G&I
Genomics & Informatics

Original article

eISSN 2234-0742
Genomics Inform 2021;19(2):e19
https://doi.org/10.5808/gi.21024

Received: April 8, 2021
Accepted: April 21, 2021

*Corresponding author:
E-mail: choii@kangwon.ac.kr

Analysis of genome variants in dwarf soybean lines obtained in F6 derived from cross of normal parents (cultivated and wild soybean)

Neha Samir Roy[1], Yong-Wook Ban[2], Hana Yoo[1], Rahul Vasudeo Ramekar[1], Eun Ju Cheong[2], Nam-Il Park[3], Jong Kuk Na[4], Kyong-Cheul Park[1], Ik-Young Choi[1*]

[1]Department of Agriculture and Life Industry, Kangwon National University, Chuncheon 24341, Korea
[2]Department of Forest Environmental System, Kangwon National University, Chuncheon 24341, Korea
[3]Department of Plant Science, Gangneung-Wonju National University, Gangneung 25457, Korea
[4]Department of Controlled Agriculture, Kangwon National University, Chuncheon 24341, Korea

Plant height is an important component of plant architecture and significantly affects crop breeding practices and yield. We studied DNA variations derived from F5 recombinant inbred lines (RILs) with 96.8% homozygous genotypes. Here, we report DNA variations between the normal and dwarf members of four lines harvested from a single seed parent in an F6 RIL population derived from a cross between *Glycine max* var. Peking and *Glycine soja* IT182936. Whole genome sequencing was carried out, and the DNA variations in the whole genome were compared between the normal and dwarf samples. We found a large number of DNA variations in both the dwarf and semi-dwarf lines, with one single nucleotide polymorphism (SNP) per at least 3.68 kb in the dwarf lines and 1 SNP per 11.13 kb of the whole genome. This value is 2.18 times higher than the expected DNA variation in the F6 population. A total of 186 SNPs and 241 SNPs were discovered in the coding regions of the dwarf lines 1282 and 1303, respectively, and we discovered 33 homogeneous nonsynonymous SNPs that occurred at the same loci in each set of dwarf and normal soybean. Of them, five SNPs were in the same positions between lines 1282 and 1303. Our results provide important information for improving our understanding of the genetics of soybean plant height and crop breeding. These polymorphisms could be useful genetic resources for plant breeders, geneticists, and biologists for future molecular biology and breeding projects.

Keywords: dwarf, RIL population, SNP, soybean, whole genome sequencing, wild type

Introduction

Soybean is one of the most important leguminous crops worldwide due to its use in human food and oil production. Currently, the United States, Brazil, and Argentina account for more than 80% of the worldwide production of soybean [1]. In Southeast Asian countries, particularly Korea, China, and Japan, soybean is used in multiple life stages as a rich source of protein, and it is considered one of the five major grains [2]. Plant height is an important trait that has a direct impact on yield and lodging resistance. Extremely tall

plants can be affected by lodging, which may reduce yield and quality [3]. Dwarfism in crops has played a major role in the "Green Revolution," in which semi-dwarf varieties were chosen for further cultivation, first in wheat and then in rice [4]. Many studies of plant height inheritance have successfully cloned dwarf genes [5-7]. In soybean, some high-yielding and lodging-resistant dwarf varieties have been developed [8].

The rapid development of next-generation sequencing (NGS) technology and instruments has supported quick and efficient genomics research [9,10]. The key advantage of NGS is that it can produce a large amount of data at low cost, and it is currently being applied to a number of plants [11-14]. Single nucleotide polymorphisms (SNPs) are ubiquitous in genomes and have emerged as a marker of choice, especially in sequenced plants [15,16]. Plants adapt to different environments by various mechanisms, one of which is allelic variation. The identification of these variations is the first step in the in-depth study of the genes and alleles involved in plant evolution and environmental adaptation.

A previously reported study, where the wild soybean *Glycine soja* was sequenced and compared to the *Glycine max* reference, found 2.5 megabases (Mb) of substituted sequences, 4.6 kilobases (kb) of indels, 32.4 Mb of deletions and 8.3 Mb of new sequences in a total of 915.5 Mb of genome sequence [14]. Although a great deal of information is available from whole genome sequencing, resequencing strategies have become an important tool to study allelic variations. There have been studies in other plants, such as rice [17], maize [18], *Arabidopsis* [13], and sorghum [19], as well as resequencing studies in soybean, where both wild and commercial varieties have been analyzed [20-22]. In this study, we compared two inbred lines obtained from a cross between *G. max* var. Peking and *G. soja* IT182936 in the F6 generation. A few lines segregated for a dwarf phenotype and continued to show the same phenotype in the next generation. Resequencing analysis revealed many SNPs and indels in both genic and non-genic regions, which are explained in this study.

Methods

Plant materials

Recombinant inbred lines (RILs) were developed from a cross between *G. max* var. Peking and *G. soja* IT182936. The soybeans used in this experiment were harvested in Chuncheon city (Gangwon-do, South Korea). All the plants were grown in field condition. Two RILs exhibiting normal and dwarf phenotype from F6 generation were selected for whole genome variant analysis and designated 1282NF6 and 1303NF6 for normal plants and 1282DF6 and 1303DF6 for dwarf plants. Two more RILs, 1214 and 1290, exhibiting semi-dwarf phenotypes, were also selected for comparison with the dwarf lines. Three leaves were collected from each plant before flowering, frozen immediately in liquid nitrogen and stored at –80°C.

DNA isolation and Illumina sequencing

Genomic DNA was isolated from the leaf tissues using the modified CTAB method [23]. DNA purification was carried out using QIAquick Purification Kit (28104, Qiagen, Beijing, China). Adaptor ligation and DNA clustering preparations were done by Solexa sequencing using Illumina HiSeq 4000 sequencing platform according to the manufactures' protocol by the National Instrumentation Center for Environmental Management (NICEM) at Seoul National University. The sequencing libraries were prepared by random fragmentation of the DNA sample, followed by 5′ and 3′ adaptor ligation. Low-quality reads (ratio of reads that have phred quality score of < 20), reads with adaptor sequences, and duplicated reads were eliminated. The remaining high-quality data were used for mapping. We used the published genome sequence of *G. max* version 1 as a reference [24].

Identification of DNA variations in normal and dwarf lines

The raw Illumina sequencing data were filtered and compared to characterize the genotype of normal and dwarf samples using the Bowtie2 (v2.3.4.3) aligner [25]. The SNPs were qualified by GATK (version 2.3.9 Lite) [26] and biallelic filtering. GATK filtering was performed with the options MQ0 ≥ 4 && ((MQ0/(1.0*DP)) > 0.1), QUAL < 30, QD < 5.0, and FS > 200.0 [27]. The biological function of each SNP and indel locus were identified using SnpEff software [28]. Paired-end reads were mapped against the TAIR10 reference genome sequence [29]. The DNA variations common to the two sets of dwarf lines were discovered by comparing the SNPs discovered at the sample loci between the normal and dwarf samples.

Results

In silico mapping of resequencing reads to reference and variant calls

The parental lines (*G. max* var. Peking and *G. soja* IT182936) did not exhibit dwarfism, but few plants in F3 generation appeared dwarf (Fig. 1). Although F3 dwarf lines didn't produce any seeds, there were five dwarf lines in next generation (F3). Two out of five produced seeds and continued to produce dwarf phenotype in their next generation. Two of such samples, 1282 normal and dwarf (labeled 1282NF6 and 1282DF6) and 1303 normal and dwarf (labeled 1303NF6 and 1303DF6) from F6 generation were

https://doi.org/10.5808/gi.21024

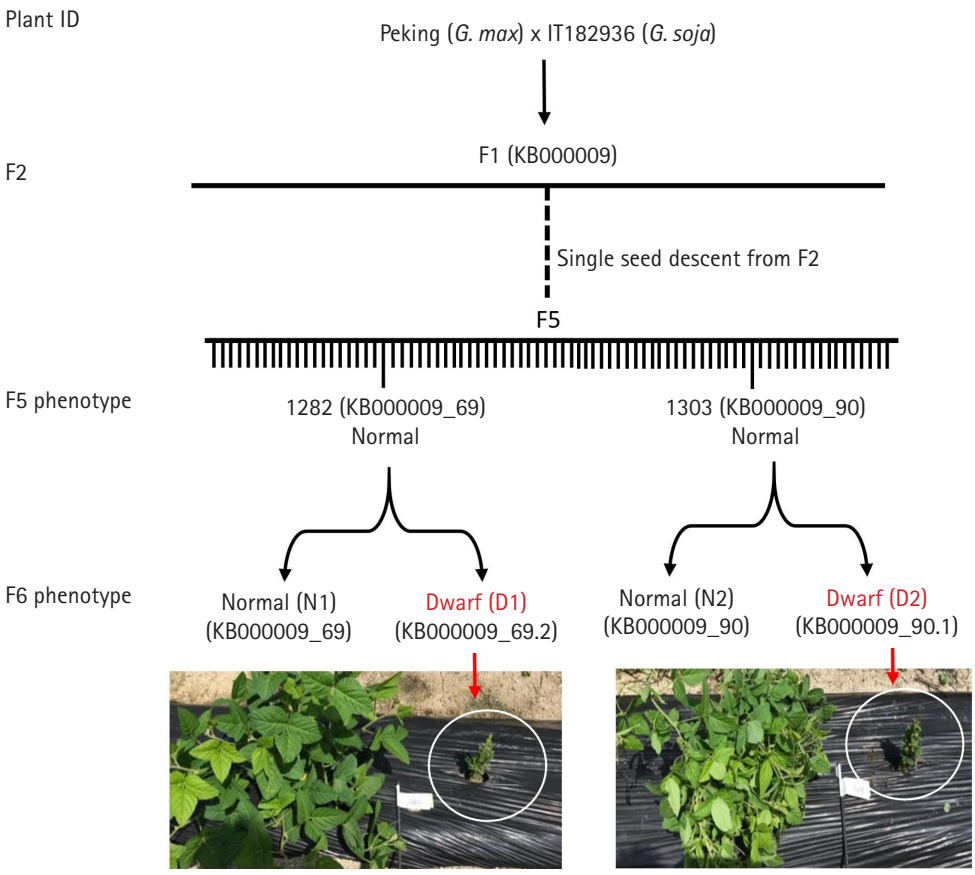

Fig. 1. Schematic representation of dwarf recombinant inbred lines (RIL) development. *Glycine max* var. Peking and *G. soja* var. IT182936 were crossed to develop RIL lines. From F2 few lines exhibited dwarf phenotype and continued to appear dwarf in their successive generation. Two such lines from the F6 generation were chosen for analysis.

Table 1. Variant calling statistics as compared to reference genome

Raw variants (SNP + INDEL)	SNP		INDEL	
	Raw variants	Filtered	Raw variants	Filtered
5,597,100	4,645,717	4,108,601	944,240	904,602

SNP, single nucleotide polymorphism.

chosen and used for genomics variation analysis using Illumina sequencing method. Additional two samples that exhibited semi-dwarf phenotype were also used to analysis viz. (labeled 1214NF6 and 1214DF6 and 1290NF6 and 1290DF6). The total number of reads obtained for 1282NF6 was 159,196,424, which accounted for 24 G bps. The GC content was 35.65% with a Q20 value of 94.82%. Likewise, the total reads for 1282DF6 were 172,215,288, accounting for 26 Gbps, and in line 1303, the numbers of total reads for the normal and dwarf samples were 150,712,004 and 139,484,320, accounting for 22 Gbps and 21 Gbps, respectively (Supplementary Table 1). The GC content for all the lines was above 35%, and the Q20 average percentage was above 94%. More than 93%, on average, was mapped to the reference genome. There were 5,597,100 variant calls in both genotypes (Table 1). The raw data was deposited in NCBI SRA database with an accession number PRJNA665611.

To estimate the number of variants that can be obtained in a hybrid of *G. max* and *G. soja* we need to know the total number of SNPs and indels in them. G. Ramakrishna et al. identified a total 77,339 SNPs and 451,522 indels in *G. max* whereas 215,932 SNPs and 697,295 indels in *G. soja*, with comparison to reference post-filtering [30]. Among them, the number of common variants for both species was 10,873 SNPs and 80,078 indels. So if we exclude the common SNPs and indels from both species, we would be left with 282,398 SNPs and 1,069,739 indels making the total count of variations into 1,351,137. As per the Mendelian genetics, the cross be-

tween *G. max* and *G. soja* will distribute the variants into half from each parent to its offspring F1 into 1:2:1 ratio of both parents, then subsequently to the next F2 generation and so on [31].

With respect to that, theoretically the distribution of SNP in resulting cross should be 1 SNP per 0.748 kb in F1 generation (genome size 1,013,200 kb/total variation 1,351,137) [31]. Therefore in F2 generation the distribution is 1 SNP per 1.49 kb and so on. Consequently in theory in F6 generation there should be 1 SNP per 23.99 kb. From the filtered SNPs, we obtained average 217,764 and 57,403 homozygous SNP in dwarf lines and semi-dwarf lines of F6 generation (Table 2). Implying there is 1 SNP per 3.68 kb region of dwarf lines and 11.13 kb region of semi-dwarf lines (Fig. 2).

Distribution of SNPs in coding regions of the reference genome

The genes that directly affect the growth of plants viz. plant defense, phytohormones, and photosynthesis were considered while comparing the SNPs in dwarf and normal lines. Furthermore we focused exclusively on missense SNPs on coding regions as they may cause base changes in protein sequence and alter the gene function. We observed 503 and 485 SNPs among dwarf and normal of 1,282 and 1,303, respectively, out of which 98 were common to both genotypes when comparison to reference genome. The highest number of SNPs was observed in NB-ARC (nucleotide-binding APAF-1 R proteins and CED-4) domain–containing disease resistance protein (75%), followed by disease resistance protein (TIR-NBS-LRR [toll interleukin 1 receptor nucleotide-binding site leucine-rich repeat resistance proteins] class) family (35%) and auxin-like 1 protein (27%) (Fig 3, Supplementary Table 2). The distribution was highest on chromosome (Chr) 18, followed by Chr 16, for both genotypes. When considering the individual nonsynonymous SNP distribution, we observed that Chr 16 had the maximum number of SNPs in both 1282 (74) and 1303 (67), followed by Chr 7, with 58 and 63 in 1282 and 1303, respectively (Supplementary Table 3). We discovered a total of 33

Table 2. Number of candidate variants after filtering with reference genome

Sample	Different between normal and dwarf (SNP)	Homozygous in dwarfism sample (SNP)	Different between normal and dwarf (INDEL)	Homozygous in dwarfism sample (INDEL)
1282	458,209	182,497	108,277	51,622
1303	337,001	253,032	100,267	63,184
1214	116,136	65,599	55,767	30,225
1290	111,052	57,098	56,000	29,152

SNP, single nucleotide polymorphism.

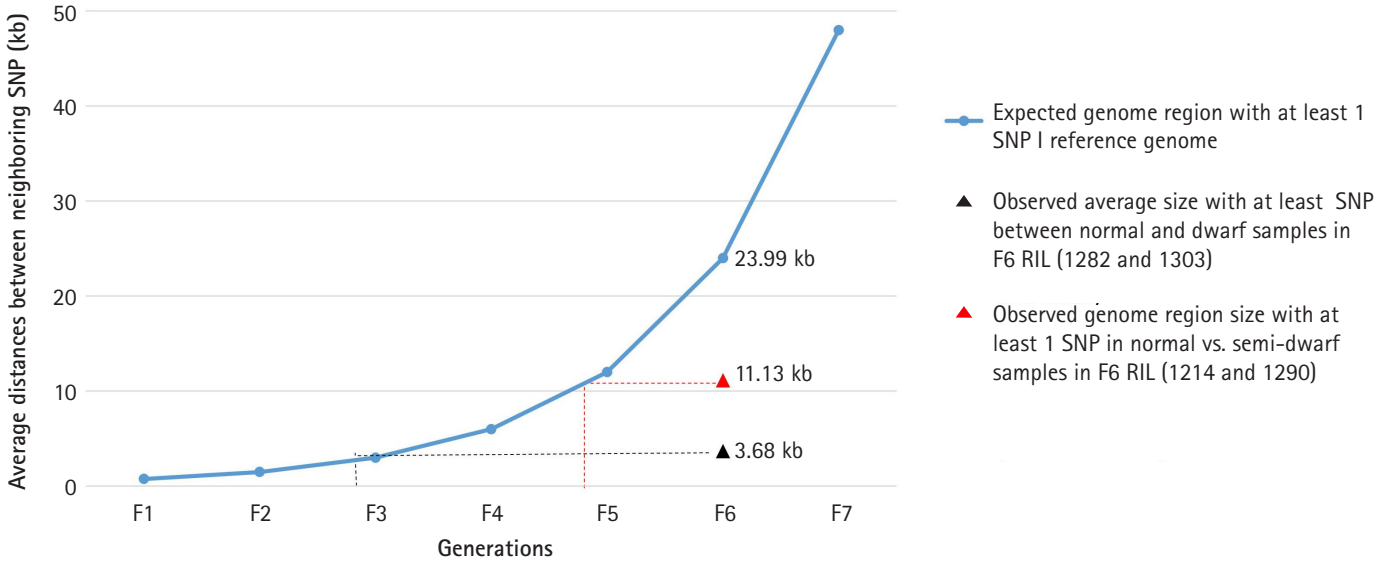

Fig. 2. Single nucleotide polymorphism (SNP) per kb of gene length in dwarf soybean lines. The frequency of SNP in dwarf lines was observed significantly higher than expected. Dwarf lines exhibited SNP on every 3.68 kb of genome whereas the expected length of the genome should be 23.99 kb. RIL, recombinant inbred lines.

https://doi.org/10.5808/gi.21024

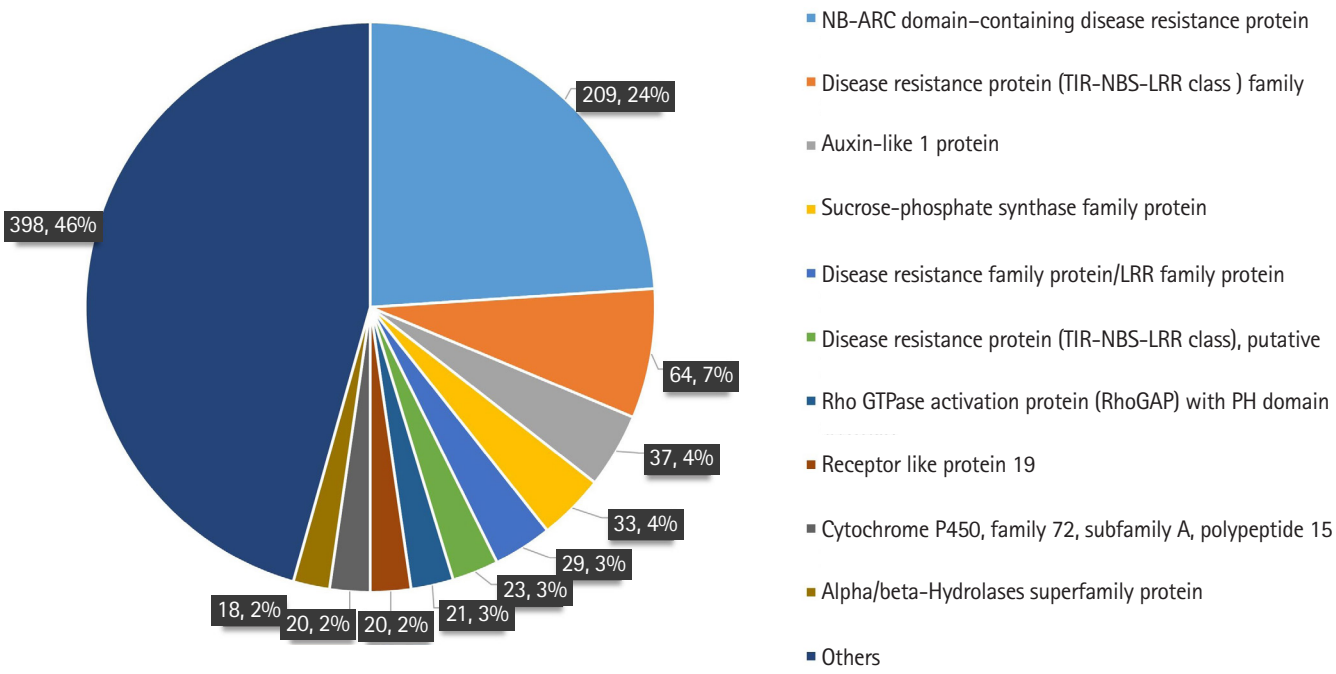

Fig. 3. Distribution of missense single nucleotide polymorphisms (SNPs) in the soybean genome in genic regions. NB-ARC, nucleotide-binding APAF-1 R proteins and CED-4; TIR-NBS-LRR, toll interleukin 1 receptor nucleotide-binding site leucine-rich repeat resistance proteins.

homogeneous nonsynonymous SNPs that occurred at the same loci in each set of dwarf and normal soybean derived from normal soybean. We then identified the homogeneous SNPs across all the dwarf samples, with the highest representation of SNPs on Chr 16 and Chr 7 in both genotypes. SNPs that were common to both genotypes were highest on Chr 4, followed by Chr 7 and Chr 15 (Supplementary Table 3).

There were three nonsense SNPs found in both genotypes: one was on Chr 2, another was on Chr 13, and the last was on Chr 20; the SNP from Chr 20 was homogeneous in the dwarf lines (Supplementary Table 4). The three nonsense SNPs obtained in the genic regions included proteins with the gene functions single-stranded DNA (ssDNA)-binding transcriptional regulator, UDP-glycosyltransferase protein (UGT) and PIF1 helicase. Twenty frameshift SNPs were common to both genotypes (Supplementary Table 4). The frameshift mutations observed in the genic regions of the normal and dwarf individuals are shown in Supplementary Table 3. Out of 20 mutated genes, five were leucine-rich receptor-like protein family genes; three were proteins of unknown function (DUF647); two were NADH-ubiquinone/plastoquinone oxidoreductase chain 4L, 2 cytochrome P450, family 78, subfamily A, polypeptide 5; and one each was MUTS homolog 6, unknown protein and RNA helicase-like 8.

Table 3. Chromosome wise distribution of SNP among normal and dwarf lines

Chromosome No.	1282 Normal and dwar	1303 Normal and dwarf	1282 + 1303 Common SNPs
1	29	18	2
2	7	54	0
6	3	0	0
7	36	71	0
8	23	17	0
9	11	11	0
10	17	8	0
11	4	6	0
12	1	2	0
13	27	28	0
14	3	1	0
15	22	0	0
19	13	11	2
20	3	1	1
Total	186	241	5

SNP, single nucleotide polymorphism.

Distribution of SNPs among normal and dwarf plants

When we considered the SNP among normal and dwarf lines and not with the reference, 426 SNPs were obtained, among which five were common to both samples (1282 and 1303) (Table 3). The

SNPs were distributed on all chromosomes except 3, 4, 5, 16, and 17. Sample 1282 had the highest number of SNPs on Chr 7 (36), followed by Chr 1 (29) and Chr 13 (27). Likewise, sample 1303 had the highest number of SNPs on Chr 7 (71), followed by Chr 2 (54) and Chr 13 (28). The gene functions of the SNPs that were common to both normal and dwarf samples in both lines (1282 and 1303) included matrixin family protein, zinc finger protein, transcription factor and cytochrome P450 protein polypeptide (Table 4).

Discussion

RIL lines were developed from *G. max* and *G. soja* and none of the parent exhibited dwarf phenotype. From F3 generation we obtained few lines that exhibited dwarf phenotype which continued to produce dwarf lines in subsequent generation. We chose two lines from F6 that had both phenotypes (dwarf and normal) to survey the variance among them. Genome variant analysis was performed in such lines from F6 generation exhibiting dwarf and semi-dwarf phenotype. We observed that the dwarf lines in this study had a higher number of SNPs and indels than the semi-dwarf lines (Table 2). This implies that higher number of variations could cause changes in gene function causing dwarf phenotype. We obtained 5M of total variation in our data (Table 1). The number although is much smaller than that obtained in a recent study in *G. soja* [32] which was more than 15M SNPs and 14M indels. The reason could be due to the fact that they have used more than 26 accessions to determine variants whereas we have used only one Wm82 genome to look for the differences. Initially, we compared the number of SNPs with respect to reference *G. max* genome where we observed a smaller number of SNPs in semi-dwarf lines and a higher number in dwarf lines. The homozygosity in F5 and F6 RILs are 94% and 97%, respectively [31]. This can explain the higher numbers of variants which may also impact the phenotype. Although the number of SNPs observed in the F6 populations in our study was greater between the dwarf and normal plants (1 SNP per 3.68 kbp for dwarf and 11.13 kb for semi-dwarf plants) (Fig. 2). This is much higher than the normal SNP frequency in the F6 generation, which is 1 per 23.99 kbp [33].

Natural variations in the genome, such as SNPs in coding regions, can alter amino acid sequences and modify the post-translation products, which may affect gene function [34,35]. Notably, disease resistance protein genes exhibited a high number of variants in our study. Similar studies have shown that changes in protein function (gain or loss of function) may contribute to dwarf phenotypes. Dwarfism due to a gain-of-function mutation in a TIR-NB-LRR protein was reported in *Arabidopsis* as one of the mechanisms underlying enhanced disease resistance [36]. Moreover, changes in these proteins cause autoimmunity in plants, and one of the useful features of autoimmune mutants is their dwarf phenotype [37]. Temperature and humidity are known to play important roles in dwarfism, although the exact mechanism is still unknown [38-42]. The RILs in our study were all grown in the same environmental conditions with the same temperature and humidity. This suggests that the dwarf phenotypes observed in our study were not due to temperature or humidity.

The SNPs were mostly in the NB-ARC domain-containing disease resistance protein, followed by disease resistance protein and auxin-like protein (Fig. 4). In *Arabidopsis*, NB-ARC mutants induced autoimmunity in plants, of which dwarfism is one of the typical phenotypes [37]. Common SNPs occurring in gene coding regions could affect the phenotype, whether or not they are combined with other genes. There is not much known about the matrixin family protein in plants, but the members of the matrix metalloproteinase family are thought to be involved in remodeling of the extracellular matrix during plant growth and development [43]. Thus, an SNP in a gene encoding a member of this protein family might have caused an alteration in protein function leading to dwarfing. Mutations in the transcription factor jumonji (jmjC) have been reported to complement plant growth defects and expression changes [44]. Overexpression of TTF-type zinc finger protein in *Arabidopsis* resulted in divergent physiological and metabolic phenotypes, some of which were significant for improved plant performance [45]. The SNPs occurring in these vital genes

Table 4. Common SNP among normal and dwarf lines

Gene	Chromosome No.	1282NF6	1282DF6	1303NF6	1303DF6	TAIR TOP hit function
GLYMA01G04370	1	G/G	A/G	G/G	A/G	Matrixin family protein
GLYMA01G06671	1	C/C	T/T	T/T	C/C	TTF-type zinc finger protein with HAT dimerization domain
GLYMA19G14700	19	A/G	A/A	A/G	A/A	Transcription factor jumonji (jmjC) domain–containing protein
GLYMA19G14700	19	G/G	G/A	G/A	A/A	Transcription factor jumonji (jmjC) domain–containing protein
GLYMA1057S00200	20	T/C	T/T	T/T	T/C	Cytochrome P450, family 76, subfamily C, polypeptide 4

SNP, single nucleotide polymorphism.

https://doi.org/10.5808/gi.21024

may have contributed to the dwarf phenotypes found in the RILs.

There were nonsense SNPs in three loci: ssDNA-binding transcriptional regulator, UGT superfamily protein, and PIF1 helicase. ssDNA-binding transcriptional regulators are known to function as positive and negative regulators in leaf senescence [46]. UGTs also act as major contributors to plant growth, including development, disease resistance, and interaction with the environment, by interacting with various substrates, such as flavonoids, terpenes, auxins, cytokinin, and many others [47]. PIF1 helicases are enzymes that are essential in DNA replication, repair, and recombination in all organisms. Likewise, frameshift mutations were found in genes with important functions, such as leucine-rich receptor-like protein family, protein of unknown function (DUF647), NADH-ubiquinone/plastoquinone oxidoreductase chain 4L, cytochrome P450 family 78-subfamily A polypeptide 5, MUTS homolog 6, unknown protein and RNA helicase-like 8. All these proteins are major regulators and contributors to plant growth and development.

The SNPs obtained in our study (missense, nonsense SNP, and frameshift mutations) were in vital genes but may or may not have impacted plant growth. We were unable to derive any particular conclusion about the dwarf phenotype based on our results, but this study will provide a basis to analyze further and evaluate which SNPs affect plant growth type. The identification of functional SNPs in genes and analysis of their effects on phenotype may lead to a better understanding of their impacts on gene function and thus support varietal improvement.

ORCID

Neha Samir Roy: https://orcid.org/0000-0002-4529-4861
Yong-Wook Ban: https://orcid.org/0000-0002-1912-0374
Hana Yoo: https://orcid.org/0000-0002-3451-4646
Rahul Vasudeo Ramekar: https://orcid.org/0000-0003-3461-3617
Eun Ju Cheong: https://orcid.org/0000-0002-2576-5435
Nam-Il-Park: https://orcid.org/0000-0001-6725-0758
Jong Kuk Na: https://orcid.org/0000-0003-2616-4890
Kyong-Cheul Park: https://orcid.org/0000-0002-3737-815X
Ik-Young Choi: https://orcid.org/0000-0003-4168-0471

Authors' Contribution

Conceptualization: KCP, NIP, IYC. Data curation: RVR, NSR, JKN. Funding acquisition: IYC. Methodology: YWB, HY, EJC, KCP. Writing - original draft: YWB, NSR, IYC. Writing - review & editing: NSR, IYC.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2017R1A2B4011198)

Supplementary Materials

Supplementary data can be found with this article online at http://www.genominfo.org.

References

1. Fried HG, Narayanan S, Fallen B. Evaluation of soybean [Glycine max (L.) Merr.] genotypes for yield, water use efficiency, and root traits. PLoS One 2019;14:e0212700.
2. Shin DH. Utilization of soybean as food stuffs in Korea. In: El-Shemy H, ed. Soybean and Nutrition. London: IntechOpen, 2011. pp. 81-110.
3. Zhang Y, Yu C, Lin J, Liu J, Liu B, Wang J, et al. OsMPH1 regulates plant height and improves grain yield in rice. PLoS One 2017;12:e0180825.
4. Milach SC, Federizzi LC. Dwarfing genes in plant improvement. Adv Agron 2001;73:35-63.
5. Kuraparthy V, Sood S, Gill BS. Genomic targeting and mapping of tiller inhibition gene (tin3) of wheat using ESTs and synteny with rice. Funct Integr Genomics 2008;8:33-42.
6. Ku L, Wei X, Zhang S, Zhang J, Guo S, Chen Y. Cloning and characterization of a putative TAC1 ortholog associated with leaf angle in maize (*Zea mays* L.). PLoS One 2011;6:e20621.
7. Dardick C, Callahan A, Horn R, Ruiz KB, Zhebentyayeva T, Hollender C, et al. PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. Plant J 2013;75:618-630.
8. Cooper RL, Martin RJ, Walker AK, Schmitthenner AF. Registration of 'Hobbit' soybean. Crop Sci 1991;31:231.
9. Egan AN, Schlueter J, Spooner DM. Applications of next-generation sequencing in plant biology. Am J Bot 2012;99:175-185.
10. Brautigam A, Gowik U. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol (Stuttg) 2010;12:831-841.

11. Qi J, Liu X, Shen D, Miao H, Xie B, Li X, et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet 2013;45:1510-1515.
12. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 2013;45:43-50.
13. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 2008;18:2024-2033.
14. Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci U S A 2010;107:22032-22037.
15. Kumar S, Banks TW, Cloutier S. SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012;2012:831460.
16. Tang W, Wu T, Ye J, Sun J, Jiang Y, Yu J, et al. SNP-based analysis of genetic diversity reveals important alleles associated with seed size in rice. BMC Plant Biol 2016;16:93.
17. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 2011;30:105-111.
18. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. SNP discovery via 454 transcriptome sequencing. Plant J 2007;51:910-918.
19. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun 2013;4:2320.
20. Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 2010;42:1053-1059.
21. Chung WH, Jeong N, Kim J, Lee WK, Lee YG, Lee SH, et al. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Res 2014;21:153-167.
22. Maldonado dos Santos JV, Valliyodan B, Joshi T, Khan SM, Liu Y, Wang J, et al. Evaluation of genetic variation among Brazilian soybean cultivars through genome resequencing. BMC Genomics 2016;17:110.
23. Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor). Genome Biol 2011;12:R114.
24. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature 2010;463:178-183.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-359.
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-1303.
27. De Summa S, Malerba G, Pinto R, Mori A, Mijatovic V, Tommasi S. GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. BMC Bioinformatics 2017;18:119.
28. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80-92.
29. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 2012;40:D1202-D1210.
30. Ramakrishna G, Kaur P, Nigam D, Chaduvula PK, Yadav S, Talukdar A, et al. Genome-wide identification and characterization of InDels and SNPs in *Glycine max* and *Glycine soja* for contrasting seed permeability traits. BMC Plant Biol 2018;18:141.
31. Singh BD. Principles of Genetics. New Delhi: Kalyani Publishers, 1992.
32. Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, et al. Pan-genome of wild and cultivated soybeans. Cell 2020;182:162-176.
33. Xie M, Chung CY, Li MW, Wong FL, Wang X, Liu A, et al. A reference-grade wild soybean genome. Nat Commun 2019;10:1216.
34. Hawkins C, Caruana J, Schiksnis E, Liu Z. Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. Sci Rep 2016;6:29017.
35. Robert F, Pelletier J. Exploring the impact of single-nucleotide polymorphisms on translation. Front Genet 2018;9:507.
36. Shirano Y, Kachroo P, Shah J, Klessig DF. A gain-of-function mutation in an Arabidopsis Toll Interleukin1 receptor-nucleotide binding site-leucine-rich repeat type R gene triggers defense responses and results in enhanced disease resistance. Plant Cell 2002;14:3149-3162.
37. van Wersch R, Li X, Zhang Y. Mighty dwarfs: Arabidopsis autoimmune mutants and their usages in genetic dissection of plant immunity. Front Plant Sci 2016;7:1717.
38. Gao M, Wang X, Wang D, Xu F, Ding X, Zhang Z, et al. Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis. Cell Host Microbe 2009;6:34-44.

https://doi.org/10.5808/gi.21024

39. Bi D, Cheng YT, Li X, Zhang Y. Activation of plant immune responses by a gain-of-function mutation in an atypical receptor-like kinase. Plant Physiol 2010;153:1771-1779.
40. Noutoshi Y, Ito T, Seki M, Nakashita H, Yoshida S, Marco Y, et al. A single amino acid insertion in the WRKY domain of the Arabidopsis TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. Plant J 2005;43:873-888.
41. Yoshioka K, Kachroo P, Tsui F, Sharma SB, Shah J, Klessig DF. Environmentally sensitive, SA-dependent defense responses in the cpr22 mutant of Arabidopsis. Plant J 2001;26:447-459.
42. Zhou F, Menke FL, Yoshioka K, Moder W, Shirano Y, Klessig DF. High humidity suppresses ssi4-mediated cell death and disease resistance upstream of MAP kinase activation, H2O2 production and defense gene expression. Plant J 2004;39:920-932.
43. Marino G, Funk C. Matrix metalloproteinases in plants: a brief overview. Physiol Plant 2012;145:196-202.
44. Audonnet L, Shen Y, Zhou DX. JMJ24 antagonizes histone H3K9 demethylase IBM1/JMJ25 function and interacts with RNAi pathways for gene silencing. Gene Expr Patterns 2017;25-26:1-7.
45. Luttgeharm KD, Chen M, Mehra A, Cahoon RE, Markham JE, Cahoon EB. Overexpression of Arabidopsis ceramide synthases differentially affects growth, sphingolipid metabolism, programmed cell death, and mycotoxin resistance. Plant Physiol 2015;169:1108-1117.
46. Miao Y, Jiang J, Ren Y, Zhao Z. The single-stranded DNA-binding protein WHIRLY1 represses WRKY53 expression and delays leaf senescence in a developmental stage-dependent manner in Arabidopsis. Plant Physiol 2013;163:746-756.
47. Yonekura-Sakakibara K, Hanada K. An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J 2011;66:182-193.