In Silico Analysis of Sequence Variation in rbcL Gene to Assess Phylogenetic Relations in Setaria Species

Akshama Singh[1], Drishti[2], Garima Negi[3] and Aparajita Mohanty[4]

[1,2,3,4]*Department of Botany Gargi College University of Delhi*
E-mail: [1]akshama.singh95@gmail.com, [2]123drishti123@gmail.com,
[3]garimanegi79@gmail.com, [4]aparajita.gargi@gmail.com

Abstract—*Setaria sp. belongs to Poaceae or the grass family with more than 100 species distributed in the tropical and temperate regions of the world. The various species of Setaria are cultivated for human food and/or as animal fodder. Gene rbcL (present on the chloroplast genome) is popularly used for phylogenetic analysis which have further implication in biodiversity conservation. The present investigation was done to analyse sequence variations of rbcL gene amongst the species of Setaria. Sixteen annotated gene sequences of rbcL were retrieved from GenBank, NCBI. Multiple Sequence Alignment (MSA) of the retrieved sequences followed by construction of phylogenetic tree was done using CLUSTAL OMEGA. Length of rbcL gene sequence varied between 1290 bp to 1990 bp in the 16 species of Setaria, that were analysed. Analysis of MSA of rbcL DNA sequence of the 16 species of Setaria showed few sites of variations in the form of mostly substitutions. There is no sequence variation between Setaria verticillata and Setaria adhaerens as analysed from MSA results. Similarly no variations were observed between Setaria barbata and Setaria megaphylla. The present analysis shows that rbcL is a very conserved gene as shown by limited sequence variation in the genus Setaria. Therefore, rbcL gene may prove better for assessing phylogenetic relatedness at genus / family level, than at species level.*

Keywords*: Phyogenetic, rbcL gene, Sequence variations, Setaria.*

1. INTRODUCTION

Setaria sp belongs to Poaceae or the grass family. It is a widespread genus and its name is derived from the Latin word *seta*,which means “bristle" or "hair", There are more than 100 species distributed in the tropical and temperate regions around the world. Several species are cultivated for food or as animal fodder, such as foxtail millet (*S. italica*) and korali (*S. pumila*), while others are invasive weeds. *Setaria viridis* is currently being developed as a genetic model system for bioenergy grasses [1]. *Setaria viridis* is used to extract anti-oxidant and anti-microbial phenolic compounds and thus medicinally important. [2]. It is a bioenergy grass and is also important for genetic improvement of cereals. *Setaria sp* has been used in studies related to enhancing the productivity of grasses under high-density planting, by engineering light responses [3]. *Setaria viridis* is a model organism to study C4 photosynthesis [4]. Genetic transformation of *Setaria* using *Agrobacterium tumifaciens* has been done as a model transformation process for sugarcane [5]. Thus this grass has economic importance and is of special interest to researchers as well making it essential to study the various aspects of *Setaria sp*.

Chloroplast DNA diversity studies are important for assessing genetic variations and phylogenetic relationships between the species. Since *Setaria sp* grows in wild and are hardy, they can be potential genepool for biotic and abiotic resistance genes. To tap the potential of *Setaria sp*, it is important to understand the phylogenetic relatedness between the grasses. Phylogenetic studies will help in breeding experiments and improve traits of related species. Maturase K (matK) and ribulose 1,5-biphosphate carboxylase/oxygenase (rbcL) genes of chloroplast genome have been successfully used to assess the phylogenetic relationships amongst plants. These are coding regions and show low levels of interspecific variations, but are in some cases sufficient to analyse the phylogenetic distances between plants [6]. Keeping this in view our objective was to assess the sequence variations in complete rbcL gene of *Setaria sp*, available in the nucleotide database of NCBI and to analyze the phylogenetic relatedness between the species.

2. MATERIALS AND METHOD

The complete rbcL gene sequences for available *Setaria sp* were retrieved from nucleotide database of NCBI (URL: http://www.ncbi.nlm.nih.gov). A list of 16 species of *Setaria* with complete rbcL gene sequence that were used for analyses, along with their sizes and Accession numbers are given in Table 1. For analyses, rbcL sequence of each of the 16 species was saved in FASTA format. The average size of rbcL gene in *Setaria sp* was calculated. The sequences were subjected to Multiple Sequence Alignment (MSA) using Clustal omega available at EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/). MSA's were performed to assess the nucleotide/s insertions, deletions and substitutions of nucleotides, within the genus of *Setaria sp*. Further the generated MSA was used to construct a phylogenetic tree and interspecific relationships were assessed.

Table 1: List of *Setaria sp*. with Accession Numbers and Size of rbcL gene of the analysed species

Name of species	Accession Number	Size of rbcL gene
Setaria adhaerens	FN870407.1	1343 bp
Setaria barbata	HE573436.1	1342 bp
Setaria faberi	KC164333.1	1290 bp
Setaria humbertiana	LN908007.1	1353 bp
Setaria incrassta	LN908008.1	1353 bp
Setaria italica	X79900.1	1990 bp
Setaria madecassa	LN908009.1	1353 bp
Setaria magna	KJ773885.1	1323 bp
Setaria megaphylla	LN908010.1	1353 bp
Setaria parviflora	FN870409.1	1353 bp
Setaria plicata	AM849386.1	1342 bp
Setaria punctata	HE575856.1	1349 bp
Setaria sphacelata	LN908011.1	1353 bp
Setaria sulcata	HE575858.1	1316 bp
Setaria verticillata	LN908012.1	1353 bp
Setaria viridis	KT289405.1	1313 bp

3. RESULTS AND DISCUSSION

Chloroplast genome is a structurally conserved molecule across vascular plants [7] (Palmer et al 1986). The coding as well as non coding regions have been frequently used in phylogenetic and population genetics studies [8,9] (Demesure et al 1995; Dumolin et al 1997). Of the coding regions, rbcL gene is known to evolve at a slow rate [10](Clegg 1993). Although rbcL is known for its conserved sequences, it has also been recommended as one of the locus alongwith matK gene for DNA barcoding in plants [11](Group CBOL Plant working et al 2009).

In the present study, sequence variations in rbcL gene of 16 species of *Setaria* have been assessed. The average size rbcL gene in *Setaria sp* was 1379.94 bp (appx 1380 bp). The smallest length (1290 bp) was observed in *S. faberi* and largest (1990 bp) was in *S. italica*. Multiple Sequence Alignment of the retrieved rbcL gene sequences revealed that the gene sequence is highly conserved over the genus *Setaria*. The limited sequence variations were due to nucleotide substitutions (Fig1). Except for the total length variation of rbcL gene in *Setaria sp*., no single base insertion or deletion was observed. There was no sequence variation between *S. verticillata* and *S, adhaerens*; *S. megaphylla S. barbata* and *S. sulcata* and also between *S. humbertiana* and *S. magna*. The phylogenetic tree showed four main clusters based on the genetic distances. However, since the sequence variations were very limited, additional coding loci are required to better understand the interspecific genetic relatedness. Also, since rbcL gene sequence could not distinguish all the species of *Setaria*, our study suggests that rbcL is highly conserved at intrageneric level (in case of *Setaria*). In recent study, rbcL gene sequence has been used successfully to assess sequence variations in the genus *Prunus* [6]. At higher taxonomic levels (eg tribe or family), rbcL gene can provided important information on sequence divergence which help in understanding the phylogenetic relationships of different taxa [12].

```
S.italica        GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTGCTTATGCAAAAACTTTCCAAGGTC
S.magna          GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.plicata        GCGCTCTACGTTTGGAGGATCTACGAATTCCCACTAGTTATTCAAAAACTTTCCAAGGTC
S.sphacelata     GTGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.incrassata     GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.parviflora     GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.sulcata        GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.megaphylla     GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.barbata        GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.faberi         GCGCTCTACGTTTGGAGGATCTACGAATTCCCACTGCTTATTCAAAAACTTTCCAAGGTC
S.viridis        GCGCTCTACGTTTGGAGGATCTACGAATTCCCACTGCTTATTCAAAAACTTTCCAAGGTC
S.punctata       GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.madecassa      GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.humbertiana    GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.verticillata   GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
S.adhaerens      GCGCTCTACGTTTGGAGGATCTACGAATTCCCCCTACTTATTCAAAAACTTTCCAAGGTC
                 * ****************************** **  **** ******************
```

Fig. 1: Multiple Sequence Alignment (a region shown) of rbcL gene sequence from 16 *Setaria sp* showing regions of sequence variations (in boxes). *denotes the conserved nucleotides

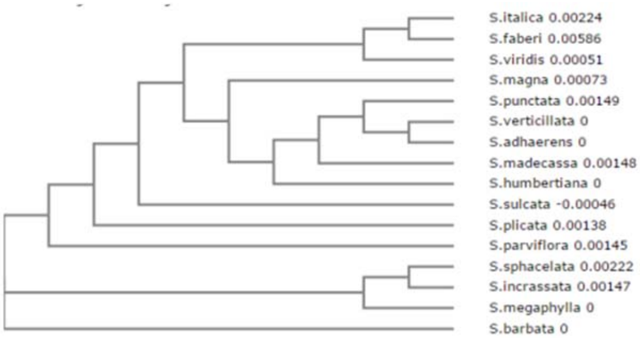

Fig. 2: Phylogenetic tree constructed using full length rbcL gene sequences of 16 *Setaria* species

4. ACKNOWLEDGEMENTS

This work was supported in by a grant from University of Delhi under the Innovation Project - GC 305.

REFERENCES

[1] Briand, L. C., Daly, J., and Wüst, J., "A unified framework for coupling measurement in objectoriented systems", *IEEE Transactions on Software Engineering*, 25, 1, January 1999, pp. 91-121.

[2] Brutnell, T. P., Bennetzen J., and Vogel, J. P., *Brachypodium distachyon* and *Setaria viridis*: model genetic systems for the grasses. *Annual review of plant biology* 66, 2015, pp. 465-485.

[3] Fan, L., Ma, J., Chen, Y. H., and Chen, X. Q., Antioxidant and antimicrobial phenolic compounds from *Setaria viridis. Chemistry of natural compounds*, 50, 2014, pp. 433-437.

[4] Warnasooriya, S. N., and Thomas P. Brutnell, T. P., Enhancing the productivity of grasses under high-density planting by engineering light responses: from model systems to feedstocks. *Journal of experimental botany* 65, 2014, pp. 2825-2834.

[5] Brutnell, T. P., Wang, L., Swartwood, K., Goldschmidt, A., Jackson, D., Zhu, X. G., Kellogg E., and Van Eck, J. (2010). *Setaria viridis*: a model for C4 photosynthesis. *The Plant Cell*, 22, 2010, pp. 2537-2544.

[6] Martins, P. K., Ribeiro, A. P., da Cunha, B. A. D. B., Kobayashi, A. K., and Molinari, H. B. C., A simple and highly efficient *Agrobacterium*-mediated transformation protocol for *Setaria viridis*. *Biotechnology Reports,* 6, 2015, pp. 41-44.

[7] Sarhan, S., F. Hamed, and Al-Youssef, W., The rbcL Gene Sequence Variations among and within *Prunus* Species." *Journal of Agricultural Science and Technology,* 18, 2016, pp. 1105-1115.

[8] Palmer, J. D., and Stein D. B., Conservation of chloroplast genome structure among vascular plants. *Current genetics,* 10, 1986, pp. 823-833.

[9] Demesure, B., Sodzi, N. and Petit, R. J., A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular ecology,* 4, 1995, pp. 129-134.

[10] Dumolin-Lapegue, S., Pemonge M. H., and Petit, R. J., An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular ecology,* 6, 1997, pp. 393-397.

[11] Group, CBOL Plant Working, et al. "A DNA barcode for land plants."*Proceedings of the National Academy of Sciences* 106, 2009, pp. 12794-12797.

[12] Cheng, Y., Zhou, K., Ma, X., Humphreys, M.M.W., Harper, J.A., Zhang, X., Yan, H. and Huang, L., Phylogenetic relationships in the *Festuca-Lolium* complex (Loliinae; Poaceae): New insights from chloroplast sequences. *Frontiers in Ecology and Evolution*, 4, 2016, p.89.