

Distribution of Genetic Variants in Korean Soybeans

Kitae Song*, Jeong Hoon Kim*, Gi Yong Yoon*, Hyo Chul Kim*, Seungho Shin*,
Won Cheol Yim*, Kyung-Hee Kim*, and Byung-Moo Lee*[†]

*Department of Life Science, Dongguk Univ.-Seoul, Seoul 100-715, Korea

ABSTRACT Next generation sequencing technologies provide opportunities to reveal the genetic variants and differentially expressed genes. The genetic variants are closely relevance to understanding of genes and phenotypic differences related to agronomic characteristics among cultivars. In this study, we conducted RNA-seq using two Korean soybean accessions, including Daewon and Hwangkeum, by using next generation sequencing against Williams 82 genome as reference. A number of variants such as single nucleotide variants (SNV), multiple nucleotide variants (MNV), insertion/deletion (InDel) and replacement, was 34,411 and 55,544 in Daewon and Hwangkeum, respectively. Among these variants, 9,611 non-synonymous variants were detected within 4,290 genes in Daewon and 13,225 non-synonymous variants were located on 5,672 genes in Hwangkeum. The distribution of non-synonymous variants and expression values of genes can serve as invaluable resource for genotyping and study of traits within genes for soybean improvements.

Keywords : soybean, genetic variants, RNA-seq

Soybean (*Glycine max*) is one of the most important crops worldwide and it is providing oil and useful protein. Despite of many advantages, soybean has not been researched in genomic details than other model plants because of its complex genome structures, high haploid chromosome number ($n=20$), and large genome size. The 20 haploid chromosomes in soybean were almost doubled such as rice ($n=12$), maize ($n=10$) and *Arabidopsis thaliana* ($n=5$). After the soybean genome sequence of Williams 82 had been established (Schmutz *et al.*, 2010), some papers have shown whole-genome resequencing studies of *Glycine max* and *Glycine soja* (Kim *et al.*, 2010; Lam *et al.*, 2010; Chung *et al.*, 2014). The reference genome allows us to find more single nucleotide polymorphisms (SNP) (Zhu *et al.*, 2003; Hyten *et al.*,

2006; Lam *et al.*, 2010; Chung *et al.*, 2014), structural variation (SV), such as copy number variation (CNV), large insertions/deletions (InDels), inversions and translocations (McHale *et al.*, 2012; Haun *et al.*, 2011). Next generation sequencing (NGS) is also an important technology in the genomic studies and molecular breeding. In addition to its availability of the soybean genome sequence, NGS technologies and bioinformatics provides many opportunities to reveal genetic diversities such as SNP and InDel among soybean accessions. These technologies are valuable resources that have facilitated whole-genome genotyping, genome-wide association studies, predictions of gene functions, and analysis of genetic diversities in plants (Lam *et al.*, 2010; Chung *et al.*, 2014; Li *et al.*, 2013).

The genome re-sequencing provides enormous number of new information on the record of SNPs. Over 3.8 million SNPs were identified through high depth of re-sequencing of 10 cultivated and 6 wild accessions (Chung *et al.*, 2014). Recently, SoySNP50k platform was developed, which comprise over 19,000 accessions (Song *et al.*, 2013). Genome-wide identification of SNPs among soybean accessions was also reported and developed as markers for breeding (Zhu *et al.*, 2003; Van *et al.*, 2005). In contrast to SNPs which have been studied extensively, other forms of highly abundant genetic variation such as InDel, have received relatively less attention. Small InDels, a type of polymorphisms are common but functionally important type of polymorphism. Especially InDels in coding region cause single or more amino acid substitutions such as frameshift mutations and affect sequence conservations, structure and function of proteins. They are easily detected by based on mismatching in common beans (Zou *et al.*, 2014). However many studies focus on single nucleotide variants (SNVs) or large structural variants (Wu *et al.*, 2010; McHale *et al.*, 2012; Lee *et al.*, 2015). DNA polymorphisms positioned within coding regions are particularly important as it has a

[†]Corresponding author: (Phone) +82-31-961-5130 (E-mail) bmlee@dongguk.edu

<Received 15 May, 2015; Accepted 18 May, 2015>

potential to change the protein function itself. Discovery of polymorphisms related to functional changes of genes is important in investigations on differences among various accessions and its main cause of diversions.

In this study, we have obtained RNA-sequencing data of 2 soybean accessions (Daewon and Hwangkeum) using NGS. The sequence data were analyzed to discover a large number of DNA polymorphisms, including single nucleotide variation (SNV), multiple nucleotide variation (MNV), insertion/deletion (InDel) and replacements. This data can be used in future studies on genotyping and marker development for molecular breeding.

MATERIALS & METHODS

Plant materials and RNA-sequencing

To detect polymorphisms, two Korean soybean accessions were analyzed in this study (Daewon and Hwangkeum). The soybean seeds were obtained from the National Institute of Crop Science, RDA, Suwon, Korea. The plants were grown until V5 stage at 25°C under 16 h light / 8 h dark conditions. For all samples, total RNA was prepared from leaf tissues using an RNeasy Plant Mini Kit (Qiagen, Germany). The RNA quality was checked for integrity before performing the RNA sequencing process using a 2100 Bioanalyzer RNA Nanochip (Agilent Technologies). After a quality check, RNA sequencing using Illumina HiSeq platform was performed by Theragen, Inc. (Korea).

Alignments and variants detection

Raw reads from Illumina Hiseq 2000 sequencing platforms were imported to CLC Genomics Workbench 7.0 (CLC Bio) and read statistics were assessed using sequencing data quality control, followed by read trimming for quality. The raw read were

trimmed with a quality score limit of 0.01. The CLC Genomics Workbench was used to map trimmed reads to the soybean genome sequence of Williams 82 from Phytozome v.9 (<http://www.phytozome.net>) used as the reference sequences. The mapping process followed defaults parameters, and global alignment setting. Consensus sequence were aligned to the reference sequence and analyzed for SNVs, MNVs, InDels and replacements. The variants were screened using the CLC Genomics Workbench 7.0 with the following parameters (minimum count: 3; minimum coverage: 10; minimum variants frequency: 35%).

RESULTS

RNA-Seq and assembly

A total of 78,994,236 reads were obtained from two accessions. Number of reads was mapped 27,534,208 and 51,460,028 in Daewon and Hwangkeum, respectively (Table 1). The total number of reads in Daewon was lower than the reads in Hwangkeum. After quality filtering and trimming, the 23,646,188 reads in Daewon (85.85%) were mapped to the reference genome (contained 56,044 genes and 177,294 transcripts), and 45,715,040 reads (88.84%) in Hwangkeum. The mapped reads uniquely covered 82.72% of the reference genome in Daewon and 86.23% reads were uniquely mapped in Hwangkeum.

Discovery of variants on transcriptome

Using the default parameters implemented in CLC genomics workbench, we identified a total 89,955 genetic variants between the reference genome and the RNA sequence of two soybean accessions. The number of variants was 34,411, comprising 29,777 SNPs, 1,311 MNVs, 1,754 insertions, 1,458 deletions and 111 replacements in Daewon; 55,544, comprising 47,418 SNPs,

Table 1. RNA sequencing data assembled to soybean reference genome.

Mapped reads	Daewon		Hwangkeum	
	Number of sequences	%	Number of sequences	%
Reads mapped in pairs	23,646,188	85.85	45,715,040	88.84
- unique fragments	11,391,332	82.72	22,186,184	86.23
- non-specifically	431,762	3.14	671,336	2.61
Reads mapped in broken pairs	2,495,225	9.06	3,288,716	6.39
Reads not mapped	1,401,795	5.09	2,456,272	4.77
Total	27,543,208	100	51,460,028	100

Table 2. Variants against reference genome.

Type	Daewon		Hwangkeum	
	Homozygous	Heterozygous	Homozygous	Heterozygous
SNV	19,946	9,831	31,543	15,875
MNV	593	718	950	1,343
Insertion	899	855	1,542	1,688
Deletion	1,080	378	1,705	691
Replacement	68	43	109	98
Total	22,586	11,825	35,849	19,695

Table 3. Variants against each chromosome of reference genome.

Chromosome	Daewon					Hwangkeum				
	SNV	MNV	Insertion	Deletion	Replacement	SNV	MNV	Insertion	Deletion	Replacement
Chr01	1,249	32	65	64	4	1,870	70	124	108	4
Chr02	1,552	58	103	94	11	2,581	148	201	174	18
Chr03	1,561	63	89	76	5	2,627	147	158	143	18
Chr04	1,039	28	60	55	2	1,742	82	139	98	5
Chr05	1,756	82	130	83	3	2,616	102	167	120	10
Chr06	1,559	63	93	78	8	2,897	154	172	144	18
Chr07	1,566	68	97	59	3	2,833	138	209	143	14
Chr08	1,872	87	121	110	3	2,865	156	217	153	10
Chr09	1,653	61	94	77	7	2,333	88	152	118	8
Chr10	1,427	63	85	61	4	2,379	110	166	108	8
Chr11	1,246	57	83	67	3	1,348	76	108	95	3
Chr12	1,265	68	98	49	7	2,018	94	170	94	9
Chr13	1,974	79	129	108	8	2,946	147	202	134	10
Chr14	1,195	85	58	61	3	1,957	111	126	76	8
Chr15	1,575	70	82	72	7	2,258	110	151	101	5
Chr16	1,405	70	71	62	8	3,019	143	163	139	18
Chr17	986	30	41	73	5	1,469	81	117	103	7
Chr18	2,062	98	107	80	7	3,488	160	188	130	14
Chr19	1,403	72	72	60	6	1,962	81	157	115	9
Chr20	1,245	64	67	61	7	2,052	88	139	96	11
scaffold_21	95	7	6	6	0	41	5	0	1	0
scaffold_27	1	0	0	0	0	1	0	0	0	0
scaffold_28	20	1	0	2	0	11	1	0	1	0
scaffold_32	16	0	0	0	0	34	1	2	1	0
scaffold_123	7	0	0	0	0	9	0	2	0	0
scaffold_137	2	1	0	0	0	0	0	0	0	0
scaffold_165	6	0	0	0	0	0	0	0	0	0
scaffold_208	18	2	1	0	0	0	0	0	0	0
scaffold_216	7	0	0	0	0	5	0	0	0	0

Table 3. Variants against each chromosome of reference genome (Continue).

Chromosome	Daewon					Hwangkeum				
	SNV	MNV	Insertion	Deletion	Replacement	SNV	MNV	Insertion	Deletion	Replacement
scaffold_263	7	0	0	0	0	0	0	0	0	0
scaffold_344	4	0	0	0	0	20	0	0	0	0
scaffold_614	2	0	0	0	0	0	0	0	0	0
scaffold_846	1	2	2	0	0	1	0	0	0	0
scaffold_2080	1	0	0	0	0	1	0	0	1	0
scaffold_22	0	0	0	0	0	15	0	0	0	0
scaffold_44	0	0	0	0	0	6	0	0	0	0
scaffold_361	0	0	0	0	0	1	0	0	0	0
scaffold_633	0	0	0	0	0	7	0	0	0	0
scaffold_843	0	0	0	0	0	1	0	0	0	0
scaffold_1241	0	0	0	0	0	3	0	0	0	0
scaffold_1742	0	0	0	0	0	2	0	0	0	0
Total	29,777	1,311	1,754	1,458	111	47,418	2,293	3,230	2,396	207

Table 4. Non-synonymous variants detected within coding regions.

Type	Daewon		Hwangkeum	
	Non-synonymous	Synonymous	Non-synonymous	Synonymous
SNV	7,908	11,517	11,522	16,030
MNV	449	204	665	285
Insertion	316	122	481	159
Deletion	378	0	511	0
Replacement	29	0	46	1
Total	9,080	11,843	13,225	16,475

2,293 MNVs, 3,230 insertions, 2,396 deletions and 207 replacements in Hwangkeum (Table 2).

The distribution of DNA polymorphisms was detected across the 20 soybean genomes and 21 scaffolds. The largest number of variants was observed in chromosome 18 of both accessions. However the smallest number of variation was observed on different chromosome, chromosome 17 of Daewon and chromosome 11 of Hwangkeum (Table 3).

The highest SNP and MNV frequencies were found on chromosome 18 in both accessions, also on the other hand the lowest frequencies of SNP and MNV were found on different chromosomes. Other variants had different distributions in whole chromosomes. A relative high frequency of InDels was observed on chromosome 8 in both accessions, in comparison to other small number of InDel frequencies in different chromosomes. The 390 variants

were found in location of the 21 scaffolds in the Williams 82 genome. The frequency of replacement showed to have a pretty even distribution among 20 chromosomes.

Annotation of variants

Some variants were detected in non-coding regions, introns and large repeat regions. These variants were excluded by amino acid change test with default parameter using CLC genomic workbench 7. After filtering out the variants, the remaining number of variants located in CDS regions was 13,225 non-synonymous variants, comprising 11,522 SNVs, 665 MNVs, 481 insertions, 511 deletions, and 46 replacements in Hwangkeum and 9,611 non-synonymous variants were detected in Daewon that comprised 7,908 SNVs, 665 MNVs (Table 4), 481 insertions, 511 deletions and 46 replacements. The 4,290 genes and 5,672

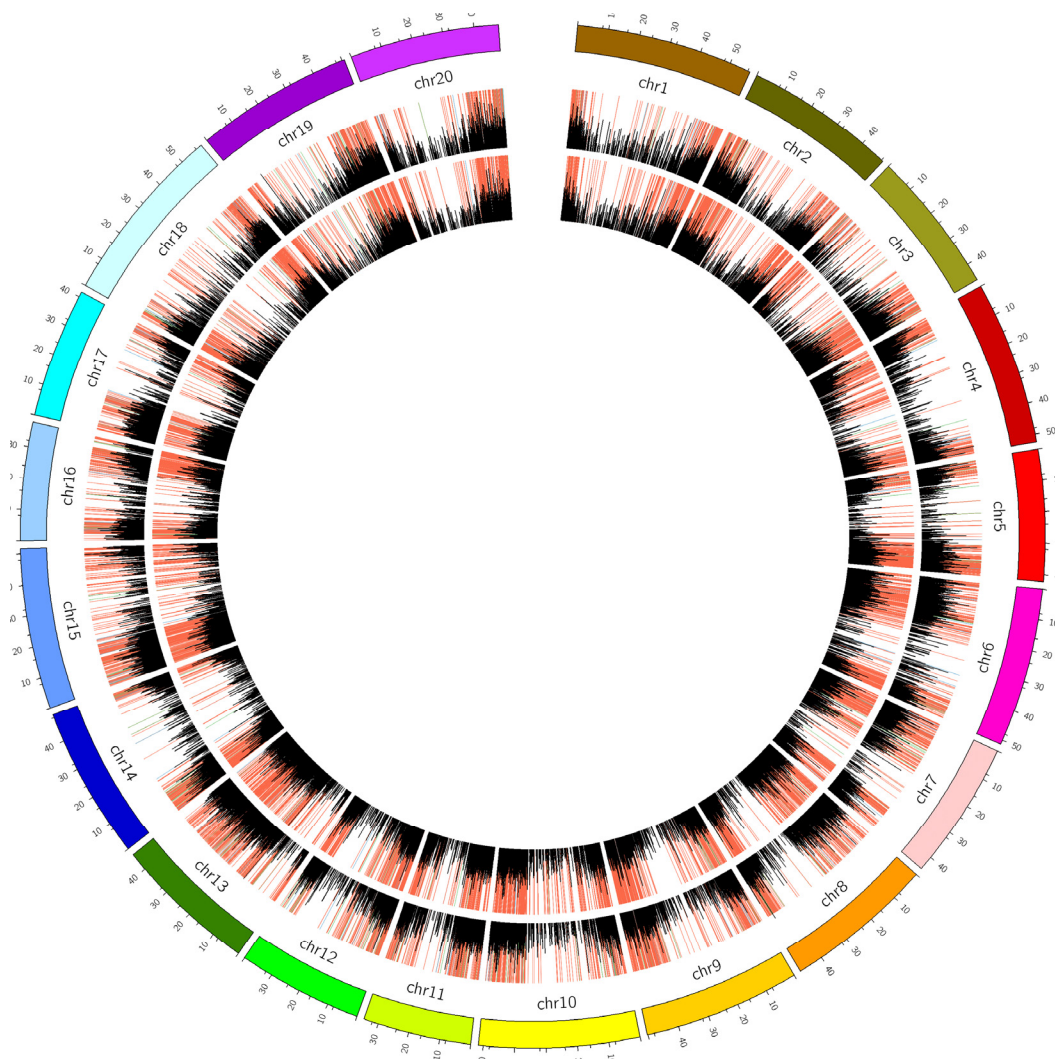


Fig. 1. A circos plot showing differentially expressed genes and non-synonymous variants from RNA-sequencing. The outermost circle represents 20 soybean chromosomes (Chr 1-20) in different colours, and inner two circles represent gene expression value from RPKM (black) and distribution of variants (SNV and MNV: red, InDel: blue, Replacement: green) in Daewon (I) and Hwangkeum (II)

genes were predicted to change protein functions through non-synonymous variants in Daewon and Hwangkeum, respectively. Additionally, we observed that non-synonymous variants were not uniformly distributed within the chromosomes. The distributions of non-synonymous variants within coding regions and expression values of each accession are shown across the 20 soybean chromosomes (Fig. 1).

DISCUSSION

This study provides non-synonymous variants within coding regions (CDS) of two Korean soybean accessions using RNA-seq.

We analyzed the transcripts and obtained a total 19,430 non-synonymous SNPs, 1,114 MNVs, 1,686 InDels and 75 replacements against Williams 82 genome as a reference. These non-synonymous variants and expression data will possibly provide information for markers development and functional genetics. The reads from each Korean soybean accession were mapped on to reference genome. The mapping ratio of reads was shown highly coverage in each accession, which varied from 85.82 % in Daewon to 88.84 % in Hwangkeum. This coverage is similar to the result of whole genome sequence in various soybean accession studies (Zou *et al.*, 2014; Chung *et al.*, 2014).

Short reads was used to map on reference genome in RNA-seq.

The short reads were possible to assemble large part of DNA sequence, especially the heterochromatic regions around centromeres and other highly repetitive regions. The plant genomes contain large repetitive sequences that more over 35% regions of genome are observed in rice and *Arabidopsis* (*Arabidopsis* Genome Initiative 2000; International Rice Genome sequencing project 2005). Soybean genome also has 57% heterochromatic and large repetitive regions in its genome (Schmutz *et al.*, 2010). To selectively choose the non-synonymous variants on CDS, these variants within repetitive regions were removed according to other genotyping studies in soybean (Wu *et al.*, 2010; Zou *et al.*, 2014). After the basic variants were detected with default parameter, we classified the variants into two types, homozygous and heterozygous. A number of variants was 34,411 comprising 22,586 homozygous types and 9,831 heterozygous types in Daewon. The 35,849 homozygous variants and 19,695 heterozygous variants were shown in Hwangkeum. Because soybeans have been reproduced by self-pollination, the detected variants were expected to be predominantly homozygous variants. But the homozygous variants were found only 65% in both accessions. According to Pinson and Rutger (1993), genetic doubling event, fusion of genotypically different cells in chimeric callus, and abnormal meiosis that resulted in heterozygous diploid microspores could occur.

We also found that variants present in 4,290 genes and 5,672 genes in Daewon and Hwangkeum, respectively are likely to have a major impact on gene function. Similar results were found in soybean genome according to resequencing of wild and cultivated soybeans (4,600 genes, approximately) (Lam *et al.*, 2010; Chung *et al.*, 2014). The variants were classified as non-synonymous and synonymous by amino acid change from CLC genomics workbench 7. The distribution of non-synonymous variants was similar to genome-wide variants. The SNVs and MNV were observed with highest frequency in chromosome 18 of both accessions. But the other variants had different distributions in whole chromosomes. The ratio of non-synonymous to synonymous SNP was 1.45 and 1.39 in Daewon and Hwangkeum, respectively (Table 4). This ratio is similar to ratio shown in soybean (1.40) which is higher than those in *Arabidopsis*, maize, and sorghum and rice (Clark *et al.*, 2007; Hufford *et al.*, 2012; Lam *et al.*, 2010; McNally *et al.*, 2009; Nelson *et al.*, 2011).

Except partial regions of chromosome 4, 9 and 18, both accessions showed similar patterns of gene expression and non-synonymous polymorphisms against reference genome. In

comparison to Hwangkeum, Daewon had lower variants density regions in chromosome 4. Hwangkeum had lower density of variants in chromosome 9 and 18. Large regions of repetitive centromeric and pericentromeric regions in each chromosome had low expression genes and non-synonymously variants. Found in other studies, centromeric regions showed low expressions of genes, in contrast with telomere regions with highly gene expression (Lee *et al.*, 2013). The distribution of variants corresponds with non-centromeric regions, which is similar to the result from SNP genotyping research (Lee *et al.*, 2015) (Fig. 1).

In conclusion, this study reveals large number of polymorphism and non-synonymous variants via RNA-seq of two Korean soybean accessions. The variants included SNVs, MNVs, InDels and replacements were detected in genome and coding regions. The genome-wide variants and non-synonymous variants will serve as valuable resources for genetic and genomic study. In addition, expression profiles and non-synonymous variants within coding regions can be used for identification of agronomic traits with analysis of functional relevance. We believe that this information will be help to find genes associated with useful agronomic trait in future studies.

ACKNOWLEDGEMENTS

This work was carried out with the support of “Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ00924302)” Rural Development Administration, Republic of Korea.

REFERENCES

- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 408(6814) : 796-815.
- Chung, W.-H., N. Jeong, J. Kim, W. K. Lee, Y. -G. Lee, S.-H. Lee, W. Yoon, J.-H. Kim, I.-Y. Choi, H.-K. Choi, J.-K. Moon, N. Kim, and S.-C. Jeong. 2014. Population Structure and Domestication Revealed by High-Depth Resequencing of Korean Cultivated and Wild Soybean Genomes. *DNA Res.* 21(2) : 153-167.
- Clark, R. M., G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller, P. Shinn, N. Warthmann, T. T. Hu, G. Fu, D. A. Hinds, H. Chen, K. A. Frazer, D. H. Huson, B. Schölkopf, M. Nordborg, G. Rättsch, J. R. Ecker, and D. Weigel. 2007. Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*. *Science*. 317(5836) : 338-342.

- Haun, W. J., D. L. Hyten, W. W. Xu, D. J. Gerhardt, T. J. Albert, T. Richmond, J. A. Jeddelloh, G. Jia, N. M. Springer, C. P. Vance, and R. M. Stupar. 2011. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol.* 155(2) : 645-655.
- Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, R. A. Cartwright, R. J. Elshire, J. C. Glaubitz, K. E. Guill, S. M. Kaeppeler, J. Lai, P. L. Morrell, L. M. Shannon, C. Song, N. M. Springer, R. A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M. D. McMullen, D. Ware, E. S. Buckler, S. Yang, and J. Ross-Ibarra. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44(7) : 808-811.
- Hyten, D. L., Q. Song, Y. Zhu, I.-Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker, and P. B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. U.S.A.* 103(45) : 16666-16671.
- International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature.* 436(11) : 793-800.
- Kim, M. Y., S. Lee, K. Van, T.-H. Kim, S.-C. Jeong, I.-Y. Choi, D.-S. Kim, Y.-S. Lee, D. Park, J. Ma, W.-Y. Kim, B.-C. Kim, S. Park, K.-A. Lee, D. H. Kim, K. H. Kim, J. H. Shin, Y. E. Jang, K. D. Kim, W. X. Liu, T. Chaisan, Y. J. Kang, Y.-H. Lee, K.-H. Kim, J.-K. Moon, J. Schmutz, S. A. Jackson, J. Bhak, and S.-H. Lee. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. U.S.A.* 107(51) : 22032-22037.
- Lam, H.-M., X. Xu, X. Liu, W. Chen, G. Yang, F.-L. Wong, M.-W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J. Wang, G. Shao, J. Wang, S. S.-M. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42(12) : 1053-1059.
- Lee, W. K., N. Kim, J. Kim, J.-K. Moon, N. Jeong, I.-Y. Choi, S. C. Kim, W.-H. Chung, H. S. Kim, S.-H. Lee, and S.-C. Jeong. 2013. Dynamic genetic features of chromosomes revealed by comparison of soybean genetic and sequence-based physical maps. *Theor. Appl. Genet.* 126(4) : 1103-1119.
- Lee, Y.-G., N. Jeong, J. H. Kim, K. Lee, K. H. Kim, A. Pirani, B.-K. Ha, S.-T. Kang, B.-S. Park, J.-K. Moon, N. Kim, and S.-C. Jeong. 2015. Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* 81(4) : 625-636.
- Li, Y., S. Zhao, J. Ma, D. Li, L. Yan, J. Li, X. Qi, X. Guo, L. Zhang, W. He, R. Chang, Q. Liang, Y. Guo, C. Ye, X. Wang, Y. Tao, R. Guan, J. Wang, Y. Liu, L. Jin, X. Zhang, Z. Liu, L. Zhang, J. Chen, K. Wang, R. Nielsen, R. Li, P. Chen, W. Li, J. C. Reif, M. Purugganan, J. Wang, M. Zhang, J. Wang, and L. Qiu. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics.* 14 : 579.
- McHale, L. K., W. J. Haun, W. W. Xu, P. B. Bhaskar, J. E. Anderson, D. L. Hyten, D. J. Gerhardt, J. A. Jeddelloh, and R. M. Stupar. 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol.* 159(4) : 1295-1308.
- McNally, K. L., K. L. Childs, R. Bohnert, R. M. Davidson, K. Zhao, V. J. Ulat, G. Zeller, R. M. Clark, D. R. Hoen, T. E. Bureau, R. Stokowski, D. G. Ballinger, K. A. Frazer, D. R. Cox, B. Padhukasahasram, C. D. Bustamante, D. Weigel, D. J. Mackill, R. M. Bruskiwich, G. Rättsch, C. R. Buell, H. Leung, and J. E. Leach. 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* 106(30) : 12273-12278.
- Nelson, J. C., S. Wang, Y. Wu, X. Li, G. Antony, F. F. White, and J. Yu. 2011. Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC Genomics.* 12 : 352.
- Pinson, S. R. M. and J. N. Rutger. 1993. Heterozygous diploid plants regenerated from anther culture of F1 rice plants. *In Vitro Cell. Dev. Biol. Plant.* 29(4) : 174-179.
- Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker, and S. A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. *Nature.* 463(7278) : 178-183.
- Song, Q., D. L. Hyten, G. Jia, C. V. Quigley, E. W. Fickus, R. L. Nelson, and P. B. Cregan. 2013. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *Plos One.* 8(1) : e54985.
- Subbaiyan, G. K., D. L. E. Waters, S. K. Katiyar, A. R. Sadananda, S. Vaddadi, and R. J. Henry. 2012. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J.* 10(6) : 623-634.
- Van, K., E.-Y. Hwang, M. Y. Kim, H. J. Park, S.-H. Lee, and P. B. Cregan. 2005. Discovery of SNPs in Soybean Genotypes Frequently Used as the Parents of Mapping Populations in the United States and Korea. *J. Hered.* 96(5) : 529-535.
- Wu, X., C. Ren, T. Joshi, T. Vuong, D. Xu, and H. T. Nguyen. 2010. SNP discovery by high-throughput sequencing in soybean. *BMC Genomics.* 11 : 469.
- Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, N. D. Young, and P. B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. *Genetics.* 163(3) : 1123-1134.
- Zou, X., C. Shi, R. S. Austin, D. Merico, S. Munholland, F. Marsolais, A. Navabi, W. L. Crosby, K. P. Pauls, K. Yu, and Y. Cui. 2014. Genome-wide single nucleotide polymorphism and Insertion-Deletion discovery through next-generation sequencing of reduced representation libraries in common bean. *Mol. Breeding.* 33(4) : 769-778.