KASPar: AN EFFICIENT AND COST-EFFECTIVE TECHNOLOGY FOR SNP GENOTYPING IN COTTON

Ramesh Buyyarapu
Siva Kumpatla
Navin Elango
Wei Chen
Thomas W. Greene
Dow Agrosciences
Indianapolis, IN

Abstract

Cotton is the world's most cultivated natural fiber and oil seed crop. Lack of reference genome sequence and tetraploid nature of cultivated cotton pose numerous challenges to develop SNP markers. While *in silico* SNP detection algorithms such as AutoSNP and QualitySNP can provide candidate SNP information based on SNP quality and confidence scores, validation of these SNPs is still an expensive process. In order to reduce the cost of SNP validation and genotyping in cotton, we have utilized KASPar (KBiosciences, Hoddesdon, UK), a cost-effective single-plex SNP genotyping technology. KASPar assays were designed for 150 candidate SNPs using Primer Picker software (http://www.kbioscience.co.uk/primerpicker/). The Assay consists of three unlabeled interrogation primers that include two allele specific primers to target SNP and one common primer. KASPar data obtained from a FRET capable plate reader with relevant filters for FAM and VIC fluorescence were plotted to determine the genotype clusters. The genotyping results and efficacy of this cost-effective SNP genotyping system for marker development and marker-assisted selection in crops are discussed.

Introduction

Cotton (*Gossypium* spp) has been the genus of interest for several years to study polyploidization. *G. hirsutum* (A_tD_{t1}) and *G. barbadense* (A_tD_{t2}) are widely cultivated allotetraploids that account for 90% of world production, and are known for higher productivity, and higher fiber quality respectively. The new world tetraploid species arose ~1-2 million years ago (mya) through the hybridization of an old world taxon of the 'A genome' cytogenetic group related to the present-day species *G. herbaceum* and *G. arboreum* (2n=2x=26), with a taxon of the 'D-genome' group related to the new world species *G. raimondii* Ulbrich and *G. gossipioides* L. (2n=2x=26) (Wendel and Cronn, 2003).

Germplasm diversity is the focus of plant breeders as they rely on genetic variation between parents to create unique cultivars that are superior to their parents. An understanding of the genetic and genomic relationships of cotton species and cultivars is critical for further utilization of diversity in the development of improved cultivars that combine favorable characteristics such as fiber quality and yield traits, insect and pathogen resistance, and tolerance to environmental stresses. Molecular marker technology provides opportunities for genome characterization and crop improvement through germplasm diagnostics, introgression and marker assisted selection (MAS). Several molecular marker technologies have been used to study the genetic diversity and relationships of *Gossypium* species (Zhang et al., 2008). However, cotton crop improvement is limited by its narrow genetic base and limited variation among the cultivated cotton cultivars.

To generate sufficiently dense genetic maps for complex trait mapping, efforts are currently being focused on identifying the most common type of DNA sequence variation, single nucleotide polymorphisms (SNPs). SNPs are single-base variations at a unique physical location. SNPs are increasingly becoming the markers of choice in genetic analysis and are used routinely as markers in agricultural breeding programs (Gupta; et al., 2001). Tetraploid nature of the cultivated cotton species and lack of reference genome sequences present numerous challenges for SNP marker development and genotyping. In this study, we have identified 150 *in silico* SNPs from publicly available sequences and used KASPar, an efficient and cost-effective technology to genotype these markers.

Materials and Methods

Using the EST sequence information from cultivated tetraploids (*G. hirsutum* and *G. barbadense*) and their progenitors (*G. arboreum* and *G. raimondii*) we have identified candidate contigs based on intraspecific and

interspecific variation at single nucleotide level. Initially 150 allele-specific primers suitable to KASPar assays were designed using PrimerPicker software. KASPar assays were conducted on a panel of 23 genotypes that included five different species, i.e. *G. hirsutum* (15), *G. barbadense* (5), *G. arboreum* (1), *G. herbaceum* (1) and *G. raimondii* (1) using Bio-Rad Tetrad2 thermal cyclers as per manufacturer's recommended PCR conditions for 40 cycles. KlusterCaller software was used to make genotype calls using the raw fluorescence data derived from TECAN, FRET capable plate reader for FAM (485/520nm) and VIC (520/570nm) fluorophores.

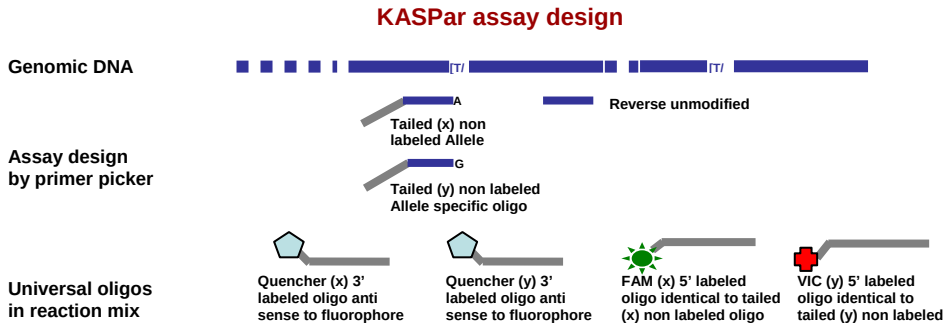


Fig 1. KASPar Assay design and chemistry

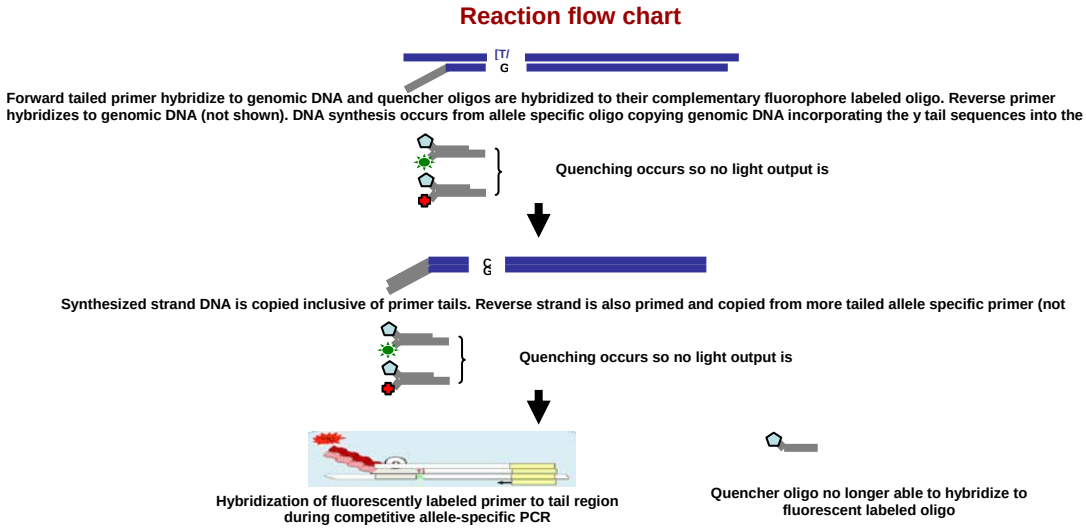


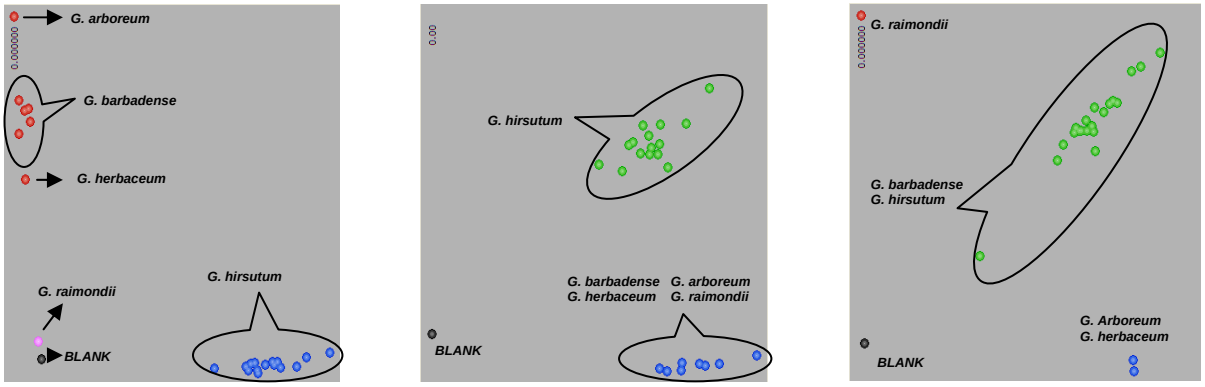
Fig 2. KASPar Assay flowchart

Results and Discussion

Recent advances in genotyping technologies facilitated rapid increase in the use of SNP markers for crop improvement programs. High-throughput SNP genotyping platforms such as Illumina GoldenGate assays are currently used in marker development and application projects for various crops. While the high throughput SNP genotyping platforms are very useful for rapid development of SNP markers, they are not economical for projects such as *in silico* SNP validation, gene-specific SNP assays, marker saturation in the regions of interest and marker application projects that utilize smaller number of SNP markers on varying number of samples. KBiosciences' PCR SNP genotyping system (KASPar®), a homogeneous fluorescent endpoint genotyping system, is very attractive and cost-effective platform for this type of applications. This technology utilizes a unique form of competitive allele-specific PCR to determine the allele at a specific locus within genomic DNA for SNP genotyping. It uses an

innovative fluorescent reporting system in assay mix that comprises two fluorescently-labeled primers and complementary quencher oligos to hybridize one another in free solution to form a fluorescent quenched pair. Unlabeled allele-specific (forward) and a common reverse primer(s) are ligated with tails complementary to fluorescently labeled oligos in assay mix at the 5' end (Robinson, Fig. 1). These unlabeled primers start PCR in the target regions in initial cycles and upon introduction of complementary sequences, fluorescent primers in assay mix lead the PCR process to generate a measurable signal that can be read by FRET (Fluorescent Resonance Energy Transfer) capable plate readers (Robinson, Fig. 2). These data are assigned allele-specific clusters using KlusterCaller (KBiosciences) software. Despite the challenging tetraploid nature of the cotton for SNP validation, we were able to successfully use KASPar assays to distinguish true, hemi and paralogous SNPs (Fig. 3).

SNP genotyping results revealed 48% polymorphism among the five species used in this study. A total of 37.5% markers were polymorphic in two tetraploid species i.e., *G. hirsutum* vs. *G. barbadense*. SNP markers were categorized into true SNPs (single locus); hemi SNPs (multi locus) and paralogous SNPs (derived from A and D genomes) (Fig. 3). These SNPs had been annotated to discover their putative functional significance and these involve transcription factors, leucine rich-repeat genes, membrane proteins, and other protein kinases. These SNP markers will be further used for mapping purposes.



True SNP Markers **Hemi SNP Markers** **Paralogous SNP Markers**

Fig 3. True, Hemi and Paralogous SNPs in Cotton

Summary

Tetraploid nature, highly complex and repetitive genome of cultivated cotton species poses significant challenges for SNP marker development. These complications usually result in high number of false positive SNPs, especially when they are developed from sequences that were not thoroughly characterized. In this study, we used KASPar genotyping technology to screen 150 *in silico* SNP markers across a panel of 23 genotypes. KASPar assays were clearly able to distinguish true, hemi and paralogous SNPs in cotton. Thus, KASPar assays can be used as an efficient and cost-effective way of validating *in silico* SNPs as well as for other applications in cotton.

References

Gupta, P.K., J.K. Roy., and Prasad. M. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. . Curr Sci. 80:524-535.

Philip Robinson. US20070117108A1: Detection system for PCR assay.

Wendel, J.F., and R.C. Cronn. 2003. Polyploidy and the evolutionary history of cotton. Advances in Agronomy 78:139-186.

Zhang, H.-B., Y. Li, B. Wang, and P.W. Chee. 2008. Recent Advances in Cotton Genomics. International Journal of Plant Genomics 2008: 742304.