COMPARATIVE ASSESSMENT OF GENETIC DIVERSITY IN CYTOPLASMIC AND NUCLEAR GENOME IN IMPROVED GOSSYPIUM HIRSUTUM AND G. BARBADENSE COTTON LINES FROM DIVERSE GEOGRAPHICAL LOCATION

M. Ayubov
T. Norov
Z. Buriev
I. Abdurakhmonov
Center of Genomics and Bioinformatics, Academy of Sciences of Uzbekistan
Tashkent, Uzbekistan
S. Saha
D. Deng
J.N. Jenkins
USDA-ARS
Mississippi State, MS
B. Scheffler
USDA-ARS
Stoneville, MS

Abstract

Cotton is the leading textile fiber and the second most important oilseed crop in the world generating over $100 billion in USA national economy and directly support about 500,000 jobs. Upland cotton (*Gossypium hirsutum* L., AD_1, 2n = 52) is the most widely grown cotton species worldwide because of its superior yield potential, whereas *G. barbadense* L. (AD_2, 2n = 52), the only other cultivated tetraploid species, grown in some areas because of the price advantage due to superior fiber quality. The genetic improvement of any crop species depends on the extent of genetic variation for desirable alleles and the accurate characterization of the variability among germplasm accessions. Breeders normally select genotypes based on morphological characters, primarily regulated by nuclear genome. However, cytoplasmic genome, including mitochondria and chloroplast genomes, plays also an important role to perform many biological functions. The narrow genetic base of cultivated cotton germplasm is the primary impediment in its genetic improvement. The characteristics and mode of genetic inheritance of cytoplasm and nuclear genome provide complementary valuable information to study gene flow, evolution, and population history in any cotton breeding program. Unfortunately limited information is available on the genetic variation especially on cytoplasmic genome of the improved cotton lines of *G. hirsutum* and *G. barbadense*. The objectives of this research was to: 1) develop cytoplasmic DNA markers as a tool for genetic studies and 2) use the DNA markers as a tool to detect genetic variation in the cytoplasmic and nuclear genome in a set of improved *G. hirsutum* and *G. barbadense* lines.

We used 61 SSR primer pairs associated with important fiber specific traits of the nuclear genome and 49 primers specific to the cytoplasmic genome to screen 20 *G. hirsutum* and 74 *G. barbadense* improved lines from diverse geographical locations. We used the overall methods of standard technique of PCR and gel methods using an ABI 3730xl with a 96-capillary system for molecular analysis. The nuclear SSR primer pairs were selected based on the previous studies considering their presence covering almost the whole genome and association with important fiber traits. We scored the molecular marker data as a dominant marker to avoid ambiguous scoring for allelic relationship without pedigree data of the cotton lines in this experiment. The marker data were analyzed to estimate genetic similarity between cultivars based on the simple matching coefficient (SI) and constructed phylogenetic trees using JMP Genomics software (SAS^{TM}). Using a cost-effective data mining strategy we detected 27 CPSSR and 22 indel specific primer pairs from the cotton chloroplast genome using NCBI database. Results from the hybrids (F_1) of the reciprocal crosses between *G. hirsutm* (TM-1) and *G. barbadense* (Pima 3-79) confirmed the association of these markers specific only to the cotton cytoplasmic genome. We also observed that cytoplasmic indel markers are more polymorphic compared to the cytoplasmic SSR markers among the lines. The overall results from the dendrogram revealed that the selected 94 lines could broadly be divided into two broad groups of *G. hirsutum* and *G. barbadense*. However, the dendrogram results also showed that two accessions of *G. hirsutum* and two accessions of *G. barb*adense clustered with different species group respectively suggesting the interspecific introgression of gene pool in the development of these lines. The overall results suggested that the genetic diversity in this group of *G. hirsutum* is narrower in both cytoplasmic and nuclear genome compared to this group of *G. barbadense*. The average coefficient of similarity based on IBD value (identical alleles) in the cytoplasmic genome is 0.48 (S.E. ±

0.005). However, the nuclear genome average IBD value is 0.44 (S.E. ± 0.005) suggesting the genetic diversity in the cytoplasmic genome is little more conserved compared to the nuclear genome among the lines. The average IBD value is 0.42 (S.E. ± 0.005) combining both of the cytoplasmic and nuclear markers among the lines. Results from the dendogram pattern showed that the accessions in some cases from the similar breeding sources or geographic locations clustered together suggesting the use of similar in house gene pool in the breeding program. Genetic variation for desirable alleles for fiber traits and the accurate characterization of the variability in the targeted lines of interest are the foundation for any successful breeding program. This research provided for the first time a genetic tool to study cotton cytoplasmic genome and a report comparing the genetic diversity in the cytoplasmic and nuclear genomes of improved *G. barbadense* and *G. hirsutum* lines. Our research will help breeders to develop a breeding strategy maximizing the effects of genetic diversity to improve cotton lines.