Full Length Research Paper

Genetic diversity assessment of *Diplocyclos palmatus* (L.) C. Jeffrey from India using internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA

M. Ajmal Ali[1]*, Fahad M. A. Al-Hemaid[1], Joongku Lee[2], R. K. Choudhary[2], Naif A. Al-Harbi[1,3] and Soo-Yong Kim[2]

[1]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh- 11451, Saudi Arabia.
[2]International Biological Material Research Centre, Korea Research Institute of Bioscience and Biotechnology, Daejeon- 305-806, South Korea.
[3]Addiriyah Chair for Environmental Studies, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

Internal transcribed spacer (ITS) region of nuclear ribosomal DNA from 20 populations of *Diplocyclos palmatus* (L.) C. Jeffrey belonging to five different geographical locations (Bihar, Jharkhand, Maharashtra, Madhya Pradesh, and Tamil Nadu) of India were sequenced. Analysis of nucleotide sequences reveals polymorphism among the populations [π = 0.01482, θw = 0.01180 (0.00236) and total variance 3.260]. AMOVA was used to partition the genetic diversity and tested whether there is any hierarchy of ITS sequence variation among individuals. The genetic differentiation between the populations is high (ΦST = 0.591). The maximum likelihood tree reveals the evolution of *D. palmatus* under reproductive isolation and under different environmental conditions which may be most probably due to long distance distribution, and possibility of genetic exchange among the populations of *D. palmatus* distributed in Bihar and Jharkhnad.

Key words: *Diplocyclos palmatus* (L.) C. Jeffrey, genetic diversity, internal transcribed spacer (ITS), nrDNA.

INTRODUCTION

Diplocyclos (Endl.) Post and Kuntze [Family Cucurbitaceae, Tribe Benincaseae (Schaefer and Renner, 2011)] is a small genus of four species; three [*Diplocyclos leiocarpus* (Hook. f.) C. Jeffrey, *Diplocyclos schliebenii* (Harms) C. Jeffrey, *Diplocyclos tenuis* C. Jeffrey] confined to tropical Africa and one [*Diplocyclos palmatus* (L.) C. Jeffrey] extends from tropical Africa to Malaysia (www.tropicos.org). In India the genus *Diplocyclos* is only represented by *D. palmatus* (commonly known as *Sivalingi* or *Pachguria*) which is growing wild on bushes, trees and hedges. The plant is a weak stemmed, branched tendril climber; leaves simple, alternate, 5-lobed, hairy above, pale and smooth beneath; flowers yellow, small, unisexual, male in small fascicles and females solitary; fruits quite conspicuous in the field due to bunches of globose green with white striped (or become bright red with white striped when ripe), smooth with 1 to 2 small seeds. Medicinally it is useful for skin diseases, inflammations and general debility (Chakravarty, 1982; Kirtikar and Basu, 1975; Ali and Pandey, 2007).

---

*Corresponding author. E-mail: majmalali@rediffmail.com. Tel: 966-75834. Fax: 966-75833.

Abbreviation: ITS, Internal transcribed spacer.

The nuclear ribosomal transcription unit (NRTU) is comprised of 18S, 5.8S and 28S genes, two internal transcribed spacers (ITS-1 and ITS-2), and an intergenic spacer (IGS). After transcription, the NRTU is processed to produce mature rRNAs that are key components of cytoplasmic ribosomes. NRTU are found in hundreds to thousands of tandem copies and usually several NRTU clusters are present within plant genomes. The conserved regions (18S and 28S genes) of NRTU are used to infer phylogenetic relationships at higher taxonomy levels, whereas the more rapidly evolving segments (ITS and IGS) are used for studies at the genic or population levels (Soltis and Soltis, 1998; Alvarez and Wendel, 2003). For over a decade, sequences of internal transcribed spacers (ITS) of NRTUs have been widely used to infer phylogenetic relationships, genetic diversity and to unravel evolution in a wide range of complexes in plants (Alvarez and Wendel, 2003; Baldwin and Markos, 1998; Baldwin et al., 1995; Hershkovitz et al., 1999; Kelch and Baldwin, 2003; Lee et al., 2002). Although, NRTUs are found in thousands of copies within a genome, intra-genomic diversity is generally low (Baldwin et al., 1995). This homogeneity among NRTUs is attributed to concerted evolution (Baldwin et al., 1995; Ainouche and Bayer, 1997), a process that acts through gene conversion and unequal crossing over. Despite the fact that homogenization is a norm among NRTUs in a genome, extensive intra-individual and intra-specific variation has been observed in various plant species (Campbell et al., 1997; Hughes et al., 2002). The accumulating evidence suggests that intra-individual variation of nuclear ribosomal ITS regions should not be considered as exceptional (Feliner et al., 2004). Because of the influence of concerted evolution, the occurrence of ancestral polymorphisms is not the most likely ultimate cause for intra-genomic variability in this marker. Instead, a more frequent origin is the merging of different ITS copies within the same genome as a consequence of gene flow. Once the two copies meet, the fate of the polymorphism depends on genetic, reproductive and population-level factors: Specifically, the number and location of ribosomal loci (on the same or different chromosomes), the occurrence of polyploidy and/or apomixes (Hershkovitz et al., 1999; Campbell et al., 1997; Buckler et al., 1997), and the relative abundance of different ITS copies in the breeding populations (Feliner et al., 2004).

D. palmatus shows morphological intermediate among the population. Because of its medicinal properties the plant is being over explored by the local people from the wild. An important prerequisite for development of an effective conservation strategy is the proper evaluation of the distribution and study at the level of genetic variation (Milligan et al., 1994). A perusal of literature reveals that phylogeny and classification of the family Cucurbitaceae have been the focus of several studies (Schaefer and Renner, 2011; Ali et al., 2009; Kocyan et al., 2007; Zhang et al., 2006; Jeffrey, 2005; Decker-Walters et al., 2004; Chung et al., 2003; Jobst et al., 1998); however, the information on population structure and genetic variation of *D. palmatus* is lacking. Hence, the main objectives of the present study were to utilize the nucleotide data of the ITS region to evaluate the degree of differentiation among the populations of *D. palmatus* from India.

MATERIALS AND METHODS

Plant materials

Leaf samples of *D. palmatus* were collected from different geographical region of India. All the collected voucher specimens have been deposited in the Tilka Manhji Bhagalpur University Herbarium (BHAG), Bihar, India. For comparison, the sequence of closely related species *Coccinia grandis* (L.) Viogt. from our earlier study (Ali et al., 2009) was included in the analysis.

DNA extraction

Leaves were dried in silica gel prior to DNA extraction. Total genomic DNA was extracted by following the 2X CTAB method (Doyle and Doyle, 1987) or using the DNeasy Plant Mini kit (QIAGEN Inc., Crawley, West Sussex, UK). Total genomic DNA was extracted from similar amounts of silica dried tissue (~10 to 50 mg dry mass) as well as from herbarium specimens following the cetyltrimethyl ammonium bromide (CTAB) procedure (Doyle and Doyle, 1987). After precipitation with isopropanol and subsequent centrifugation, the DNA pellet was washed with 70% ethanol, dried at 37°C, and resuspended in TRIS-EDTA (TE) buffer.

Amplification of ITS region

ITS sequences of nuclear ribosomal DNA were amplified using primers of White et al. (1990) ITS1F (5'-GTCCACTGAACCTT ATCATTTAG-3') and ITS4R (5'-TCCTCCGCTTATTGATATGC-3') via the polymerase chain reaction (PCR) using the AccuPower HF PCR PreMix (Bioneer, Daejeon, South Korea). One round of amplification consisting of denaturation at 94°C for 5 min followed by 40 cycles of denaturation at 94°C for 1 min, annealing at 49°C for 1 min and extension at 72°C for 1 min, with a final extension step of 72°C for 5 min. The PCR products were purified using SolGent PCR Purification Kit-Ultra (SolGent, Daejeon, South Korea) prior to sequencing.

DNA Sequencing

The purified fragments were directly sequenced using dye terminator chemistry following the manufacturer's protocol. The sequencing reaction was performed in a 10 μl final volume with the BigDye Terminator cycle sequencing kit (Perkin-Elmer, Applied Biosystems). Cycle sequencing was conducted using same primers used in amplification and BigDye vers. 3 reagents and an ABI PRISM 3100 DNA Analyzer (Perkin-Elmer, Applied Biosystems). Cycling conditions included an initial denaturing set at 94°C for 5 min, followed by 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. Sequenced product was precipitated with 17 μl of

Table 1. Sampling location of *Diplocyclos palmatus* and GenBank accession number.

Population	Population code/ Voucher	GenBank accession number
	Dp bgp1	GQ183041
	Dp bgp2	JN834058
	Dp bgp3	JN834059
Bihar	Dp bgp4	JN834060
	Dp kis1	JN834061
	Dp kis2	JN834062
	Dp kis3	JN834063
	Dp pkr1	JN834064
Jharkhand	Dp pkr2	JN834065
	Dp pkr3	JN834066
	Dp klp1	JN834067
Maharashtra	Dp klp2	JN834068
	Dp klp3	JN834069
	Dp bhp1	JN834070
Madhya Pradesh	Dp bhp2	JN834071
	Dp bhp3	JN834072
	Dp tn1	JN834073
Tamil Nadu	Dp tn2	JN834074
	Dp tn3	JN834075
	Dp tn4	JN834076

deionized sterile water, 3 µl of 3 M NaOAc, and 70 µl of 95% EtOH. Polyacrylamide gel electrophoresis was conducted with Long Ranger Single packs (FMC BioProducts) and an ABI 3100 automated DNA sequencer (Perkin-Elmer, Applied Biosystems). Each sample was sequenced in the sense and antisense direction. The sequences were analyzed with ABI sequence navigator software (Perkin-Elmer/Applied Biosystems). The sequencing was done through commercial service of Macrogen Inc. (South Korea). The sequences were analyzed with ABI Sequence Navigator software (Perkin-Elmer/Applied Biosystems). Nucleotide sequences of both DNA strands were obtained and compared the forward and reverse sequence to ensure accuracy.

Data analysis

Sequence alignment

Sequence alignments were performed using ClustalX version 1.81 (Thompson et al., 1997). Sequence alignments were subsequently adjusted manually using BioEdit (Hall, 1999). Insertion-deletions (Indels) were scored as single characters when we had confidence in positional homology. The boundaries between the ITS1, 5.8S, and ITS2 were determined by comparisons with earlier published sequences (Jobst et al., 1998). Gaps were treated as missing data in phylogenetic analyses. All sequences generated in the present study were deposited in GenBank and GenBank accession numbers are included in Table 1.

Sequence diversity

Nucleotide polymorphism, as measured by θw (Watterson, 1975) and diversity, as measured by π (Nei, 1978) were calculated using DnaSP v4.5 (Rozas and Rozas, 1999). Analysis of molecular variance (AMOVA) was performed using GenAlEx 6.1 (Peakall and Smouse, 2006) to assess genotypic variations across all the populations studied. This analysis, apart from partitioning of total genetic variation into within-group and among-group variation components, also provided a measure of intergroup genetic distance as proportion of the total variation residing between populations. The significance of the analysis was tested using 999 random permutations.

Phylogenetic analyses

The phylogenetic analysis of aligned sequences was performed by maximum likelihood (ML) using MEGA5 (Tamura et al., 2007). Phylogenetic analysis inferred using the maximum likelihood method was based on the Tamura-Nei model (Tamura and Nei, 1993). Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model (Kimura, 1980). For estimating ML values, a user-specified topology was used. Substitution pattern and rates were estimated under the

Table 2. Maximum composite likelihood estimate of the pattern of nucleotide substitution.

Nucleotide	A	T	C	G
A	-	*5.53*	*10.45*	**9.77**
T	*5.42*	-	**16.04**	*8.23*
C	*5.42*	**8.48**	-	*8.23*
G	**6.44**	*5.53*	*10.45*	-

Each entry shows the probability of substitution from one base (row) to another base (column) (Tamura et al., 2004). Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics.
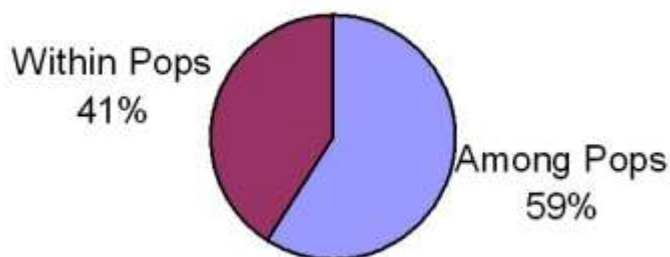
Figure 1. Percentage of ITS variance within and among population of *Diplocyclos palmatus*

Tamura and Nei, (1993) model (+Gamma) (Tamura and Nei, 1993). A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+G]).

RESULTS AND DISCUSSION

Nucleotide sequences and intraspecific divergence

The amplified region of ITS1-5.8S-ITS2 in *D. palmatus* was found 599 base pairs (bp) [ITS1- 216, 5.8S- 164, ITS2- 219]. Outgroup spacer lengths were 593 bp. Data matrix has a total number of 606 characters of which invariable sites were 534 bp, variable sites were 52 (total number of mutations were 53), singleton variable sites were 28 and parsimony informative sites were 24. Insertions and deletions (indels) were necessary to align the sequences. The substitution probabilities are given in Table 2. The nucleotide frequencies were 0.183 (A), 0.187 (T/U), 0.353 (C), and 0.278 (G). The transition/ transversion rate ratios were k_1 = 1.187 (purines) and k_2 = 1.534 (pyrimidines). The overall transition/ transversion bias (R) was found 0.874.

The absence of other variable regions in the nuclear DNA of plants that could provide useful markers at both intra-family (Baldwin and Markos, 1998) and intra-genomic level differentiation (Feliner et al., 2004), makes ITS ostensibly the best marker for phylogenetic studies.

Polymorphism was observed among the populations [π = 0.01482, θw = 0.01180 (0.00236) and total variance 3.260]. Variability within the nuclear ribosomal transcription unit (NRTU) usually depends upon number of gene copies, rates of mutation, concerted evolution, number and chromosomal location of NRTU clusters, and proportion of sexual and asexual reproduction (Dover et al., 1993). Polymorphism may arise when concerted evolution is not fast enough to homogenize repeats in face of high rates of mutation (Appels and Honeycutt, 1986) or by loss of sexual recombination (Campbell et al., 1997).

AMOVA was used to partition the genetic diversity of *D. palmatus* and tested whether there is any hierarchy of ITS sequence variation among individuals (Figures 1 and 2 and Table 3). The genetic differentiation between the populations is high (ΦST = 0.591). {(Nei, 1978) classified GST > 0.15 as high, ΦST and GST both denote fixation index and are comparable)}.

Phylogenetic reconstruction

Phylogenetic analysis inferred using the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993) resulted into the ML tree with the highest log likelihood (-1102.8302) is shown in Figure 3. Substitution pattern and rates estimated under the Kimura (1980) 2-parameter model (Kimura, 1980) shows the nucleotide frequencies A = 20.63%, T/U = 16.65%, C
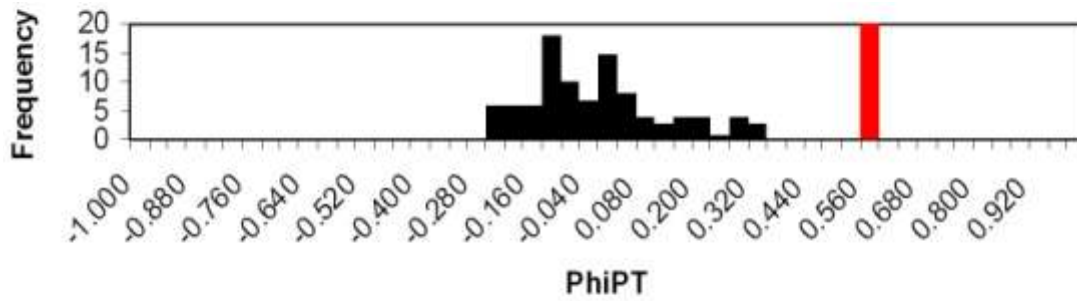
Figure 2. Frequency distribution of random PhiPT vs observed PhiPT among population of *Diplocyclos palmatus*

Table 3. Hierarchical analysis of molecular variance (AMOVA) within/among *D. palmatus* populations.

Source of variation	d.f.	SSD	Estimated variance	Total variance (%)	ΦST	p-value
Among population	4	4.233	0.233	59		
Within population	15	2.417	0.161	41	0.591	0.198
Total	19	6.650	0.394			

d.f.: Degrees of freedom; SSD: Sum of squared deviations; ΦST: Fixation index; p-value: The probability of having a more extreme variance component than the observed.
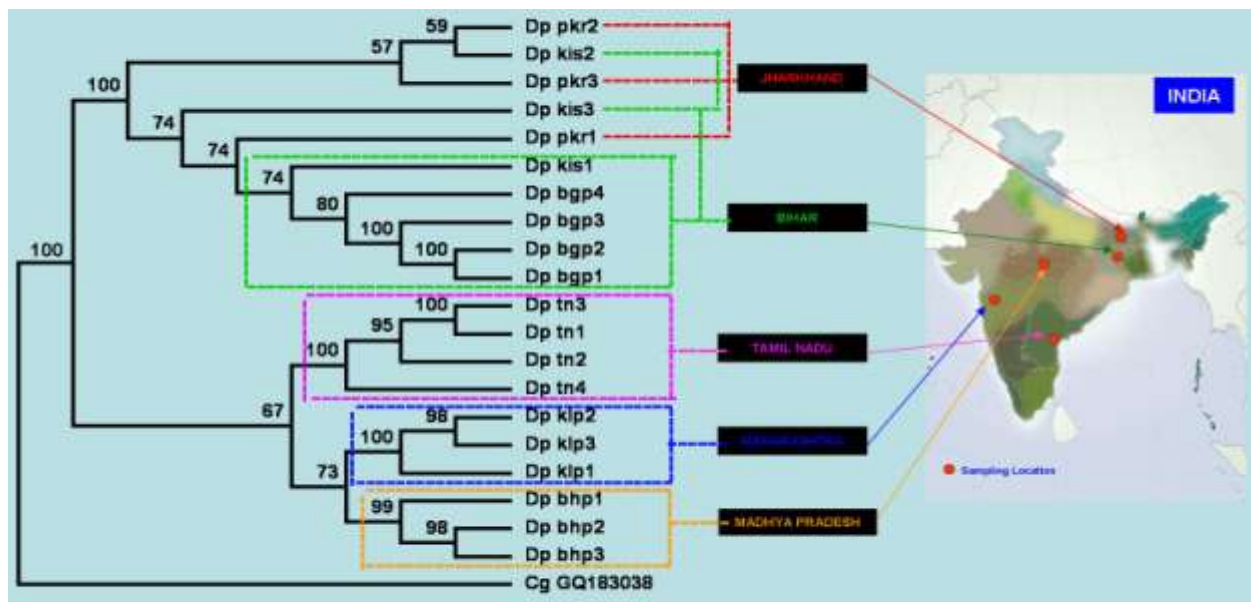

Figure 3. Evolutionary relationships of *Diplocyclos palmatus* inferred using ML method implemented in MEGA5. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Numbers on the branches indicate bootstrap support under 100 bootstrap replicates.

= 35.04%, and G = 27.67%. The maximum Log likelihood for the computation of estimating ML values was -336.940. The estimated value of the shape parameter for the discrete gamma distribution was 200.0. Substitution pattern and rates estimated under the Tamura and Nei, (1993) model (+Gamma) (Tamura and Nei, 1993) reveals mean evolutionary rates 0.90, 0.96, 1.00, 1.04, 1.10 substitutions per site. The nucleotide frequencies was A = 20.63%, T/U = 16.65%, C = 35.04%, and G = 27.67%. The maximum Log likelihood for this computation was –

336.941.

The maximum likelihood tree reveals that all the sampled population of *D. palmatus* grouped together in a single clade (100% bootstrap support). The ML tree shows that the populations of different geographic location sampled in the present study grouped according to their geographic locations. It was interesting to note that population collected from the geographic location of Jharkhand nested within the clade of Bihar. This indicates that the possibility of genetic exchange among the populations distributed in Bihar and Jharkhnad which might have evolved under reproductive isolation and under different environmental conditions.

ACKNOWLEDGEMENTS

REFERENCES

Ali MA, Pandey AK (2007). Cucurbitaceae of eastern Bihar, Plant Systematics Research Centre (TMBU) Publication, Bhagalpur, India.

Ali MA, Pandey AK, Lee J (2009). Relationships among subfamily Cucurbitoideae (Family Cucurbitaceae) from India inferred from ITS sequences of nrDNA. Phytomorphology, 59(3&4): 127-140.

Ainouche ML, Bayer RJ (1997). On the origins of the tetraploid *Bromus* species (section Bromus, Poaceae): insights from the internal transcribed spacer sequences of nuclear ribosomal DNA. Genome, 40: 730-743.

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylo. Evol. 29: 417-434.

Appels R, Honeycutt RL (1986). rDNA: Evolution over a billion years. In: *DNA systematics* edited by S Dutton (CRC Press, Boca Raton, FL, USA), pp. 81-135.

Baldwin BG, Markos S (1998). Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylo. Evol. 10: 449-463.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Miss. Bot. Gard. 82: 247–277.

Buckler ES, Ippolito A, Holtsford TP (1997). The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetic, 145: 821-832.

Campbell CS, Wojciechowski MF, Baldwin BG, Alice LA, Donoghue MJ (1997). Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier agamic* complex (Rosaceae). Mol. Biol. Evol. 14: 81-90.

Chakravarty HL (1982). Fascicles of Flora of India Fascicle II Cucurbitaceae. Botanical Survey of India, Calcutta, India.

Chung SM, Deena S, Walters D, Staub JE (2003). Genetic relationships within the Cucurbitaceae assessed by consensus chloroplast simple sequence repeat (cc SSR) marker. Can. J. Bot. 81: 814-832.

Decker-Walter DS, Chung S, Staub JE (2004). Plastid sequence evolution: a new pattern of nucleotide substitutions in the Cucurbitaceae. J. Mol. Evol. 58: 606-614.

Dover GA, Linares AR, Bowen T, Hancock JM (1993). Detection and quantification of concerted evolution and molecular drive. Methods Enzymol. 224: 525-541.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. pp. 1911-1915.

Feliner GN, Larena BG, Aguilar JF (2004). Fine scale geographic structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). Ann. Bot. 93: 189-200.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid Symp. Ser. 41: 95-98.

Hershkovitz MA, Zimmer EA, Hahn WJ (1999). Ribosomal DNA sequences and angiosperm systematics. In: Hollingsworth PM, Bateman RM, Gornall RJ (eds.) Molecular Systematics and Plant Evolution, Taylor & Francis: London, UK, pp. 268-326.

Hughes CE, Bailey CD, Harris SA (2002). Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: Insights into polyploid origins and nrDNA polymorphism. Am. J. Bot. 89: 1057-1073.

Jeffrey C (2005). A new system of Cucurbitaceae. Bot. Zhurn. 90: 332–335.

Jobst J, King K, Hemleben V (1998). Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among the species of the family Cucurbitaceae. Mol. Phylo. Evol. 9: 204-219.

Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111-120.

Kelch DG, Baldwin BG (2003). Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae-Compositae) based on ITS and ETS rDNA sequence data. Mol. Ecol., 12: 141-151.

Kirtikar K, Basu BD (1975). Indian Medicinal Plants. Vol II, pp. 1106-1115. Bisehn Singh Mahendra Pal Singh, Dehra Dun, India (Rep. ed.).

Kocyan A, Zhang LB, Schaefer H, Renner SS (2007). A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. Mol. Phylo. Evol. 44:553–577.

Lee JY, Mummenho VK, Bowman JL (2002). Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). Proc. Natl. Acad. Sci. USA 99: 16835-16840.

Milligan, BG, Leebens-Mack J, Strand AE (1994). Conservation genetics: beyond the maintenance of marker diversity. Mol. Ecol. 3: 423-435.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetic, 89: 583-590.

Peakall R, Smouse PE (2006). GENALEX 6: Genetic analysis in Excel, population genetic software for teaching and research. Mol. Ecol. Notes, 6: 288-295.

Rozas J, Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinfor. 15:174-175.

Schaefer H, Renner SS (2011) Phylogenetic relationships in the order Cucurbitales and a new classification of the gourd family (Cucurbitaceae). Taxon 60(1): 122-138.

Soltis D, Soltis P (1998). Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis D, Soltis P, Doyle J (eds.) Molecular Systematics of Plants II. DNA Sequencing, Kluwer Academic Publishers: Boston, MA, USA, pp. 1-42.

Tamura K, Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512-526.

Tamura, K., Nei, M. & Kumar, S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings National Academy of Sciences (USA) 101: 11030-11035.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4. Mol. Biol. Evol. 24: 1596-1599.

Thompson, JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins GD (1997). The ClustalX windows interface: flexible strategies for

Multiple sequence alignment aided by quality analysis tools. Nucleic Acid Res. 24: 4876-4882.

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. 7: 256-276.

White TJ, Bruns T, Lee S, Taylor J (1990). Amplifcation and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: A Guide to Methods and Applications, Innis M, Gelfand D, Sninsky J, White T (eds.). Academic Press, San Diego, CA, USA. pp. 315-322.

Zhang LB, Mark PS, Alexander K, Renner SS (2006). Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system evolution. Mol. Phylo. Evol. 39: 305–322.