Pak. J. Bot., 45(SI): 185-190, January, 2013.

PHYLOGENETIC RELATIONSHIP OF PAKISTANI SPECIES OF *CAREX* L. BASED ON *MATK* GENE SEQUENCE VARIATION

ZAHID ULLAH[1,2*], MUSHTAQ AHMAD[1] AND RICHARD I. MILNE[2]

[1]*Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan*
[2]*Institute of Molecular Plant Sciences, University of Edinburgh, UK*
Corresponding author's e-mail: zahidmatta@gmail.com

Abstract

Among the largest and taxonomically complex genera of plants, relationship with in *Carex* at subgeneric and sectional levels is unclear. For this purpose partial DNA sequences of chloroplast matK gene were generated for 11 species of *Carex* L. and two outgroup taxa i-e *Kobresia schoenoides* and *Kobresia laxa* from Pakistan. The sequenced fragments varied from 753 base pairs (bp) to 1360 bp in length, with 15.08% variable and (45) 3.29% parsimony informative sites when the outgroups were included. The aligned sequences were analyzed by maximum parsimony (MP), maximum likelihood (ML), neighbor-joining (NJ) and UPGMA methods using MEGA5. matK was found useful in resolving relationship among subgenera and species of *Carex*. All the species of subgen. *Carex* form a clear clade with 98% boostrap support and taxa of subgen. *Vignea* form a clade with 99% bootstrap support. *C. microglochin* of the Indocarex appeared as sister to the rest of *Carex* clades. Position of *C. filicina* (Primocarex) group cannot be resolved from the present data. *C. canescens* is sister to the rest of species in subgen. *Vignea*, while *C. foliosa* and *C. divulsa* are monophyletic in this group. The out group *Kobresia* do not form a clear sister group to *Carex*, supporting the view that *Kobresia* may be included in *Carex* in the future. In the sub-genus *Carex*, *C. atrofusca* is sister to the rest of the group, while C. *pseudocyperus,* and *C. songorica* forming monophyletic group. Based on these results, delimitation of the species is discussed. The implications of phylogeny for habit preferences and inflorescence structure has also been discussed.

Introduction

The genus *Carex* L. (Cariceae Pax., Cyperaceae Juss.) is among the largest angiosperm genera (Frodin, 2004) with *c.* 2000 species (Reznicek, 1990; Mabberley, 1997) and truly cosmopolitan distribution (Good, 1974) with centres of diversity in the temperate region of northern hemisphere (Reznicek, 1990; Starr *et al*., 2009). In Pakistan it is represented by 62 species (Kukkonen, 2001) covering all the 4 subgenera and predominantly inhabiting northern cold temperate and alpine region of the country. The genus *Carex* is traditionally divided into 4 subgenera (*Carex, Primocarex, Indocarex* and *Vignea*) following Kukenthal (1909) or five subgenera (*Psyllophora* (= *Primocarex*), *Vignea*, *Vigneastra* (= *Indocarex*), *Carex (Eucarex)* and *Kreczetoviczia* (Egorova, 1999) and 70 or more sections. This classification is primarily based on the morphology of inflorescence and position of male flowers in the spikes. This splitting based on inflorescence architecture may not be in line with phylogenetic relationship with in the genus (Yen & Olmstead, 2000). The *Carex* is characterized by unisexual (monoecious) flowers and a utricle or perigynium (bract) that subtends and wholly or partially cover the female flowers (Starr *et al*., 2004). The perigynia encompasses enormous diversity in colour, size, shape and degree of closure and is extensively utilized in species circumscription in *Carex*. In *Primocarex* the inflorescence is a terminal gynecandrous spike (unispicate), while *Vignea* possess bisexual spikes and two stigmas (Yen and Olmstead, 2000). *Indocarex* always has bisexual branched spikes and 3 stigmas, whereas subgenus *Carex* (Eucarex) usually, has unisexual spikes, 2 or 3 stigmas and tubular cladoprophylls (Reznicek, 1990; Yen & Olmstead, 2000).

Carex species are often the dominant component of most ecosystems due to their enormous diversity, ecologically more significant and are often indicators of the habitats. *Carex* form the dominant component in the alpine ecosystems across Pakistan, along with *Kobresia*. Despite of its global importance *Carex* is largely ignored and taxonomically poorly studied due to its reduced morphology (superficial resemblance), problematic taxonomy, greater species diversity in local floras and apparently low economic significance (Starr *et al*., 2009). As a result a large proportion of specimens remain unidentified in herbaria (Starr *et al*., 2009). The application of molecular information, in particular DNA sequencing has contributed significantly in addressing questions in systematics. The genus *Carex* has been the focus of several molecular works, mostly dealing with subgeneric and sectional circumscription (Waterway *et al*., 1997; Starr *et al*., 1999, 2004; Yen & Olmstead, 2000; Roalson *et al*., 2001; Ford & Naczi, 2001; Starr & Ford, 2009) or DNA barcoding (Starr *et al*., 2009; Clerc-Blain *et al*., 2010).

Globally there is an effort towards using molecular data to elucidate phylogenetic relationship of difficult group (Shinwari *et al*., 1994 & 1994a). Various genes were used as a candidate to find support for the classification done on morphoplogical evidences (Shinwari 1995 & 2002).

In this study, we use the chloroplast matK gene to infer the systematic relationship between species and subgenera in Pakistani *Carex* taxa and to correlate the results with morphological circumscription. The matK gene is *c.* 1500-1550 base pairs (bp) located within the trnK intron (Hilu & Liang, 1997). The main function of matK seems to be the splicing of group -II intron (Ems *et al*., 1995) as it is the only ORF in group II intron (Hilu *et al*., 1999). The *matK* is known for high rate of substitutions at all codons positions (Hilu & Liang, 1996; Muller *et al*., 2006) and thus evolves more rapidly than other chloroplast genes. matK also exhibits relatively high insertion/deletions at the 3' end (Hilu & Alice, 1999).

Materials and Methods

Taxon sampling: We partially sequenced matK gene from 11 species of the genus *Carex* L. representing all the four sub-genera and two out group taxa *Kobresia schoenoides* and *Kobresia laxa.* We sampled one species each for subgenera *Primocarex* and *Indocarex*, four species for *Vignea* and five species for *Eucarex* (*Carex*). Vouchers information are given in (Table 1), the sequences are already submitted to the gene bank. Sampling was done from wild populations of the concerned taxa. Voucher specimens were placed in the Herbarium of Pakistan (ISL), QAU Islamabad.

Table 1. *Carex* and *Kobresia* (outgroup) taxa sampled and voucher information.

Taxon	Locality	Voucher specimen
Kobresia schoenoides (C. A. Mey.) Steud.	Deosai	ISL 126966
Kobresia laxa Nees	Kachora, Skardu	ISL126962
Carex songorica Kar. & Kir.	Barapani Deosai	ISL 126938
Carex pseudocyperus Linnaeus	Deosai	ISL 126983
Carex atrofusca Schkuhr	Batakondi Naran	ISL 126974
Carex infuscata Nees	Lalazar Naran	ISL 126949
Carex melanantha C. A. Mey.	Deosai	ISL 126914
Carex canescens Linn.	Naran	ISL 126979
Carex otrubae Podpera	Deosai	ISL 126912
Carex foliosa D. Don	Miandam Swat	ISL 126975
Carex divulsa Stokes	Naran	ISL 126921
Carex filicina Wahlenb.	Miandam Swat	ISL 126970
Carex microglochin Wahlenb.	Deosai	ISL 126981

DNA Isolation: Total genomic DNA was extracted from silica-gel dried leaves of a single individual using the cetyl trimethylammonium bromide (CTAB) method of (Doyle & Doyle, 1987) with 1 µl beta-mercaptoethanol per 1 ml of CTAB, and 2% PVP (polyvinylpyrolidine). Frozen leaves were ground in liquid N_2, incubated for 1h at 65°C in 300µL of 2% CTAB solution with 1% β - mercaptoethanol, and subsequently cleaned with chloroform : isoamyl alcohol (24 : 1). DNA was precipitated with isopropanol, then washed with 80% ethanol, and resuspended in 6µL 1x TE buffer. UV spectrophotometer was used for DNA quantification, and concentrations were adjusted to 50ng/µl for PCR purpose.

PCR Amplification and Sequencing: In order to amplify the whole matK gene two primers MG1-F (**5′** CTACTGCAGAACTAGTCGGATGGAGTAGAT) and MG15-R (**5′** ATCTGGGTTGCTAACTCAATG) (Liang & Hilu, 1996) were used (Fig. 1). Using the amplified *trn*K gene as a template, two primers, S5-1F (**5′** ACCCTGTTCTGACCATATTG) and 9R (**5′** TACGAGCTAAAGTTCTAGC), internal to the *mat*K coding region (Fig. 1) were designed to sequence matK.

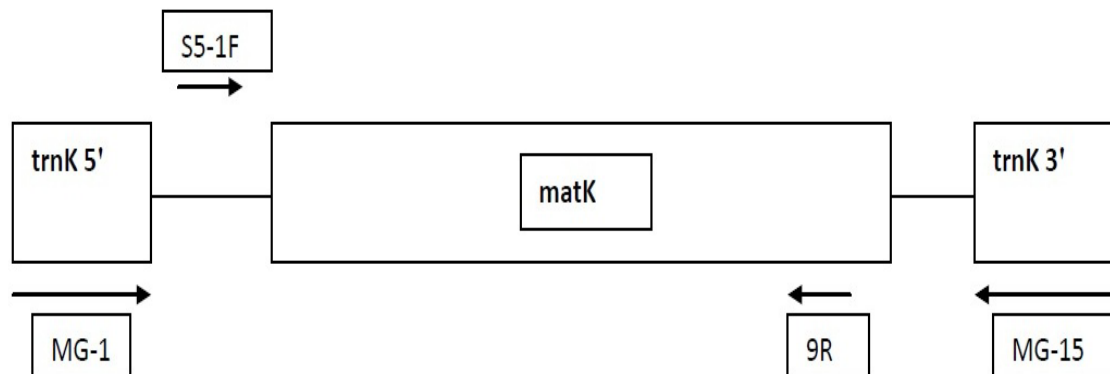

Fig. 1. Diagram showing matK gene nested between trnK introns. The arrows indicate the positions of PCR and Sequencing primers (adapted from Hilu *et al*., 1999).

Each PCR amplification used 25µl PCR reaction mixture, containing 2.5µl 10x PCR buffer (BioLine), 0.2µl of 25mM of each dNTPs, 2µl of 25mM magnesium chloride (BioLine), 0.2µl of 5U/µl Taq DNA Polymerase (BioLine, England) and, 17µl of ddH_2O, 1.25µl of each primer, and 5µl of DNA template. The PCR amplification conditions used were initial denaturation at 94°C for 3 min, followed by 35 cycles of 40 sec at 94°C, primer annealing at 51 to 58°C for 1 minute, DNA strand extension at 72°C for 1min 40 sec ,a free treatment of one cycle at 72°C (5 min) and a termination step of 4°C for 24 min.

About 7ul of each PCR product was run on 0.8% agarose gel in a 0.5X TBE buffer, stained with ethedium bromide. Successful PCR products were purified using QIAquick cleanup kits (QIAGEN) following manufacturers protocol. For sequencing the products were sent to the Sanger GenePool Sequencing Lab, Univeristy of Edinburgh, UK. The sequences were generated according to their standard protocol *(*www.genepool.bio.ed.ac.uk/sanger/protocols.html*) using primers S5-1F and 9R, as mentioned above.*

Sequence Alignment and Phylogenetic analysis: The DNA sequences were aligned with ClustalW v. 1.4 (Thompson *et al*., 1994) using the multiple sequence alignment option in Bio Edit program (Hall, 1999). Gaps were considered as missing information and a 5th character. Two species of *Kobresia* were used as outgroup for this analysis. The phylogenetic analysis was performed using MEGA5 (Tamura *et al*., 2011). Maximum Parsimony (MP) was used to infer evolutionary relationship. The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Felsenstein, 1985) with search level 2. Bootstrap analyses were used to assess the robustness of the trees with 1000 replicates for parsimony analysis. The sequences were also analyzed with Neighbor-Joining (NJ) method using P-distance option in Kimura (1980) 2-parameter method. Codon positions included were 1st+2nd+3rd+Noncoding. UPGMA algorithm (Sneath & Sokal, 1973) has also been used with bootstrap test (1000 replicates) using the p-distance method (Nei & Kumar, 2000).

Results and Discussion

Sampling was designed to include all the genera recognized in the tribe and much of the morphological and geographic variation (Table 1). The open reading frame (ORF) of matK gene studied in the present species ranges from 1360bp in *Carex sanguinea* to 753bp in *Carex otrubae*. In most of the taxa however matK sequences of about 1350bp were generated. Average (G + C) content was -32%. Pairwise divergence of sequences ranged from 15 to 27.5% between the *Kobresia* and *Carex*, and 8.6% within *Carex*. Of the 1366 sequences that were aligned for analysis 206 were variable sites, while 1157 were conserved sites and 45 sites were parsimony informative. The estimated Transition/ Transversion bias (*R*) is 1.85. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model. The nucleotide frequencies are A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a user-specified toplogy was used (Table 2). The maximum Log likelihood for this computation was -1543.446. Codon positions included erest+2nd+3rd+Noncoding.

Table 2. Pattern of nucleotide substitution.

	A	T	C	G
A	-	6.7	2.4	**9.94**
T	4.98	-	**9.39**	2.26
C	4.98	**26.16**	-	2.26
G	**21.83**	6.7	2.4	-

Table 2. Each entry shows the probability of substitution (r) from one base (row) to another base (column)[1]. For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in *italics*. The nucleotide frequencies are 30.45% (A), 40.98% (T/U), 13.86% (C), and 14.71% (G). The transition/transversion rate ratios are $k_1 = 4.387$ (purines) and $k_2 = 3.906$ (pyrimidines). The overall transition/transversion bias is $R = 1.705$, where $R = [A*G*k_1 + T*C*k_2]/[(A+G)*(T+C)]$.

One of the 72 strict concensus trees from parsimony analysis of the combined datamatrix is illustrated in Fig. 1 and 2. In the first PM tree the outgroup *Kobresia schoenoides* appears as sister to the rest of species and clearly rooting the tree. However the second outgroup *Kobresia laxa* is not resolved and forming a clade with *C. filicina* and in Fig. 2 it is nested in the clade which consists of *K. Schoenoides, C. filicina* and *C. microglochin.* The tree shows that the subgenera *Carex* and *Vignea* are clearly resolved, having 97 and 99% boostrap support. Both the genera are monophyletic (Waterway *et al*., 2009), with in the *Vignea C. canescens* seems to be sister to the other 3 species *C. otrubae, C. foliosa and C. divulsa.* The latter 2 are clearly monophyletic. Both these species are closely morphologically related as well. Within the *Carex* clade, *C. infuscata* appears to be paraphylitic and form sister to the other four species with a strong 97% boostrap support. *C. melanantha and C. pseudocyperus* form a monophyletic clade and so are *C. songorica and C. atrofusca.* While the overall topology of trees in this study is similar to that presented by (Waterway & Starr, 2007). Species of this clade possess androgynous spikes and persistent rachillae bearing male flowers, which are primitive characters (Egorova, 1999) and probably adaptation for marshy habitats.

At the subgeneric level, all subgenera, except subgenus *Vignea,* appear polyphyletic in the UPGMA analysis Fig. 4, this has also been demonstrated by (Yen & Olmstead, 2000). Here the outgroup *K. Laxa* form the base on which the tree is rooted. The UPGMA tree forms three clade i-e the *Carex* clade which has a 79% boostrap value, with in this clade C. sonogarica and C. pseudocyperus are clearly monophyletic with (99% boostrap). *C. infuscata* is sister to the rest of the species in the clade. Vignea form the second clade with 84% bootstrap, where C. foliosa and C. divulsa are monophyletic and the other two seems to be paraphylitic. The third clade is formed by the outgroup *K. Schoenoides* and subgenera *Indocarex* and *Primocarex.* All the species in this clade are polyphyletic.

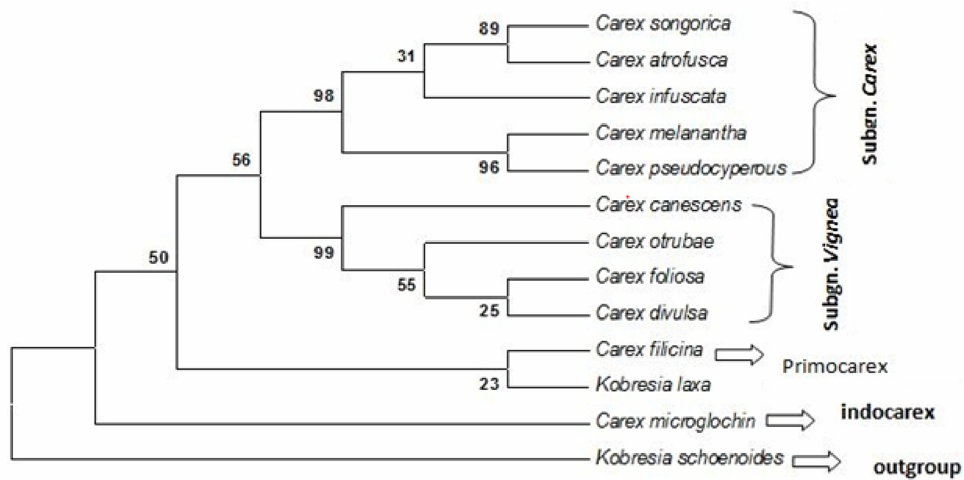

Fig. 2. Strict consensus parsimonius tree. Bootstrap values are given above branches.

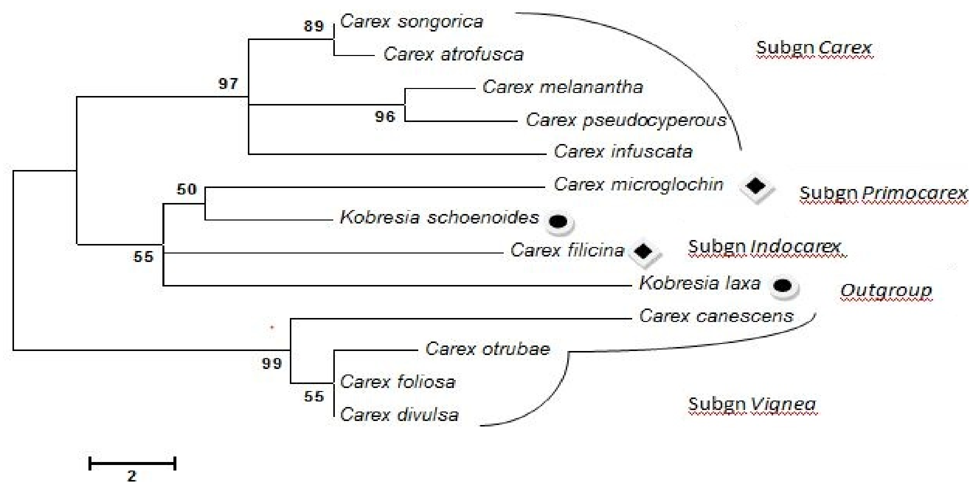

Fig. 3. Maximum parsimony analysis of *Carex* and *Kobresia* species showing the bootstrap consensus tree inferred from 1000 replicates. The MP tree was obtained using the Min-mini heuristic algorithm. Numbers above branches are bootstrap percentages.

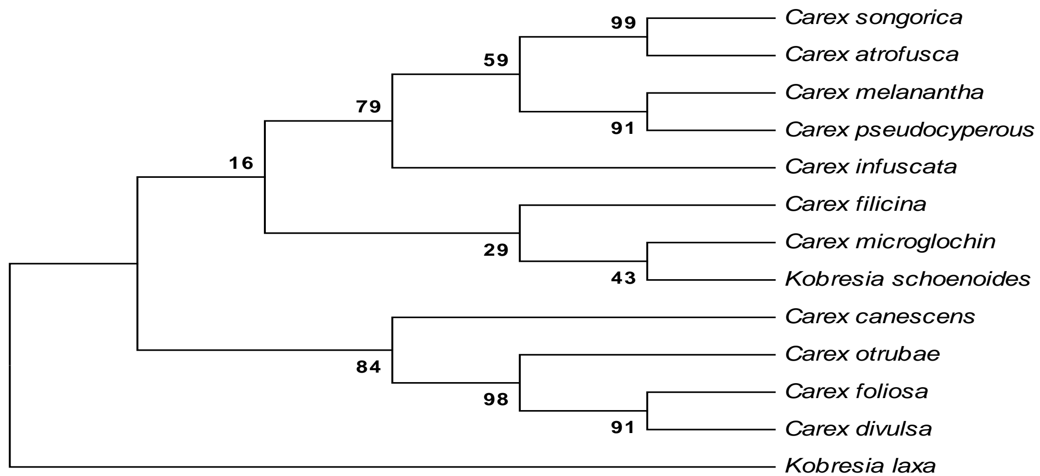

Fig. 4. UPGMA dendrogram of Carex and outgroup taxa. The optimal tree with the sum of branch = 0.09347675 is shown. The evolutionary distances were computed using the p-distance. Bootstrap values are given above branches.

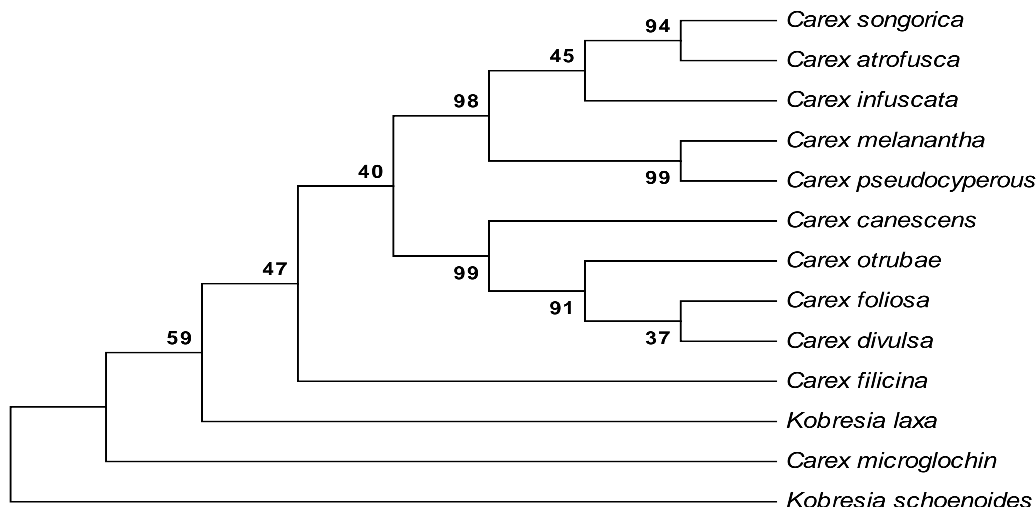

Fig. 5. Neighbor-Joining (NJ), bootstrap consensus tree inferred from 1000 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the p-distance method. Bootstrap values are given above branches.

Results show that all the subgenera form monophyletic clades and thus represent natural groups which share same common pattern of evolution. Our analyses reveal consistent findings with other recent cladistic analyses of *Carex* (Yen & olmstead, 2000; Roalson *et al*., 2001; Hipp *et al*., 2006; Waterway *et al*., 2009). Overall results are in accordance with the morphological circumscription of the taxa studied as reported elsewhere too (Shinwari & Shinwari, 2011).

Conclusion

Molecular data of DNA sequences (matK gene) is useful in addressing the systematic problems and classification and circumscription of subgenera and species in *Carex,* and as such matK is gaining the worldwide recognition of a universal plant barcode in the era ahead.

Acknowledgements

We are grateful to Mr. Adil Hussain (PhD Scholar) at the University of Edinburgh UK for his help in data analysis. This research was supported by Higher Education Commission of Pakistan, which is highly appreciated and acknowledged.

References

Clerc-Blain, J.L., J.R. and R.D. Bull and J.M. Saarela. 2010. A regional approach to plant DNA barcoding provides high species resolution of sedges (Carex and Kobresia, Cyperaceae) in the Canadian Arctic Archipelago. *Mol Ecol Resour*. 10(1): 69-91.

Doyle, J.J. & J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.,* 19: 11-15.

Egorova, T.V. 1999. The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). St. Petersburg State Chemical-Pharmaceutical Academy and Missouri Botanical Garden, St. Louis, USA.

Ems, S.C., C.W. Morden, C.K. Dixon, K.H. Wolef, C.W. Depamphilis and J.D. Palmer. 1995. Transcription, splicing and editing of plastid RNAs inthe nonphotosynthetic plant *Epifagus virginiana*. *Pl. Mol. Biol*., 29: 721-733.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Ford, B.A. and R.F.C. Naczi. 2001. Genetic diversity in the Carex jamesii complex (Cyperaceae: sect. Phyllostachyae) with insights into the evolution and origin of the newly described species Carex timida. *Sida*, 19: 885-897.

Frodin, D.G. 2004. History and concepts of big plant genera. *Taxon*, 53: 753-776.

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

Hilu, K.W. and L.A. Alice. 1999. Evolutionary Implications of matK Indels in Poaceae. *Am. J. Bot.* 86(12): 1735-1741.

Hilu, K.W., L.A. Alice and H. Liang. 1999. Phylogeny of Poaceae inferred from matK sequences. *Ann Missou Bot Gard*., 86(4): 835-851.

Hipp, A.L., A.A. Reznicek, P.E. Rothrock and J.A. Webers. 2006. Phylogeny and Classification of Carex sect. Ovales (Cyperaceae). *Int. J. Plant Sci*. 167(5): 1029-1048.

Khidir, W.H. and H. Liang. 1997. The matK gene: sequence variation and application in Plant Systematics. *Am. J. Bot.,* 84(6): 830-839.

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

Kukenthal, G. 1909. Cyperaceae–Caricoideae. In: (Ed.): A. Engler, Das Pflanzenreich, IV. 20 (Heft 38), 1-824. Wilhelm Englemann, Leipzig, Germany.

Kukkonen, I. 2001. Flora of Pakistan, No. 206. Cyperaceae. University of Karachi, Pakistan, and Missouri Botanical Press, St. Louis, Missouri, USA. 277 p.

Liang, H and K.W. Hilu. 1996. Application of the *matK* gene sequences to grass systematics. *Can. J. Bot.* 74: 125-134.

Mabberley, D.J. 1997. The plant-book, 2nd Ed., Cambridge University Press, Cambridge, p. 706.

Muller, K.F., T. Borsch, K.W. Hilu. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting matK, trnT-F and rbcL in basal angiosperms. *Mol Phylog Evol.* 41: 99-117.

Nei, M. and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Olmstead, R.G. and J.D. Palmer. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.,* 81: 1205-1224.

Reznicek, A.A. 1990. Evolution in sedges (Carex, Cyperaceae). *Can. J. Bot*., 68: 1409-1432.

Roalson, E.H., J.T. Columbus and E.A. Friar. 2001. Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and *trnT-L-F* (cpDNA) region sequences: Assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. *Syst. Bot*. 26: 318-341.

Shinwari, Z.K. 1995 Congruence between morphology and molecular phylogeneties in *Prosartes* (Liliaceae). *Pak. J. Bot*., 27(2): 361-369.

Shinwari, Z.K. 2002 Sequence divergence of *rbc*L gene and Phylogenetic relationships in Liliales. *Pak. J. Bot*., 34(2): 191-204.

Shinwari, Z.K. and Shehla Shinwari. 2010. Molecular data and phylogeny of family smilacaceae. *Pak. J. Bot.,* Special Issue (S.I. Ali Festschrift) 42: 111-116.

Shinwari, Z.K., R. Terauchi and S. Kawano 1994a. Phylogenetic relationships among genera in the Liliaceae-Asparagoideae-Polygonatae sensu lato inferred from *rbc*L gene sequence data. *PI. Systematic & Evolution*, 192: 263-277.

Shinwari, Z.K., R. Terauchi, F.H. Utech and S. Kawano. 1994. Recognition of the New World *Disporum* Section *Prosartes as Prosartes* (Liliaceae) based on the sequence data of the *rbc*L gene. Taxon, 43(3): 353-366.

Sneath P.H.A. and R.R. Sokal .1973. *Numerical Taxonomy*. Freeman, San Francisco.

Starr J.R. and B.A. Ford. 2009. Phylogeny and evolution in *Cariceae* (*Cyperaceae*): current knowledge and future directions. *Bot. Rev.* 75: 110-137.

Starr, J.R, S.A. Harris and D.A. Simpson. 2004. Phylogeny of the unispicate taxa in *Cyperaceae* tribe *Cariceae* I: Generic relationships and evolutionary scenarios. *Syst. Bot*., 29: 528-544.

Starr, J.R., R.F.C. Naczi and B.N. Chouinard. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, *Cyperaceae*). *Mol. Ecol. Resour.,* 151-163.

Starr, J.R., R.J. Bayer and B.A. Ford. 1999. The phylogenetic position of Carex section Phyllostachys and its implications for phylogeny and subgeneric circumscription in Carex (Cyperaceae). *Am. J. Bot*., 86: 563-577.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Bio. Evo.*, (In Press).

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

Waterway, M.J. and J.R. Starr. 2007. Phylogenetic relationships in the tribe *Cariceae* (*Cyperaceae*) based on nested analyses of three molecular data sets. *Aliso*, 23: 165-192.

Waterway, M.J., E. Mcintire and R.G. Olmstead. 1997. Molecular evidence suggests that *Carex* section *Limosae* is not monophyletic. *Am. J. Bot*., 84: 243.

Yen, A.C. and R.G. Olmstead. 2000. Molecular systematic of *Cyperaceae* tribe *Cariceae* based on two chloroplast DNA regions: *ndhF* and *trnL* intron-intergenic spacer. *Syst. Bot*., 25: 479-494.

Yen, A.C. and R.G. Olmstead. 2000. Molecular systematics of *Cyperaceae* tribe *Cariceae* based on two chloroplast DNA regions: ndhF and trnL intron-intergenic spacer. *Syst. Bot*., 25: 479-494.

(Received for publication 1 September 2012)