DeepScience
Multidisciplinary Research: Innovations in Science, Arts and Commerce
Published in October 2025 by Deep Science Publishing (USA | UK | India | Turkey)
ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

A Comparative Analysis of different Genes with regard To Phylogenetic Utilization

Dr. Insan Ara Rahman
Department of Botany, Pandit Deendayal Upadhyaya Adarsha Mahavidyalaya,Amjonga, Goalpara, Assam
email: insanara.rahman13@gmail.com

Abstract

Phylogenetic analysis is a very important aspect in biological sciences since it reveals the closest wild relatives of many crop plants for genetic and molecular studies which can be used in comparative studies and transfer of useful genes for crop improvement. In this study DNA sequence data from nuclear ITS1-5.8S-ITS2 region and chloroplast matK gene of seven different wild species of Solanum commonly found in Meghalaya are used. Molecular characterization of the sequences have been made on basis of conserve sites, variable sites, GC content, indels, transitions, transversions, parsimony informative sites etc. These are the parameters which indicate the evolutionary rate, so comparison of these parameters provides valuable information regarding efficiency of these regions in phylogenetic study. The result shows that both the two regions are useful for phylogenetic study in lower level, but the comparison indicates that matK is more phylogenetically informative in this case.

Keywords: *Solanum*, *ITS*, *matK*, Phylogeny

1. Introduction

According to most systematists and evolutionary biologists phylogeny should be the central underpinning of research in much of biology. Zuckerandle and Pauling (1965) suggested usage of macromolecular sequence data in phylogenetics and also gave the idea that phylogeny may be deduced from mutations. So in present days various genomic regions are exploited by different systematists to infer phylogeny according to their suitability to phylogenetic research. In case of plant phylogenetics different nuclear genes are used and among them the most extensively used are 18S rDNA and *rrn* ITS regions. This is because of its ubiquitous presences across the tree of life. It also exhibits similar phylogenetic signals at equivalent taxonomic ranks (Hillis and Dixon, 1991; Embley *et al*., 1994; D. Soltis and Soltis, 1995). The first major effort to apply 18SrDNA data to angiosperm phylogeny was undertaken in the late 1980s and based on direct RNA sequencing of portions of both the 18S and 26S regions (Hamby and Zimmer, 1988, 1992; Zimmer *et al.*, 1989). Subsequent studies (Nickrent and Franchina, 1990; Boulter and Gilory, 1992; Bharathan and Zimmer,

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

1995) contributed a small subunit database, with the trend toward complete 18S DNA sequences. ITS sequences have been utilized for reconstruction of phylogenetic relationship in angiosperms (Baldwin *et al*., 1995), algae (Bakker *et al*., 1995; Coleman *et al*., 1994). ITS sequences have also been used to infer phylogeny at lower taxonomic levels in a diverse array of organisms.

Among the chloroplast genes the commonly employed are *rbc*L (ribulose 1,5-bisphosphate carboxylase), *rps*4 (encoding a chloroplast ribosomal protein), *ndh*F (which possibly encodes a dehydrogenase) and *mat*K (encoding a maturase), *atp*B etc. Chase *et al*., 1993 has provided a database for *rbc*L, Nadot *et al*., 1994 used *rps*4 in Poaceae to infer phylogeny, *ndh*F used on Scrophulariacea by Olmstead and Reeves, 1995, and in Solanaceae by Olmstead and Sweere, 1995, *atp*B used on Lardizabalaceae by Hoot *et al.*,1995.

Both nuclear DNA Internal Transcribed Spacer region (ITS1-5.8S-ITS2) and chloroplast DNA *mat*K region are used to infer phylogeny among lower taxonomic groups. In the present study an effort has been taken to characterize Internal Transcribed Spacer region (ITS1-5.8S-ITS2) of the 18S-5.8S-28S nuclear ribosomal cistron and chloroplast *mat*K gene region, as well as comparison of their perspective for phylogenetic utility in context to some *Solanum spp.* of Meghalaya.

2. Materials And Method

Plant material and taxon sampling

Solanum aculeatissimum, *Solanum sisymbriifolium*, *Solanum aethiopicum*, *Solanum kurzii*, *Solanum clavatum*, *Solanum nigrum* and *Solanum torvum* were collected from forests around Shillong, Meghalaya (North- East India) (Table 1). Voucher specimens are lodged in the Herbarium of Botany Department, North Eastern Hill University.

Species	Specimen voucher[a]	ITS	*mat*K
Solanum khasianum Clarke	NEHU-11928	KC535792	KC535798
Solanum sisymbriifolium Lamk.	NEHU-11929	KC535789	KC535799

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

Solanum gilo Req. ex Dunal	NEHU-11931	KC535795	KC535801
Solanum kurzii Brace ex Prain	NEHU-11934	KC535794	KC535800
Solanum clavatum Rusby	NEHU-11933	KC535790	KC535796
Solanum nigrum L.	NEHU-11930	KC535791	KC535797
Solanum torvum Sw.	NEHU-11932	KC535793	KC535802

Table1. Specimens voucher, accession numbers of deposited sequences of all the seven *Solanum* spp. collected. ([a] Specimen vouchers deposited at the Herbarium, Department of Botany, North Eastern Hill University.)

DNA extraction, amplification and sequencing

Genomic DNA was isolated from fresh leaves using CTAB method (Doyle and Doyle, 1987). Primers were designed targeting two regions of the genome- nuclear ITS region and chloroplast region *matK* (Table 2).

Primer name	Sequence
SITSF	5’AAACCTGCACAGCAGAACGAC3’
SITSR	5’GGTCGCGGTCGGAGCGCG 3’
SmtKF	5’CACAAACTAGACGAAGCTC3’
SmtKR	5’TATGCACTTGCTCAGGATC3’

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

Smtk1	5’ AATATATTTCTATGGAAAAAG3’
Smtk2	5’ ATCAAAGGATCCTTGAATAAC3’

Table 2. The primers designed and used in this study and their sequences.

Amplification of *rrn* ITS, and *matK*

Each reaction mix contained 2.5 µL of 10X PCR assay buffer, 2.5 µL of $MgCl_2$ (25mM), 3 µL each of the individual dNTP (1.25 mM), 1.5 µL of each primer pair (5 pM), 0.3 µL of *Taq* polymerase (3 Units/ µL), and the final volume was made up to 25 µL by adding ultra pure water., Polymerase chain reaction was carried out targeting the genes using the respective primer pairs (Table 2). The thermalcycler was programmed with the following parameters: premelt at 94 ^{0}C for 5 min followed by 35 cycles consisting of denaturation at 94 ^{0}C for 30 sec, annealing for 1 min at 65^{0}C and 64^{0}C for *rrn* ITS and *matK* respectively and extension at 72 ^{0}C for 1 min. The main programme was followed by a final extension step of 72 ^{0}C for 10 mins. Internal primer pairs namely Smtk1, Smtk2 were used for sequencing purpose of *matK* .Amplified PCR products were purified using Himedia Quick gel extraction kit. Sequencing was carried out by 3130 Genetic Analyzer (Applied Biosystems) in North Eastern Hill University.

Sequence alignment and Sequence characteristics

Sequences of both the two regions were subjected to Multiple Sequence Alignment using the CLUSTAL X (2.0) program (Thompson et al., 1997) with default settings.

Sequence characteristics of both the studied regions were calculated before performing phylogenetic analyses by using both MEGA version 5 (Tamura et al., 2011) and Seqstate v.1.21 (Müller, 2005). The two data sets were analyzed separately as shown in the table 3.

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

RESULTS AND DISCUSSIONS

The sequences of both ITS and *mat*K of all the species of *Solanum* were aligned with Multiple Sequence Alignment using CLUSTAL X program separately. The numbers of characters for each regions, the percentage of conserved sites, variable sites, parsimony informative characters, GC content, transitions, transversions, indels etc. were analysed and are described in Table 3, as the analyses of these characters are very important in phylogenetic point of view. Both the two regions namely ITS and *matK* showed variable sequence lengths. The sequence length of ITS for all the seven species studied ranged from 575-619bp, while *mat*K sequence length ranged from 1282-1302bp. They have an aligned sequence length of 641 and 1356 characters for ITS and *mat*K respectively. ITS has lowest percentage of conserved sites (51.48%) than *mat*K (85.69%) and thus ITS has the most variable region (46.49%) generating highest number of parsimony informative sites (17.32%) than that of *matK* (10.46%) which provides 1.80% of parsimony informative sites. The % GC content is higher in case of ITS (61.2%) than *mat*K (32.4%). Among the two regions studied total transitions and transversions were higher in case of ITS than that of *mat*K. The highest number of indels were recorded in ITS (19.81%) followed by *mat*K (8.0%). Average number of base substitutions per site is recorded higher in case of *mat*K as seen in table 3 which is 0.058 than ITS where average number of base substitution per site is 0.03. Retention index and consistency index are two important criterion on basis of which the intensity of homoplasy can be assumed in a given array of DNA sequence. So in our study it has been seen that (Table 3) *mat*K is occupying higher position in case of both retention index (0.85) and consistency index (0.84) than ITS where retention index is 0.58 and consistency index is 0.65.

Dataset characteristics	*ITS*	*matK*
Genome	Nuclear	Chloroplast
Range of raw length	575-619	1282-1302
Aligned length	641	1356
Conserved sites (%)	51.48	85.69

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

Variable sites (%)	46.49	10.46
Parsimony informative sites (%)	17.32	1.80
Transitions	71	15
Transversions	47	19
Indels (%)	19.81	8
Retention index (RI)	0.58	0.85
Consistency index (CI)	0.65	0.84

Table 3. Sequence information and comparison of data sets from two nuclear regions and two chloroplast regions.

To infer phylogenetic relationship at different taxonomic levels both chloroplast DNA (cpDNA) and nuclear ribosomal DNA (nrDNA) are used (reviewed by Palmer *et al*., 1988; Hamby and Zimmer, 1992). Nuclear ribosomal ITS and chloroplast *matK* are two regions which are extensively used to infer phylogenetic relationship at lower taxonomic levels. ITS is a noncoding region and that is why it is prone to rapid evolutionary changes but though cp *matK* is a coding region it is the most rapidly evolving coding region found so far within chloroplast genome (Neuhaus and Link, 1987; Olmstead and Palmer, 1994). So in the present study both the two regions were amplified and sequenced for all the seven species of *Solanum* so that a comparison can be made in respect to their phylogenetic utility. From the results it has been seen that ITS contains more variable sites than *matK*. Insertions and deletions (indels) are also more common in case of ITS. As expected ITS provide more parsimony informative characters as was also reported by Shaw *et al*., 2005. The nuclear ITS data set had higher percentage of variable sites together with higher parsimony informative characters than *mat*K, which suggests its higher level of homoplasy. Alvarez and Wendel (2003) also reported that ITS shows higher level of homoplasy

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

as compared to other DNA sequence data sets. This is due to orthology/paralogy conflation, compensatory base changes, problem in alignment due to indel accumulation, sequence errors, and some combinations of these phenomena. The chloroplast *mat*K gene was found to be more informative having both a high percentage of Parsimony informative characters as well as high Consistency and Retention indices which suggests a low level of homoplasy. Mort *et al*., 2007 also reported similar type of results suggesting that homoplasy in ITS was relatively high compared to the cpDNA loci. According to Alvarez and Wendel (2003), despite the universal usage of ITS sequence data in plant phylogenetic studies, its complex and unpredictable evolutionary behavior reduce its utility for phylogenetic analysis. Thus from the above results and discussions it can be concluded that *mat*K is more efficient to resolve phylogenetic relationship compared to that by ITS in this case.

REFERENCES

[1] Alvarez I and Wendel JF (2003). Ribsomal ITS sequences and plant phylogenetic inference, *Mol. Phyl. Evol.* 29: 417-434.
[2] 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. Missouri Bot. Gard* 82: 247–277.
[3] Bharathan G and EA Zimmer (1995).Earlybranching events in monocotyledons–partial18S Ribosomal DNA sequences. *In:* PJ Rudall, P JCribb, DF Cutler and CJ Humphries (eds.). *Monocotyledons:systematics and evolution*, pp 81–107. Royal Botanic Gardens (Kew), UK.
[4] Boulter D and Gilory JS (1992). Partial sequences of 18S ribosomal RNA of two genera from each of six flowering plant families, *Phytochem.* 31(4):1243-1246.
[5] Chase MW, Soltis DE, Olmstead RG, Morgan D, Les D H, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer J D, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedroen M,Gaut BS,Jansen RK,Kim KJ,Wimpee CF,Smith JF, Furnier G R, Strauss S H, Xiang QY, Plunkett G M, Soltis P S, Swensen S, Williams S E, Gadek P A, Quinn C J, Eguiarte L E, Golenberg E, Learn G H Jr, Graham SW, Barrett SC,Dayanandan S, and Albert VA (1993). DNA sequence phylogenetics of seed plants: an analysis of the plastid gene *rbcL, Ann. Missouri Bot. Gard* 80: 528-580.
[6] Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue, *Phytochem Bull.* 19:11–15.
[7] Embley MT, Hirt RP and Williams DM (1994). Biodiversity at the molecular level: the domains, kingdoms, and phyla of life, *Phil. Trans. Royal Soc*. B345 (1311): 21-31.
[8] Hamby RK and Zimmer EA (1988). Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae), *Plant Sys. Evol*. 160: 29-37
[9] Hamby RK, and Zimmer EA (1992). Ribosomal RNA as a to be presented, genetic tool in plant systematic, *In*: Soltis PS, Soltis DE, and Doyle JJ, (eds). *Molecular Systematics of Plants* , pp 50– 91.

ISBN: 978-93-7185-157-2 | E-ISBN: 978-93-7185-737-6 | DOI: https://doi.org/10.70593/978-93-7185-737-6

[10] Hillis DM and Dixon MT (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference, *Quar. Rev. Bio.* 66: 411-453.

[11] Mort M, Archibald JK, Randle CP, Levsen ND, O'leary TR, Topalov K, Wiegand CM, and Crawford DJ (2007). Inferring Phylogeny At Low Taxonomic Levels: Utility of rapidly evolving CpDNA and nuclear ITS Loci, *Am. J. Bot*. 94(2): 173–183.

[12] Müller K (2005) SeqState: primer design and sequence statistics for phylogenetic DNA datasets, *Appl. Bioinform*. 4: 65–69.

[13] Nadot, Bajon SR, and Lejeune B (1994). The chloroplast gene rps4 as a tool for the study of Poaceae phylogeny, *Plant Sys. Evol*.191: 27-38.

[14] Neuhaus H and Link G (1987). The chloroplast tRNA LYS(UUU)gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide, *Curr. Genet*. 11: 251-257.

[15] Nickrent DL and Franchina CR (1990). Phylogenetic relationships of the Santalales and relatives, *J. Mol. Evol*.31: 294-301.

[16] Olmstead RG and Palmer JD (1994) Chloroplast DNA systematics: A review of methods and data analysis, *Am. J. Bot*. 81:1205–1224.

[17] Olmstead RG and Reeves PA (1995) Evidence for the polyphyly of the Scrophulariaceae based on chloroplast rbcL and ndhF sequences, *Ann Missouri Bot Gard.* 82:176–193.

[18] Olmstead RG and Sweere, JA (1994). Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae, Syst Biol 43: 467-481.

[19] Palmer JD, Jansen RK, Michaels HJ, Chase MK, and Manhart JR (1988). Chloroplast DNA variation and plant phylogeny, *Ann Missouri Bot Gard* 75: 1180- 1206.

[20] Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Twinder C, Schilling EE, and Small RL (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis, *Am. J. Bot*. 92: 142–166.

[21] Soltis PS and Soltis DE (1995). Molecular evolution of 18S rDNA in angiosperms: implication for character weighting in phylogenetic analysis, *In*: Soltis DE, Soltis PS and Doyle JJ (eds), *Molecular systematics of plnts II: DNA sequencing*, pp 188-210, Kluwer Academic Publishers, Boston.

[22] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, *Mol. Bio. Evo*. 28(10): 2731-2739.

[23] Thompson JD, Gibson TJ, Plewniak F, Mongin FJ, Higgins DG (1997). Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res*. 25: 4876–4882.

[24] Zimmer EA, Hamby RK, Arnold ML, Leblance DA and Theriot EC (1989). Ribosomal RNA phylogenies and flowering plant evolution, *In* Fernhilm B, Bremer K and Jornvall J, (eds). *The hierarchy of life*, pp- 205-214, Elsevier Science Publishers, Amsterdam.

[25] Zuckerandle E and Pauling L (1965). Molecules as documents of evolutionary history, *J Theoret Biol* 8:357-366.