

Participating of CICR in the Cotton Genome Sequencing

- V.N. Waghmare

The first draft sequence of the cotton genome was unveiled in a scientific paper published in the world's highest rated scientific journal 'Nature' on 20th December by a team led by Dr. Andrew H Paterson, University of Georgia, USA. Dr. Vijay N. Waghmare, Principal Scientist, Central Institute for Cotton Research, Nagpur was associated in the research on genome sequencing. The paper was published by an international consortium of 74 scientists from 31 institutions across 11 countries. The team presented a high-quality Gold Standard draft assembly of the smallest cotton D genome (*Gossypium raimondii*) sequence. The team compared the draft D genome sequence to sequences of three other *Gossypium* species including that of *Gossypium hirsutum*, upland American cotton. The results of the study, allowed the researchers to discover how the fibre-producing plant evolved over a period of several million years, from wild form to the presently domesticated species.

The cotton, genus *Gossypium*, consists of 50 species of which four are cultivated for their spinnable fibre and the remaining 46 are wild species. Of the four cultivated species two are diploids ($2n=26$ chromosomes) viz *G. arboreum* and *G. herbaceum* and two are tetraploids ($2n=52$ chromosomes) viz *G. hirsutum* and *G. barbadense*. Tetraploid cotton, *Gossypium hirsutum* is the most widely cultivated (90%) among all *Gossypium* species. The cultivated diploid cotton has 'A' genome while the tetraploid species has 'A' and 'D' genome together, i.e., AD genome. In the tetraploid cotton, A genome comes from *Gossypium herbaceum*, a diploid fibre producing species, and D genome is contributed by *Gossypium raimondii* a wild diploid species that does not produce spinnable fibres. The haploid genome size of diploid cottons ($2n=2x=26$) varies from about 760 Mb (*G. raimondii*) in the D genome to 2,500 Mb in the K genome. The *Gossypium raimondii*'s D genome is the smallest with fewer repetitive elements than other *Gossypium* genomes, was chosen for sequencing.

Features of *Gossypium raimondii* draft genome: Across 13 chromosomes the genome size totaling 737.8 Mb (737.8 million base pairs), about 350 Mb (47%) consisted of a gene-rich region. A total of 37,505 genes and 77,267 protein-coding transcripts were annotated. The study revealed *G. raimondii* to be among the most complex of flowering plant genomes, experiencing atleast a 5-6 fold ploidy increase (duplication) of its genetic complement while formation and divergence of AD genome conferred 30-36 fold increase (duplication) from an ancestral flowering plant.

The cotton genome sequence will be invaluable to the cotton researchers. The finding will help expedite the development of robust and innovative cotton varieties to increase yields, fibre quality and can help optimize the use of inputs and, thus, contribute to reducing environmental impact of growing cotton, Dr. Paterson said, "The cotton genome data will help accelerate the study of gene function, particularly cellulose biosynthesis, understanding of which is fundamental to improved biofuels production," said Dr. Jeremy Schmutz, head of the DOE JPI Plant Program, USA, who led the effort to sequence and assemble the genome. In addition, the unique structure of the cotton fibre makes it useful in bioremediation, and accelerated cotton crop improvement also promises to improve water efficiency and reduce pesticide use". Dr. Don Jones, Director of Agricultural Research at Cotton Incorporated, said "this sequence is a cornerstone that will help advance our knowledge so we more thoroughly understand the biology that leads to enhanced yield, improved fibre quality, and better stress tolerance, all improvements that will benefit growers in the not-too-distant future".

The cotton genome sequence data has been made freely available in public domain with open access and researchers across the continent can use it in their research to improve sustainability of cotton production.