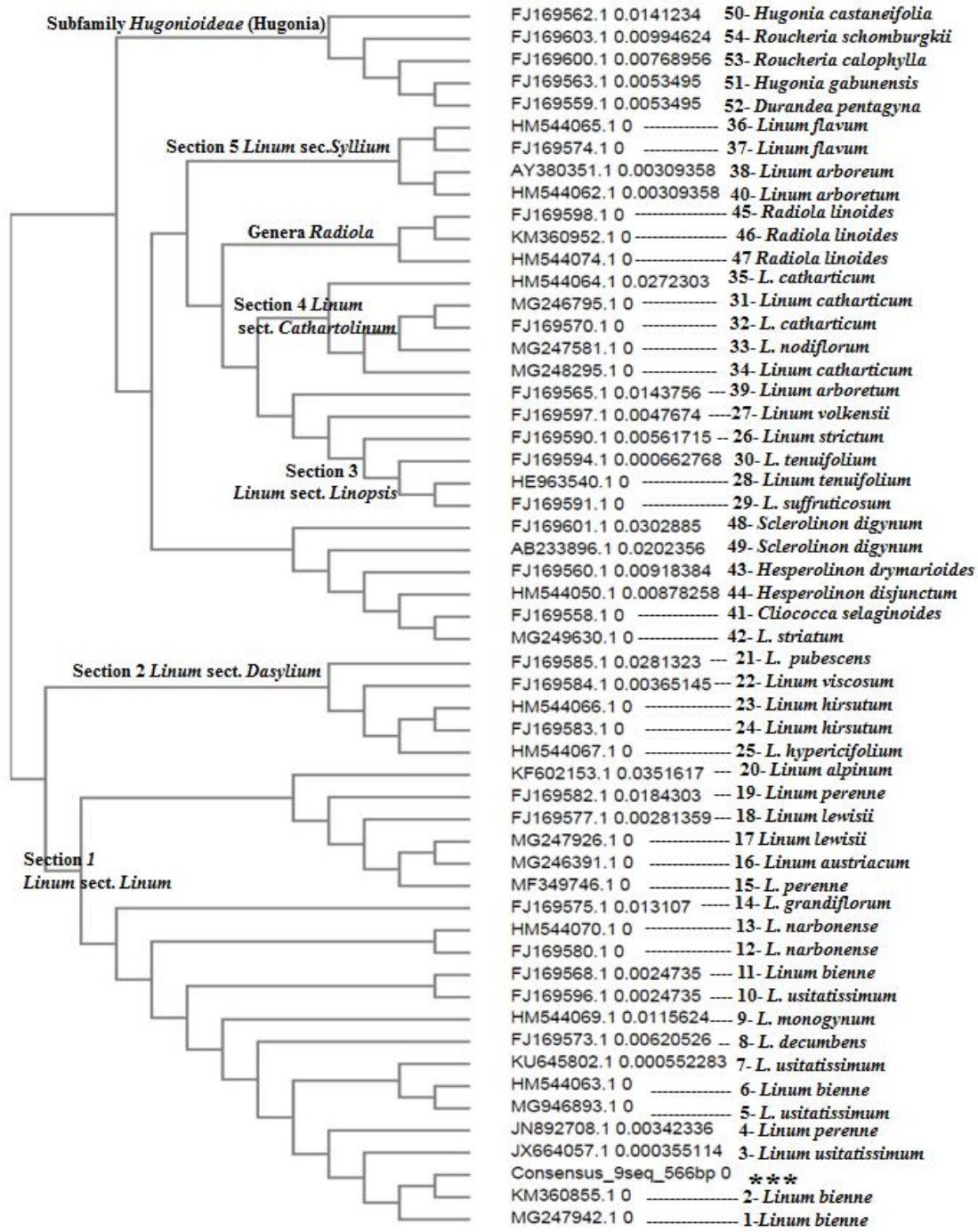

Fig. 3: The phylogenetic relationships of 54 genotypes of the flax germplasm based on nucleotide sequence of *rbcL* genes and consensus *rbcL* genes partial nucleotide sequences of the nine cultured flax genotypes (*Linum usitatissimum* L.). Software packages, (https://www.ebi.ac.uk/Tools/msa/clustalo/) were used to do multiple sequences alignment of nucleotide sequences and generate the phylogenetic tree. (***) The consensus 9 flax sequences (566bp) of current study.

Our analysis was in agreement with McDill *et al*., (2009); McDill and Simson, (2011); Sahare and Srinivasu, (2012); Vohra and Khera, (2013); Techen, *et al*., (2014) and Mishra *et al*., (2016). The identification of the seven Egyptian flax cultivars and the other two foreign flax cultivars in this study for the genus *Linum*, family Linaceae at the molecular level, using *rbcL* partial regions through *rbcL*

reference sequences in NCBI database were discriminated and confirmed. Our result of phylogenetic analysis was agreed with McDill *et al*., (2009), who analyzed sequences of *rbcL* from 48 species of *Linaceae.* The rest of the subfamily forms two major lineages: a blue-flowered clade (*Linum* sections *Linum* and *Dasylinum*) and a yellow-flowered clade (*Linum sects. Linopsis, Syllinum,* and *Cathartolinum,* and the genera *Cliococca, Hesperolinon, Radiola,* and *Sclerolinon*). Phylogenetic analyses of 44 species representing all eight genera of subfamily *Linoideae* and 37 species of *Linum*, with data from the chloroplast (*ndh*F, *trn*L-F, *trn*K 3′ intron) and the nuclear ITS, with *Hugonia* (*Linaceae* subf. *Hugonioideae*) as out-group were presented. McDill and Simson, (2011) found that *Linaceae* and both subfamilies, the temperate *Linoideae* and the tropical *Hugonioideae*, are monophyletic in likelihood- and parsimony-based analyses, although the monophyly of *Hugonioideae* was not well supported.

Flax (*Linum usitatissimum* L.) is an ancient crop that grown commercially for several purposes as a source of fiber, oil and medicinally relevant compounds. DNA barcoding is a power technique for the quick identification and phylogenetic relationship analysis whereby a standardized region of DNA is used for the identification of a species or a taxonomic group of organisms. In plants, DNA barcoding good technique for phylogenetic analysis, classification into wild and cultivated genotypes. It is able to discriminate the closely related species and identify new cryptic species as well. For the success of DNA barcode, the barcode loci must have sufficient information to differentiate unambiguously between closely related plant species and discover new cryptic species. For flax plant species identification, *rbcL* gene sequence region, have been used as successful DNA barcodes. Further, development of single or multilocus marker strategy and its application has provided new vistas to the DNA barcode-based plant identification for flax seeds varieties market. This result using *rbcL* partial gene as a single locus marker could be helpful in the identification, authentication, and to monitor the adulteration of seed varieties under national and international trade.

References

APG II. (The Angiosperm Phylogeny Group) 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society, 141: 399–436. doi: 10.1046/j.1095-8339.2003.t01-1-00158.x

APG III. (The Angiosperm Phylogeny Group) 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105–121. doi:10.1111/j.1095-8339.2009.00996.x.

APG IV. (The Angiosperm Phylogeny Group) 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society, 181, 1-20. https://academic.oup.com/botlinnean/article-abstract/181/1/1/2416499

Bafeel, S.O., I.A. Arif, M.A. Bakir, A.A. Al Homaidan, A.H. Al Farhan and H.A. Khan, 2012. DNA barcoding of arid wild plants using *rbcL* gene sequences. Genetics and Molecular Research 11 (3): 1934-1941.

Bell, K.L., V.M. Loeffer and B.J. Brosi, 2017. An *rbcL* reference library to aid the identification of plant species mixtures by DNA metabarcoding. Application in Plant Science, 5(3): 1600110; http://www.bioone.org/loi/apps.

Bousquet, J., S.H. Strauss, A.H. Doerksen and R.A. Price, 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proc. Natl Acad. Sci. USA, 89: 7844–7848.

Burgess, K.S., A.J. Fazekas, P.R. Kesankurti, S.W. graham, B.C. Husband, S.G. Newmaster, D.M. Percy, M. Hajibabaei and S.C.H. Barrett, 2011. Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. Methods in Ecology and Evolution, (2): 333–340, doi: 10.1111/j.2041-210X.2011.00092.x.

CBOL (The Consortium for the Barcode of Life Plant) Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. USA 106: 12794-12967.

Chase, M.W., D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les, B.D. Mishler and *et al.*, 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Annals of the Missouri Botanical Garden, 80: 528–580.

Chase, M.W., N. Salamin, M. Wilkinson, J.M. Dunwell, R.P. Kesanakurthi, N. Haider and V. Savolainen, 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans., Ser. B 360: 1889–1895.

Chase, M.W., R.S. Cowan, P.M. Hollingsworth, C. Van den Berg, S. Madriñán and *et al*., 2007. A proposal for a standardised protocol to barcode all land plants. Taxon, 56(2): 295–299.

Chase, M.W., S. Zmarzty, M.D. Lledó, K.J. Wurdack, S.M. Swensen and M.F. Fay, 2002. When in doubt, put in *Flacourtiaceae*: a molecular phylogenetic analysis based on plastid *rbcL* DNA sequences. Kew Bulletin 57: 141–181.

China Plant BOL Group, Li, D.Z., L.M. Gao, H.T. Li and *et al*., 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. PNAS, (108), 49: 19641–19646.

Davis, C.C., C.O. Webb, K.J. Wurdack, C.A. Jaramillo and M.J. Donoghue, 2005. Explosive radiation of *Malpighiales* supports a mid-Cretaceous origin of modern tropical rain forests. The American Naturalist 165: E36–E65.

De Vere, N., T.C.G. Rich, S.A. Trinder and C. Long, 2015. In: DNA barcoding for plants. Methods in Molecular Biology*,* 1245: 101-118. DOI https://doi.org/10.1007/978-1-4939-1966-6

Gaut, B.S., S.V. Muse, W.D. Clark and M.T. Clegg, 1992. Relative rates of nucleotide substitutions at the *rbcL* locus of monocotyledonous plants. J. Mol. Evol., 35: 292–303.

Gielly, L. and P. Taberlet, 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. Mol. Biol. Evol., 11: 769–777.

Gill, K.S., 1987. In: Flax. Indian Council of Agricultural Research, New Delhi, India. 386 p.

Hebert, P.D., A. Cywinska, S.L. Ball and J.R. de Waard, 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci., 270:313–321.

Hebert, P.D.N. and T.R. Gregory, 2005. The promise of DNA barcoding for taxonomy. Syst. Biol., 54: 852–859.

Hickey, M. and C. King, 1988. In: 100 families of flowering plants. 2nd edition. University Press, Cambridge. Page 306.

Kress, W.J., 2017. Plant DNA barcodes: Applications today and in the future: Plant DNA barcode applications. Journal of Systematics and Evolution. 55(4): 291–307. doi: 10.1111/jse.12254.

Kress, W.J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen, 2005. Use of DNA barcodes to identify flowering plants. PNAS, 102(23): 8369-8374; https://doi.org/ 10.1073/pnas.0503123102.

Levin, R.A., W.L. Wagner, P.C. Hoch, M. Nepokroeff, J.C. Pires, E.A. Zimmer and K.J. Sytsma, 2003. Family-level relationships of *Onagraceae* based on chloroplast *rbcL* and *ndhF* data. American Journal of Botany, 90: 107-115.

Maloukh, L., A. Kumarappan, M. Jarrar, J. Salehi, H. El-wakil and T.V.R. Lakshmi, 2017. Discriminatory power of *rbcL* barcode locus for authentication of some of United Arab Emirates (UAE) native plants. 3 Biotech, (7): 144. DOI 10.1007/s13205-017-0746-1.

McDill, J., M. Repplinger, B.B. Simpson and W.J. Kadereit, 2009. The Phylogeny of *Linum* and *Linaceae* Subfamily *Linoideae*, with implications for their systematics, biogeography, and evolution of heterostyly. Systematic Botany, 34(2): 386–405.

McDill, J.R. and B.B. Simpson, 2011. Molecular phylogenetics of *Linaceae* with complete generic sampling and data from two plastid genes. Botanical Journal of the Linnean Society, 165: 64–83.

Mishra, P., A. Kumar, A. Nagireddy, D.N. Mani, A, K. Shukla, R. Tiwari and V. Sundaresan, 2016. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnology Journal, 14(1): 8-21. doi: 10.1111/pbi.12419.

Morgan, D.R. and D.E. Soltis, 1993. Phylogenetic relationships among members of *Saxifragaceae* sensu *lato* based on *rbcL* sequence data. Ann. Mo. Bot. Gard., 80: 631–660.

Newmaster, S.G., A.J. Fazekas and S. Ragupathy, 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. Canad. J. Bot., 84: 335–341.

Ockendon, D.J. and S.M. Walters, 1968. *Linum* L. In: Tutin T.G., Heywood V. H., Burges N.A., Moore D.M., Valentine D.H., Walters S.M. & Webb D.A. (eds), Flora Europaea, Cambridge University Press, Cambridge, 2: 206–211.

Olmstead, R.G., H.J. Michaels, K.M. Scott and J.D. Palmer, 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. Annals of the Missouri Botanical Garden 79: 249–265.

Power, K., N, Saarinen, J. Chen and L. Thompson, 2004. Interactive effects of flaxseed and soy in human breast cancer. In: Proc 60th Flax Institute of the United States, Fargo, ND: 80–85

Przybylski, R., 2001. Flax oil and high linolenic oils. In Bailey's industrial oil fat products. 6th edition. Edited by Shahidi F. Hoboken: John Wiley and Sons. Inc., 281-301.

Rogers, C.M., 1982. The systematics of *Linum s*ect. *Linopsis* (*Linaceae*). Plant Systematics and Evolution
140: 225-234.

Sahare, P. and T. Srinivasu, 2012. Barcoding for authentic identification of medicinal plants. Int. J. Eng. Sci., 1: 33–36.

Schori, M. and A.M. Showalter, 2011. DNA barcoding as a means for identifying medicinal plants of Pakistan. Pak. J. Bot., 43: 1-4.

Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson and D.G. Higgins, 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7: 539, DOI: 10.1038/msb.2011.75.

Soltis, D.E., P.S. Soltis, M.W. Chase, M.E. Mort, D.C. Albach, M. Zanis and *et al*., 2000. Angiosperm phylogeny inferred from 18S, *rbcL*, and *atpB* sequences. Botanical Journal of the Linnean Society 133: 381–461.

Taberlet, P., L. Gielly, G. Pautou and J. Bouvet, 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol., 17: 1105–1109.

Techen, N., I. Parveen, Z. Pan and A.I. Khan, 2014. DNA barcoding of medicinal plant material for identification. Current Opinion in Biotechnology, 25: 103-110.

Thompson, J.D., D.G. Higgins and T.J. Gibson, 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22(22): 4673–4680.

Thompson, L., J. Chen, E. Hui, J. Mann and T. Ip, 2004. Interactive effects of flaxseed and tamoxifen in human breast cancer. In: Proc 60th Flax Institute of the United States, Fargo, ND: 86–90.

Van Zeist, W. and J.A.H. Bakker-Heeres, 1975. Evidence for linseed cultivation before 6000 BC. J. Archaeological Sci., 2: 215-219.

Vohra, P. and K.S. Khera, 2013. DNA barcoding: current advance and future prospects - a review. AJBLS, 3: 185–189.

Westcott, N.A. and A.S. Muir, 2003. Flax seed lignin in disease prevention and health promotion. Phytochemistry Rev., 2(3): 401-417.

Wurdack, K.J. and C.C. Davis, 2009. Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. American Journal of Botany 96: 1551–1570.