Exploring targeted genetic diversity in the core Indian germplasm of cotton using the Oxford Nanopore platform

Babita Singh# *, Shatrujeet Pandey#, Priti Prasad, and Samir Sawant

Molecular Biology and Biotechnology Division, CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow-226001, India.

Abstract

Global climate changes pose enormous challenges for plant breeders to maintain and further enhance yield in varying environments. Cotton is one of the most economically important crops in the world. Among the four cultivated cotton species (*Gossypium hirsutum*, *G. arboreum*, *G. herbaceum*, and *G. barbadense*), *G. hirsutum* contributes 90% to global production. Despite the larger area under cotton cultivation in India, productivity still needs significant improvement compared to other cotton-growing countries. In India, cotton is cultivated in three distinct agro-climatic zones: the North, Central, and South. These zones have several significant constraints and stress conditions like poor soil health, biotic (insects, pests, etc.), and abiotic stresses (drought, salinity, etc.). During the reproductive growth stage, yield reduction and fiber quality compromises are inescapable when biotic and abiotic stress conditions override the plant's protective mechanisms.

Next-generation sequencing and functional genomics studies on economically essential plants assume greater importance due to their ability to unravel the complexities of key traits of interest and harness such traits for crop breeding and improvement. Our group has identified a few critical genes that may play a vital role in cotton fiber development and drought tolerance in *Gossypium hirsutum*. We are exploring genetic variations of these potential genes through targeted sequencing of 320 core sets of *G. hirsutum* using Oxford Nanopore technology. We thus sequenced entire genomic loci of 24 targeted genes using ~1.5 Kb overlapping amplicon. Therefore, 104 amplicons encompassing 24 genes were sequenced in 320 genotypes to generate 120 Gb of raw data, which was an average of 700X coverage of raw data for each gene. Our data analysis identifies high-confidence SNP variations in the targeted genes. We have also performed a multi-location trial of 320 core germplasm of *Gossypium hirsutum* at three locations and collected high-quality phenotypic data related to agronomically important traits in cotton. Thus, associations of SNPs in targeted genes and phenotypic data are expected to provide insights into the genetic basis of fiber yield-related traits and stress tolerance in cotton, which will also help us develop SNP markers for molecular-assisted breeding of high-yield and stress resistance varieties under climate changes.

Objectives

- ➢ Targeted sequencing of potential genes associated to cotton fiber development and drought tolerance in the core Indian germplasm of *Gossypium hirsutum.*
- ➢ Analysis of genic diversity and their association to fiber development and drought tolerance.

Introduction

- Natural mutations have played crucial role in the domestication of many economical plants.
- Present work is focus on allele mining of many selected genes associated to some important trait of Indian cotton.
- Selected locus are related to root and shoot architecture, flowering, fiber yield and quality.

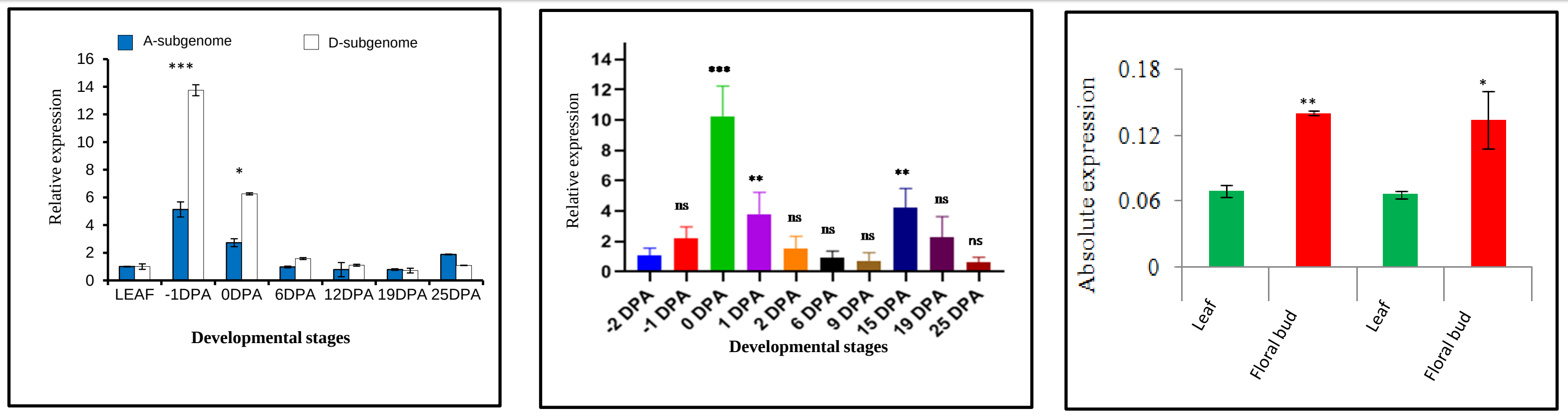
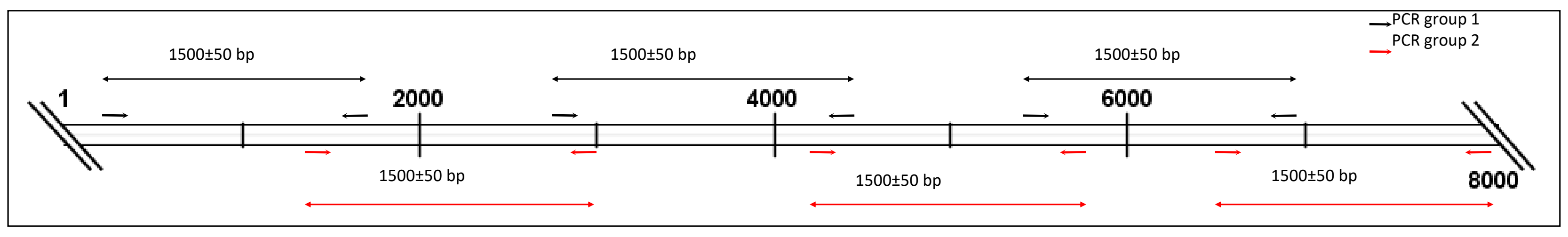


Figure1: Expression profiles of few genes selected for analysis of genetic diversity.

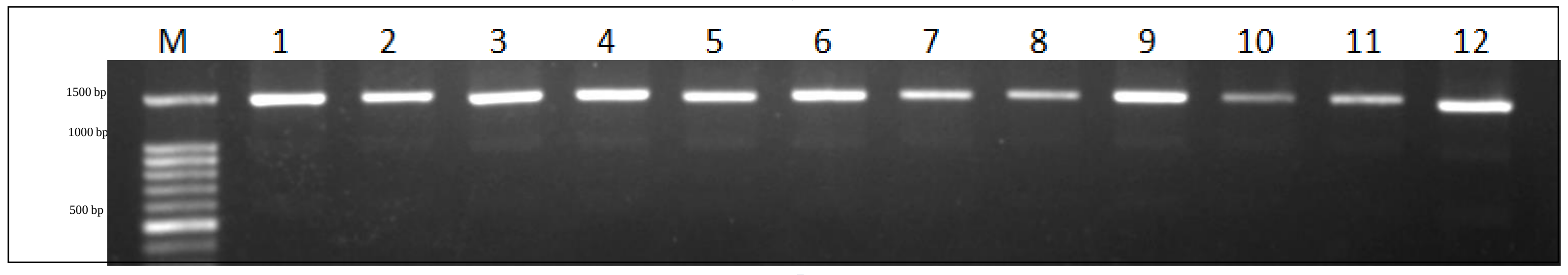
Methodology

Strategy to design the primer sets for targeted amplicon sequencing.



Genomic DNA isolation from all genotypes

Multiplex PCR using group 1 & 2 primer sets



PCR clean-up and Nanopore library preparation using Native Barcoding kit with amplicons

Targeted sequencing using PromethION Oxford Nanopore Sequencer

Data analysis using GATK pipeline and Inhouse script

Correlation analysis of variants with phenotypic data to identify the potential SNPs/INDELs that may play vital role in cotton fiber yield and drought tolerance

Results

Overview of sequencing status in cotton genotypes

	Genetic diversity
No. of genotype sequenced	249
No. of Amplicon for genes	3-8/gene
Size of amplicon	~1500 bp
Total coverage per genes	700X
Max. no. of variations in gene	33
Min. no. of variations in gene	1
No. of genotypes with variations	182

Identification of total number of variants in the genotypes and its distribution

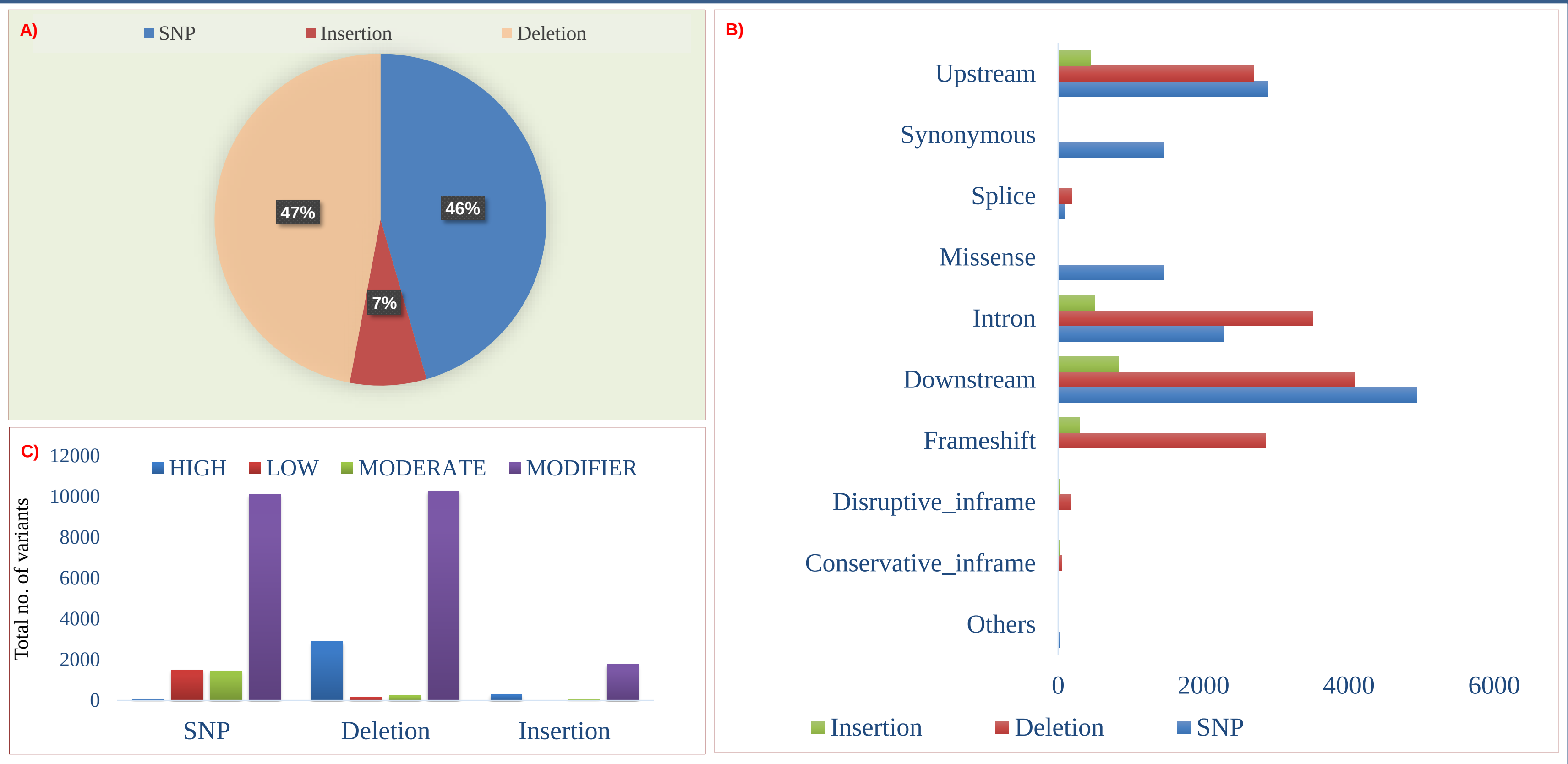


Figure 2: Targeted analysis of all 24 genes in core Indian germplasm. (A) Total number of variants, (B) Distribution of variants, and (C) Impact of variants on genes.

Multi-location field trials of Indian germplasm of *Gossypium hirsutum*

Location 1 (L1)

Location 2 (L2)

Location 3 (L3)

Correlation between genetic variants with cotton fiber and drought related-traits

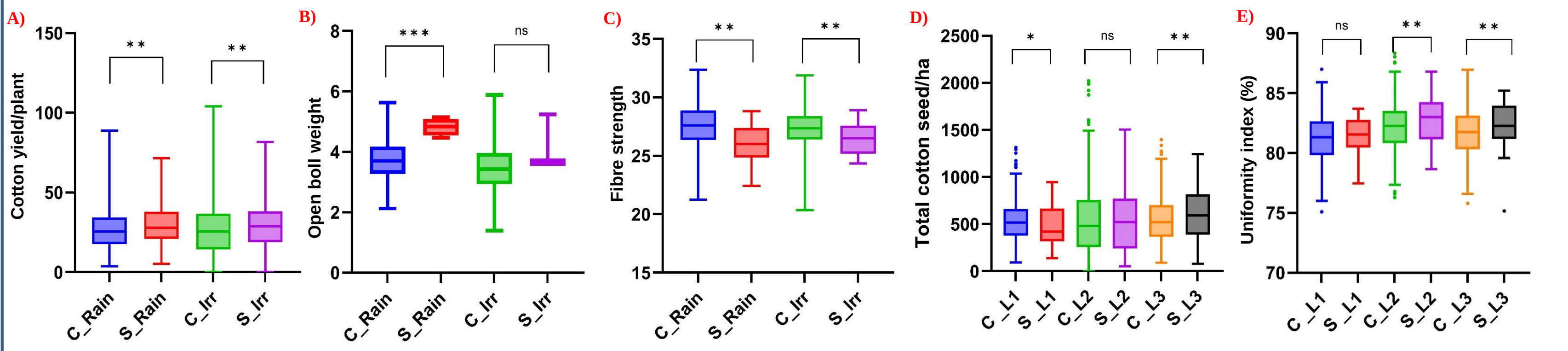


Figure 3: Correlation between SNPs and phenotypic data of multiple location. [C_Rain: control rain fed, S_Rain: SNPs rain fed, C_Irr: control irrigated, S_Irr: SNPs irrigated, C:control, S: SNPs.] (Asterisks indicate significant differences between the wild type and a given transgenic(one-way ANOVA; *, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$).)

Conclusions

- Associations of SNPs in targeted genes and phenotypic data revealed its involvement in fiber development and drought stress.
- Targeted sequencing is a potential technique to identified high coverage confidence variants at selected genes.

References

- Kumar et al., (2018) Role of GhHDA5 in H3K9 deacetylation and fiber initiation in *Gossypium hirsutum* (The Plant Journal).
- Karst et al., (2021) High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing (Nature methods).
- https://primalscheme.com/#protocol

Further Information

Dr. Samir V Sawant
(Chief Scientist)
CSIR-NBRI, Lucknow, INDIA
samirsawant@rediffmail.com
+91- 522-2297941

Dr. Babita Singh
Post-Doctoral Fellow
CSIR-NBRI, Lucknow, INDIA
babita.singh009@gmail.com

Dr. Shatrujeet Pandey
Post-Doctoral Fellow
CSIR-NBRI, Lucknow, INDIA

Dr. Priti Prasad
Scientist
Redcliffe, INDIA

Acknowledgement

The authors thanks to CSIR for providing fellowship, infrastructure and facilities at CSIR-NBRI, Lucknow to carry out the research work.