

DEVELOPMENT AND APPLICATION OF 1536-PLEX SINGLE NUCLEOTIDE POLYMORPHISM MARKER CHIP FOR GENOME WIDE SCANNING OF INDONESIAN RICE GERMPLASM

Pengembangan dan Penerapan 1536-Plex Single Nucleotide Polymorfism Chip Marker untuk Membaca Genom Plasma Nutfah Padi Indonesia

Dwinita W. Utami, I. Rosdianti, P. Lestari, D. Satyawan, H. Rijzaani and I. M. Tasma

*Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development
Jalan Tentara Pelajar No. 3A, Bogor 16111, West Java, Indonesia, Phone +62 251 8337975, Fax. +62 251 8338820*

E-mail: bb_biogen@litbang.deptan.go.id

Corresponding author: dnitawu@windowslive.com

Submitted 3 January 2013; Accepted 17 September 2013

ABSTRACT

A successful molecular breeding program requires detailed and comprehensive understanding of the diversity of rice germplasm and genetic base of target traits. The objective of this research was to develop the high throughput 1536-SNP chip linked to heading date and yield component traits and used it for genotyping the diverse Indonesian rice germplasm. The genotype data obtained could be used for diversity analysis and genome wide association mapping study. A 1536-SNP genome wide assay was developed using the Illumina's GoldenGate technology. The SNP markers were selected in the rice genome regions containing heading date and yield component genes or regions where the quantitative trait loci (QTLs) of the two traits were mapped. The developed custom SNP chips were then used for genotyping 467 rice accessions showing diversity in heading dates and yield components. The assay can reliably be used for diversity analysis and mapping genes associated with heading date and yield component traits. For 1536-SNP BIO-RiceOPA-1 custom chip designed, a total of 34.832 SNPs distributed in rice genome particularly in the region of heading date and yield component genes or QTLs were identified. A total of 1536-SNP were selected and confirmed to be used for genotyping analysis. Analysis performance and quality of 1536-SNP BIO-RiceOPA1 showed that 60% (918/1536) of total SNP markers had a good differentiating power in scanning the rice accessions tested ($MAF \geq 0.2$). The 1536-SNP genome wide assay Illumina's GoldenGate designed was useful for diversity analysis and could be used as SNP marker for large scale genotyping in rice molecular breeding involving Indica-Indica, Indica-Japonica and Indica-Tropical Japonica crosses.

[**Keywords:** 1536 SNP-chip, genome scanning, rice germplasm, Indonesia]

ABSTRAK

Keberhasilan program pemuliaan padi berbasis molekuler memerlukan pemahaman komprehensif dan rinci tentang keragaman plasma nutfah dan aspek genetik dari sifat yang

ditargetkan. Penelitian ini bertujuan mengembangkan high-throughput 1536 SNP-chip yang terkait dengan karakter umur genjah dan komponen hasil unggul dan menggunakannya untuk analisis genotipe pada plasma nutfah padi Indonesia. Data genotipe yang diperoleh selain digunakan untuk analisis keragaman juga untuk analisis pemetaan berdasarkan uji asosiasi. 1536-SNP chip dari Illumina dikembangkan menggunakan teknologi GoldenGate assay. Marka SNP didesain dari hasil seleksi 34.832 SNP yang tersebar pada genom padi yang mengandung gen umur genjah dan komponen hasil unggul atau pada QTL yang terkait dengan kedua sifat tersebut. Custom 1536-SNP BIO-Rice OPA-1 chip yang didesain kemudian digunakan untuk analisis genotipe 467 aksesori plasma nutfah padi yang memiliki keragaman sifat umur dan komponen hasil. Hasil analisis terhadap keragaman kualitas dari 1536-SNP BIO-RiceOPA1 menunjukkan 60% (918/1536) dari total SNP memiliki tingkat diferensiasi yang baik dalam membaca keragaman genetik plasma nutfah padi ($MAF \geq 0.2$). Analisis genom menggunakan 1536-SNP dari Illumina dengan teknologi GoldenGate assay bermanfaat untuk analisis keragaman genetik plasma nutfah padi Indonesia dan dapat digunakan sebagai marka SNP dalam pemuliaan berbasis molekuler yang melibatkan subspecies padi Indica-Indica, Indica-Japonica, dan Indica-Tropical Japonica.

[**Kata kunci:** 1536 SNP-chip; scan genom, plasma nutfah padi, Indonesia]

INTRODUCTION

Rice is one of the most important food crops in the world. With a broad geographical adaptation and wide genetic diversity, rice is an excellent model crop for studying genetic and domestication process.

Single nucleotide polymorphisms (SNPs) are abundant and evenly distributed throughout the genome of most plant species (Yan *et al.* 2009). SNP discovery has begun to make large pools of SNPs to take advantage of rice genome information. While

high density SNP-chip detection system will be useful for a wide range of rice breeding programs and applications. A high throughput SNP platform will automate genotyping and allele calling which allow more efficient breeding strategies that were not feasible using previous gel-based genotyping system (Chen *et al.* 2011).

The successful implementation of MAS strategies is dependent on having an efficient and robust genotyping system in place (Collard *et al.* 2008). For many years, simple sequence repeat (SSR) markers have been the marker system of choice due to their high polymorphism rates, simple polymerase chain reaction-based system and the ability to detect on the gel electrophoresis system. SSRs have proven useful for diversity analysis (Ram *et al.* 2009; Zheng *et al.* 2011), QTL mapping (Wang *et al.* 2007; Wan *et al.* 2008) and marker-assisted breeding (Karakousis *et al.* 2003; Collard *et al.* 2008). However, the markers have difficulties in scoring precise allele sizes, running in high throughput system and reducing costs through multiplexing. Routine integration of markers into modern breeding programs requires high throughput genotyping platforms that can handle a large number of samples at a low cost (Aumann *et al.* 2005). Thus, a new generation of markers based on SNPs is now rapidly overtaking SSRs because of the new SNP genotyping platforms that offer multiplexed sets of markers for different applications. SNPs have the potential to greatly increase the speed and reduce the cost of molecular-marker genotyping, which makes it feasible to 'mainstream' MAS into conventional breeding programs.

The value of having a large pool of available SNPs is that a subset containing the most informative and useful SNPs can be selected for different applications and sets of germplasm. For certain applications, it may be important to select evenly a subset of SNP markers that are polymorphic between targeted germplasm groups, while others may use trait-specific functional SNPs that are diagnostic of desirable alleles. A key step in this process is validating SNPs performance with specific marker assays, since not every SNP will be able to be converted into a reliable marker across different genotyping systems. The objective of this research was to develop the high throughput 1536-SNP chip linked to heading date and yield component traits and used it for genotyping the diverse Indonesian rice germplasm. The genotype data obtained would be useful for diversity analysis and genome wide association mapping study.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

Four hundred and sixty seven rice accessions were used in this study. These included 29 accessions of released varieties, 34 accessions of near isogenic lines (NIL), 136 accessions of local varieties, 11 accessions of wild species and 162 accessions of breeding lines for SNP genotyping and characterization. Rice leaves were collected from a single plant of derived lines. The Thermo Scientific King Fisher Plant DNA kits were used for genomic DNA extraction (Qiagen 2011). DNA concentration was counted using a Nano Drop spectrophotometer (Thermo Scientific 2011). A minimum of 15 μL genomic DNA (50 $\text{ng } \mu\text{L}^{-1}$) was required for the GoldenGate assay. DNA was stored in TE buffer (10 mM Tris, pH 7.5: 1 mM EDTA). At least 10% of samples was duplicated within the samples to act as quality control (QC). The DNA concentration was standardized using dilution into 50 $\text{ng } \mu\text{L}^{-1}$ for final concentration. DNA purity was determined by using the A_{260}/A_{280} ratio of 1.8-2.0 (Sambrook dan Russell 2001).

Selection of Subsets of SNPs for Marker Application

SNPs were selected from several sources of rice SNP databases, including SNP Consortium (www.ricesnp.org), Illumina 1536 SNP chip on 395 *O. sativa* accessions (Zhao *et al.* 2010), IRFGC-Perlegen SNP (TIGR Pseudomolecules v5), and SNPs of the rice diversity projects developed by Susan McCouch Lab (www.ricediversity.org). The 1536 custom designed SNPs were selected based on genetic map position of genes or QTL contributed to heading date and yield component characters. The trait distributions from which the SNP markers taken in the rice chromosomes are shown in Table 1.

SNP Genotyping Using the GoldenGate Reader

The current study selected subsets of previously validated and new custom designed SNPs for running as 96- and 1536-SNP multiplexed sets. The SNP sets were designed for the Illumina GoldenGate assay, which uses locus and alleles-specific oligos with cy3/cy5 labelling to detect SNP alleles at each locus.

Table 1. Genetic map position of the genes/QTL target of rice for designing the custom SNP-chip.

Chr	Number of QTLs/genes	Traits ¹⁾	Trait mapped location (Mb)	Number of SNPs	
				Surveyed	Selected
1	4	YC, HD	20,9-41,5	14,423	469
2	2	YC, HD	17,5-34,7	1,776	102
3	5	YC, HD	1,7-34,0	5,009	215
4	3	MD	23,1-25,5	3,214	69
5	3	YC, HD	24,3-28,7	1,920	112
6	6	YC, HD	0,64-30,8	2,680	194
7	4	HD, MD	23,1-29,3	2,111	101
8	3	YC	18,9-24,3	1,041	57
9	2	YC, HD	11,8-20,5	965	47
10	1	HD	12,9-15,8	969	45
11	2	YC	1,5-30,6	77	64
12	1	MD	0,97-2,4	647	61
Total				2,690	1536

¹⁾YC = yield component; HD = heading date; MD = maturing date

These custom Oligo Pool Assay (OPA) sets were then run on the Illumina platform which consists of an iSCAN reader with autoloader and GenomeStudio analysis software which can be used with a variety of chemistries for genotyping (Illumina Product Guide 2009). The genotyping is carried out in large multiplexes of 1536 SNPs, in multiples of 96 SNPs, using Illumina custom SNP panels. The GoldenGate assay is an allele-specific oligo hybridization, ligation and extension assay followed by universal PCR amplification, allowing that no amplification bias can occur. These amplification products were then bound to the 3 µM microbeads in 32-sample bead chips and alleles were read by fluorescent readout using the iSCAN reader. The GenomeStudio software from Illumina was used for allele clustering based on the ratio of the cy3/cy5 signal intensities to call the three genotype classes. The resulting SNP calls were then reformatted for subsequent data analysis for SNP visualization using diversity analysis with Power Marker program (<http://statgen.ncsu.edu/powermarker>) and population structure using STRUCTURE (<http://pritch.bsd.ichicago.edu/structure.html>).

RESULTS AND DISCUSSION

1536 SNP-Chip Developed

The summary of custom SNP designed were done based on GoldenGate Illumina assay detection systems, as a result, 1536 SNPs were selected and developed as a 1536 SNP-chip. Hereafter this custom designed was termed as BIO-RiceOPA-1. The dis-

tribution of SNP marker of the 1536 SNPs and led to develop as the BIO-Rice OPA-1 is shown in Figure 1.

SNP Performance and Quality

The system using GoldenGate assay and iScan technology could genotype 96 samples and 1536 SNP markers in a single plate. All SNP genotyping data generated from Bead Array system were scored using the Illumina GenomeStudio genotyping software with a no call threshold of 0.25. Scoring of SNP genotyping data using the GenomeStudio software generally produced three clear clusters denoting the AA homozygote, BB homozygote and AB heterozygote. Panels with the normal genoplot usually showed three clear clusters, AA (homozygote), AB (heterozygote) and BB (homozygote) (Fig. 2A and B). In addition, the SNPs showing different cluster patterns were also observed (Fig. 2C and 2D). In this case, the genotypes of the rice samples could not be clearly clustered into the three definite clusters. The phenomenon is commonly observed in genotyping studies using SNP chip technology due to the nature of the SNPs studies in that particular rice samples.

Allele Frequency Estimates in Diverse Rice Germplasm

Allele frequency was calculated for characterizing the differentiation patterns of genetic diversity in the rice samples tested. Based on the full genotype data set, the minor allele frequency (MAF) and cluster separa-

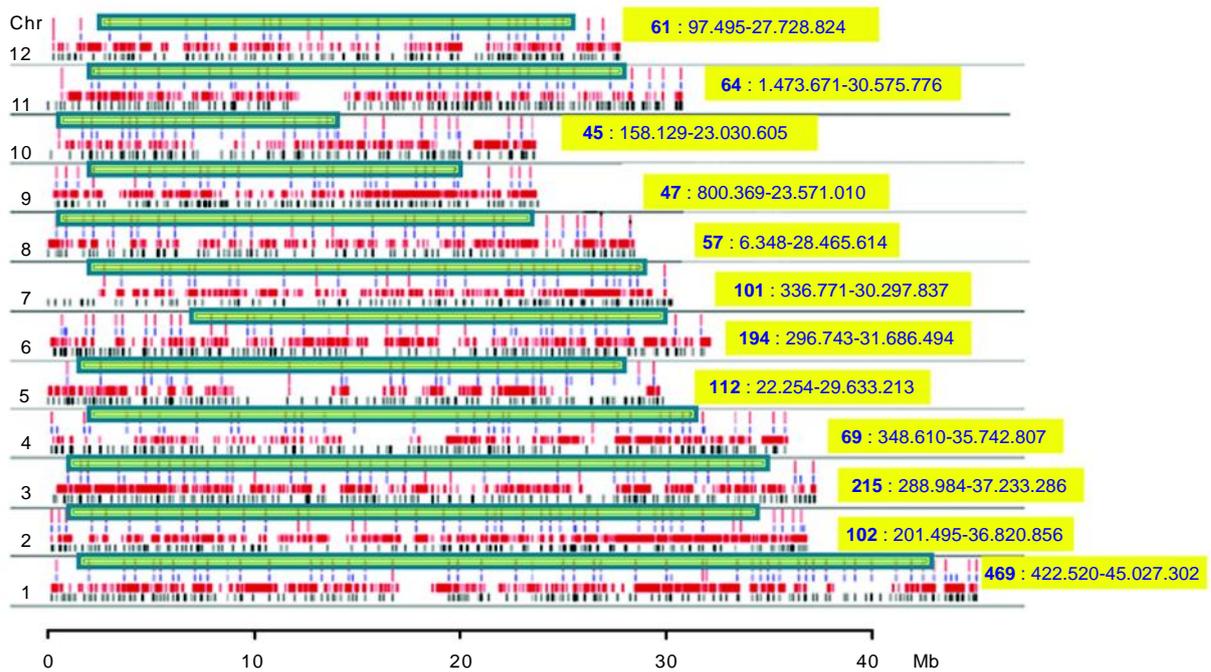


Fig. 1. Distribution of SNPs along the 12 chromosomes of rice. The yellow highlight along chromosomes represents 1536 SNP chips designed based on genetic map of yield component and heading date genes/QTL. References for designed SNP-chip: first row: 1536-plex (Zhao *et al.* 2010), second row: 44K (Tung *et al.* 2010), third row: 384-plex (Feltus *et al.* 2004), fourth row: 384-plex (Thomson *et al.* 2011).

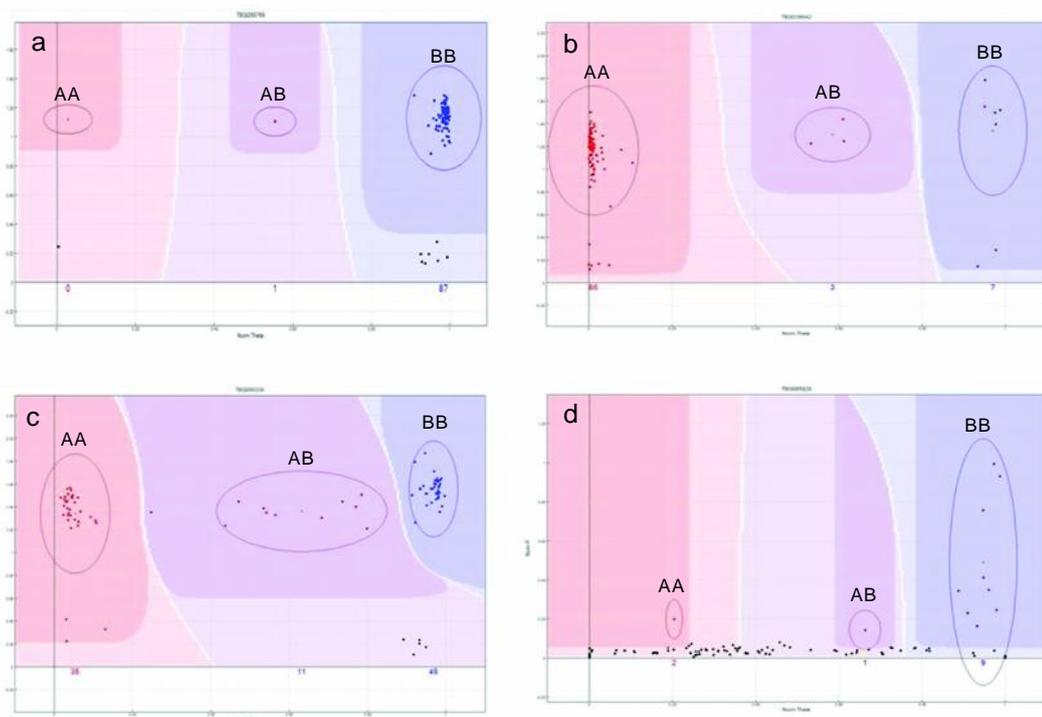


Fig. 2. Examples of SNP cluster patterns with the GoldenGate assay of 1536 SNPs genotyped in 467 rice accession using Genome Studio genotyping software; a and b = the clearly genoplotted differentiate homozygote and heterozygote clusters, c and d = the ambiguously genoplotted undefined homozygote and heterozygote clusters.

tion of the 1536 SNPs used are shown on Figure 3 and 4.

The distribution of MAF was observed in 10 continued classes from 0.05 to 0.5 with different numbers of SNPs in each MAF class. A total of 14.3% (219/1536) of the SNPs had a MAF of less than 0.15. It indicates that SNP markers with low MAF scores may not be informative for most diversity analysis. Alleles with very low frequencies generally have very little impact on large scale diversity studies and have a low probability of being polymorphic in mapping studies. Therefore, markers with higher MAF score should be valuable for screening a diverse source of rice germplasm. However, markers with low MAF score may be highly valuable in allele mining (Yan *et al.* 2009). Figure 3 shows that 918 (60%) of SNP markers showed $MAF \geq 0.2$ which were considered to

have a good differentiating power in distinguishing the rice accessions tested. Figure 3 also shows that 224 (14.6%) SNPs demonstrated $MAF > 0.4$. These SNPs are excellent to be used for differentiating the rice accessions under study.

The cluster separation (displayed as a GeneTrain scores) provided by the GenomeStudio software was used to describe the separation of the three genotypic classes of each SNP marker used in this study. Figure 4 shows that the cluster separation scores were distributed between 0.1 and 1.0. Around 98.5% of the SNPs tested could be called successfully with most of the cluster separation scores of greater than 0.3% and less than 2% with separation scores of 0.1-0.2 (23 SNP markers) for the total of 467 rice accessions tested. Therefore, more than half SNP markers were well separated.

The MAF distribution and cluster separation of 1536 SNPs resulted from this study were comparable with MAF and cluster separation of rice and maize, which showed distribution scores between 0.05-0.5 for MAF and 0.3-1.0 for cluster separation, respectively (Yan *et al.* 2009; Chen *et al.* 2011). The reliable genotypic classes measured in this study suggests a high stringent SNP selection which is relevant with previous report in other plants (Akhunov *et al.* 2009; Grattapaglia *et al.* 2011). The diverse SNPs assay performed well both within and across rice varietal groups not with standing the high nucleotide variation is pursued.

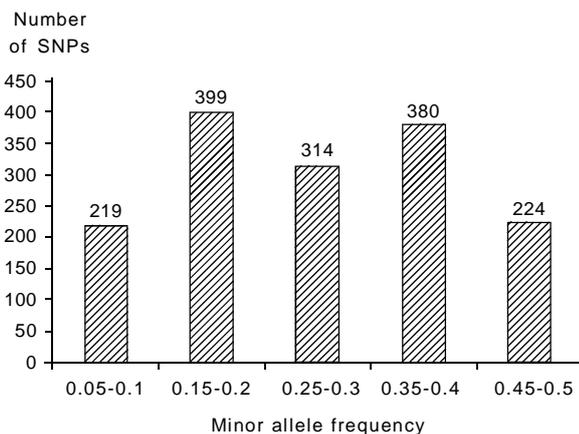


Fig. 3. Minor allelic frequency distribution in the 467 diverse rice germplasms based on a GoldenGate assay of 1536 SNPs.

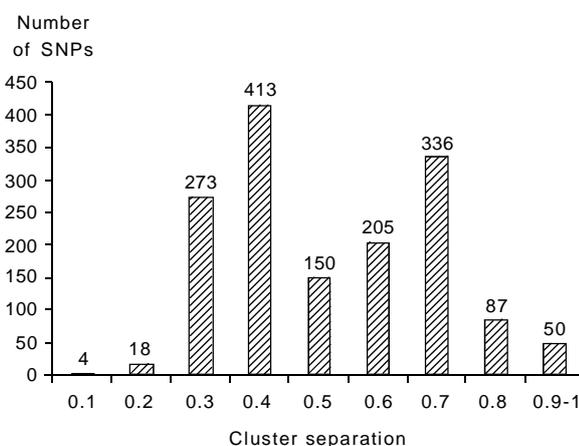


Fig. 4. Cluster separation distribution in the 467 diverse rice germplasms based a GoldenGate analysis using 1536 SNP chips.

Diversity and Population Structure Analysis

The most straight forward use of these SNP sets is to characterize the relatedness of a set of rice germplasm through a genetic analysis. Figure 5 shows the clustering and heat map of the 467 rice accessions based on the GoldenGate assay using the 1536 SNP chips. There were enormous variations of the genome profiling on the 467 rice accessions used in this study. The homozygote genome profiles were shown on double haploid rice lines, while the wide hybridizations rice lines samples demonstrated varied genome profiling indicating that they have heterozygote genome profiles.

To figure out the genomic distribution of the 284 individual subpopulation differences, the Bayesian cluster estimation of population structure was done using the STRUCTURE software. Ten replicates were performed for each value of K, the number of clusters considered. Each run used a burn-in period of 10.000 iterations. The best replicate giving the maximum

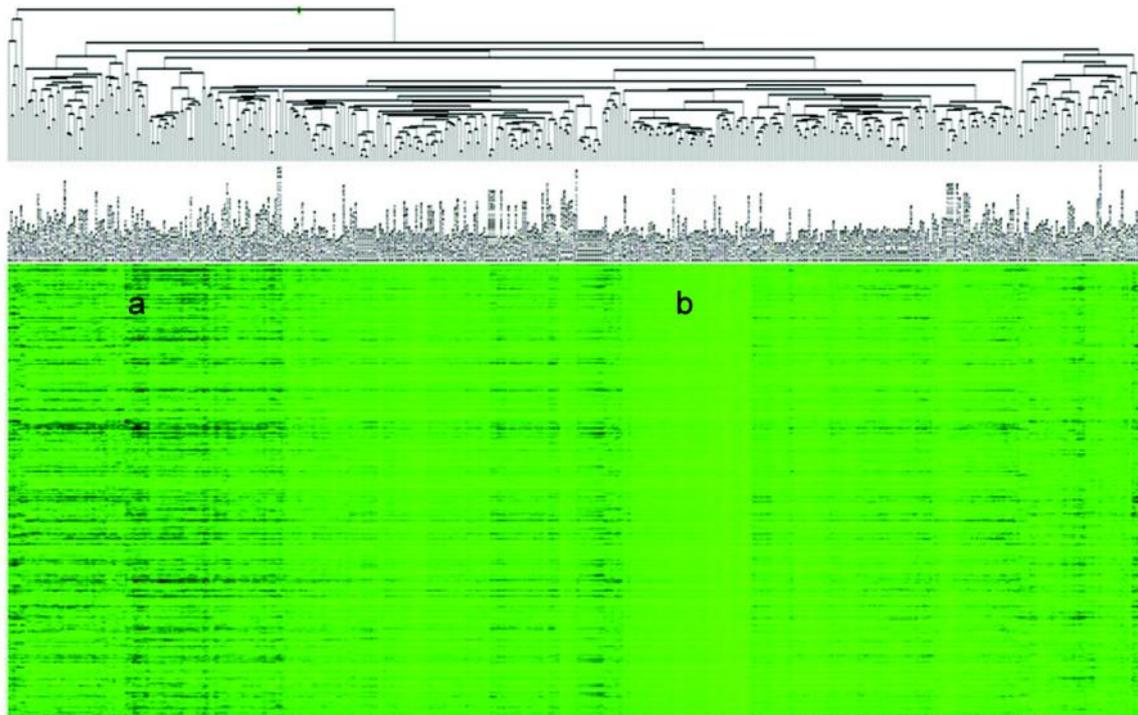


Fig. 5. Clustering and genome profile of 467 accessions of rice germplasm on 1536 SNPs; a=heterozygote profile genome on wide hybridization rice line samples, b= homozygote profile genome on double haploid rice line samples.

likelihood was chosen as the final for $K = 6$. The result of this analysis areas are shown in Figure 6. The genetic relationship among the accessions in the 284-individual subpopulation showed that the samples were separated in two main different subpopulations (Fig. 6a), i.e. subpopulation containing Japonica-Tropical Japonica varietal groups and that containing Indica-based varietal groups. The genetic divergence observed in this study between the Indica and Japonica groups led to conclusion that these subspecies may represent independent domestication events (Sweeny and McCouch 2007).

In the first cluster, the close genetic relationship between the Japonica and Tropical Japonica subpopulations was indicated by the fact that they shared alleles on the domestication traits, in this case particularly in yield component and heading date traits. This result was comparable with the studies of Zhao *et al.* (2010) who analysed the genomic diversity of 395 *O. sativa* using 1536 SNP from the high quality MBML intersection data in the *Oryza* SNP project (McNally *et al.* 2009). They found that Japonica and Tropical Japonica groups are the outcomes of

selection events from a single genetic pool that have been adapted to different climate conditions (Garris *et al.* 2005).

The second cluster is the subpopulation dominated by the Indica genetic background. Most of the Indonesian elit rice varieties were developed based on Indica genetic background. Indica has been introgressed into many other different varietal groups, including Japonica, Tropical Japonica and wild relatives, *Oryza glaberrima* and *Oryza rufipogon*. *O. glaberrima* is the African domesticated rice from the *O. barthii* (formally named *O. breviligulata*) ancestors (Vaughan *et al.* 2003), while *O. rufipogon* is the ancestor of *O. sativa* (Sweeny and McCouch 2007). They shared the same clusters because they have the same domestication traits (e.g. yield component, heading date and biotic and abiotic stresses). The population structure analysis of 284 individual rice accessions supported the same grouping as shown in Figure 6. It showed two main subpopulations, Indica (pink color) and Japonica (blue color) and one more block is Tropical Japonica (light blue color).

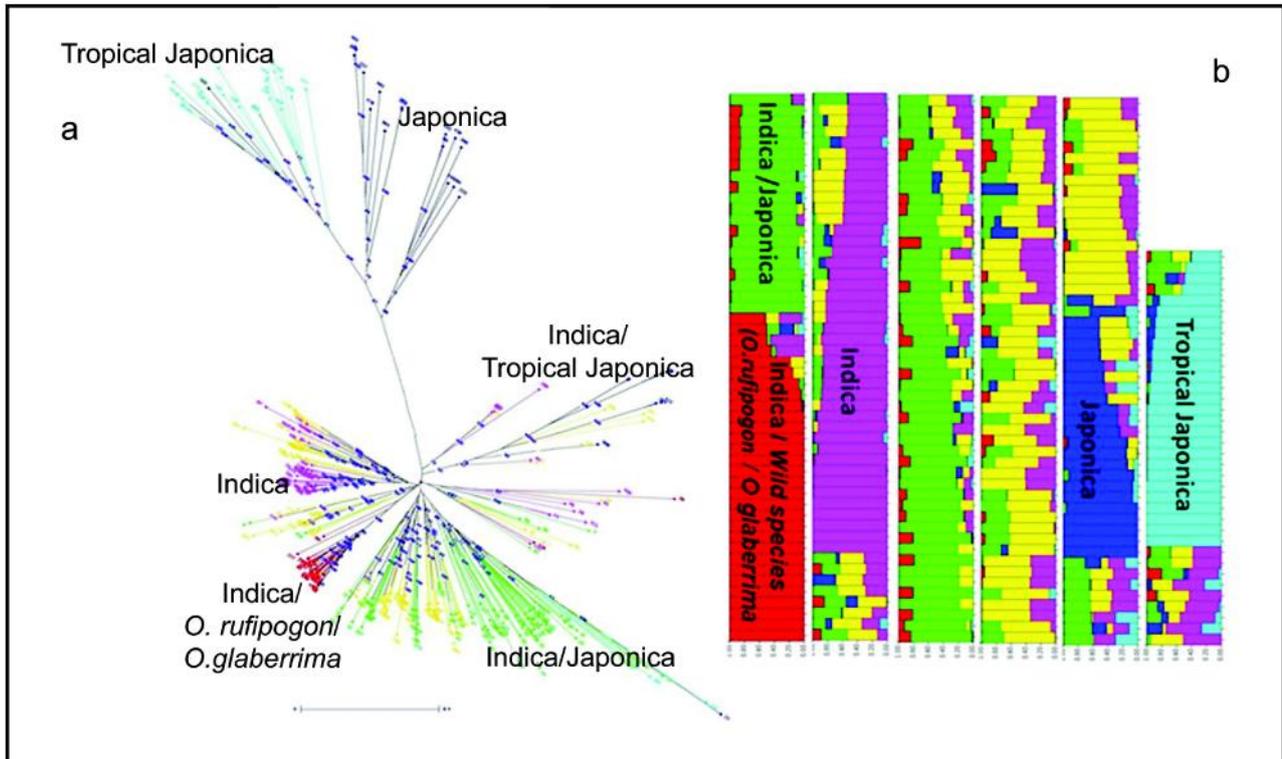


Fig. 6. Population structure of 284 accessions of rice germplasm based on Illumina-1536 SNP GoldenGate platform assay. a = phylogenetic tree of 284 accessions of rice germplasm, b = population structure estimated from STRUCTURE output for $K = 6$.

CONCLUSION

A 1536-SNP BIO-RiceOPA-1 custom chip was successfully designed based on selected SNP markers located in the region of rice heading date and yield component QTL. About 60% (918/1536) of SNP markers in the chip had a good differentiating power in scanning the 467 rice accessions with $MAF > 0.2$. The 1536-SNP chip is useful for large scale rice genotyping analysis for molecular breeding involving Indica-Indica, Indica-Japonica and Indica-Tropical Japonica genetic background.

ACKNOWLEDGEMENTS

This research was supported by the National Indonesian Budget of 2011-2012 Fiscal Year of the Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development. We are grateful to Dr. Michael J. Thomson for supporting sharing-custom design of 1536 SNP-chip.

REFERENCES

- Akhunov, E., C. Nicolet, and J. Dvorak. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* 119: 507-517.
- Aumann, Y., E. Manisterski, and Z. Yakhini. 2005. Designing optimally multiplexed SNP genotyping assays. *J. Computer and System Sci.* 70: 399-417.
- Chen, H., H. He, Y. Zou, W. Chen, R. Yu, X. Liu, Y. Yang, Y.M. Gao, J.L. Xu, L.M. Fan, Y. Li, Z.K. Li, and X.W. Deng. 2011. Development and application of set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 123: 869-879. DOI 10.1007/s00122-011-1635-5.
- Collard, B. and D. Mackill. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B.* 363: 557-572.
- Feltus, F.A., J. Wan, S.R. Schulze, J.C. Estill, N. Jiang, and A.H. Paterson. 2004. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Res.* 14: 1812-1819.
- Garris, A.J., T.H. Tai, J. Coburn, S. Kresovich, and S. McCouch. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631-1638.

- Grattapaglia, D., O.B. Silva-Junior, M. Kirst, B.M. de Lima, D.A. Faria, and G.J. Pappas. 2011. High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biol.* 11: 65.
- Illumina Product Guide. 2009. SNP genotyping and copy number analysis. Illumina Inc, San Diego, USA.
- Karakousis, A., A.R. Barr, K.J. Chalmers, G.A. Ablett, T.A. Holton, R.J. Henry, P. Lim, and P. Langridge. 2003. Potential of SSR markers for plant breeding and variety identification in Australian barley germplasm. *Aust. J. Agric. Res.* 54(12): 1197-1210.
- McNally, K.L., K.L. Childs, R. Bohert, R.M. Davidson, K. Zhao, B.J. Ulat, G.G. Zeller, R.M. Clark, D.R. Hoen, and T.E. Bureau. 2009. Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* 106: 12273-12278.
- Qiagen. 2011. TissueLyser Handbook for High-Throughput Disruption of Biological Samples: Purification of DNA from plant tissues (mini protocol). Second Ed. p. 3.
- Ram, G.G., V. Thiruvengadan, and K.K. Vinod. 2009. Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. *J. Appl. Genet.* 48(4): 337-345.
- Sambrook, J. and D.W. Russell 2001. *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor Laboratory Press.
- Sweeny, M. and S. McCouch. 2007. The complex history of domestication of rice. *Ann. Bot.* 100: 951-957.
- Thermo Scientific. 2011. KingFisher Plant DNA Kit: Instruction manual. Thermo Fisher Scientific, Inc., Finland.
- Thomson, M.J., K. Zhao, M. Wright, K.L. McNally, H. Leung, and S.R. McCouch. 2011. Development and application of 96- and 384-plex single nucleotide polymorphism (SNP) marker sets for diversity analysis, mapping and marker-assisted selection in rice. *Proceeding of Second Africa Rice Congress: Innovation and Partnerships to Realize Africa's Rice Potential*. Bamako, Mali, 22-26 March 2011.
- Tung, C.W., K. Zhao, M.K. Wright, M.L. Ali, J. Jung, J. Kimball, W. Tyagi, M.J. Thomson, K. McNally, H. Leung, H. Kim, S.N. Ahn, A. Reynolds, B. Scheffler, G. Eizenga, A. McClung, C. Bustamante, and S.R. McCouch. 2010. Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* sp.). *Rice* 3: 205-217.
- Vaughan, D.A., H. Morishima, and K. Kadowaki. 2003. Diversity in the *Oryza* genus. *Curr. Opin. Plant Biol.* 6: 139-146.
- Wang, J., X. Wan, H. Li, W.H. Pfeiffer, J. Crouch, and J. Wan. 2007. Application of identified QTL-marker associations in rice quality improvement through a design-breeding approach. *J. Theor. App. Genet.* 115: 87-100.
- Wan, X., J. Weng, H. Zhai, J. Wang, C. Lei, X. Liu, T. Guo, L. Jiang, N. Su, and J. Wan. 2008. Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. *Genetics* 179(4): 2239-2252.
- Yan, J., X. Yang, T. Shah, H.S. Villeda, J. Li, M. Warburton, Y. Zhou, H.J. Crouch, and Y. Xu. 2009. High-throughput SNP genotyping with the GoldenGate assay in maize. *J. Mol. Breed.* 25: 441-451. DOI 10.1007/s11032-009-9343-2.
- Zheng, L.Y., X.S. Guo, B. He, L.J. Sun, Y. Peng, S.S. Dong, T.F. Liu, S. Jiang, R. Ramachandran, C.M. Liu, and H.C. Jing. 2011. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biology*, 12: R114.
- Zhao, K., M. Wright, K.J. Kimball, G. Eizenga, A. McClung, M. Kovach, W. Tyagi, L. Ali, C.W. Tung, and A. Reynolds. 2010. Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLOS One* 5: e10780.