February 2008

High-throughput genotyping with the GoldenGate assay in the complex genome of soybean

David L. Hyten
U.S. Department of Agriculture, Agricultural Research Service

Qijian Song
University of Maryland, College Park, MD

Ik-Young Choi
Seoul National University, Seoul 151-921, South Korea

Mun-Sup Yoon
Genetic Resources Division, National Institute of Agricultural Biotechnology, Rural Development Administration, Suwon 441-707, South Korea

James E. Specht
University of Nebraska - Lincoln, jspecht1@unl.edu

See next page for additional authors

Hyten, David L.; Song, Qijian; Choi, Ik-Young; Yoon, Mun-Sup; Specht, James E.; Matukumalli, Lakshimi; Nelson, Randall L.; Shoemaker, Randy C.; Young, Nevin D.; and Cregan, Perry B., "High-throughput genotyping with the GoldenGate assay in the complex genome of soybean" (2008). *Agronomy & Horticulture -- Faculty Publications*. 128.
http://digitalcommons.unl.edu/agronomyfacpub/128

Authors
David L. Hyten, Qijian Song, Ik-Young Choi, Mun-Sup Yoon, James E. Specht, Lakshimi Matukumalli, Randall L. Nelson, Randy C. Shoemaker, Nevin D. Young, and Perry B. Cregan

Theor Appl Genet (2008) 116:945–952
DOI 10.1007/s00122-008-0726-2

ORIGINAL PAPER

High-throughput genotyping with the GoldenGate assay in the complex genome of soybean

David L. Hyten · Qijian Song · Ik-Young Choi · Mun-Sup Yoon · James E. Specht · Lakshmi K. Matukumalli · Randall L. Nelson · Randy C. Shoemaker · Nevin D. Young · Perry B. Cregan

Received: 19 September 2007 / Accepted: 28 January 2008 / Published online: 16 February 2008

Abstract Large numbers of single nucleotide polymorphism (SNP) markers are now available for a number of crop species. However, the high-throughput methods for multiplexing SNP assays are untested in complex genomes, such as soybean, that have a high proportion of paralogous genes. The Illumina GoldenGate assay is capable of multiplexing from 96 to 1,536 SNPs in a single reaction over a 3-day period. We tested the GoldenGate assay in soybean to determine the success rate of converting verified SNPs into working assays. A custom 384-SNP GoldenGate assay was designed using SNPs that had been discovered through the resequencing of five diverse accessions that are the parents of three recombinant inbred line (RIL) mapping populations. The 384 SNPs that were selected for this custom assay were predicted to segregate in one or more of the RIL mapping populations. Allelic data were successfully generated for 89% of the SNP loci (342 of the 384) when it was used in the three RIL mapping populations, indicating that the complex nature of the soybean genome had little impact on conversion of the discovered SNPs into usable assays. In addition, 80% of the 342 mapped SNPs had a minor allele frequency >10% when this assay was used on a diverse sample of Asian landrace germplasm accessions.

Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval of a product to the exclusion of others that may be suitable.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0726-2) contains supplementary material, which is available to authorized users.

D. L. Hyten · Q. Song · I.-Y. Choi · M.-S. Yoon · P. B. Cregan (✉)
Soybean Genomics and Improvement Laboratory,
U.S. Department of Agriculture, Agricultural Research Service,
10300 Baltimore Ave., Bldg 006, Rm 100,
Beltsville, MD 20705, USA
e-mail: Perry.Cregan@ars.usda.gov

Q. Song
Department Plant Science and Landscape Architecture,
University of Maryland, College Park, MD 20742, USA

I.-Y. Choi
National Instrumentation Center for Environmental Management,
Seoul National University, Seoul 151-921, South Korea

M.-S. Yoon
Genetic Resources Division, National Institute of Agricultural
Biotechnology, Rural Development Administration,
Suwon 441-707, South Korea

J. E. Specht
Department of Agronomy and Horticulture,
University of Nebraska Lincoln, Lincoln, NE 68583, USA

L. K. Matukumalli
Bovine Functional Genomics Laboratory,
U.S. Department of Agriculture, Agricultural Research Service,
Beltsville, MD 20705, USA

R. L. Nelson
Soybean/Maize Germplasm, Pathology, and Genetics Research
Unit and Department of Crop Sciences,
U.S. Department of Agriculture, Agricultural Research Service,
University of Illinois, Urbana, IL 61801, USA

R. C. Shoemaker
Department of Agronomy, U.S. Department of Agriculture,
Agricultural Research Service, Iowa State University,
Ames, IA 50011, USA

N. D. Young
Department of Plant Pathology,
University of Minnesota, St. Paul, MN 55108, USA

The high success rate of the GoldenGate assay makes this a useful technique for quickly creating high density genetic maps in species where SNP markers are rapidly becoming available.

Introduction

Single nucleotide polymorphisms (SNPs) can be used as molecular markers for a variety of tasks in crop improvement including quantitative trait locus (QTL) discovery, assessment of genetic diversity, association analysis, and marker-assisted selection. SNPs have two main advantages over other molecular markers; they are the most abundant form of genetic variation within genomes (Zhu et al. 2003), and a wide array of technologies have now been developed for high throughput SNP analysis (Fan et al. 2006a). Only recently has there been large scale SNP discovery efforts in a number of plant species to ascertain the level of DNA sequence variation and the effects that evolutionary events have had on that variation (Choi et al. 2007; Hyten et al. 2006; Schmid et al. 2003; Tenaillon et al. 2002).

An initial phase of a large-scale soybean SNP discovery effort resulted in the identification of 5,551 SNPs (Choi et al. 2007). Since then, approximately 12,000 additional soybean SNPs have been identified (unpublished data). The consensus genetic linkage map created by Choi et al. (2007) contained 1,158 sequence tagged sites (STS), but these represented only 2,928 of the 5,551 SNPs and were mapped using low-throughput technologies involving limited multiplexing (i.e., 5–10 SNPs in a single reaction). The genetic mapping of the ~2,600 remaining SNPs, along with ~12,000 newly discovered SNPs, will require higher-throughput technologies to more quickly achieve the desired high-density soybean genetic map. Such a map is not only critical for assembly of the whole genome shotgun sequence but is also needed for the fine mapping and cloning of agronomically important genes. While newer technologies have been reported to have far greater multiplex potential, they remain untested in species with highly duplicated genomes such as soybean.

The diploid soybean is believed to be an ancient tetraploid (Hymowitz 2004). High levels of paralogous genomic sequence cause difficulties, not only in SNP discovery, but also in SNP detection. For example, Choi et al. (2007) found that 23% of the primers designed to 3′ expressed sequence tags (EST) did not produce robust STS, due to amplification of two or more loci. For the robust SNP-containing STS, Choi et al. (2007) redesigned PCR primers and single base extension primers for SNP detection using the Sequenom Mass Spectrometer MassARRAY™ technology. Of these assays, 33% did not yield a useable co-dominant assay, most likely due to paralogous sequence that interfered with the amplification of a single PCR product (unpublished data from Choi et al. 2007).

The Illumina GoldenGate assay is capable of multiplexing from 96 to 1,536 SNPs in a single reaction over a 3-day period (Fan et al. 2003). Recently, a custom GoldenGate assay was designed for barley (*Hordeum vulgare*) in which 91% of the SNP assays generated reliable allelic discrimination (Rostoks et al. 2006). However, the barley genome contains only 16% paralogous genes compared to the 32% paralogous genes known to be present in soybean (Blanc and Wolfe 2004). Many other plant species, such as maize (*Zea mays*), cotton (*Gossypium hirsutum* and *G. arboretum*), and wheat (*Triticum aestivum*) also have a high number of paralogous genes similar to soybean (Blanc and Wolfe 2004), which might limit the high multiplexing capacity of the GoldenGate assay. Our objective was to determine if the GoldenGate assay will successfully function in a highly duplicated plant genome such as soybean.

Materials and methods

Plant materials

Mapping populations: Three mapping populations were used to position newly identified SNPs on the consensus genetic map (Song et al. 2004). The University of Utah “Minsoy” x “Noir 1” (MN) and Minsoy x “Archer” (MA) populations were described by Cregan et al. (1999a) and Mansur et al. (1995; 1996). The University of Minnesota “Evans” x “Peking” (EP) population was described by Concibido et al. (1997). Genomic DNA was obtained from the two parents and 89 recombinant inbred lines (RILs) of each of the three mapping populations from bulked leaf tissue of 20–50 plants of each line as described by Keim et al. (1988). The MN and MA populations had been used in creating the current integrated genetic linkage map (Choi et al. 2007).

Diverse landraces: a group of 96 diverse landrace accessions were selected that consisted of plant introductions from China, Korea, and Japan (Supplementary Table 1). The landraces represent a range of geographic origin, morphological descriptors, and maturity classes to maximize the diversity sampled. Pure line seeds of all genotypes were obtained from the USDA Soybean Germplasm Collection (USDA-ARS, Univ. of Illinois, Urbana, IL). DNA was extracted from bulked leaf tissue of 20–50 plants of each accession as described by Keim et al. (1988).

SSR mapping

The EP population had not previously been used in map integration efforts, and its map was insufficiently populated with markers to ensure accurate positioning of all mapped SNPs. Therefore, the Evans and Peking parents were first screened with the SSR markers mapped by Song et al. (2004) to identify a set of polymorphic markers spaced at about 10 cM intervals across each linkage group. The EP RILs were then genotyped for these SSRs, as described by Cregan et al. (1999a). SSR allele size differences were determined on a non-denaturing polyacrylamide gel as described by Wang et al. (2003), or with a 2% agarose gel.

SNP discovery and mapping

The SNPs chosen for the design of the 384 custom GoldenGate assay were selected from three sources of sequences previously determined to contain SNPs that were segregating in one or more of the RIL mapping populations described previously. The sources included gene or EST sequences, BAC-end sequences, and targeted BAC subclone sequences. Gene sequences containing SNPs were selected from the Beltsville Agricultural Research Center Soybean SNP database, found on the web at http://bfgl.anri.barc.usda.gov/soybean/ (Choi et al. 2007). BAC-end sequences for SNP discovery were obtained from sequences deposited in GenBank (http://www.ncbi.nlm.nih.gov), as a result of the Williams 82 physical map project (http://www.soybase.org). PCR primers were designed for these BAC-end sequences using Primer3 (Rozen and Skaletsky 2000) to amplify products in the 600–800 bp range. PCR primers were initially used to amplify Williams 82 DNA followed by DNA sequence analysis of the resulting amplicon on an ABI 3730 DNA Analyzer as described by Choi et al. (2007). When high quality sequence data were obtained, the STS primers were then used to amplify and sequence genomic DNA of the five parents of the three mapping populations: Archer, Minsoy, Noir 1, Evans, and Peking. The sequence of the Williams 82 and the five parental genotype amplicons were then analyzed using the SNP-PHAGE software (Matukumalli et al. 2006) for the presence of SNPs which would segregate in one or more of the three RIL mapping populations. To obtain targeted BAC subclone sequences for SNP discovery, a Williams 82 BAC library (Marek and Shoemaker 1997) was screened with SSR markers within regions known to have verified QTL, as described by Cregan et al. (1999b). The selected BACs were then subcloned and the insert isolated, also as described by Cregan et al. (1999b). Sequencing of the insert was performed on the ABI 3730 DNA Analyzer. PCR primers were designed to the sequenced BAC subclone and SNP discovery was performed as described previously for the BAC-end sequences. All STS information along with PCR primer sequences for the 384 selected STS can be found in Supplementary Table 2.

SNP-containing sequences were screened with RepeatMasker software (http://www.repeatmasker.org), using the three repeat databases obtained from Dr. Randy Shoemaker of the USDA-ARS, Ames, IA, Dr. Scott Jackson of Purdue Univ., West Lafayette, IN (http://www.soymap.org/); and Dr. Gary Stacy of the Univ. of Missouri, Columbia, MO (http://www.soybeangenome.org/). Repeats contained within the SNP-containing sequences were replaced with lowercase letters prior to submission to Illumina Inc. to undergo a preliminary design phase of the custom oligo pool all (OPA), which contains the allele-specific oligos and the locus-specific oligos for all SNPs included in the assay. A designability rank score was given to each SNP by Illumina, with the score ranging from 0 to 1.0, where a rank score of <0.4 has a low success rate, 0.4 to <0.6 has a moderate success rate, and >0.6 has a high success rate for the conversion of a SNP into a successful GoldenGate assay. A total of 384 SNPs with a designability rank score of 0.4 or higher were selected to be included in the OPA, except when multiple SNPs were located on the same STS, in which case only the SNP with the highest designability rank score was selected (the others were not used).

The GoldenGate assay was performed as per the manufacturer’s protocol and as described in Fan et al. (2003). Briefly, a total of 5 μl of 50 ng/μl in 10 mM Tris-HCL pH 8.0, 1 mM EDTA of genomic DNA was used to make single-use DNA. The single-use DNA underwent an allele specific oligonucleotide hybridization, which involves three oligos at each of the 384 different SNP loci, thus comprising 1,152 custom oligos in the OPA. At each SNP locus, two of the oligos are allele-specific oligos that are complementary to the genomic sequence directly adjacent to the SNP being assayed except they differ at the 3′ base in order to be complementary with one of the two SNP alleles and each oligo has a universal primer site attached at the 5′ end with one allele for each SNP having universal primer site sequence 1 which is 5′-ACTTCGTCAGTAACGGAC-3′ and the other allele for each SNP having universal primer site sequence 2 which is 5′-GAGTCGAGGTCATATCGT-3′. The third oligo is a locus-specific oligo which hybridizes to the complementary sequence located between 1 and 20 bases downstream of the target SNP and has attached to the 5′ end of the oligo universal primer site sequence 3 which is 5′-GTCTGCCTATAGTGAGTC-3′ and located between the universal primer 3 and the complementary genomic sequence is an “IllumiCode” sequence which is unique to each SNP locus. The hybridization is followed by an extension and ligation step connecting one of the allele-specific oligos with the locus-specific oligo. This step is followed by universal PCR for all 384 loci with three universal primers which are complementary to the three universal primer site sequences described previously. Universal primer 1 is labeled with *Cy3* and universal primer 2 is labeled with *Cy5*. After amplification, the products are hybridized to a sentrix array matrix (SAM) for detection. The internal IllumiCode sequence is specific for each SNP locus and thus binds only to its complementary sequence attached to a bead on the SAM. The hybridized SAM is then analyzed on the Illumina BeadStation 500 G (Illumina, San Diego, CA). The automatic allele calling for each locus is accomplished with the GenCall software (Illumina, San Diego, CA). A genotype that is homozygous for one or the other SNP alleles will display a signal in either the *Cy3* or *Cy5* channel, whereas a genotype that is heterozygous will display a signal in both channels. All GenCall data were manually checked and re-scored if any errors in calling the homozygous or heterozygous clusters were evident.

The soybean SNP markers were first mapped within each of the three mapping populations, and then all markers (SNPs and SSRs) were positionally integrated into one common integrated linkage map using JoinMap 3.0 (Van Ooijen and Voorrips 2001). Genetic distances were calculated using the Kosambi mapping function.

Results

The Illumina custom GoldenGate assay (hereafter termed SoyOPA-1) consisted of 384 SNPs from 380 separate STS, with the latter derived from 204 random gene sequences, 140 BAC-end sequences, and 36 targeted BAC subclone sequences. All SNPs chosen for SoyOPA-1 had been verified through forward and reverse sequencing of PCR amplicons from the five diverse genotypes which serve as mapping parents for three RIL mapping populations. The SoyOPA-1 was evaluated by assessing its performance in terms of mapping these 384 SNPs in the three RIL populations.

GenCall software output

The cluster separation score provided by the GenCall software for the three RIL mapping populations used was not a good indicator of the true cluster separation between the two homozygous classes. This is due to the GenCall cluster separation score being calculated on the degree of separation between the two homozygous clusters from the heterozygous cluster and not the degree of separation between the two homozygous clusters. Since our mapping populations were RIL populations that contain few to no heterozygotes, a cluster estimation based on the degree of separation between the two homozygous clusters would seem to be more informative on how well the GoldenGate assays functioned in soybean. Therefore, we calculated our own cluster separation score using the average separation between the two homozygote clusters based on the normalized theta value ($(2/\pi)\mathrm{Tan}^{-1}$ (*Cy5*/*Cy3*)) and converted this value to a 0.5–5 scale (Fig. 1). As shown in Fig. 2, the majority of SoyOPA-1 SNPs had good cluster separation. Because the genotypes scored in the three mapping populations were mostly inbreds, cluster separation values as low as 0.5 to 2.0 were still scorable. A total of 342 of the 384 assays (89%) had cluster separations of 0.5 or greater and thus, were classified as successful assays.

Genetic mapping of SNP loci

The MN and MA populations were used previously to position SNPs (via other assay technologies) into the current soybean integrated linkage map (Choi et al. 2007). Thus, of the 342 successfully mapped SNPs, 256 SNPs were polymorphic in the MN and/or MA populations (Fig. 3) and were mapped with the map data from Choi et al. (2007) to add 256 new markers to the current integrated linkage map.

The EP population had not been used previously in the construction of the consensus soybean genome map and contained 86 SNPs that were uniquely polymorphic to this population (Fig. 3). To facilitate a better integration of the linkage map formed from the EP population, 166 SSR markers with map positions distributed across the genome (and segregating in at least one of the four RIL mapping populations used by Choi et al. 2007) were genotyped in the EP population. These 166 SSR markers, along with the 96 SNP markers that also were in common between the EP population and either one or both of the MN and MA populations (Fig. 3), aided in the integration of the EP population linkage map into the current consensus linkage map. The use of the three mapping populations allowed 334 SNPs to be integrated into the current consensus map which can be found at Beltsville Agricultural Research Center Soybean SNP database located on the web at http://bfgl.anri.barc.usda.gov/soybean/. The map positions from this consensus map can also be found in the Supplementary Table 2 for the 334 integrated assays.

Allele frequency estimates in diverse germplasm

While we were able to successfully map 334 SNPs using SoyOPA-1, knowing the allele frequencies in diverse germplasm would help determine their usefulness in future efforts aimed at QTL mapping, marker-assisted selection, or association analysis. The six mapping parent accessions Archer, Minsoy, Noir 1, Evans, Peking, and PI 209332 were previously found to possess 93% of the common SNPs (minor allele frequency >10%) that were present in a diverse sampling of 25 genotypes (Zhu et al. 2003). However, it is not known how well five of the six genotypes used in the current three RIL mapping populations would serve to identify SNPs with high minor allele frequencies in

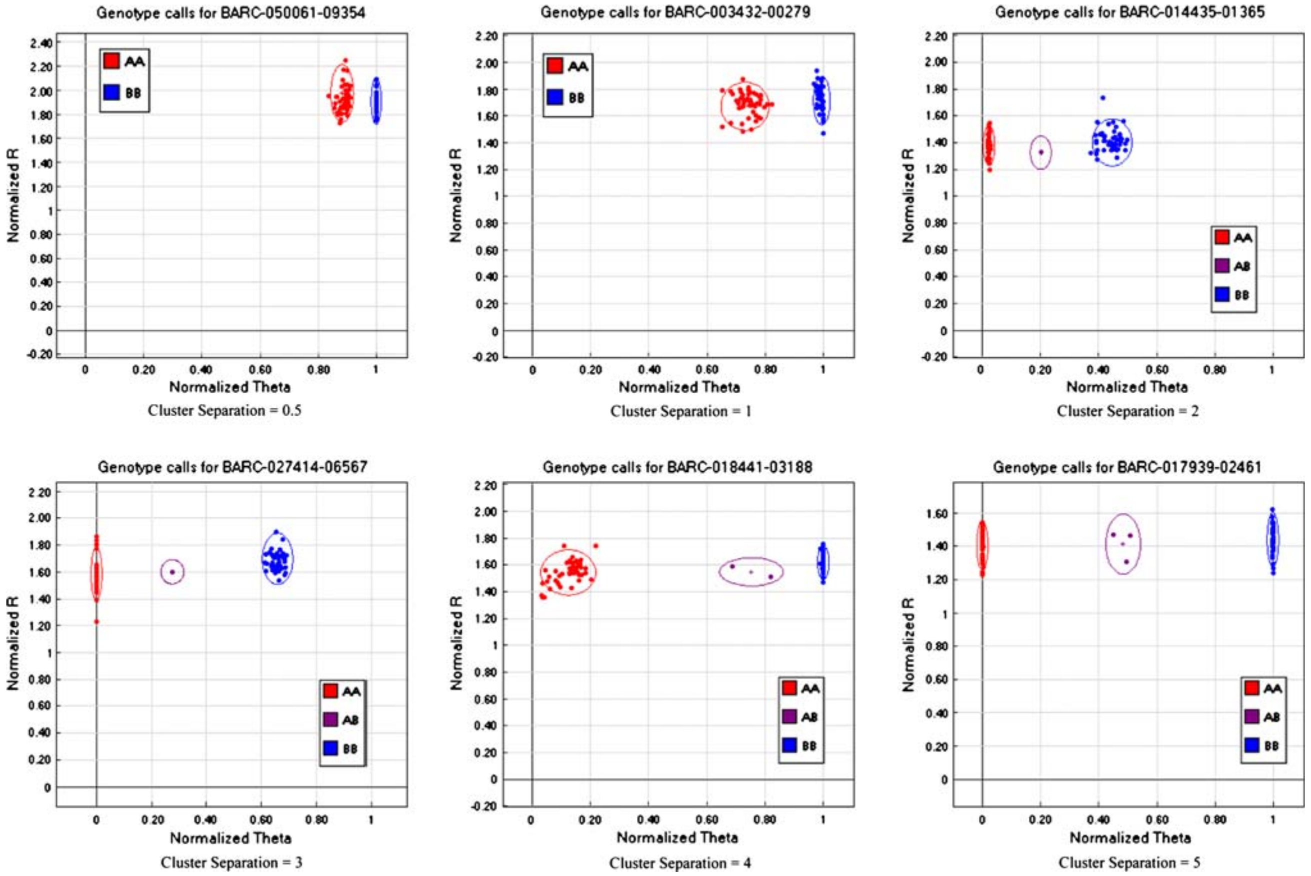

Fig. 1 Cluster separation using the average separation between the two homozygote clusters on the normalized *theta* value and converting this value to a 0.5–5 scale (0.5 = 0.1 normalized theta value separation, 1 = 0.2 normalized theta value separation, 2 = 0.4 normalized theta value separation, 3 = 0.6 normalized theta value separation, 4 = 0.8 normalized theta value separation, and 5 = 1.0 normalized theta value separation). The normalized *R* (*y* axis) is the normalized sum of intensities of the two channels (Cy3 and Cy5) and normalized theta (*x* axis) is $((2/\pi)\mathrm{Tan}^{-1}(Cy5/Cy3))$ where a normalized theta value nearest 0 is a homozygote for allele A and a theta value nearest 1 is a homozygote for allele B (Fan et al. 2006b)

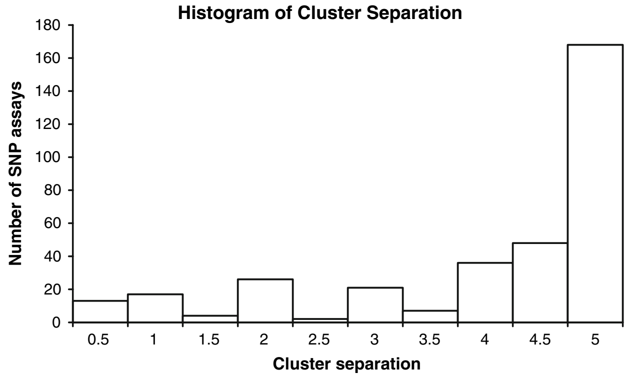

Fig. 2 Histogram of the homozygote cluster separations (in 0.5 scale increments) obtained in the assays of the 342 SNPs that were mapped with Soy OPA-1

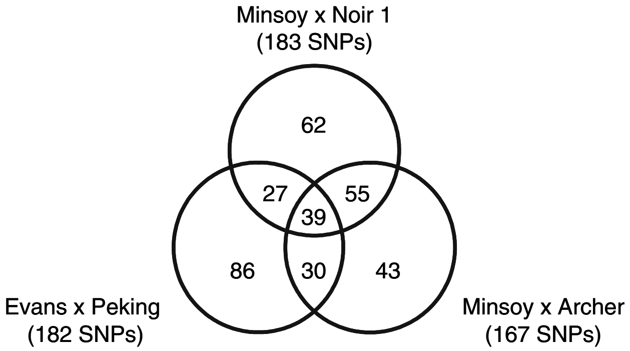

Fig. 3 A Venn diagram illustrating the number of SNP loci mapped in each, and shared between and among, the three recombinant inbred line mapping populations

a more diverse set of landraces. To answer this question, SoyOPA-1 was used to genotype 96 diverse landrace accessions. Figure 4 indicates a uniform distribution of minor allele frequencies in the 0.0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4 and 0.4–0.5 classes, with only 72 (21%) of the 342 loci falling into a class wherein the minor allele frequency was less than 0.1. This uniform distribution of allele frequencies was expected due to the ascertainment bias of initially discovering SNPs in a small sample of five genotypes (Hartl and Clark 2007).

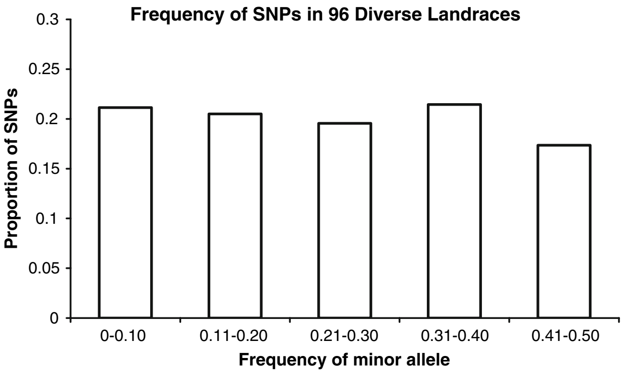

Fig. 4 Minor SNP allele frequency distribution in 96 diverse Asian landraces of the 342 SNP loci mapped with Soy OPA-1

Discussion

We have shown that a 384-SNP GoldenGate assay can be used successfully for SNP genotyping in soybean despite the high number of paralogous genes present in soybean. In total, SNP allelic data were obtained for 342 of the 384 SNPs in SoyOPA-1 and those SNPs were successfully mapped in one or more of the three populations (Fig. 3). This 89% success rate of the GoldenGate assay is comparable with the 90% success rate previously reported in barley (Rostoks et al. 2006), indicating that the greater degree of paralogous gene sequence in soybean will not be a factor in the design and use of future soybean GoldenGate assays. However, in a population with a large number of heterozygotes, SNPs with a cluster separation of 2 and below are unlikely to be scorable, i.e., the heterozygous cluster will not always be readily distinguishable from one or both of the homozygous classes. Of the 342 successful assays, 60 had cluster separations of <2. Therefore, in F_2 or backcross populations with high proportions of heterozygotes, the efficiency of converting a SNP into a GoldenGate assay would have decreased from 89 to 73%. Consequently, only GoldenGate-validated SNPs with good cluster separation (>2.0) should be selected for developing a custom OPA for F_2 or backcross mapping populations.

The compression of the two homozygous clusters is most likely due to soybean having a highly duplicated genome and the GoldenGate assay being sensitive to the number of copies (target locus + paralog/homoeolog) actually being assayed. An example of this is provided in data from the assay of SNP BARC_007900_00197 in the Minsoy x Noir 1 and Evans x Peking populations (Fig. 5). The SNP, BARC_007900_00197, was found to be polymorphic between the parents of both populations through resequencing and should map to the same location in both populations. In the Minsoy x Noir 1 population, the AA homozygous cluster has a normalized theta of 0, the BB homozygous cluster has a normalized theta of 1.0, and the AB heterozygous cluster (only one RIL) has a normalized theta value of approximately 0.5. The assay for the same SNP in the Evans x Peking population shows compression of the BB homozygous cluster toward the AA cluster. The normalized theta value for the BB homozygous cluster is 0.6 and the AB heterozygous cluster has a normalized theta value of 0.25. When both populations are combined, the Evans x Peking BB homozygous cluster is clustering with the Minsoy x Noir 1 AB genotype. This suggests that in the Evans x Peking population there is an additional locus with identical sequence complementary to the genome specific oligos (identical or nearly identical paralog) being interrogated by the GoldenGate assay that was designed for SNP BARC_007900_00197. The allele for the nearly identical paralog matches the AA allele at the allele specific base hence increasing the background signal when the RIL genotype contains a BB allele at the BARC_007900_00197 SNP allele moving the allele cluster to 0.6. Most likely the paralog is not present in Minsoy or Noir 1 or there is sequence variation in the Minsoy x Noir 1 paralog that has interfered with the hybridization of the SNP BARC_007900_00197 GoldenGate assay oligos to the paralog. Despite this compression of the homozygous clusters, BARC_007900_00197 still had enough cluster separation to be successfully mapped in the Evans x Peking population and to map to the same map location as it did in the Minsoy x Noir 1 population suggesting that the GoldenGate assay is capable of mapping SNPs occurring on a sequence that has a nearly identical paralog somewhere else in the genome.

BARC_007900_00197 is a good example of why controls with known SNP alleles need to be run with the GoldenGate assay in a complex genome. In Fig. 5 Minsoy and Evans both have the AA allele while Peking and Noir 1 have the BB allele. A cross between Peking and Noir 1 would generate in the RIL population two clusters for BARC_007900_00197 at theta values of 0.6 and 1.0 mimicking a true polymorphic SNP with some cluster compression. In fact, without knowing that BARC_007900_00197 is not polymorphic between Peking and Noir 1 the SNP assay would be scored in this hypothetical population and the nearly identical paralog would be the locus mapped. In this 384 OPA we did not encounter this scenario since all SNPs were originally discovered through resequencing and were only scored and mapped if they had been originally determined to be polymorphic through the resequencing.

The higher multiplex custom assay of 1,536 SNPs uses the same chemistry as the 384-plex reactions. The large number of SNPs that can be analyzed using the GoldenGate assay would certainly accelerate mapping a current backlog of 14,000+ identified, but as yet unmapped soybean SNPs. This assay will also be helpful in conducting association

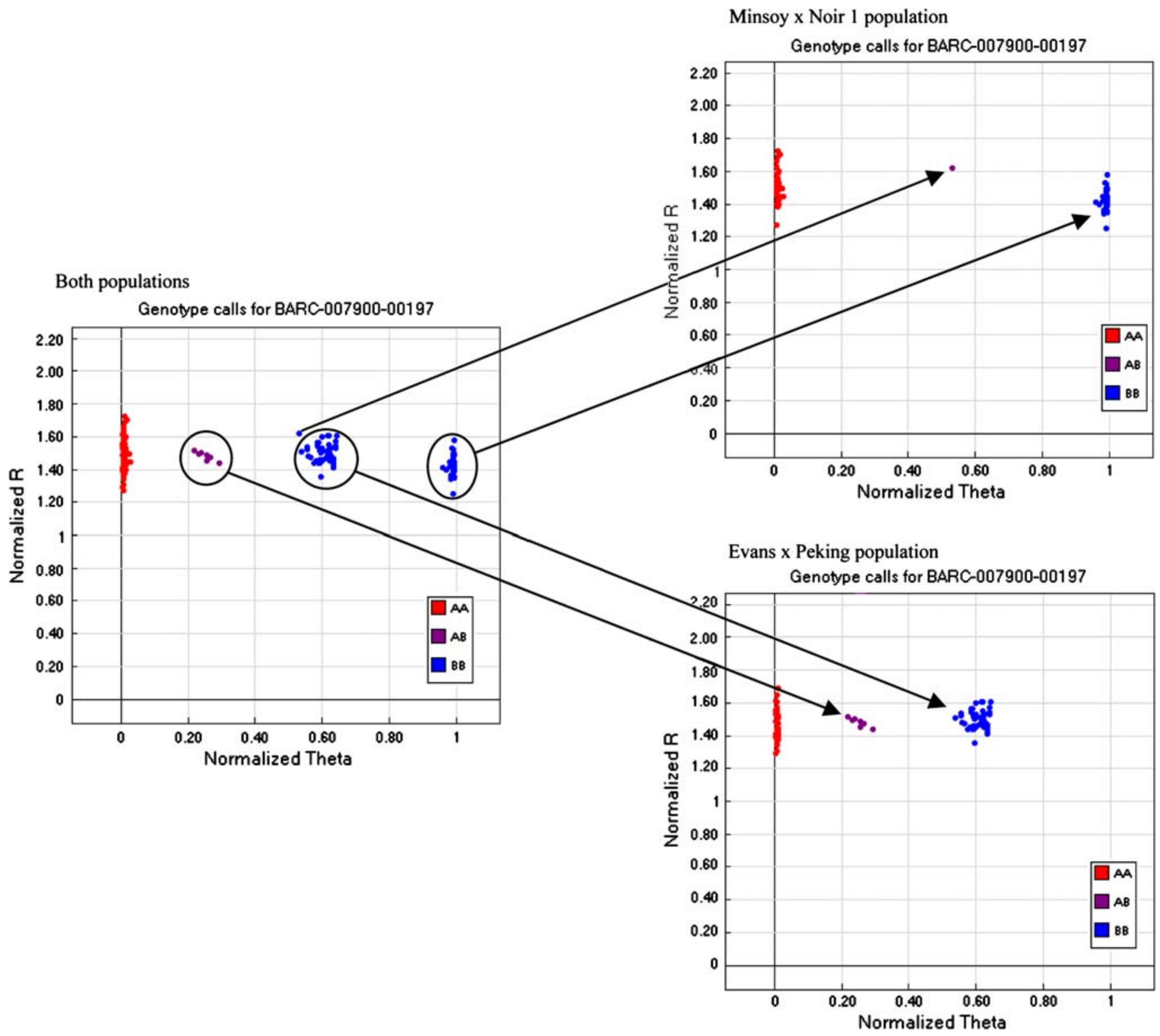

Fig. 5 Example of cluster compression with the GoldenGate assay of SNP BARC_007900_00197 in the Minsoy x Noir 1 and Evans x Peking populations. The normalized R (y axis) is the normalized sum of intensities of the two channels (Cy3 and Cy5) and normalized theta (x axis) is $((2/\pi)\mathrm{Tan}^{-1}\,(Cy5/Cy3))$ where a normalized theta value nearest 0 is a homozygote for allele A and a theta value nearest 1 is homozygous for allele B (Fan et al. 2006b)

analyses for fine mapping QTL. Conventional QTL analysis in mapping populations has a resolution limited to about 10–20 cM. In contrast, association analysis may be used in soybean to fine map these QTL with much greater resolution (Hyten et al. 2007). Still, even with a candidate QTL approach for association analysis, a large number of SNPs may be needed to efficiently fine map QTL (Hyten et al. 2007). The high multiplex capacity of the GoldenGate assay allows the analysis of sufficient loci to provide the SNP density likely needed to successfully fine map QTLs known to exist in a given 10–20 cM region using association analysis. However, the Illumina platform will also make conventional QTL discovery analyses much more efficient. We have currently designed two new custom 1,536 SNP GoldenGate assays. Along with the 342 successful assays in SoyOPA-1 and preliminary data from the additional two sets of 1,536 assays suggesting a very similar assay success rate, there will be well over 3,000 successful GoldenGate assays from which to select 1,536 SNP loci to create a universal SoyOPA linkage panel for QTL discovery. For this panel, SNP loci will be selected based upon (1) equidistant map position within each of the 20 consensus linkage groups, (2) a cluster separation >2, and (3) a high minor allele frequency, not only with respect to the set of 96 exotic landraces, but also with regard to a set of 96 elite cultivars released between 1990 and 2000.

This universal SoyOPA linkage panel will be quite useful for future QTL discovery research.

Acknowledgments We thank Alicia Bertles, Tad Sonstegard, and the BFGL-ANRI BARC East DNA Sequencing Facility for assistance with the genomic STS sequencing. This work was partially supported by United Soybean Board Projects 5212 and 6212. The support of the United Soybean Board is greatly appreciated.

References

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I, Young ND, Shoemaker RC, Van Tassel CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Concibido VC, Lange DA, Denny RL, Orf JH, Young ND (1997) Genome mapping of soybean cyst nematode resistance genes in ‘Peking’, PI 90763 and PI 88788 using DNA markers. Crop Sci 37:258–264

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999a) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Shoemaker RC, Matthews BF, Jarvik T, Young ND (1999b) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919–928

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78

Fan JB, Chee MS, Gunderson KL (2006a) Highly parallel genomic assays. Nat Rev Genet 7:632–644

Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, Lebruska LL, Laurent M, Shen R, Barker D (2006b) Illumina universal bead arrays. Methods Enzymol 410:57–73

Hartl DL, Clark AG (2007) Principles of population genetics, 4th edn. Sinauer Associates, Sunderland

Hymowitz T (2004) Speciation and cytogenetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, Wis., pp 97–136

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

Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944

Keim P, Olson TC, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. Soybean Genet Newsl 15:150–152

Mansur LM, Orf JH (1995) Agronomic performance of soybean recombinant inbreds in northern USA and Chile. Crop Sci 35:422–425

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40:420–427

Matukumalli LK, Grefenstette JJ, Hyten DL, Choi IY, Cregan PB, Van Tassell CP (2006) SNP-PHAGE–High throughput SNP discovery pipeline. BMC Bioinformatics 7:468

Rostoks N, Ramsay L, Mackenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Res 13:1250–1257

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS (2002) Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp. mays L.). Genetics 162:1401–1413

Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Plant Research Internation, Wageningen

Wang D, Shi J, Carlson SR, Cregan PB, Ward RW, Diers BW (2003) A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci 43:1828–1832

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134