A phylogenetic analysis of the *Cucurbitaceae*: evidences from *mat*k nucleotide sequences

Uma Reddy B.

Department of Bioinformatics, Gulbarga University, Gulbarga, India, E. mail: drumareddy11@yahoo.co.in, Ph: +91-9449329266

Abstract- MaturaseK gene (*mat*K) of chloroplast is highly conserved in plant systematics which is involved in Group-II intron splicing. The size of the gene is approximately 1500 base pairs in length, located with in the intron of trnK. In the present study, matK gene from Cucurbitaceae was taken for the analysis of variants, parsimony site, patterns, transition/ transversion rate and phylogeny. The family Cucurbitaceae comprises over 120 genera and 825 species with medicinal values. The matk gene sequences have been withdrawn from GenBank and used for the analysis. The multiple sequence alignments were performed by Genebee-ClustalW. The phylogenetic analyses were carried out by MEGA (Molecular Evolutionary Genetics Analysis) software package based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Maximum Parsimony (MP) methods. The results indicated that the Trichosanthes, Austrobryonia, Momordica, Neoalsomitra, Gomphogyne, Sicyos genera showed polyphyletic condition. The overall invariants between the 88 taxa are 28.5%, average composition of nucleotides among taxa are found to be approximately, A-0.30, T-0.37, G-0.15, C-0.17 and the ratio of transition and transversion is 0.824. Phylogenetic tree was designed to identify the ideal regions that could be used for defining the inter and intra-generic relationships. From this study, it could be concluded that, the chloroplast matK gene sequences can unambiguously resolve the relationships, as well as provided a good indication of major supra-generic groupings among the selected members of the family Cucurbitaceae.

Key Words: matK gene, polyphyletic, Cucurbitaceae, Phylogenetic relationships, MEGA

INTRODUCTION

Advancement in molecular biology and DNA techniques has enabled to characterize the genomes of various organisms rapidly. Analyses of the DNA sequences of are providing various species information about their taxonomy, gene makeup and utilizations. The matK gene is one of the most rapidly evolving plastid-coding regions [1]. It is approximately 1550 base pairs (bp) long and encodes a maturase involved in splicing type -II introns from RNA transcripts [2]. The matK gene sequence has been widely employed as a powerful tool in examining inter- and intragenous phylogeny due to its high substitution rates [3 -4]. Recent studies have shown the usefulness of this gene for resolving intergeneric or interspecific relationships among flowering plants. example, Malphigiaceae [5], Poaceae Cornaceae [7], Orchidaceae [8-9].

Cucurbitaceae is one of the largest families of flowering plants, comprising over 120 genera and 825 species [10]. The family is economically very important being the major source of food and forage and its great diversity (ranked as fifth largest family in flowering plants) has also attracted much interest in ecological as well as systematic studies [11]. The main objective to elucidate the systematic positions of members of the family Cucurbitaceae also aimed to generate parallel data sets for representatives of all genera strengthen the morphology based classification.

METHODOLOGY

Data collection: The entire coding region of *mat*K sequences of 88 different species belonging to 54 genera of *Cucurbitaceae* are presented in the table-1. Generic and species information were obtained from taxonomy database of National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nlh.gov/Genbank/index.html).

Sequence analysis: The data analysis was done for the 88 different species belonging to 54 genera of *Cucurbitaceae* for which the sequences are available in GenBank to find the interspecific and intergeneric relationship among the selected genera. Multiple sequence alignment was performed by using Genebee-ClustalW service (http://www.genebee.mus.userviceschitee reduced.htm) to findout the basic sequence statistics including nucleotide frequencies, transition/ transversion (Ts/Tv) ratio and variability in different regions of sequences.

Phylogenetic analysis: The phylogenetic tree was constructed on the basis of *mat*K gene sequences ranging between (1051 to 1393) of 88 taxa of 54 genera belong to the family *Cucurbitaceae*, based on UPGMA and MP methods using MEGA software [12]. Tree statistics included the consistency index (CI) [13], retention index (RC) [14] and rescaled consistency index (RCI) for all sites [14]. Branch lengths (ACCTRAN optimization with equal weights) and the level of support for branches of the phylogenetic trees was evaluated with the bootstrap analysis [15] to verify the strength of the branches based on 100 replicates, using

branch and bound search. Bootstrap percentages are described as high (85-100%), moderate (75 – 84%) or low (50 – 74%) [16]. The number of nucleotide substitutions per site was estimated by Kimura's two parameter method [17] with the DNADIST program.

RESULTS AND DISCUSSION

The study presented here is a preliminary attempt to employ plastid DNA sequence of a few selective genera in Cucurbitaceae to address interfamilial relationships; it is evidenced from the Fig-1. The results of the multiple sequence alignment studies demonstrated that, 695 variable sites and 257 parsimony sites. The overall invariants between the 88 taxa are 28.5%, average nucleotide composition in 88 taxa found to be approximately, A-0.30, T-0.37, G-0.15, C-0.17 and transition and transversion ratio is 0.824. Phylogenetic tree was designed to identify the ideal regions that could be used for defining the inter and intra-generic relationships. Phylogenetic tree of family Cucurbitaceae of 88 taxa were inferred using UPGMA (MP) method. The length of the tree was found to be 18473. The consistency index is 0.161641, the retention index is 0.657458 and the rescaled consistency index for all sites is 0.106272. All positions containing gaps and missing data were eliminated from the dataset. A total of 1000 residues were used from the sequences of 88 selected taxa to construct a phylogenetic tree, out of which 257 are invariants and potentially informative. The family tree obtained from the UPGMA method was split into fourteen well supported clades and also noticed that, the few genera selected namely, Trichosanthes. Austrobryonia. Momordica. Neoalsomitra. Gomphogyne, Sicvos showed polyphyletic condition, indicating that the members of these genera are evolved several times during evolutionary processes.

The UPGMA method of phylogenetic analysis strongly suggests that, the Clade - I consists of two subclades, which are sisterly origin and majority of the taxa present in this clade consists of 1146 nucleotide residues. Subclade-I consists of six species, of which, five belongs to tribe Trichosantheae, namely, Trichosanthes ovigera. T. cucumerina, T. bracteata, T. subvelutina, Hodgsonia heteroclite and are commonly sharing certain features like pollen striate, smooth or verrucose, flower longer, stamens united with triplicate thecae. The Acanthosicyos naudianianus and Cucumeropsis manii belongs and tribe Melothrieae Benincaeseae respectively, and sharing certain common features like tendrils proximally 2 -7 fid or simple, seeds unwinged and reticulated pollen. Where as, the members of subclade-II of clade -I are belongs strictly to the tribe Zanonieae of sub family Zanonioideae, sharing commonly the small

and striate pollen grains, tendrils distally bifid, seeds mostly winged.

In case of clade-II, only *Gomphogyne cissiformis* belongs to the tribe Zanoniae showing the typical character is tendrils distally bifid, seeds mostly winged. Where as, remaining two taxa, *Sicyos hillebrandii* and *S. baderoa* belongs to tribe Sicyeae sharing common features like trichomatous nectary, 4 to 10 – colporate pollen grains, ovules solitary, pendulous, pollen spinulose, polyporate. They are closely related and this relationship is strongly supported by high bootstrap values (98%).

The clade-IV is also divided into 2 subclades, the subclade-II consists of total four taxa, of which Luffa graveolens and L. echinata belongs to tribe Benincaseae, sharing the common features like ovules many, horizontal, pollen reticulate, triporate and their association is strongly supported by very high bootstrap, values, 97%. Where as, Trichosanthes schlechteri and Gymnopetalum chinense belongs to subtribe Trichosanthinae of tribe Trichosantheae, are sisterly originated with the Luffa species, and the relationship among these two tribal members is supported by very low bootstrap values i.e., 44%. These Trichosantheae tribal members sharing commonly pollen striate, smooth or verrucose, flower longer stamens united with triplicate thecae characters.

In case of clade-V, the *Tricyclandra leandril* and Trochomeria macrocarpa belongs to tribe Trichosantheae and Melothrieae respectively, sharing some of the common features like, styles united into a single column, receptacle tube relatively longer, alike in male and female flowers. Other four members of the same clade namely, Austrobryonia centralis, Polyclathra cucumerina. Scopellaria marginata. Schizocarpum palmeri belongs to the subfamily Cucurbitoideae, exhibiting common features are styles united into a single column, ovules many, horizontal or 1 to few ascending pollen, spinose pentaporate.

In clade-VI, out of three species, Sicyosperma gracile and Rytidostylis ciliate are belongs to subtribe Cyclantherinae of tribe Sicyeae and exhibiting strictly some common characters are trichomatous nectary, 4 to 10 colporate pollen grains, ovules solitary, pendulous, pollen spinulose, polyporate and is strongly supported by high bootstrap values 96%. Where as, Nothoalsomitra suberosa belongs to subtribe Benincasinae of tribe Benincaseae and is distantly related with above two species. This is also confirmed by poor bootstrap values only 13%.

Similarly, clade –VII consists of two subclades, subclade-I comprises *Melothrianthus smilacifolius, Trochomeriopsis diversifolia, Psiguria umbrosa* belongs strictly to the tribe Melothrieae and sharing features commonly are

pollens usually reticulate, flowers small, stamens usually free with simple thecae. However, subclade-2 consists of Cogniauxia trilobata and Peponopsis adhaerens belons Benincaseae and Cucurbiteae respectively, and sharing certain features commonly are petals without a fringe or ventral scales, receptacle tube usually relatively short, if long in male flowers then short in female flowers. Clade-VIII consists of total six taxa and are belongs to the tribes Benincaseae, Joliffieae and Melothtrieae of subfamily Cucurbitoideae, sharing common features are styles united into a single column, tendrils proximally 2-7 fid or simple seeds unwinged.

In case of clade-IX, all the members are strictly belongs to the tribe Zanonieae of subfamily Zanonioideae and sharing common characters are small striate pollen grains, tendrils distally bifid, seeds mostly winged. The members of the clade-XII belong to the subfamily Cucurbitoideae and exhibiting commonly styles united into a single column. In clade-XIII, Sechiopsis tetraptera, Sechium edule, Sicyos angulatus, are closely related and strictly belongs to tribe Sicyeae and sharing common features like, trichomatous nectary, 4 to 10 colporate pollen grains, ovules solitary, pendulous, pollen spinulose, polyporate. The another member Selysia prunifera belongs to tribe Cucurbiteae is also originated sisterly with above three Sicyeae tribal members, and their relationship is strongly supported by high bootstrap values (905%) and sharing some of its characters are receptacle tube usually relatively short, if long of male flowers then short in female flowers. Tendrils proximally 2 -7 fid or simple, seeds unwinged.

In case of clade-XIV, consists of total ten species strictly belongs to subfamily Cucurbitoideae, of which Trichosanthes ovigera and T. kirilowii belongs to subtribe Trichosanthinae of tribe Trichosantheae, sharing pollen striate, smooth, flowers larger, stamens united with triplicate thecae. The members of the same clade like Siraitia grosvenorii, Thladiantha davidii. Momordica foetida, M. calantha belongs to tribe Joliffieae sharing some common characters are petals fringed or with ventral scales, in addition to receptacle tube usually relatively short, if long in male flowers then short in female flowers. The members of the Austrobryonia pilbarensis, A. micrantha and A. argillicola exhibiting highest level of similarity (both) morphologically, location and sequence similarity.

In assessing the results presented here, are compared with the most recent system of classification of the family *Cucurbitaceae* as given by Charles Jeffrey as of 1990 [18 - 19]. This new technology affords the opportunity to apply methods of molecular analysis, especially DNA sequencing to Plant systematics.

The *mat*K gene is an ORF (open reading frame) embedded in the intron of the *trn*K gene coding the transfer RNA for lysine (tRNALys^{UUU}). The exon of the *trn*K gene consists of only 70 base nucleotides and is divided into two parts by a long intron of about 2500 bp in length [20]. Based on the phylogenetic analysis, the *mat*K gene showed higher support for individual clades, subclades and better resolution within the family *Cucurbitaceae* than that obtained from *rbc*L [11]. The plastid gene *mat*K has proven useful in addressing Phylogenetic relationships at a variety of taxonomic levels within a number of taxonomic groups.

CONCLUSION

Overall results of the present studies clearly indicated that, the amount of sequence divergence exhibited by the matK is sufficient and appropriate for addressing the evolutionary relationships within the family Cucurbitaceae. Here matK gene sequences provided valuable information for phylogenetic relationships among the different taxa included here. The members clustered into different clades and subclades based on sequence similarity are also strongly evolutionary relationships correlating the exhibited suprageneric at generic, interfamilial levels as the classification made based on the morphological similarities as well as traditional methods in the family Cucurbitaceae. Thereby, it is confirmed that, these molecular analysis further strengthening the traditional method of systematic classification. Because of the few sequences analyzed, the subfamilial relationships within the Cucurbitaceae could not be inferred conclusively that, matK gene is a good candidate for DNA barcoding of Cucrbitaceae family members. It is hoped that, the results of this study and its interpretation will stimulate future research.

REFERENCES

- [1] Wolfe K.H. (1991) Protein coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries and rates of molecular evolution. In K. Vasil [ed.], Cell culture and somatic cell genetics of plants, 7 Bl, Academic Press, San Diego, California, USA, 467-482.
- [2] Wolfe K.H., Morden C.W. and Palmer J.D. (1992) Proc Natl Acad Sci, USA, 89, 10648 -10652.
- [3] Cao H., Sasaki Y., Fushimi H. and Komatsu K. (2001) Biol Pharm Bull, 24, 1389-1394.
- [4] Zhu S., Fushimi H., Cai S. and Komatsu K. (2003) *Planta Med*, 69, 647 653.
- [5] Cameron K.M., Chase M.W., Anderson W.R. and Hillis H.G. (2001) Am J Bot, 88, 1847-1862.

- [6] Liang H.P. and Hilu K.W. (1996) Can J Bot, 74, 125 – 134.
- [7] Xiang Q.Y., Soltis D.E. and Soltis, P.S. (1998) *Am J Bot*, 85, 285-297.
- [8] Goldman D.H., Freudenstein J.V., Kores P.J., Molvray M., Jarrell D.C., Whitten W.M., Cameron K.M., Jansen R.K. and Chase M.W. (2001) Syst Bot, 26, 670-695.
- [9] Salazar G.A., Chase M.W., Arenas M.A.S. and Ingrouille M. (2003). Am J Bot, 90, 777-795.
- [10] Mabberley D.I. (1987) The Plant Book. Cam. Univ Press, Cambridge, New York.
- [11] Uma Reddy B. (2009) Int J Bioinf, ISSN: 0974 6439, 2(1), 54-59.
- [12] Sudhir Kumar., Masatoshi Nei., Joel Dudley. and Koichiro Tamura. (2008) Brief in Bioinf, 9(4), 299-308.
- [13] Kluge A.G. and Farris J.S. (1969) *Syst Bot*, 18, 1-22.
- [14] Farris J.S. (1989) Cladistics, 5, 417 419.

- [15] Felsenstein J. (1985) *Evolution*, 39, 783 791.
- [16] Kenneth Cameron M., Mark Chase W., Mark Whitten W., Paul Kores J., David Jarrell C., Victor Albert A., Tomosia Yukawa., Harold Hillis G. and Douglas Goldman H. (1999) Am J Bot, 86(2), 208 – 224.
- [17] Kimura M. (1980) *J Mol Evol*, 16, 111-120.
- [18] Jeffrey C. (1990) Systematics of Cucurbitaceae: an overview. In D.M. Bates, R.W. Robinson and C. Jeffrey (eds.), Biology and utilization of the Cucurbitaceae, Cornell University Press, Ithaca, New York, USA. 3-9
- [19] Vasishta P.C. (2001) Taxonomy of Angiosperms. Published by G.S. Sharma, Proprietor, R. Chand and Company, 1 – Ansari Road, Daryaganj, New Delhi, 32.
- [20] Sugita M. and Shinozaki K. (1985) Proc Natl Acad Sci, USA, 82, 3557 – 3561.

Table: List of the matK nucleotide sequences for analysis of phylogenetic relationships in the family Cucurbitaceae.

SI. No	GenBank	No of	I	ionships in the family Cucurbitaceae
31. NO	Accession Number	base pairs	TAXA USED	REFFERENCES
01	DQ491025	1219	Helmontia leptantha	Kocyan et al., (2007)
02	DQ491024	1234	Zombitsia lucorum	Kocyan et al., (2007)
03	DQ491023	1206	Tricyclandra landrii	Kocyan et al., (2007)
04	DQ491022	1231	Thladiantha hookeri	Kocyan et al., (2007)
05	DQ491021	1264	Telfairia pedata	Kocyan et al., (2007)
06	DQ491020	1263	Telfairia occidentalis	Kocyan et al., (2007)
07	DQ491019	1232	Momordica charantia	Kocyan et al., (2007)
80	EU155611	1146	Trichosanthes subvelutina	Schaefer et al., (2008)
09	EU155610	1149	Trichosanthes schlechteri	Schaefer et al., (2008)
10	EU155609	1146	Trichosanthes cucumerina	Schaefer et al., (2008)
11	EU155608	1146	Trichosanthes bracteata	Schaefer et al., (2008)
12	EU155607	1146	Hodgsoniaheteroclita	Schaefer et al., (2008)
13	EU155606	1149	Gymnopetalum chinense	Schaefer et al., (2008)
14	EF066331	1128	Ampelosycios scandens	Kocyan <i>et al.</i> , (2007)
15	DQ521608	1251	Ampelosycios humblotii	Kocyan et al., (2007)
16	EU590121	1227	Papuasicyosm papuanus	Schaefer et al., (2009)
17	EU541412	1256	Zehneria anomala	Schaefer et al., (2009)
18	EU436422	1143	Zanonia indica	Schaefer et al., (2009)
19	EU436420	1149	Trochomerina polymorpha	Schaefer et al., (2009)
20	EU436419	1122	Sicyos hillebrandii	Schaefer et al., (2009)
21	EU436418	1149	Puthalicia egladulosa	Schaefer et al., (2009)
22	EU436417	1137	Pseudosicydium acariianthum	Schaefer et al., (2009)
23	EU436416	1194	Neoalsomitra trifoliolata	Schaefer et al., (2009)
24	EU436415	1194	Neoalsomitra stephensiana	Schaefer et al., (2009)
25	EU436414	1047	Neoalsomitra stephensiana	Schaefer et al, (2009)
26	EU436413	1146	Neoalsomitra integrifoliola	Schaefer et al., (2009)
27	EU436412	1146	Neoalsomitra angustipetala	Schaefer et al., (2009)
28	EU436411	1125	Muelleragia jeffreyana	Schaefer et al., (2009)
29	EU436407	1134	Hemsleya aff ambilis	Schaefer et al., (2009)
30	EU436406	1122	Gomphogyne cissiformis	Schaefer et al., (2009)
31	EU436409	1149	Luffa graveolens	Schaefer et al., (2009)
32	EU436408	1149	Luffa echinata	Schaefer et al., (2009)
33	EU436405	1146	Gamphogyne cirromitrata	Schaefer et al., (2009)
34	EU436402	1146	Cucumeropsis manii	Schaefer et al., (2009)
35	EU436401	1152	Corallocarpus triangularis	Schaefer et al., (2009)
36	EU436400	1064	Cogniauxia trilobata	Schaefer et al., (2009)
37	EU436399	1159	Alanthosicyos naudinianus	Schaefer et al., (2009)
38	EU436398	1149	Apodanthera sagittifolia	Schaefer et al., (2009)
39	EU436397	1149	Anangia microsepala	Schaefer et al., (2009)
40	EU037012	1158	Trichosanthes ovigera	Schaefer et al., (2008)
41	EU037005	1393	Trichosanthes villosa	Schaefer et al., (2008)
42	EU037001	1175	Trichosanthes amara	Schaefer et al., (2008)
43	EF487562	1175	Neoachmandra cunnighamii	Schaefer et al., (2008)
44	EF487561	1144	Momordica colchinensis	Schaefer et al., (2008)
45	EF487560	1154	Austrobryonia pilbaren	Schaefer et al., (2008)
46	EF487559	1154	Austrobryonia micrantha	Schaefer et al., (2008)
47	EF487557	1148	Austrobryonia centralis	Schaefer et al., (2008)
48	EF487556	1154	Austrobryonia argillicola	Schaefer et al., (2008)
49	DQ536755	1103	Zygosicyos tripartitus	Kocyan <i>et al.</i> , (2007)
50	DQ536754	1156	Zehneria bodinieri	Kocyan <i>et al.</i> , (2007)
51	DQ536753	1157	Neoachmandra japonica	Kocyan <i>et al.</i> , (2007)
52	DQ536752	1157	Neoachmandra indica	Kocyan <i>et al.</i> , (2007)
53	DQ536751	1145	Scopellaria marginata	Kocyan <i>et al.</i> , (2007)
54	DQ536750	1151	Xerosicyos pubescens	Kocyan <i>et al.</i> , (2007)
∪ ∓	2 0000,00			, (2007)

56	DQ536745	1148	Trochomerina macrocarpa	Kocyan et al., (2007)
57	DQ536740	1051	Thladiantha dubia	Kocyan <i>et al.</i> , (2007)
58	DQ536688	1154	Trichosanthes pubera	Kocyan <i>et al.</i> , (2007)
59	DQ536742	1154	Trichosanthes kirilowii	Kocyan <i>et al.</i> , (2007)
60	DQ536739	1154	Thladiantha davdii	Kocyan <i>et al.</i> , (2007)
61	DQ536737	1119	Solena heterophylla	Kocyan et al., (2007)
62	DQ536736	1154	Siraitia grosvenorii	Kocyan <i>et al.</i> , (2007)
63	DQ536735	1151	Siolmatra brasilinsis	Kocyan et al., (2007)
64	DQ536734	1090	Sicyosperma gracile	Kocyan <i>et al.</i> , (2007)
65	DQ536733	1122	Sicyos baderoa	Kocyan et al., (2007)
66	DQ536732	1130	Sicyos angulatus	Kocyan et al., (2007)
67	DQ536731	1142	Sicydium tamnifolium	Kocyan et al., (2007)
68	DQ536730	1142	Siydium diffusum	Kocyan et al., (2007)
69	DQ536729	1154	Sciana odorifera	Kocyan <i>et al.</i> , (2007)
70	DQ536728	1130	Selysia prunifera	Kocyan et al., (2007)
71	DQ536727	1130	Sechium edule	Kocyan et al., (2007)
72	DQ536726	1130	Sechiopsis tetraptera	Kocyan et al., (2007)
73	DQ536725	1145	Schizocarpum palmeri	Kocyan et al., (2007)
74	DQ536724	1090	Rytidostylis cliata	Kocyan <i>et al.</i> , (2007)
75	DQ536717	1147	Polyclathra cucumerina	Kocyan <i>et al.</i> , (2007)
76	DQ536719	1157	Praecitrullus fistulosus	Kocyan et al., (2007)
77	DQ536721	1150	Psiguria umbrosa	Kocyan <i>et al.</i> , (2007)
78	DQ536722	1121	Pteropepon parodii	Kocyan et al., (2007)
79	DQ536716	1037	Peponopsis adharens	Kocyan et al., (2007)
80	DQ536715	1157	Peponium vogelii	Kocyan <i>et al.</i> , (2007)
81	DQ536714	1157	Peponium caledonicum	Kocyan et al., (2007)
82	DQ536707	1153	Neoalsomitra clavigera	Kocyan et al., (2007)
83	DQ536708	1122	Neoalsomitra schefferiuana	Kocyan et al., (2007)
84	DQ536709	1129	Neoalsomitra suberosa	Kocyan et al., (2007)
85	DQ536703	1154	Momordica foetida	Kocyan et al., (2007)
86	DQ536702	1154	Momordica calantha	Kocyan et al., (2007)
87	DQ536700	1162	Melothrianthus smilacifolius	Kocyan et al., (2007)
88	DQ536698	1157	Marah fabaceus	Kocyan <i>et al.</i> , (2007)

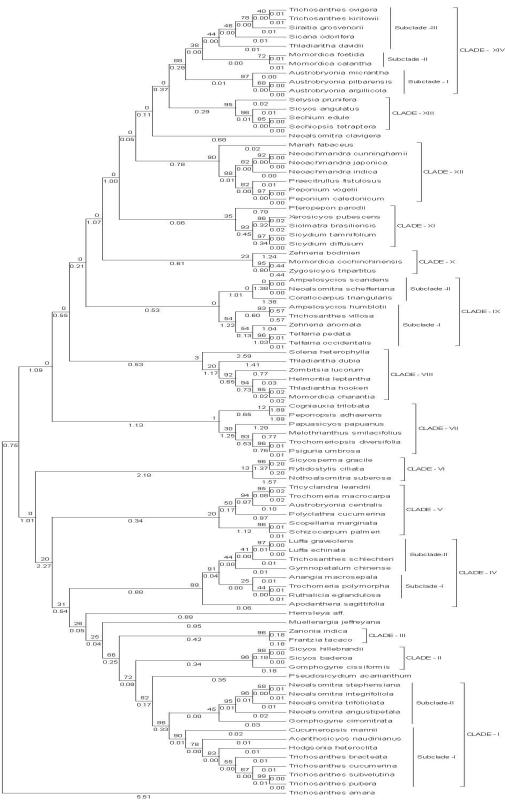


Fig -1 Phylogenetic tree from matK sequences representing 88 taxa of family *Cucurbitaceae*, were inferred using UPGMA method. Bootstrap percentages and branch lengths are indicated above and below the branches, respectively.

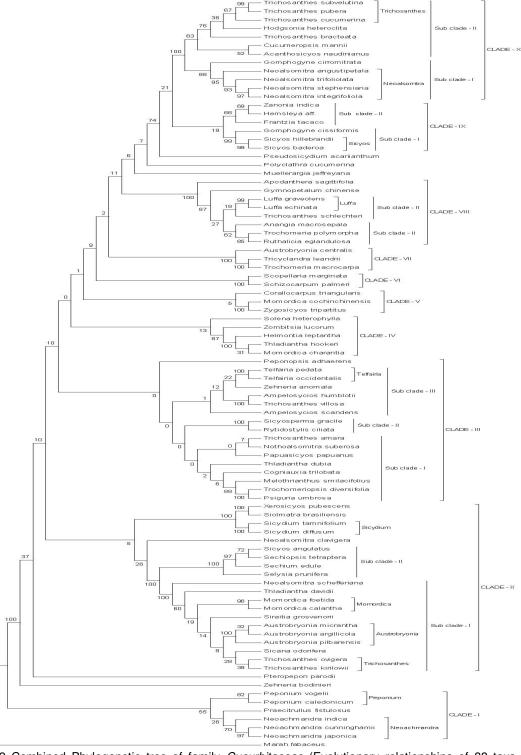


Fig -2 Combined Phylogenetic tree of family *Cucurbitaceae* (Evolutionary relationships of 88 taxa were inferred using Maximum Parsimony method. The length of the tree is 18473 is shown. The consistency index is 0.161641, the retention index is 0.657458 and the rescaled consistency index for all sites is 0.106272, . All positions containing gaps and missing data were eliminated from the dataset. There were a total of 1000 residue were used to construct a phylogenetic tree, out of which 257 are invariants and potentially informative