Utilization of Natural Diversity in Upland Cotton (*G. hirsutum*) Germplasm Collection for Pyramiding Genes via Marker-assisted Selection Program

Ibrokhim Y. Abdurakhmonov*, Zabardast T. Buriev, Shukhrat E. Shermatov, Fakhriddin N. Kushanov, Abdusalom Makamov, Umid Shopulatov, Ozod Turaev, Tohir Norov, Chinpulat Akhmedov, Mukhammadjon Mirzaakhmedov, Abdusattor Abdukarimov

Center of Genomic Technologies, Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan,Yuqori Yuz, Kibray region Tashkent 111226, Republic of Uzbekistan, Phone: (998)712-69-18-32; Fax: (998) 712- 64-23-90; e-mails: genomics@uzsci.net*

Abstract

Cotton is the world's leading cash crop, but it lags behind other major crops for marker-assisted breeding that underlies a need for characterization, tagging, and utilization of existing natural polymorphisms in cotton germplasm collections. Previously, we conducted molecular genetic analyses in a global set of ~1000 *G. hirsutum* accessions from Uzbek cotton germplasm collection, representing, at least, 37 cotton growing countries and 8 breeding ecotypes as well as wild landrace stock accessions. The important fiber quality (fiber length and strength, micronaire, uniformity, reflectance, elongation and ect.) traits were measured in two distinct environments of Uzbekistan and Mexico. This study allowed us to design an "association mapping" (mixed liner model-MLM) study to find biologically meaningful marker-trait associations for important fiber quality traits in Upland cotton that accounts for population confounding effects. Several SSR markers associated with main fiber quality traits along with donor accessions were selected to be used for marker-assisted selection (MAS) programs. In this study, utilizing our previous results of association mapping in Uzbek cotton germplasm resources, with specific objective of introducing and enriching the currently-applied traditional breeding approaches with more efficient modern MAS tools in Uzbekistan, we began marker-assisted selection efforts using molecular markers associated with important fiber traits. For this purpose, we selected 1) major fiber quality trait – associated SSR markers and 2) donor genotypes that were identified in our previous studies. We selected 23 major (micronaire, fiber strength and length, and elongation) fiber trait-associated DNA markers as a tool to control the transferring of QTL loci during a genetic hybridization. We selected 37 (11 wild race stocks and 26 variety accessions from diverse ecotypes) donor cotton genotypes that bear important QTLs for fiber traits. These donor genotypes were crossed with 9 commercial cultivars of Uzbekistan (as recipients) in various combination with objective of improving one or more of fiber characteristics of these recipients. These 9 parental recipient genomes preliminarily were screened with our DNA-marker panel to compare with 37 donor genotypes. The polymorphic states of marker bands between donor and recipient genotypes were recorded. The subsequent generation of hybrid plants from each crossing combination were tested using DNA-markers at the seedling stage and hybrids bearing DNA-marker bands from donor plants were selected for further backcross breeding. Testing the major fiber quality traits using HVI in trait-associated marker-band-bearing hybrids revealed that mobilization of the specific marker bands from donors really have positively improved the trait of interest in recipient genotypes. Currently, we developed a second generation of recurrent parent backcrossed hybrids (F_1BC_2), bearing novel marker bands and having superior fiber quality compared to original recipient parent (lacking trait-associated SSR bands). These results showed the functionality of the trait-associated SSR markers detected in our association mapping efforts in diverse set of Upland cotton germplasm. Using these effective molecular markers as a breeding tool, we aim to pyramid major fiber quality traits into single genotype of several commercial Upland cotton cultivars of Uzbekistan.

Introduction

The goal of many breeding programs is to mobilize a gene, or genes, from a donor parent into an elite parent, usually through conventional breeding methods. Although traditional breeding methods are the best and efficient tool with working on single gene-controlled, organoleptic and qualitative traits, they are less efficient, costly and time consuming in regards to breeding of complex quantitative traits controlled by multiple genes. DNA-based molecular markers and results of quantitative trait loci (QTL) mapping are being used extensively to identify and track regions of the genome in introgression programs, in order to identify individuals that have genome compositions that are significantly better than would be expected (Tanksley et al., 1989; Abdurakhmonov, 2002). This is referred as a marker-assisted selection (MAS). DNA markers linked to genomic regions of interest enable breeders to select individuals on the basis of genotype rather than phenotype. This is very helpful if trait of interest is complex and time-consuming to score, as it is the case with all quantitative traits. Using molecular markers in marker-assisted selection of crops may revolutionize the process of elite cultivar development reducing field-tests in early crop breeding and cutting the requisite time in less then half. DNA-markers are very informative to select individuals or lines with crossovers very near targeted gene. This way breeder can remove “linkage drag” that frequently comes from a donor parent (Zeven et al., 1983; Abdurakhmonov, 2002). DNA-markers can also be used to select for individuals having a minimal donor parent germplasm not linked to trait of interest since such chromosomal segments are often associated with undesirable traits (Young and Tanksley, 1989; Abdurakhmonov, 2002). Moreover, desirable alleles in wild crop relatives that having less overall phenotypes also can be identified using DNA markers. Then such transgressive loci of wild species can be selected and used to create new cultivars with superior phenotypes introducing useful variation to agricultural crops (deVicente and Tanksley, 1993).

To carry out a marker-assisted selection a sufficient number of polymorphic markers must be identified throughout the genome for the whole genome to be assayed. Little map-based information is required; a marker need only be scored as informative between the parents used in the cross, and this can then be used to score a segregating population for the presence or absence of that genetic marker. The benefits obtained from genetic selections can be maximized by increasing genetic pools, so that individuals with exceptional genotypes can be identified. Likewise, expanding the number of markers employed will proportionally increase the confidence in the estimate of genome composition (Abdurakhmonov, 2002). In order to apply molecular breeding to large-scale breeding programs, automation technologies must be introduced. Molecular marker analysis in large genetic pools requires fast assays that can be automated based on currently available DNA amplification-based technologies. The steps in amplification-based assays that have been automated include DNA extraction, purification, and quantitation, DNA amplification and analysis, as well as data acquisition and analysis. Thus, the utility of molecular markers extends throughout all phases of plant breeding programs, from trait identification to trait introgression. The economic benefits have been most evident in marker assisted selection programs, but this application requires large-scale efforts and automation strategies. The development of high-density DNA marker maps will increase the efficiency of quantitative trait mapping and thereby facilitate the introgression of more complex traits (Abdurakhmonov, 2002).

Modern genomics based marker-assisted selection technology is being applied in many crops (e.g. see), but being the world’s leading cash crop, cotton lags behind other major crops for marker-assisted selection (MAS) due to limited polymorphisms and 'a genetic bottleneck' through historic domestication. A challenging problem for cotton fiber quality improvement is that the Upland cotton germplasm base originated from the cross of a small number of ancestral elite lines (Abdalla et al., 2001). Continuous selection among crosses of genetically related elite cultivars has led to a narrow genetic base and erosion of the important cotton gene pools for agronomic and fiber traits in Upland cotton (Bowman et al., 1996; Bowman et al., 2003; Van Esbroeck et al., 1998; Van Esbroeck et al., 1999). Although the shallow elite gene pools have provided some initial genetic gains, they have been accompanied by a decreased genetic diversity within the elite gene pool, increased genetic uniformity in improved cotton lines, and erosion of exotic genetic resources in the germplasm of Upland cotton. Hence, cotton researchers and producers worldwide are concerned with the narrow genetic base of cultivated cotton germplasm that caused recent cotton yield and quality declines. These declines, however, is largely due to challenges and the lack of innovative tools to effectively exploit the genetic diversity of *Gossypium* species because the amplitude of genetic diversity of cotton genus is exclusively wide. This requires broadening the genetic diversity of cultivar germplasm to meet future sustainable cotton production through characterization, tagging, and utilization of existing natural polymorphisms in cotton germplasm collections. This could be done effectively and within a short-time only by using modern genomics technologies such as gene characterization and genetic mapping, which are the prerequisites for successful genetic engineering and MAS.

Although a huge genomics resources are developed and advances in cotton genomics are made (Chen et al., 2007; Zhang et al., 2008; Abdurakhmonov, 2007) there is no report for successful utilization of MAS for complex traits like fiber quality in cotton. Molecular markers, found to be associated with the traits in traditional QTL mapping experiments and within two genotype backgrounds, are most often time fail to be useful in consequent breeding efforts. To increase the reliability and power of trait-associated molecular markers, it requires performing a genetic mapping in a large sample of genotypes (e.g. germplasm collections covering many meiotic events to fix useful gene allele) segregating for trait(s) of interest. This underlies turning the gene-tagging efforts from bi-parental crosses to germplasm collections, and from traditional linkage mapping to linkage disequilibrium (LD)- based association study to provide the most effective utilization of *ex situ* conserved natural genetic diversity of worldwide cotton germplasm resources (Abdurakhmonov and Abdukarimov, 2008; Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009).

In order to address these issues, for the past decade, within the frame of multi-institutional international collaborations, we have established one of the leading cotton genomics and biotechnology program in Uzbekistan, organized a modern genomics research facility and prepared young generation of cutting-edge genomics scientists and molecular breeders to bridge up molecular and traditional breeding efforts in Uzbekistan. To better assess, understand and exploit a molecular diversity of Upland cotton genome, we conducted molecular genetic analyses in a global set of ~1000 *Gossypium hirsutum* L. (so called Upland cotton) accessions, one of the widely grown allotetraploid cotton species, from Uzbek cotton germplasm collection. This global set represented, at least, 37 cotton growing countries and 8 breeding ecotypes as well as wild landrace stock accessions. The important fiber quality (fiber length and strength, micronaire, uniformity, reflectance, elongation and ect.) traits were measured in two distinct environments of Uzbekistan and Mexico (Abdurakhmonov et al., 2006; Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009; Abdurakhmonov et al., 2010). This study allowed us to quantify the linkage disequilibrium level in Upland cotton germplasm genome and to design an “association mapping” study to find biologically meaningful marker-trait associations (Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009; Abdurakhmonov et al., 2010) for important fiber quality traits that accounts for population confounding effects. Several SSR markers associated with main fiber quality traits along with donor accessions were identified and selected for MAS programs. In this paper, we report our initial effort to utilize MAS technology to improve fiber quality traits in Upland cultivars of Uzbekistan. Our efforts should be useful for broadening the genetic diversity level of commercialized cultivars, accelerating breeding efforts and quality of cotton cultivar improvement through introgression and pyramiding of novel haplotypes of “still underutilized” fiber quality trait associated QTLs in cotton genome.

Materials and Methods

Based on our previous analyses of a global set of ~1000 Upland cotton germplasm accessions and association mapping results (Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009; Abdurakhmonov et al., 2010), we have selected several unique, genetically very distant, superior quality accessions as donor lines. At the same time, we selected several Upland cultivars as recipient genotypes that were commercialized in Uzbekistan and are being widely grown by Uzbek farmers, aiming to improve one of more fiber quality traits of theses commercialized cultivars through introducing novel QTLs from donor lines via MAS. Several fiber quality trait associated SSR markers from our association mapping efforts in wild landrace stock Upland cotton germplasm (Abdurakhmonov et al., 2008) as well as in cultivar germplasm (Abdurakhmonov et al., 2009) were selected for MAS program, based on LD information and their trait association values in two very distinct environments. Polymorphic states of selected markers between donor and recipient genotypes were studied and recorded for monitoring of the consequent genetic hybridization. DNA isolation, PCR-amplification and checking polymorphisms were conducted according methodology described in our previous publications (Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009).

Based on polymorphic states, we created a platform and hybridization scheme to mobilize superior fiber quality QTLs from donor to recipient genotypes by means of monitoring with fiber trait associated molecular markers. We also selected several additional SSR markers flanking the targeted LD haplotype blocks, associated with fiber quality traits, in order to use them in verification of transferring the minimal genomic regions associated with the trait of interest, minimizing the linkage drag. Different combination of genetic crosses were made between selected donor and recipient genotypes in greenhouse condition. F_0 seeds were collected and germinated to get F_1 plants. Immediately at the seedling stage, F_1 plants were screened using pre-determined SSR markers detecting the QTL of our interest and only trait associated marker-band-bearing hybrids were selected from seedlings for future growing. Using recurrent parent backcrossing approach, we further investigated transfer and

fixation of novel QTLs in recipient genotypes, continuously monitored with fiber trait-associated SSR markers from pre-determined marker panel. The fiber quality characteristics in some of BC generations were tested using High Volume Instrumentation (HVI) test at the fiber testing center “SIFAT”, Tashkent, Uzbekistan.

Results and Discussions

One of the main objective of cotton breeding programs worldwide to improve middle fibered Upland cotton (*G. hirsutum*) fiber quality close to long-staple Pima cotton(*G. barbadense*) fiber quality, bringing more income for farmers (Figure 1). However, improvement of Upland cotton fiber quality through inter-species genetic hybridization with Pima cottons is challenging due to wide genetic segregation and genetic distortion of consequent hybrid generations that has a minimal success. To overcome this genetic obstacle, one of the best and wise alternative approach is to investigate, understand and exploit the genetic potential of *ex situ* conserved Upland germplasm resources that would help to improve fiber quality characteristics of Upland cultivars meeting the global fiber market needs. Besides, targeting fiber quality improvement in Upland cotton is vital because historically cultivar development has focused more on yield than fiber quality. Fiber quality has become a major issue because of: 1) the technological changes in the textile industry 's demand for high quality fibers for maximizing production efficiency and quality, and 2) the occurrence of price discount for unfavorable fiber quality.

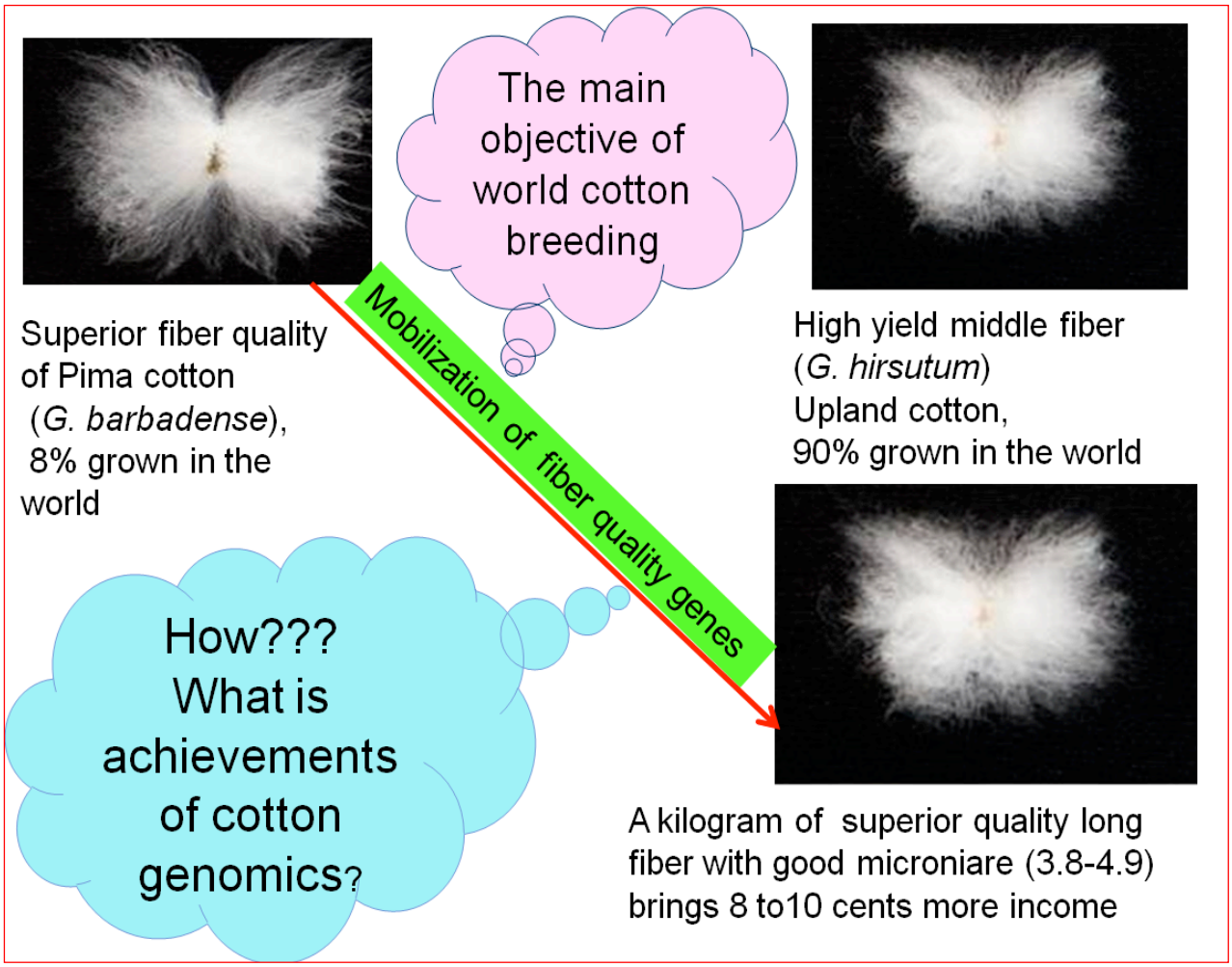

Figure 1. The main objective of cotton breeding programs worldwide

Therefore, in this study, we initiated the first MAS efforts using molecular markers associated with important fiber traits in Uzbekistan. We applied fiber-trait associated molecular markers (Figure 2), identified and validated in our large-scale association mapping efforts across a diverse set of germplasm (Figure 3), broadly covering many meiotic events (Abdurakhmonov et al., 2008; Abdurakhmonov et al., 2009), to improve fiber quality of Uzbek cultivars using MAS.

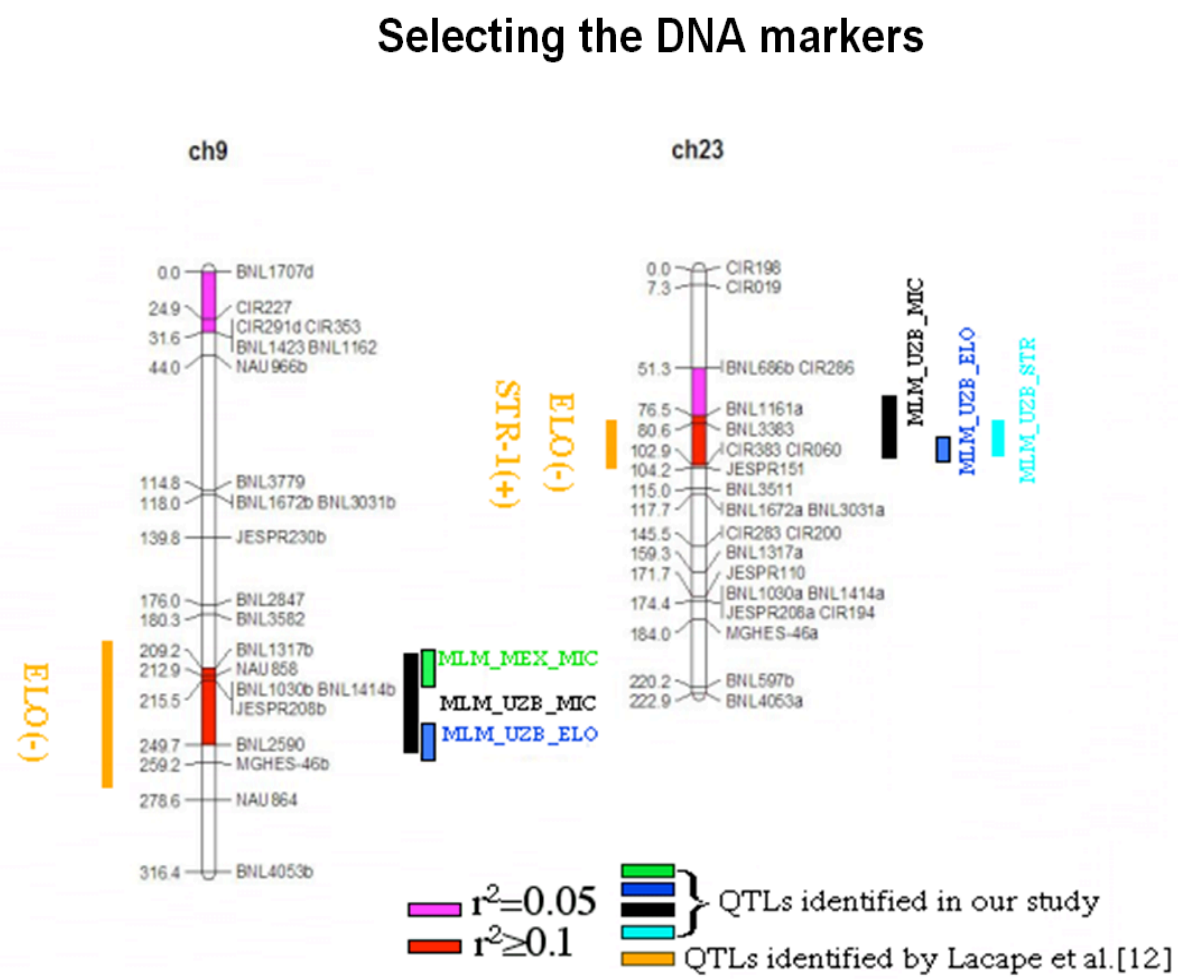

Figure 2. Example of selecting SSR marker for MAS efforts

For this purpose, we selected 1) major fiber quality trait – associated SSR markers and 2) donor genotypes that were identified in our previous studies. We selected 23 major (micronaire, fiber strength and length, and elongation) fiber trait-associated DNA markers as a tool to control the transferring of QTL loci during a genetic hybridization (Figure 2). Thirty seven donor cotton genotypes, including 11 wild race stocks and 26 variety accessions from diverse ecotypes, bearing important QTLs for fiber traits, were selected based on genetic diversity estimates shown in Figure 3. These donor genotypes were crossed with 9 commercial cultivars of Uzbekistan (as recipients) in various combination with objective of improving one or more of fiber characteristics of these recipients. These 9 parental recipient genomes preliminarily were screened with our DNA-marker panel to compare with 37 donor genotypes. The polymorphic states of marker bands between donor and recipient genotypes were recorded.

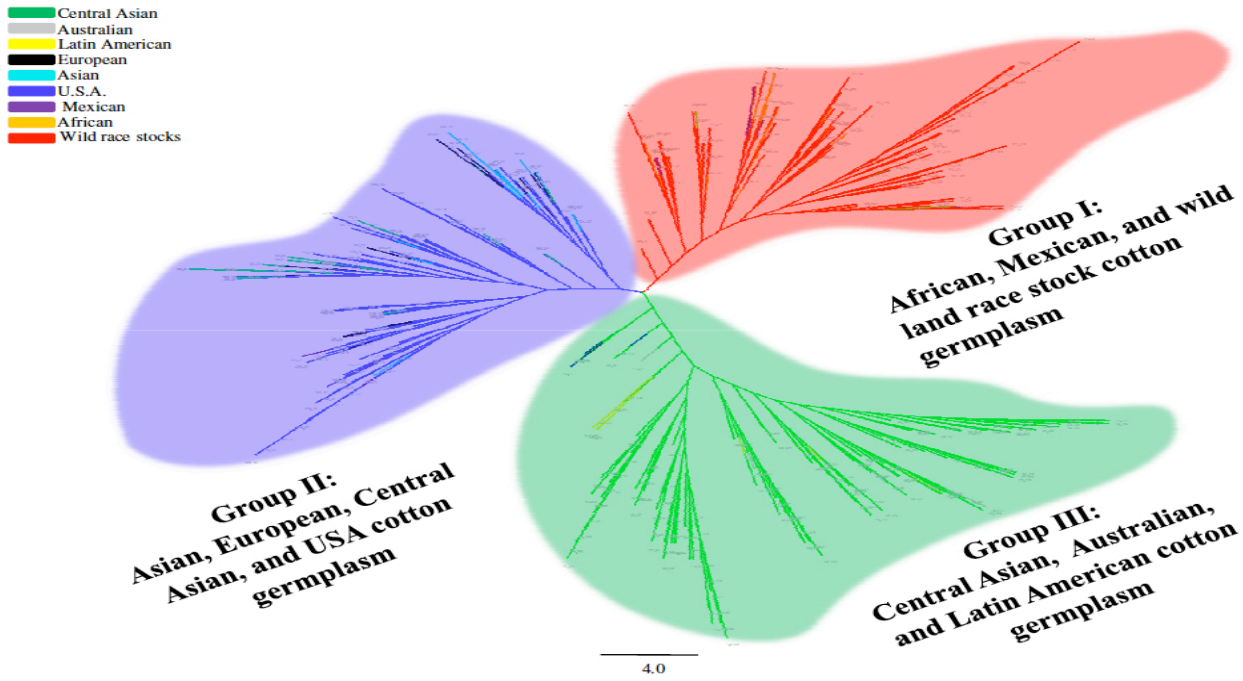

Figure 3. Molecular diversity analysis of a global set of Upland cotton germplasm using SSR markers

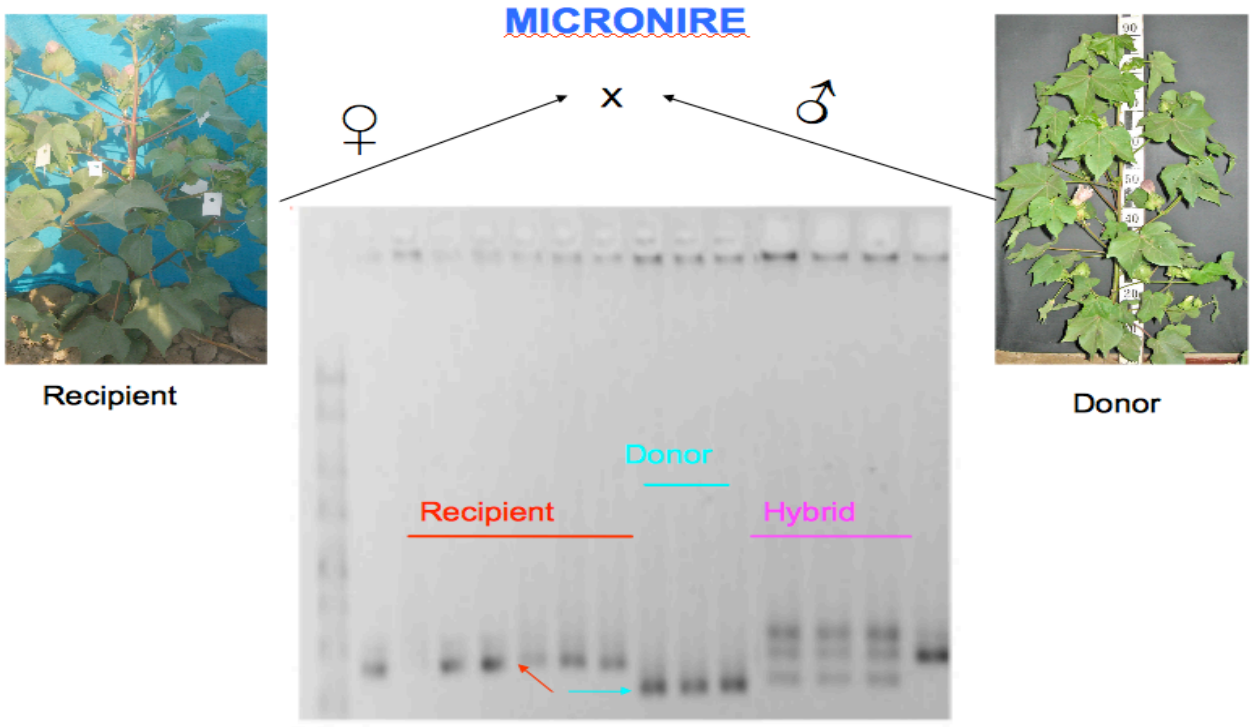

Figure 4. Example of breeding for marker assisted backcrossing for mobilization of micronaire trait from donor to recipient.

The subsequent generation of hybrid plants from each crossing combination were tested using DNA-markers at the seedling stage and hybrids bearing DNA-marker bands from donor plants were selected for further backcross breeding (Figure 4). Testing the major fiber quality traits using HVI in trait-associated marker-band-bearing hybrids revealed that mobilization of the specific marker bands from donors really have positively improved the trait of interest in recipient genotypes. Currently, we developed a second generation of recurrent parent backcrossed hybrids (F_1BC_2), bearing novel marker bands and having superior fiber quality compared to original recipient parent (lacking trait-associated SSR bands). These results showed the functionality of the trait-associated SSR markers detected in our association mapping efforts in diverse set of Upland cotton germplasm. Using these effective molecular markers as a breeding tool, we aim to pyramid (Figure 5) major fiber quality traits into single genotype of several commercial Upland cotton cultivars of Uzbekistan.

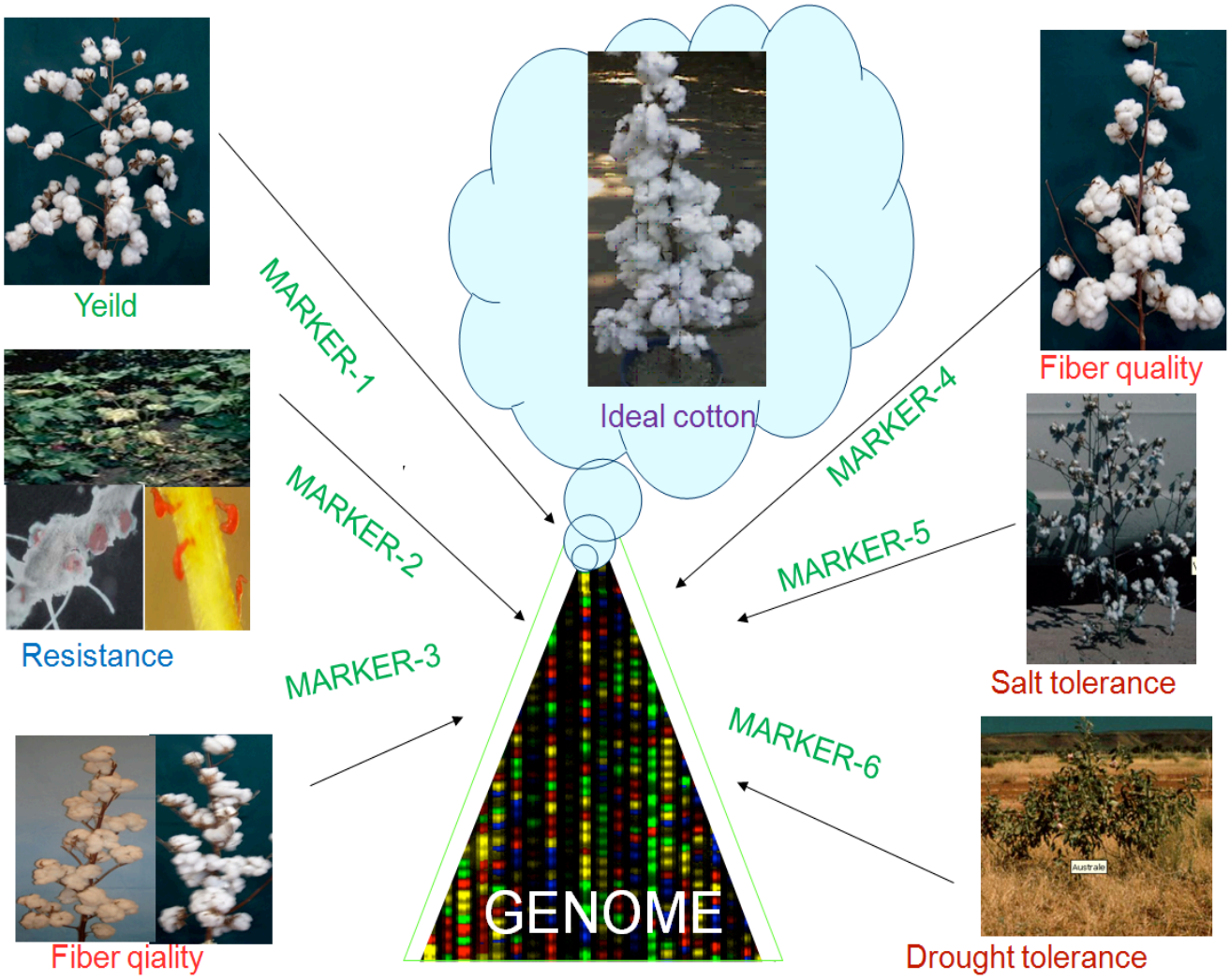

Figure 5. Pyramiding of important quantitative trait loci using MAS efforts

Conclusions

In conclusion, our initial efforts toward utilization of MAS technology to improve one or more fiber quality traits of several commercialized Upland cultivars of Uzbekistan, based on a global analysis of molecular diversity and association mapping of the main fiber quality traits in Upland cotton genome, suggest the possibility of development a successful innovative MAS tools to enrich current traditional breeding programs in the country. Our efforts will help to rapid introgression of novel polymorphisms, broadening the genetic diversity of cotton cultivars and accelerating the breeding efforts to develop superior cotton cultivars for future sustainable cotton production in Uzbekistan. To achieve the ultimate goal, we are extensively working on getting more higher generation of recurrent parent backcrossed MAS hybrids. Future steps also include to 1) select the best, stable generation of MAS lines with introgression of minimal genomic regions for targeted QTLs using molecular markers, 2) pyramid several fiber quality QTLs into single genotype via multiple crosses between stabilized higher generation MAS hybrids of the same genotype background and continuous monitoring with trait associated molecular markers, 3) increase seeds of the best and genetically stable MAS lines, and 4) test them in different cotton growing environments of Uzbekistan. Our efforts also promotes and underlies the preparation of young generation of cutting-edge genomics scientists and molecular breeders to bridge up molecular and traditional breeding efforts in Uzbekistan. The development of simple, easy-to-use MAS platform for current cotton breeders, creation of an infrastructure to perform MAS programs and education of traditional breeders on these new genomic technologies will be a vital part of our efforts. If successful, our efforts will be a platform to demonstrate usefulness of development modern science and technologies in developing countries in order to not only understand a modern genomics science of crops, but also to successfully apply it for production of economic value.

Acknowledgments

This research funded by Academy of Uzbek Sciences applied research grant # FA-F10-T74. We thank to Dr. Rafiq Chaudhry, International Cotton Advisory Committee as well as organizing committee of the 5th meeting of Asian Cotton Research and Development Network, for an opportunity to present these results in this conference.

References

Abdalla, A.M., O.U.K. Reddy, K.M. El-Zik, and A.E. Pepper. 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor. Appl. Genet. 102:222–229.

Abdurakhmonov I.Y., S. Saha, J.N. Jenkins, Z.T. Buriev, S.E. Shermatov, B.E. Scheffler, A.E. Pepper, J.Z. Yu, R.J. Kohel, and A. Abdukarimov. 2009. Linkage disequilibrium based association mapping of fiber quality traits in G. *hirsutum* L. variety germplasm. Genetica 136:401-417.

Abdurakhmonov, I.Y. 2002. Molecular cloning of new DNA-markers for marker-assisted selection of cotton. PhD dissertation. Institute of Genetics and Plant Experimental Biology, Academy of Uzbek Sciences, Tashkent, Uzbekistan.

Abdurakhmonov, I.Y. 2007. Exploiting genetic diversity. In: Ethridge D (ed). Plenary Presentations and Papers. Proceedings of World Cotton Research Conference-4; 2007 Sept 10-14; Lubbock, TX USA.

Abdurakhmonov, I.Y., and A. Abdukarimov. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genom. 2008:Article 574927.

Abdurakhmonov, I.Y., Z.T. Buriev, S.E. Shermatov, A. Abdukarimov, S. Saha, J.N. Jenkins, R.J. Kohel, J.Z. Yu, A.E. Pepper. 2010. Molecular diversity and population structure analysis in a global set of G. hirsutum exotic and variety germplasm resources and association mapping f the main fiber quality traits. S10. pg.22. Proceedings of International Cotton Genome Initiative conference, Canberra, Australia, 2010.

Abdurakhmonov, I.Y., R.J. Kohel, J.Z. Yu, A.E. Pepper, A.A. Abdullaev, F.N. Kushanov, I.B. Salakhutdinov, Z.T. Buriev, S. Saha, B.E. Scheffler, J.N. Jenkins, and A. Abdukarimov. 2008. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 98: 478-487.

Abdurakhmonov, I.Y., Z.T. Buriev, I.B. Salakhuddinov, S.M. Rizaeva, A.T. Adylova, S.E. Shermatov, A. Adukarimov, R.J. Kohel, J.Z. Yu, A.E. Pepper, S. Saha, and J.N. Jenkins. 2006. Characterization of *G. hirsutum* wild and variety accessions from Uzbek cotton germplasm collection for morphological and fiber quality traits and database development. p. 5306. In Proc. Cotton Beltwide Conf., San Antonio, TX. 3–6 Jan. 2006.

Bowman, D.T., O. L. May, and D. S. Calhoun.1996. Genetic base of upland cotton cultivars released between 1970 and 1990. Crop Sci 36:577-581.

Bowman, D.T., O. L. May, and J. B. Creech. 2003. Genetic uniformity of the U.S. upland cotton crop since the introduction of transgenic cottons. Crop Sci 43:515-518.

Chen, Z.J., B. E. Scheffler, E. Dennis, B. A. Triplett, T. Zhang, W. Guo et al., 2007. Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol 145:1303-1310

deVicente, M. C, S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics. 134:585–96

Tanksley, S. D., N. D. Young, A. H. Paterson, M. W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–64

Van Esbroeck, G. A., D.T. Bowman, D. S. Calhoun, and O. L. May. 1998. Changes in the genetic diversity of cotton in the USA from 1970 to 1995. Crop Sci. 38:33-37.

Van Esbroeck, G. A., D.T. Bowman, O. L. May and D. S. Calhoun. 1999. Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. Crop Sci. 39:323-328.

Young, N. D., S. D. Tanksley. 1989. RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353–59.

Zeven, A. C., D.R. Knott, R. Johnson. 1983. Investigation of linkage drag in near isogenic lines of wheat by testing for seedling reaction to races of stem rust, leaf rust and yellow rust. Euphytica 32:319–27

Zhang, H. B., Y. Li, B. Wang, P. W. Chee. 2008. Recent advances in cotton genomics. Int J Plant Genomics 2008:742304.