EXECUTIVE SUMMARY OF RESEARCH COMPLETION REPORT

MARKER DEVELOPMENT IN SOYBEAN AND REDBEAN USING NEXT GENERATION SEQUENCING AND STUDY ON SOYBEAN SEED-BORNE DISEASE

PUJI LESTARI

IDB Ref. 26/IND/D32

Home Institute:

Indonesian Center for Agricultural Biotechnology and
Genetic Resources Research and Development

Sponsored by

ISLAMIC DEVELOPMENT BANK

Host Institute:

SEOUL NATIONAL UNIVERSITY

THE REPUBLIC OF KOREA

2013

MARKER DEVELOPMENT IN SOYBEAN AND REDBEAN USING NEXT GENERATION SEQUENCING

AND

STUDY ON SOYBEAN SEED-BORNE DISEASE

Post Doctoral Program supported by Islamic Development Bank

and

Under the supervision and funding project of Prof Suk-Ha Lee, Department of Plant Science, Seoul National University

By

Puji Lestari

IDB Ref. 26/IND/D32

December 2011-December 2012

Permanent address:

Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (ICABIOGRAD)

Jl. Tentara Pelajar No. 3A, Cimanggu, Bogor 16111, West Java, Indonesia

Telp:+62-251-8337975, Fax. +62-251-8338820

E-mail:lestari_71@yahoo.com

SUMMARY

This report contains five chapters which describe research studies on soybean and redbean. These focus on works related to next generation sequencing technology in parallel with bioinformatics and microbiology. The chapters are as follows: 1) Resequencing whole genome of Indonesian soybean cultivar; 2) Redbean SNP marker development by resequencing; 3) Gene divergence analysis of recent segmental duplication for prediction of a major QTL associated with seed protein in soybean; 4) Characterization of *Fusarium solani*, a causal agents of soybean sudden death syndrome' and 5) Molecular characterization of novel antifungal bacteria against soybean seed-borne disease. All works reported here were conducted as team in Lab. Crops Genomics, Seoul National University. Importantly, some results in these studies will be linked with research program of Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development and tried to be applied.

RESEQUENCING WHOLE GENOME OF INDONESIAN SOYBEAN CULTIVAR

Bioinformatics analysis assisted by Yangjae Kang

SUMMARY

Whole genome of three soybean cultivars were sequenced using shotgun paired-end library and Illumina Hi-Seq. High quality of sequences ware obtained. With 500 bp insert, Anjasmoro, Wilis and Tanggamus produced 10,116,235,800 bp, 11,445,032,200 bp, and 13,651,431,600 bp with mapping coverage of 8.89, 8.86 and 13.8 respectively. These genomes sequences showed high percentage when mapped on reference genome (William82) ranging from 95-97%. Sequence variations of three cultivars against Willia82 especially SNP and indels (insertions/deletions) were detected in coding region across total chromosomes. Around 1000 indwells were identified based on the alignment of Ajasmoro and Wilis on William82, and more indwells (around 1700 indels) on Tanggamus with William82. Various indwells ranging from one to dozens were observed among three cultivars genome. Simple sequence repeat (SSR) appeared to be various among indels which showed as two bases (CT, GT, GA etc), three bases (AAG, CCT, GAT, GGA, ATT, TTC, GGT, AAC, TGA, TCC, GTT, CGG, CCA, CAA, TAA etc), four bases (TCCC, GTTG etc), six bases (GATTTG, CAACAG etc) etc.

In coding region, abundant SNPs were identified, accounted for 389,943; 344,202; and 775,645 SNP for Anjasmoro, Wilis and Tanggamus, respectively. Among base substitution, transition (A/G, C/T) seemed more dominant than transversion (T/G, G/C, A/T, A/C) on three cultivars. However, only few sequence variations among three cultivars themselves, indicating close genetic relationship of them. Indels and SNPs in coding regions could be valuable marker. Thus among SNP detected on three cultivars against William82, approximately 9000 flanking sequences of 500 BP containing SNP in the middle was extracted. Screening these SNPs was needed to find potensial SNP associated with genes in soybean. These raw sequences will be further analyzed together with whole genome sequences of other soybean cultivars resequenced by ICABIOGRAD for the need of SNP chip for Golden Gate assay in tropical soybean.

REDBEAN SNP MARKER DEVELOPMENT BY RESEQUENCING

Lab work assisted by: Kwang-soo Han, Bioinfomatics: Yangjae Kang

SUMMARY

Whole genome of two redbean, Kyungwon and wild redbean were sequenced using shotgun pair-end library and Hi-Seq2000 (Illumina). No reference genome of redbean is available, thus de novo assembly was conducted on Kyungwon redbean. Approximately 18.8 and 17.66 Gb were produced by Kyungwon and wild redbean with high quality, respectively. Ratio after trimming on both accession showed 95 and 97%. Based on alignment of contigs longer than 10kb, abundant SNPs were discovered. All possible sequence variations especially indels and SNPs were called using SAMTools program. To decrease the false positive SNPs and indels, thus minimum and maximum mapped read depth to the Kyungwon contigs of 5 and 100, respectively; sequencing quality of 100. Various length of indels between Kyungwon and wild redbean are presented. Range of indwells between 1 to 5 bp showed normal distribution. Total of 1,565,699 SNPs were detected between Kyungwon and wild redbean. Among base substitution type, transition appeared higher (929,432 bp) than transversion (636,267 bp). Six types of bi-allelic SNP contribute transition with 2 possibilities and transversion with 4 possibilities. Base change of A/C (463,560 bp) and C/T (465,872 bp) showed comparable number.Transversion of G/C (149,942 bp) seemed lower between two accessions than other base change types (T/G, A/T, A/C). Around 213,758 SNPs were identified in coding sequence which showed higher non-synonymous (57.2%) than synonymous base substitution (42.8%) Although the percentage of sequencing errors is less than 1% in NGS platforms, sequencing error is still one of the potential obstacle in SNP false-calling. The most of false-calling was produced from indel in platforms, Illumina Solexa, SOLiD and Roche 454 (Harismendy et al. 2009)

SNPs in coding regions showing amino acid change, especially altering gene function with phenotype could be valuable genome-based markers (Van et al. 2004).

High quality SNP positioned in coding region were selected and decreased for validation. A subset of 96 candidate SNPs (SNP1 to SNP96) from 96 different contigs were randomly selected. Afterward, 96 primers were designed based on the SNPs using Primer3. Target redbean DNA sequence of 96 randomly selected contig were successfully amplified in PCR using gradient PCR This gradient PCR is able to find optimum melting temperature during amplification. Of total SNPs, 96 SNPs in 96 contigs revealed high rate of validation based on the well matching between SNP using NGS and Sanger method. This study demonstrated that most of alleles of redbean accessions origin from three countries (South Korea, Japan, China) belonged to Kyungwon, not wild redbean. However some primers designed based on these SNPs showed less quality on some accessions, therefore further work for sequencing or redesign primers are needed. The validated SNP primers could be used as marker applicable on redbean. This work will be continued by team of Lab. Crop Genomics.

GENE DIVERGENCE ANALYSIS OF RECENT SEGMENTAL DUPLICATION FOR PREDICTION OF A MAJOR QTL ASSOCIATED WITH SEED PROTEIN IN SOYBEAN

Puji Lestari, Jay Ern Lee, Yang Jae Kang, Kyujung Van, and Suk-Ha Lee

SUMMARY

A second round of soybean segmental duplication occurred approximately 13 million years ago, which had several effects on evolutionarily duplicated loci and genes. Since soybean seed protein is a major source of animal feed, breeding efforts have been directed toward enhancing seed protein content in soybean. The availability of soybean genome and related genomic feature elements facilitates comprehending of quantitative trait loci (QTL) for seed protein content in concordance with their homeologous regions. After recently duplicated segments involving seed protein content were compared, it revealed that the appearance or disappearance of QTLs related to protein and/or oil was occurred by increased divergence after recent duplication. Also, the analysis of gene and sequence divergence between recently duplicated segments of a major QTL for seed protein

content on Chr 20 showed the homeologous regions between Chr 20 and Chr 10 which shared 27 out of 81 genes. Interestingly, Glyma20g17960 and Glyma10g24060, the GTP-binding family protein gene, showed highly divergent gene structure. Our study indicates that the divergence of loci related to protein/oil content after the recent segmental duplication may allow the possibility of identifying candidate genes for agronomic traits including seed protein content.

CHARACTERIZATION OF *Fusarium solani* (Mart), A CAUSAL AGENTS OF SOYBEAN SUDDEN DEATH SYNDROME

Work team: Puji Lestari and Suli Sun

SUMMARY

This report described characterization of indigenous *F. solani* isolated from soybean and their evaluation of SDS resistance on soybean. Eleven out of 57 fungal isolates belonging to *Fusarium* genus were identified as *F. solani* based on inter transcribed spacer (ITS) region. Morphological characteristics and translation elongation factor (EF-1α) confirmed that ten among the eleven isolates were *F. solani* which are consistent with ITS identification. The ten isolates genetically closed to other *F. solani* species complex causing soybean SDS, and their ascospore and mycelia structures demonstrated to persist for longer defence. Virulence varied among isolates based on disease severity ranging from moderately to highly resistant Korean cultivars. Hyphae found in xylem and phloem tissues of discoloration root and basal stem with foliar SDS symptom indicates that fungal infection into the tissues contributes in foliar symptom expression. A clue of contrasting SDS severity between some cultivars pairs could be beneficial to identify QTL for SDS and also provide genetic resources of resistance to develop new breeding lines/cultivars. The high aggressiveness of isolates SSLP15 and SSLP1 could be novel SDS pathogens that prospectively assist for evaluation of soybean germplasm to SDS resistance.

MOLECULAR CHARACTERIZATION OF NOVEL ANTIFUNGAL BACTERIA AGAINTS SOYBEAN SEED-BORNE DISEASE

Writing of manuscript (report): Puji Lestari, Lab work: Suli Sun

SUMMARY

Bacteria isolated from soybean roots or rhizosphere produced metabolites that could be inhibitory to survival. A primary selection was performed from the antagonism test plates to observe the confluent growth of bacteria inhibited development of fungal mycelia. Pure bacterial cultures were tested for fungal antagonism with respect to fungi growing alone. As demonstrated in this study, seed-borne bacteria isolated from soybean plants with disease symptom have various antagonistic activity, medium to high levels of antifungal activity against soybean seed-borne diseases. Of the 130 isolates screened for fungal mycelial growth, 47 and 49 strains showed PIRG of 70-80% on *P. longicolla* and *F. solani*, respectively. And, 70 strains with PIRG higher than 80% are against both *P. longicolla* and *F. solani.* Surprisingly, bacterial strains which have high PIRG on *P. longicolla* also showed high effectivenessagainst*F. solani* growth. These mycelial plugs failed to further growth when plated on PDA plated. This indicates that these bacterial strains have high potency as antifungal against both PSD and SDS in soybean.

All seed-borne bacterial strains were subjected for molecular identification and amplified the expected fragment size of the 16S rDNA sequence.Sequencing of PCR product of the each strains allowed to classify them to the related known species. Most strains belonged to Bacillus genus in several species, especially *Bacillus* sp, *Bacillus subtilis* and *B. amyloliquefaciens, B. tequilensis,* and *B. siamensis*.The sequences of the strains exhibited close similarity to Bacillus species ranging from 94 to 100% according to Blast analysis. Some strains with low portions revealed high homology with other species, including *Paenibacillus jamilae, P. polymixa,* and *Pantoea rodasii. Bacillus* spp and *Paenibacillus* spp, as Gram positive bacteria are well known as antifungal agents. *B. subtillis*and *B. amyloliquefaciens*has been progressively studies as biocontrol for many plants phytopathogens (Yu et al 2002, Al-Obaidy 2010; Zhang et al 2012). A wide range of antifungal activities toward different kinds of pathogenic fungi have been shown by Paenibacillus spp (Lee et al

2008; Al-Obaidy 2010). Thus, these bacterial strains identified in our this study with antifungal activity on fungal causing PSD and SDS might represent good alternative or the most potent biological control agent toward seed-borne diseases in soybean.