Plant Genetics, Genomics, and Biotechnology 2 (1): 90–94, 2014
doi: 10.5147/pggb.2014.0144

Introduction of High Throughput and Cost Effective SNP Genotyping Platforms in Soybean

Jiazheng Yuan, Zixiang Wen, Cuihua Gu, and Dechun Wang*

Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing MI, 48824 USA

Received: September 1, 2014 / Accepted: November 8, 2014

Abstract

We presented here the application of two in-plate SNP (single nucleotide polymorphism) genotyping platforms for soybean plants [*Glycine max* (L.) Merr.], KASP® (Kompetitive Allele Specific PCR genotyping, LGC Genomics) and TaqMan® (Life Technologies) respectively. These two systems offer us an ability to determine the genotypes of 384 individual samples accurately and efficiently by allele specific PCR in a single plate using typical PCR conditions. Both of the systems require small quantity of genomic DNA obtained from a simple DNA extraction. The genomic sequences containing target SNPs can easily be used as a basic blueprint to design the probes and primers of KASP® and TaqMan® assays whether the sequences are obtained from the genome sequence of soybean William 82 (Wm82.a2.v1), Illumina Soy50k SNPs, or parallel resequencing. Moreover, we listed the pros and cons of the two systems and explained the principles behind the platforms. The high call rate and clear clustering separation of the SNPs can be readily obtained from these platforms without conducting any assay optimization processes. These platforms can routinely be performed on 96/384-well plate format with or without an automation procedure. Therefore, these platforms are especially suitable for the SNP genotyping on a particular trait with a large sample size, gene fine mapping, and marker assisted selection. Further, they require little hands-on experience and achieve per-site and per-individual costs below that of current SSR, AFLP, RFLP, and SNP chip technologies. The platforms can be used for genotyping on a wide range of organisms due to their simplicity and flexibility of handling. Meanwhile, we also especially presented some of the advantages using KASP® SNP genotyping pipeline, which was cost effective in the selection of allele specific assay and therefore, efficiently facilitated the soybean genotyping across large numbers (thousands or more) of individual lines for a great range of markers (hundreds to thousands) in our laboratory.

Keywords: Plant SNP genotyping, TaqMan®, KASP®, high throughput, cost effective.

Introduction

A cost effective and high throughput platform that is able to efficiently and accurately determine genotypes of plants is important for genetic studies. There are many platforms available for plant genotyping and the cost and throughput of the genotyping can vary greatly from one to another. Previous genotyping methods using SSR, RAPD, AFLP, and RFLP etc. were costly and labor intense. Currently, in-silico chip platform such as Illumina HD infinium system appears to be high throughput and has been gradually applied for plant genotyping. The microarray-derived SNP genotyping platform requires substantial financial input on the instrument and prior knowledge of genome sequences. Once chips manufactured, they will only be suitable for known polymorphic sites of nucleotides. Recent developments in GBS (genotyping by sequencing) and other whole genome sequencing platforms have generated massively parallel sequences. However, the cost of sequencing and the complexity of data processing limit their adoption for large scale of selective genotyping and marker-assisted selection. Thus, with the target SNPs identified, a high-throughput and cost effective platform will be

* Corresponding author: wangdech@msu.edu

needed to use these SNPs in the massive scale genotyping and marker-assisted selection for genetic studies and plant breeding. It is postulated that the selection in crops such as soybean, either natural or through breeding would lead to increase in the abundance of beneficial sequences and purge deleterious sequences from the population. Hence, the selections on the target genomic regions in breeding programs often reduce genetic diversity and therefore, fewer sequence variants will be observed. Hyten et al., (2006) showed that cultivated soybean has an average SNP abundance of 1/1000 bp compared to 1/425 bp of its wild counterpart [*Glycine soja* (Sieb and Zucc.)]. Nevertheless, SNPs are yet distributed abundantly throughout plant genome and are the most promising molecular markers in the genome. With the availability of next-gen whole genome sequencing, a large number of polymorphic SNP markers have been identified by re-sequencing different genotypes in soybean. Most SNPs do not appear to cause any effect on gene expression or function. However, they are also difficult to be distinguished from those sequence variations at their vicinity that may have a function. Moreover, SNPs and sequence structural variants as the root of genetic variation among individuals and populations may alter pathogen accessibility and host resistance during the pathogenesis of pathogens. If the SNP occurs in the coding region, it may also cause alternative splicing, mRNA instability, or nonsynonymous changes in the sequence of amino acids and therefore, results in alternation of gene function or earlier termination of translation (Richard and Beckman, 1995). Because SNPs are the most abundance type of genetic polymorphism and it is evolutionarily stable from generation to generation, SNPs represent the ultimate form of molecular marker (Rafalski 2002; Edwards et al., 2007). Thus, there is an increasing need of a high throughput and low-cost SNP genotyping platforms based on study objectives, samples numbers, and judicious allocation of resources. In this report, we presented two high throughput and affordable SNP genotyping platforms, TaqMan® and KASP® in soybean plants. These two systems are also compared in terms of efficiency, sensitivity, cost, and reliability for gene fine mapping and marker assisted breeding in soybeans (Table 1).

Materials and Methods

DNA Extraction

Genomic DNA was extracted from young leaves of soybean plants. One piece of young trifoliate tissue was manually picked and placed into a 1 mL deep 96-well plate (VWR, Batavia, IL or USA Scientific, Ocala, FL). These samples were freeze–dried and then ground by 4 mm glass beads (Fisher Scientific, Pittsburg, PA) using a shaker. DNA was extracted from the powdered leaf material using CTAB (cetyltrimethyl ammonium bromide) method (Hoisington et al. 1994). 400 uL of CTAB solution was added into each well of the plate containing the lyophilized powder of plant tissue and then the plate was covered by a cap mat (VWR or USA Scientific). The plate was gently mixed before placed in a water bath at 65°C for 60 minutes with gentle shaking at least three times. After the plate cooled down, 400 uL of Chloroform solution (1:24 Chloroform:Isoamyl alcohol) was added and mixed gently. The plate was then spinning at 2700 g for 15 minutes at room temperature on a centrifuge. The supernatant was transferred into a new 0.5 mL well plate (VWR or USA Scientific). In each well of the plate, 250 uL of 190 proof chilled alcohol was added, covered with a new cap mat, and mixed gently for 3 minutes. The plate was then placed into a -20 °C freezer for 1-2 hours for DNA precipitation (This step is optional). The plate was centrifuged at 2700 g for 5 minutes to precipitate the DNA. The supernatant was discarded and the DNA pellets were kept at room temperate to air-dry for 30 minutes. DNA samples were dissolved into 200 uL of 0.1X TE buffer or ddH2O. The DNA was quantified using NonaDrop® (Thermal Scientific, Wilmington, DE) or other spectrophotometers. In our experiments, the DNA samples were further diluted with dd H2O at 1:4 ratio before SNP assays. Note that the total amount of DNA quantity of 1-20 ng and 3-30 ng is required for the PCR reactions of TaqMan® and KASP®, respectively in order to conduct the assay successfully.

PCR of TaqMan® Assay

For TaqMan® SNPs, the allele specific assays (probes plus primers) were designed and synthesized by Life Technologies (Foster City, CA, USA). The master mix (MM) was also ordered from this company. The TaqMan® SNP PCR reaction was carried out in a 384-well plate with a minimum total volume of 3 uL/well on the LightCycler 480 instrument (Roche Applied Science, Indianapolis, IN, USA). The PCR reaction mixture was consisted of 1-20 ng of genomic DNA, 0.15 uL of 20X TaqMan® assay, and 1.5 uL of 2X genotyping master mix containing a modified Taq DNA polymerase, reaction buffer, MgCl2 and dNTPs. After 10 minutes (min) pre-incubation at 95 °C, 45 PCR cycles were conducted with 10 second (s) of denaturation at 95 °C, 30 s of annealing at 60 °C, and 10 s extension at 72 °C. A final melting cycle was performed by raising the temperature to 95 °C for 10 s, lowering the temperature to 45 °C for 30 s and then increasing the temperature to 83 °C with continuous fluorescent acquisition. The final step was to hold the PCR plate at 40 °C on the LightCycler 480. The genotype calling was performed using the Roche Applied Science software version 1.5.0.

PCR of KASP® Assay

The KASP® SNP PCR reaction was conducted on the BioRad C1000 thermal cyclers (BioRad, Hercules, CA) in a 384-well plate with total volume of 3 uL/well. The PCR reaction mixture contained 5-30 ng of genomic DNA, 0.05 uL of KASP® assay, and 1.5 uL of 2X KASP® master mix. After 10 min pre-incubation at 95 °C, 10 PCR cycles were carried out with 20 s of denaturation at 95 °C, 60 s of annealing at 65 °C with -0.8°C touchdown every cycle for the amplification-I and then 28-34 PCR cycles were required with 20 s of denaturation at 95 °C, 60 s of annealing at 58 °C for the amplification-II. The fluorescent acquisition was programmed by holding the temperature at 37 °C for 5 s with three times of continuous fluorescent acquisition on the LightCycler 480.

Table 1. SNP genotyping features and relative costs of two platforms, TaqMan® and KASP®, respectively (Life Technologies and LGC Genomics). The cost of allele specific assay and master mix per data point was calculated based on the ratios of posted prices of the reagents from two companies and individual customer may get different quotes from the companies on the reagents.

Platform	TaqMan®	KASPar®
Sample capacity	Single plex	Single plex
Labelling method	FRET	FRET
FRET location	Probe	Master mix
Allele specific	Probe	Forward primers
Reproducibility	High	High
Accurate genotyping	High	High
End point data acquizition temperature	Anneling TM	37 °C
High throughput	Yes	Yes
Reaction volume (uL)	3	3
Easy data interpretation	Yes	Yes
Total DNA (ng)	1-20	3-30
length of sequence required	50-200 bp	80-200 bp
Allele specific assay cost/data point ($)	0.238	0.005
Master Mix cost/data point ($)	0.143	0.405

Results and Discussion

The allelic specificity of TaqMan® SNP genotyping platform was provided by two probes. For each SNP assay of TaqMan® platform, forward and reverse primers and two TaqMan® probes specific for each allele were synthesized prior to the PCR reaction based on the total 100-300 bp DNA sequences containing target biallelic SNPS. In the fluorophore detection, probes were made of a FRET (fluorescence resonance energy transfer) cassette, which was labeled with a reporter dye (FAM or VIC) fused to the 5'-end of probe and the nonfluorescent quencher attached to the 3' end (De la Vega et al., 2005). The fluorescent reporters (FAM and VIC) were specific to the alleles. The probes hybridize specifically to the complementary sequences, which contained the target SNPs. During the PCR process, the alleles were detected by the corresponding fluorescence signal generated via 5'-exonuclease cleavage. The cleavage of the FRET cassette in the 5'-nuclease reaction broke the proximity of the reporter and quencher if the target sequence was complementary to the probe and hence, liberated the reporter dye, resulting in increased fluorescence during PCR cycles (Figure 1). If both FAM and VIC signals were produced at same time suggesting that the locus was a heterozygote. However, mismatches could not produce high fluorescence due to low-efficiency probe hybridization. It is relatively straightforward to design new assays and the PCR reactions can generally run on any thermo cyclers such as BioRad CFX384 Touch (BioRad, Hercules, CA) and Roche Lightcycler 480 (Roche Diagnostics, Indianapolis, IN). The main drawback for the TaqMan® SNP assay is its limited capacity to be used for SNP multiplexing and higher cost in the synthesis of allele specific assays.

In contrast to TaqMan®, the allelic specificity of KASP® SNP genotyping platform was provided by two forward primers.

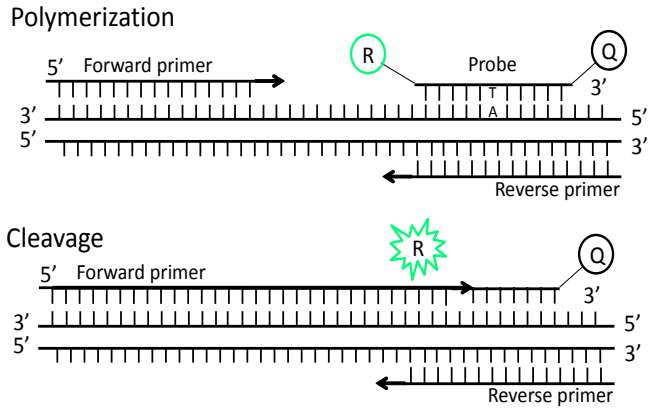

Figure 1. Allele detection of TaqMan® platform during PCR amplification. The target allele containing SNP is detected by each probe, which is complementary to one of the alleles labelled with two dyes (FAM or VIC). After perfect binding to the target sequence, a fluorescent dye (R) of the probe is released and generated a fluorescent signal because the quencher (Q) has been cleaved by 5′ endonuclease activity of Taq DNA polymerase in following PCR steps. The diagram is modified based on the information from Life Technologies (www.lifetechnologies.com).

Each forward primer was attached with the FAM or HEX compatible tail (FAM tail: 5' GAAGGTGACCAAGTTCATGCT 3'; HEX (or VIC) tail: 5' GAAGGTCGGAGTCAACGGATT 3') at 5'-end of oligos (Trick et al., 2012). The target SNPs were positioned at 3'-end of the forward primers. The master mix contained the fluorescent reporter dyes of FAM and HEX specific FRET cassette and Taq polymerase. One common reverse primer was designed using the genomic sequence containing target SNPs. The amplicoms of KASP® assay of each allele were amplified using the same reverse primer combining with two forward allele-specific

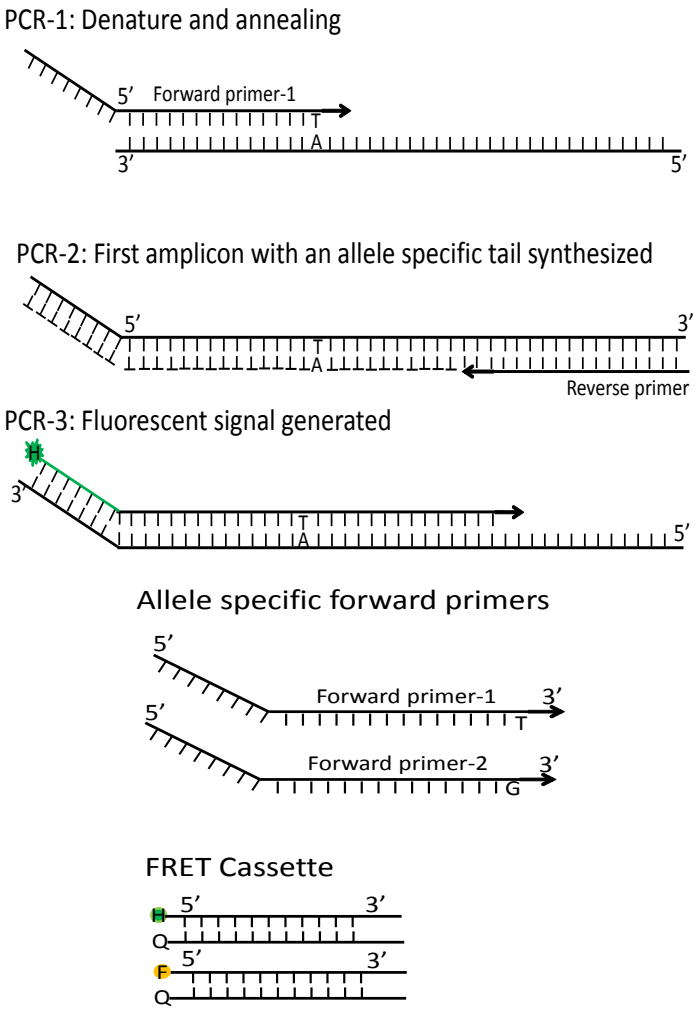

Figure 2. Major steps of KASP® PCR reaction (modified based on the information from LGC Genomics). The KASP® PCR uses two allele specific forward primers and one common reverse primer. The forward primers contain unique allele specific tails at 5' end and a SNP at 3'end of the oligos. The master mix of KASP® contains the FRET cassettes, which are complementary to the new tails. When the FRET cassettes of KASP® are intact, the approximate of reporter (FAM or HEX) and quencher prevents the emission of any fluorescence. During the following rounds of PCR, the more allele-specific tails are amplified. The FRET cassette binds to new tail and the 5′-endonuclease activity of the Taq DNA polymerase cleaves the FRET, which separates reporter and quencher dyes and the fluorescence is detected.

primers. In the master mix of KASP®, the sequences of FRET cassette were complementary to the sequences of forward primers, which were labeled with one of the fluorescent reporter dyes (FAM or HEX) (Figure 2). In KASP PCR, the proximity of the reporter dye to the quencher in the FRET Cassette of KASP® also suppressed the reporter fluorescence if no complementary binding occurred (lgcgenomics.com). The proximity of the reporter and quencher in the FRET prevented the reporter dye excited. For fluorophore detection, the fluorescent signals were accumulated with the PCR cycles increase. More signals were generated in further rounds of PCR cycles and releasing the quencher of the FRET cassette due to the increase of allele specific tails, which were complementary to FRET sequences.

For KASP® SNP assay, the allele specific forward primers and reverse primer were initially designed and synthesized by the LGC Genomics (Beverly, MA). The master mix was ordered from LGC Genomics. To reduce cost, the forward primers and reverse primer were synthesized from an oligo synthesis company such as Sigma-Aldrich (St. Louis, MO). Primer mix was created by mixing in a total volume of 100 uL with 30 uL reverse primer (100 uM) and 12 uL of each tailed forward primer (100μM). The SNP calling was using the Roche Applied Science software version 1.5.0 with the end-point genotyping function and two opposites of homozygous genotypes were grouped into two clusters near two axes and the heterozygous genotypes were clustered between the two clusters for homozygous genotypes (Figure 3). The homozygous alleles labelled HEX were grouped into Y axis while homozygous alleles reported with FAM were

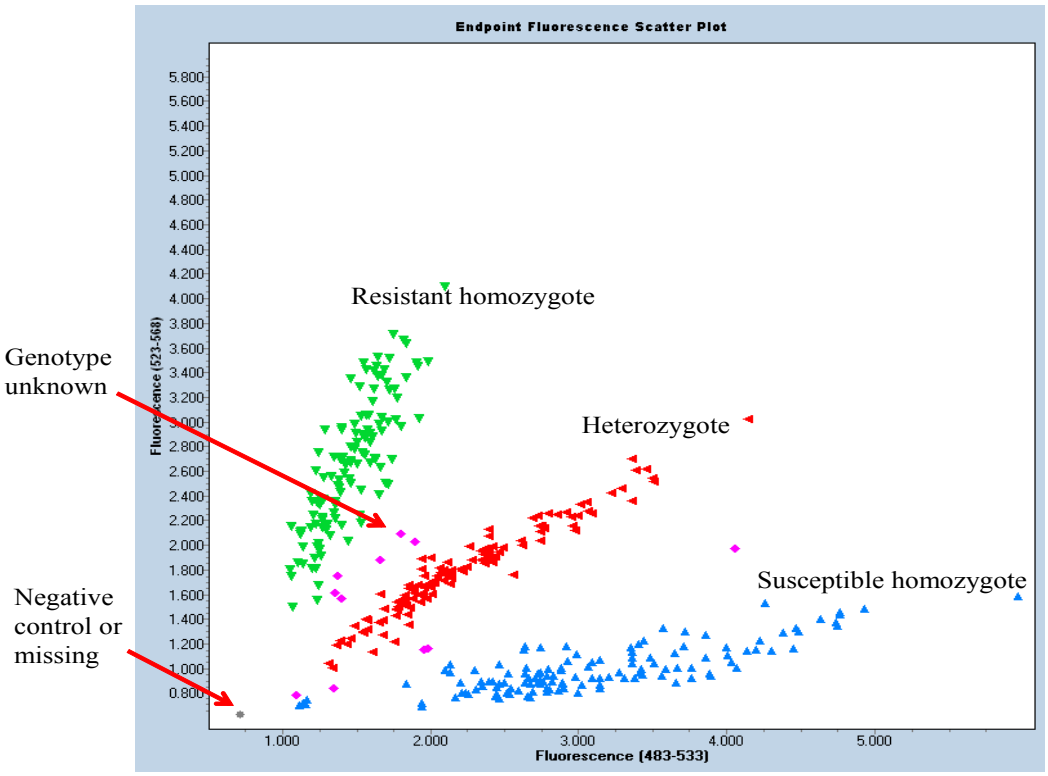

Figure 3. Clustering plot of individuals assayed for a disease resistance project in soybean using KASP® SNP genotyping platform on Roche LightCycler 480. The no-template controls or missing sample (no call) was positioned near the origin of the plot at the corner. The fluorescence of HEX on Y-axis was designed to detect the resistant homozygotes while the fluorescence of the FAM on X-axis was made to detect the susceptible homozygotes. Heterozygotes were clustered within the intermediate. The genotype unknown can be re-grouped manually using Roche's software.

clustered near X axis. A BioMek 2000 Laboratory Automation Workstation (Beckman Coulter, Inc., Fullterton, CA) was used to set up the PCR reaction for the genotyping. Note that both TaqMan® and KASP® genotyping PCRs can be conducted in any thermal cycler and then the results can be acquired using a plate reader or realtime PCR machine with a function of the end-point genotyping algorithm. Based on our estimation, the cost of each data point of KASP® master mix was higher than that of TaqMan® while the cost of TaqMan® allele specific assay for each data point was much higher than (posted prices from both companies) that of KASP® (Table 1). Thus, KASP® has provided much greater flexibility in terms of cost for SNP selection for the soybean genotyping.

Except for the higher cost on allele specific assay of TaqMan® and master mix of KASP®, we found that both in-plate platforms possessed additional advantages: low genotyping error rates, low labor cost, and exceptional tolerance of variations in quantity and quality of DNA. These properties are critical for plant selection in a breeding program because the optimization for a large number of DNA samples prior to conduct PCR reaction is not easily applicable. Based on our experience, the assays of TaqMan® and KASP® platforms can be conducted for hundreds of thousands of samples using a few dozen SNPs within a short time period. These platforms are very robust and can be easily integrated with automated liquid handling systems and no post-PCR handling is required. The endpoint genotyping algorithm allows running the PCR on various thermocyclers including a regular PCR machine and then detects the fluorophore on a plate reader or a real-time PCR machine. However, one of the drawbacks of these platforms is their limited capacity for multiplexing. Nevertheless, each PCR plate can also be subset into smaller groups for different assays in order to maximize the capacity of the equipment.

References

Hyten DL, Q Song , Y Zhu , IY Choi, RL Nelson, JM Costa, JE Specht, RC Shoemaker, PB Cregan (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103 (45): 16666-16671.

Richard I and JS Beckman (1995) How neutral are synonymous codon mutations? Nat Genet 10: 259.

De la Vega FM, KD Lazaruk, MD Rhodes, MH Wenz (2005) Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. Mutat Res 573: 111–135.

Trick M, NM Adamski, SG Mugford, CC Jiang, M Febrer, et al (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol 12 (14): 1-17.