Indian J. Genet., 75(2): 232-236 (2015)
DOI: 10.5958/0975-6906.2015.00035.8

Phylogenetic informativeness of plastid regions in inferring the species relationships among *Cucumis* species

Nidhi Shubhanand, Madhoolika Agrawal[1], K. Joseph John[2], Shrikant Sutar[3], S. R. Yadav[3], S. R. Rao[4] and K. V. Bhat*

Division of Genomic Resources, ICAR-National Bureau of Plant Genetic Resources, New Delhi 110 012; [1]Department of Botany, Banaras Hindu University, Varanasi 221 005; [2]NBPGR Regional Station, Thrissur, Kerala; [3]Botany Department, Shivaji University, Kohlapur 416 004; [4]Biotechnology Department, North Eastern Hill University, Shillong

(Received: September 2014; Revised: April 2015; Accepted: May 2015)

Abstract

Molecular approaches are playing major role in phylogeny reconstruction and species relationship studies. The uniparent inheritance, small size, non recombinant nature and extremely low rate of evolution of chloroplast genome make it an appropriate marker for phylogenetic studies. In the present study, phylogenetic informativeness of two coding regions, *rbcL* and *rpoC1* and two non-coding intergenic spacers *psbA-trnH* and *trnL-F* of plastid DNA were analysed for inferring relationships among *Cucumis* species. Maximum evolutionary divergence was shown by *trnL-F* followed by *psbA-trnH* region. The non-coding spacers evolved 1.62 to 9 times faster than coding regions of *rbcL* and *rpoC1*. In the phylogenetic analyses, all species of Asian origin showed clear divergence from *Cucumis dipsaceus* and *Cucumis prophetarum*, which were African in origin. *Mukia maderaspatana* and *Dicaelospermum ritchiei* showed very low evolutionary divergence and are reported close to each other. *C. setosus* had been confirmed as distinct species and *C. callosus* was found to be a wild progenitor of *C. melo*.

Key words: *Cucumis*, plastid DNA, species relationships

Introduction

With the advancement in genome sequencing approaches, variations at the DNA level has gained primary importance for phylogeny reconstructions. However, choosing an appropriate marker or genome region is the prime challenge of molecular phylogenetic analysis of the desired angiosperm group. The region should yield similar results as that of multigene data analysis and shorter enough to facilitate easy sequencing of large number of taxons. Extremely slow rate of nucleotide substitution in chloroplast DNA provides it with advantages in molecular phylogenetics compared to the diverse mitochondrial and nuclear genome. The small size, nonrecombinant nature, uniparent inheritance and conservative evolution of chloroplast genomes in terms of organization and sequence (Barton and Jones, 1983; Palmer, 1985) make them potentially useful tools for molecular systematic of the plants at higher as well as lower taxonomic level. In contrast to most of the multi copy nuclear genes, chloroplast genes are essentially single copy (Palmer, 1986). Around two third of plastid variations result from size differences in inverted repeat sequences. They are free from gene duplication and deletion, concerted evolution and pseudogene formation, which can severely effect evolutionary interpretations. The various protein coding regions as well as non coding sites of plastome have been used to resolve molecular phylogeny at high taxonomic levels as well as at low taxonomic levels. The large subunit of ribulose bisphosphate carboxylase gene (*rbcL*) was one of the first molecular markers to be used in phylogenetic studies of numerous plant taxa. Subsequently many other protein coding genes like *matK*, *atpB*, *rpoC1, ndhF*, *rps4* have been used for confirming the species relationships at high taxonomic levels.

*Corresponding author's e-mail: kvbhat@nbpgr.ernet.in
Published by the Indian Society of Genetics & Plant Breeding, F2, First Floor, NASC Complex, PB#11312, IARI, New Delhi 110 012
Online management by indianjournals.com

The chloroplast genomes of cucumber and melon are circular double stranded molecule of 155,527 bp (Kim et al. 2006) and 156,017 bp (Rodriguez-Moreno et al. 2011) respectively. A comparative analysis of the melon and cucumber chloroplast genomes can provide useful information about the evolutionary relationships among cucurbit species. In the present study we have assessed the utility of two intergenic spacers, *psbA-trnH* and *trnL-F* and two coding regions *rbcL* and *rpoC1* of chloroplast to describe relationships among various *Cucumis* species. These relationships could be utilized in conservation of genetic resources and their beneficial utilization in crop improvement programmes for the development of disease resistant and new improved varieties of *Cucumis.*

Materials and methods

Plant material

A total of 12 ingroup taxa, namely, *Cucumis sativus* L. var. *sativus* (BBL73-2000), *C. sativus* forma *hardwickii* (Royle) W.J. de Wilde & Duyfjes (IC259826), *C. hystrix* Chak. (BY6-2010), *C. silentvalleyi* (Manilal, T. Sabu & P. J. Mathew) Ghebretinsae & Thulin (IC582747), *C. maderaspatanus L.* (*Mukia maderaspatana* 1), *C. ritchiei* (C. B. Clarke) Ghebretinsae & Thulin (SUC23) *Cucumis melo* L. subsp*. agrestis* (Naudin) Pangalo (IC258167), *C. melo* subsp. *melo* L. (IC524135), *C. callosus* (Rottl.) Cogn. (IC550169), *C. setosus* Cogn. (IC583551), *Cucumis dipsaceus* Ehrenb. *ex* Spach, *C. prophetarum* L. (IC550178) were taken for this study. Leaf materials for experimentation, were collected from plants grown in the net house of National Bureau of Plant Genetic Resources (NBPGR), New Delhi and its Regional Station at Thrissur.

DNA extraction, amplification and sequencing

Total Genomic DNA was extracted from fresh leaves of seed-grown plants with commercial plant DNA extraction kits (Qiagen, Himedia, Life-Tech, Axygen), following the manufacturers' specifications.

The *psbA-trnH* region was amplified using the primers *psbA*f 5'-GTTATGCATGAACGTAATGCTC-3' and *trnH* r 5'-CGCGCATGGTGGATTCACAATC-3' of Kim et al. (1999). The *trnL-trnF* region was amplified using *trnL* 5'-GGT TCA AGT CCC TCT ATC CC-3' and *trnF* 5'-ATT TGA ACT GGT GAC ACG AG-3' of Taberlet et al. (1991). Similarly, *rbcLf* 5'-ATG TCA CCA CAA ACA GAA ACT AAA GCA AGT-3' and *rbcLr* 5'-GCT TTA GTC TCT GTT TGT GG-3' of Hiratsuka et al. (1989) were used for amplification of *rbcL* region and *rpoC* 1-2 f 5'-GGC AAA GAG GGA AGA TTT CG-3 and *rpoC 1*-3r 5'-TGA GAA AAC ATA AGT AAA CGG GC-3 (http://www.kew.org) were used to amplify partial sequence of *rpoC1* gene. Reaction of 25 µL contained 20 ng/µL of purified genomic DNA, 2.5 mM $MgCl_2$, 1U Taq DNA polymerase, 1x PCR buffer without $MgCl_2$, 0.25 µM of each of primers for each specific region and 200 µM of dNTPs. The thermal cycle protocol included an initial denaturing at 94^oC for 6 min; 40 cycles at 94^oC for 1 min; 55^oC for *psbA-trnH* and *trnL-F*,48^oC for *rbcL* and 54^0C for *rpoC1* for 1 min 72^oC for 1 min; ending with an extension at 72^oC for 10 min.

PCR products were purified using the PCR purification kit (Roche) following manufacturer's recommendations and then cycle sequenced using ABI Prism Big Dye Terminator cycle sequencer (Applied Biosystems, CA).

Data analysis

The sequenced DNA regions were aligned using the CLUSTALX program (Thompson et al. 1997). Genetic distances (GD) were calculated by using the Jukes and Cantor (1969) method and the phylogenetic trees were drawn by applying the neighbor-joining. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jukes-Cantor model. Initial tree for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. A maximum parsimony analysis was also performed for each region applying the Tree-Bisection-Regrafting (TBR) algorithm. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5.1 (Tamura et al. 2011).

Results and discussion

Successful DNA amplifications were obtained for all the species except *Cucumella silentvalleyi* that could not be amplified for *rbcL* region. After excluding alignment gaps and missing data sites, the sequenced fragment length of 209, 1181, 339 and 372bp were obtained for *psbA-trnH*, *rbcL*, *rpoC1* and *trnL-trnF*, respectively. Comparisons for each locus have been summarized in Table 1. The maximum evolutionary divergence was shown by *trnL-trnF* intergenic spacer

region followed by the spacer *psbA-trnH* and coding regions *rbcL* and *rpoC1*. The results indicated that non coding plastid intergenic spacer regions show high evolutionary divergence as compared to coding regions at lower taxonomic level, as reported earlier (Gielly and Taberlet, 1994). The non coding intergenic spacer regions of *psbA-trnH* and *trnL-F* evolve between 1.62 times (*psbA-trnH-rbcL*) and 9 times (*trnL-F – rpoC1*) faster than coding regions of *rbcL* and *rpoC1* gene.

Phylogenetic analyses were performed based upon the evolutionary distance matrix (Neighbor-joining, NJ) as well as character matrix (Maximum Parsimony, MP and Maximum Likelihood, ML). The topologies of the trees obtained from analyses of all the four loci, showed two major clades (A and B). *C. prophetarum* and *C. dipsaceus* (clade B) grouped together to form the first major clade with a very strong bootstrap support (*rbcL*, 97; *rpoC1*, 87; *trnL-F*, 99; *psbA-trnH*, 95). Earlier study based on chloroplast and nuclear DNA, suggests the African origin of these entities (Sebastian et al., 2010). Both of these *Cucumis* species have travelled from Africa and dispersed to Asia independently of the *C. sativus* and *C. melo* ancestor. These species now flourish well all over India. All other *Cucumis* species grouped together in the second major clade (clade A), supporting the Asian origin of these taxa . Although, it has been suggested that RNA polymerase genes can be 'lucky gene' (Logacheva et al. 2007) for phylogenetic inferences, but in present study the *rpoC1* gene is not found as useful as that of other loci for resolving species relationships. The two chloroplast spacer regions (*trnL-F, psbA-trnH*) and one coding gene (*rbcL*) resolved the relationships among the species of clade A upto different extent. The melons, *C. melo* subsp. *melo*, *C.m*. subsp. *agrestis* and *C. callosus*, grouped together in a distinct subclade (Figs. 1, 2, 3) within the major clade A (*rbcL*, ML = 100, *trnL-F*, MP = 99, *psbA-trnH*, NJ = 39). This group is best resolved by *rbcL* gene. The *C. callosus*, earlier treated as synonym of *C. melo* (Kirkbride, 1993) was found identical to *C. melo*. Normal fertility of F_1 and BC_1 generations of its cross with cultivated melon (John et al. 2013) and molecular analysis of nuclear and chloroplast DNA (Sebastian et al. 2010) indicates that *C. callosus* is wild progenitor of cultivated melon. The *rbcL* gene sequence data suggests *C. hystrix* as the closest relative of cucumber (Fig. 1). Although it resembles cucumber in its vegetative and floral morphology, the chromosome number of *C. hystrix* (2n = 24) is different from the chromosomal count of *C. sativus* (2n = 14). The previous molecular analysis has warranted *C. hystrix* as the closest wild relative of cucumber (Sebastian et al. 2010). *C. s. sativus* and *C. s. hardwickii* showed an evolutionary divergence of 0.008 and 0.011 respectively with *C. hystrix*. Although, *rbcL* gene and *psbA-trnH* spacer could not differentiate between *C. sativus* and *C. hardwickii*, the intergenic spacer *trnL-F* differentiated both these entities with an evolutionary divergence of 0.003. *C. setosus* which has been synonimized under *C. sativus*, was found closer to *Cucumella silentvalleyi* as suggested in earlier molecular studies. *C. setosus* showed an evolutionary divergence of 0.005 with *Cucumella silentvalleyii* and 0.015 with *C. sativus*. Morphological observations, chromosomal count and crossability barriers identify *C. setosus* as a distinct species endemic to western India (John et al. 2014). *Mukia maderaspatana* and *Dicaelospermum ritchiei* show a very low evolutionary divergence and are found close to each other. *Cucumella silentvalleyii*, *M. maderaspatana* and *Dicaelospermum ritchiei* are found nested within *Cucumis* as suggested by Schaefer (2007), who recircumuscribed the genus to monophyletic and merged these genera alongwith *Myrmecosicyos* C. Jeffrey and *Oreosyce* Hook.f. with *Cucumis*. These entities are renamed as *Cucumis silentvalleyi*, *Cucumis maderaspatanus* and *Cucumis ritchiei* respectively.

Table 2. A comparative account of all four plastid loci used in the present study. The regions were analysed for their phylogenetic informativness in inferring the relationships among *Cucumis* species

Comparisons	*psbA-trnH*	*trnL-F*	*rbcL*	*rpoC*
Fragment length	209	372	1181	339
Number of conserved sites	200	346	1152	336
Number of variable sites	9	26	29	03
Number of parsimony informative sites	06	14	20	02
Number of singletons	03	12	9	01
Evolutionary divergence	0.0135	0.0207	0.0083	0.0023

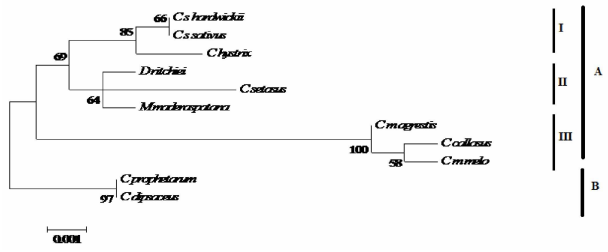

Fig. 1. ML consensus tree for 11 taxa of *Cucumis* species for the coding region *rbcL* gene. Analysis was conducted on the basis of Jukes-Cantor model. The tree with the highest log likelihood (-1903.5152) is shown here. The vertical bars on the right side indicate groups and subgroups described in the text

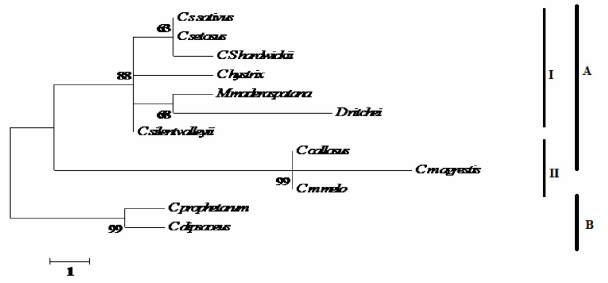

Fig. 2. MP consensus tree for *Cucumis* species for the intergenic spacer region *trnL- trnF* gene. The most parsimonious tree with length = 32 is shown. The MP tree was obtained using the Tree-Bisection-Regrafting (TBR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (100 replicates). The vertical bars on the right side indicate groups and subgroups described in the text

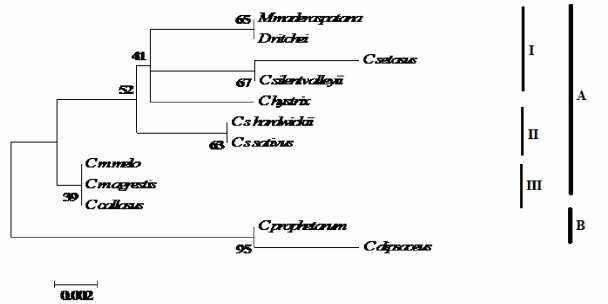

Fig. 3. NJ consensus tree with bootstrapping (1000 replicates) for *Cucumis* species for the intergenic spacer region *psbA-trnH* gene. Analysis was based on Jukes-Cantor Model. The vertical bars on the right side indicate groups and subgroups described in the text

The present study confirms the utility of multiple loci for inferring phylogenetic relationships of the genus *Cucumis.* Except the coding region of *rpoC1* gene, all other regions are found useful in resolving the finer details of phylogeny. The maximum evolutionary divergence has been shown by the non coding plastid intergenic spacer *trnL-F* however the coding region of *rbcL* gives the clear picture of evolutionary relationships. These relationships could be utilized in introducing resistance genes from wild relatives and in raising new varieties with new taste and more nutritive value. These regions can also be used in developing barcodes for correct identification of *Cucumis* species.

Acknowledgements

This work is part of the Ph.D. thesis submitted by the first author. The study was carried out with funding from the National Agricultural Innovation Project of the Indian Council of Agriculture Research.

References

Barton N. and Jones J. S. 1983. Mitochondrial DNA: new clues about evolution. Nature, **306:** 317-318.

Gielly L. and Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus rbcL sequences. Mol. Biol. Evol., **11**(5): 769-777.

Hiratsuka J. and et al. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Molecular and General Genetics, **217:** 185-194

John J. K., Khedasana R., Nissar V. A. M., Scariah S., Sutar S., Rao S. R., Nizar M. A., Latha M., Yadav S. R. and Bhat K. V. 2014. On the occurrence, distribution and taxonomy of Cucumis setosus Cogn., an endemic wild edible vegetable from India. Genet Resour Crop Evol., **61:** 345-355.

John J. K., Scariah S., Nissar V. A. M., Latha M., Gopalakrishnan S. S. R. and Bhat K. V. 2013. On the occurrence, distribution, taxonomy and genepool relatioship of *Cucumis callosus* (Rottler) Cogn., the wild progenitor of Cucumis melo L. from India. Genet Resour Crop Evol., **60:** 1037-1046.

Jukes T. H. and Cantor C. R. 1969. Evolution of protein molecules. In Munro HN, editor, *Mammalian Protein Metabolism*, 21-132, Academic Press, New York.

Kim J. S., Jung J. D., Lee J. A., Park H. W., Oh K. H., Jeong W. J., Choi D. W., Liu J. R. and Cho K. Y. 2006. Complete sequence and organization of the cucumber (*C. sativus* L. cv. Baekmibaekdadagi) chloroplast genome. Plant Cell. Rep., **25:** 334-340.

Kim S.-C., Crawford D. J., Jansen R. K. and Guerra A. S. 1999. The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribe *Sonchinae* (*Asteraceae: Lactuceae*). Pl. Syst. Evol., **215**: 85-99.

Kirkbride J. H. Jr. 1993. Biosystematic monograph of the genus *Cucumis* (Cucurbitaceae). Parkway Publishers, Boone, North Carolina.

Logacheva M. D., Penin A. A., Samigullin T. H., Vallejo-Roman C. M. and Antonov A. S. 2007. Phylogeny of Flowering Plants by the Chloroplast Genome Sequences: in Search of a "Lucky Gene". Biochemistry (Moscow), **72**(12): 1324-1330.

Palmer J. D. 1985. Comparative organization of chloroplast genomes. Ann. Rev. Genet., **19:** 325-354.

Palmer J. D. 1986. Chloroplast DNA and phylogenetic relationships. In "DNA Systematics: Evolution", Dutta, S. K. (ed.), CRC Press, Inc., Boca Raton, Florida: 63-80.

Rodriguez-Moreno L., Gonzalez V., Benjak A. and et al. 2011. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC Genomics, **12**: 424.

Royal Botanical Gardens, Kew: DNA barcoding: Phase 2 Update. 2010. http://www.kew.org

Schaefer H. 2007. *Cucumis* (Cucurbitaceae) must include *Cucumella, Dicoelospermum, Mukia, Myrmecosicyos* and *Oreosyce*: a recircumscription based on nuclear and plastid DNA data. Blumea, **52**: 165-177.

Sebastian P., Schaefer H., Telford I. R. H. and Renner S. S. 2010. Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc. Natl. Acad. Sci., **107**(32): 14269-14273.

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

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol., **28:** 2731-2739.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. and Higgins D. G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., **24:** 4876-4882.