FIRST ATTEMPT TO IDENTIFY AND MAP QTLs ASSOCIATED WITH PROMISCUOUS NODULATION IN SOYBEAN

Eric E. Agoyi^{1,2}, John B. Tumuhairwe², Godfree Chigeza³, Phinehas Tukamuhabwa², and Brian W. Diers¹

*Department of Crop Science, University of Illinois, 1101 W Peabody Drive, Urbana, IL, 61801, USA*¹

*Department of Agricultural Production, Makerere University, P.O. Box 7062, Kampala, Uganda*²

*International Institute of Tropical Agriculture, IITA-Zambia, Plot 32, Poplar Avenue, Avondale, Lusaka, Zambia*³

ABSTRACT

To inform possibility of conducting marker assisted breeding of promiscuous soybean varieties, this study used 92 F2 lines from biparental cross to identify QTLs associated with promiscuous nodulation in soybean. GBS; genotyping by sequencing platform was used to generate SNPs through the pipeline 2 in TASSEL 5.0, Bowtie2 version 2.2.8 for tag alignment, Beagle version 4.1 to impute missing SNPs and R-QTL package in R for QTL identification. Four nodulation traits were assessed *viz* number of nodules (NN), percent of effective nodules (NE), fresh weight and dry weight of nodules (NFW and NDW). Two QTLs were identified on chromosomes 10 and 13. Both QTLs were associated with NN, only QTL13 was associated NE and only QTL10 was associated with nodule 'weights. It was observed that NN, NFW and NDW shared QTL10 and NN and NE share QTL13 allowing hypothesize on the existence of pleiotropic genes in those two regions. Over dominance effect was observed for QTL10 and non additive effect for QTL13. The paper recommend investigations be pursued to validate those QTLs and set foundation for marker assisted selection of promiscuous soybean varieties. Also these findings could serve as starting point for gene cloning to better understand nodulation trait in soybean.

Keywords: QTL, nodules, pleiotropic, cloning, marker assisted

INTRODUCTION

Soybean (*Glycine max* L. Merrill) is a grain legume with high protein content making it an alternative protein source in diet formulation, globally and particularly in Africa where there is low affordability of animal proteins. The crop requires high doses of nitrogen to process these proteins mostly at pod filling stage (Machido *et al.*, 2011). It has been reported that soybean crops require up to 80 kg of nitrogen per hectare to yield a ton of grain, with approximatively 58% of this nitrogen coming from N₂ fixation (Salvagotti *et al.*, 2008), which is estimated up to 337 kg of nitrogen per hectare when inoculated with effective strains of *Bradyrhizobium*. Incidentally, soybean is a highly specific legume making symbiotic relationship mainly with *Bradyrhizobia japonicum*. This specific symbiotic association reduces on the nitrogen fixation potential of most soybean

cultivars, especially in areas where *Bradyrhizobia japonicum* is not endemic. Breeding for promiscuity in soybean is prioritized to increase biological nitrogen fixation and improve production (IITA, 1996).

Like most economically important traits, promiscuous nodulation in soybean is likely to be quantitative, evidence of this has been shown by Agoyi *et al.* (2016b) who reported predominant additive gene actions in promiscuous nodulation traits in soybean. Quantitative traits are controlled by genes located in regions known as quantitative trait loci (QTLs) (Prasanna, 2007). In dealing with quantitative traits, molecular breeding requires the mapping of QTLs associated with the traits under consideration, to enable marker assisted breeding and individual gene cloning. QTL mapping is the process of constructing a linkage map and analyzing QTLs with the aim of identifying their number, where there are located, and their effect on traits of interest (Doerge, 2002; Jonah *et al.*, 2011). QTLs analysis aims at identifying molecular markers that are closely linked to the desired allele, it is useful for marker-assisted breeding, as it provides knowledge for detecting marker-trait association (Jiang, 2013). According to Malosetti *et al.* (2011) mapping QTLs helps in dissecting the complexity of quantitative traits by identifying loci with effect on the trait. It is therefore a major step towards cloning individual genes related to economically important traits. This has not been exploited on promiscuous nodulation in soybean.

Population types used for QTLs mapping include F₂s, Backcrosses, recombinant inbred lines (RILs), and near-isogenic lines (NILs). Parental lines used to generate any of those populations must be genetically distant to show polymorphism (Semagn *et al.*, 2006).

An appropriate and cheap high-throughput genotyping platform such as next generation sequencing (NGS) is key for efficient mapping of QTLs and genomic studies. In this study we used Genotyping by sequencing (GBS), an ideal platform that shows efficiency from single gene markers to whole genome profiling (Poland and Rife, 2012). Several authors have used GBS to serve various purposes; it has been used for genome sequencing and gene mapping, genomic diversity studies, genome wide association studies (GWAS), genomic selection (GS), or improve marker resolution in number of crops including soybean (Yu *et al.*, 2015; Rincker *et al.*, 2016), maize (Romy *et al.*, 2013), wheat (van Poecke *et al.*, 2013), barley (Fu and Peterson, 2011), rice, potato (Uitdewilligen *et al.*, 2013), switch grass (Lu *et al.*, 2013), yellow mustard (Fu *et al.*, 2014) and Brassica oleracea (Liu *et al.*, 2014). GBS uses a multiplexed system for constructing reduced representation libraries for high diversity and large genome species (Elshire *et al.*, 2011). GBS generates hundreds of millions of single nucleotide polymorphisms (SNPs) for use in genetic analyses and genotyping

(Beissinger *et al.*, 2013). It is well suited for dealing with reduced sample size, as barcoded DNA from the different samples are pooled. GBS requires fewer PCR and purification steps, barcoding is efficient and easy to operate, with no limitation due to lack of reference genome (Davey *et al.*, 2011). GBS allows SNPs to be discovered and genotyped at the same time (Poland and Rife, 2012; Narum *et al.*, 2013). GBS has enabled identifying a set of 205,614 SNPs in soybean, thus providing valuable tools for soybean breeding programs (Lam *et al.*, 2010).

The present study identified and mapped QTLs associated to promiscuous nodulation in soybean, to provide basis for marker-assisted selection and nodulation genes cloning.

MATERIAL AND METHODS

MAPPING POPULATION

The parents used consisted of two soybean genotypes Wondersoya and GC 2043 that are genetically distant in their response to *Bradyrhizobium sp.* strain USDA 3456. For instance, Wondersoya is highly responsive to *B. sp* USDA 3456, while GC 2043 showed low response to that *Bradyrhizobium* strain (Agoyi *et al.*, 2016a). *Bradyrhizobium* spp which have been reported as native and widespread in African soils (Abaidoo *et al.*, 2000) are known to form nodules with promiscuous soybean genotypes (IITA, 1996).

A reciprocal cross was made between Wondersoya and GC 2043, F1 seeds were grown together with their corresponding female parent in plastic pots to generate F2 seeds and at the same time to check the successfulness of the crosses. Morphological traits such as hypocotyl color, pubescence color on stem, leaf and pod, days to flowering, flower color, maturity stage, plant height, were used to eliminate selfed individuals. Later on, seeds from parental lines and F2 seeds were inoculated with fresh culture of *Bradyrhizobium sp.* strain USDA 3456. *Bradyrhizobium sp.* strain USDA 3456 was obtained from Biofix (Kenya), purified and incubated in Soil Science Biological Nitrogen Fixation (BNF) laboratory at Makerere University. It was grown to 7.91×10^9 cells g^{-1} and formulated into an inoculum carried in steam-sterilised peat soil. Two table spoonfuls of sugar were dissolved into 300 ml of clean lukewarm water, in a soda bottle, to be used as sticker. The inoculant was mixed with the sticker and directly applied on seeds to enhance plant-rhizobium association. Inoculated seeds were planted in plastic pots filled up with steam-sterilized top soil in greenhouse at Makerere University Agricultural Research Institute of Kabanyolo (MUARIK). Prior to steam sterilization, soil samples were analyzed for nutrients content. As it was established in earlier studies that P and K are essential for biological nitrogen fixation (Giller *et al.*, 1997), the soil was pre-mixed with 0.0356 and

0.036 g of TSP (Triple Super Phosphate) and Muriate of Potash (MOP), respectively per kg of soil, representing 20 kg of phosphorus and 40 kg of potassium per hectare. Ninety-two individual plants were successfully generated.

PHENOTYPING F2 POPULATION

Six weeks after emergence, each of the 92 individual plants were evaluated for nodulation. Each plant was carefully released from the soil aggregate, root system was carefully and abundantly washed, rinsed and wrapped in tissue paper to reduce wetness. All nodules were harvested per plant and counted for number of nodules (NN), and weighed to determine their fresh weight (NFW). Then, all nodules were split opened using bistoury to assess their effectiveness. The percentage of effective nodules (NE) per plant was calculated based on the presence of brownish or pinkish pigmentation inside nodules. Thereafter, nodules were oven dried at 65°C for four days (Gwata *et al.*, 2004), and weighed to determine the total nodule dry weigh (NDW) per plant.

DNA EXTRACTION

Young trifoliate leaves were collected from F2 plants, dried in Falcon tubes containing silica gel, and DNA was extracted using the cetyl trimethylammonium bromide (CTAB) method according to Kim *et al.* (2012) with slight modifications. DNA pellet was re-suspended in 400 µL of 0.1× TE buffer, then 4 µL of RNase-A was added to remove RNA. To ensure purity of DNA, further cleaning was done adding 40 µL Chloroform: Isoamyl alcohol (24:1) and mixing all thoroughly. Each Eppendorf tube was spun for 10 min. Then, fixed volume of 30 µL was picked from each sample and added equal volume of cold Isopropanol, then spun to pellet and washed in 70% ethanol, and air dried. Later, DNA was shipped to the National Soybean Research Laboratory (NSRL), University of Illinois, Urbana-Champaign, United States.

LIBRARY PREPARATION, GENOTYPING AND SEQUENCING

GBS libraries are pools of short DNA fragments, obtained by the means of restriction enzymes, and ligated with barcoded adapters.

The work described here was carried out at the National Soybean Research Laboratory (NSRL) University of Illinois, Urbana-Champaign, United States, where all DNA was quantified using the intercalating dye (PicoGreen®; Invitrogen, Carlsbad, CA) and diluted to 25 ng/µL. A 96-plex GBS library comprising 92 soybean DNA samples, duplicate of parental lines and two negative (no DNA) controls were prepared as described in Elshire *et al.* (2011). Five microliter (5µl) of each DNA sample was pipetted into wells of ready-

to-use plate with A1 barcoded adapters at the bottom of each well, and 20µl restriction-ligation master mix was added to each well. The master mix contained two restriction enzymes; a rare cutter *HindIII*/MseI, and a common cutter *HindIII*/BfaI, and a single restriction-ligation reaction was performed.

Samples were incubated and aliquot (50 µl) of each sample was pooled and cleaned up using 100ul AMPure beads. Pools were PCR amplified using 10uM Illumina primer pair (F+R) and 2X Phusion Master Mix (NEB #M0531), temperature cycling consisted of 98°C for 30s, followed by 18 cycles (98°C for 10s, 68°C for 30s, 72°C for 30s), and thereafter 72°C for 5min, and 4° C.

PCR products were run on agarose gel to verify smear of sheared DNA fragments, and a second cleanup was done using 100ul AMPure beads. DNA7500 chip was used for Agilent Bio Analyzer in order to determine average size (bp) and concentration (ng/µl) of fragments in the libraries.

Libraries were quantitated by qPCR (quantitative PCR) and sequenced on one lane for 101 cycles from one end of the fragments on a HiSeq2500, SBS sequencing kit version 4.

DATA PROCESSING: SNPs CALLING, ALIGNMENT AND SNPs PRODUCTION

Fastq.gz files were generated from the sequencing machine and de-multiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina). The sequencing result downloaded on the SFTP server (<ftp://bdiers@ftp.biotech.illinois.edu>) consisted in Fastq.gz files with over 267 millions of Reads (Reads are 100 nucleotides in length). Fastq .gz files were processed and SNPs were called through the GBS analysis pipeline 2 as implemented in TASSEL 5.0 with modification of default settings; <http://www.maizegenetics.net/tassel/docs/TasselPipelineGBS.pdf> (Glaubitz *et al.*, 2014).

Briefly: i) tags were generated from fastq.gz files with the GBSSeqToTagDBPlugin (parameters: -KmerLength 80, -minKmerL 20, -c 10); ii) tags were extracted from database for alignment to reference genome with the TagExportToFastqPlugin (parameters: -c 1); iii) tags were aligned to the reference genome; http://soybase.org/cmap/cgibin/cmap/viewer?ref_map_set_aid=GmComposite2003;ref_map_aid=GmComposite2003_D1a;comparative_maps=1%3Dmap_aid%3DGmConsensus40_D1a;data_source=sbt_cmap. Bowtie2 version 2.2.8 was used for alignment; iv) tags were stored to physical map position in the database using the SAMToGBSdbPlugin, default parameters were used for that. v) SNPs were identified from the aligned tags with the DiscoverySNPCallerPluginV2 (parameters: mnLCov 0.05, mnMAF 0.01), vi) all discovered SNPs were scored for various parameters, including coverage, depth and genotypic statistics,

using the SNPQualityProfilerPlugin, vii) based on the estimated parameters, quality scores were calculated using an R script, the inbreeding coefficient (minF), and the minimum minor allele frequency (minMAF) were both set to 0.05, viii) database was updated with quality scores using the UpdateSNPPositionQualityPlugin, ix) data from fastq and key files were converted and a VCF genotype file was generated (parameters: -kmerLength 80, and minPosQS 3). This generated VCF file contained the SNPs which need to be cleaned through further processing steps, as there are always a high number of missing SNPs, and some SNPs might have been affected by segregation distortion. Beagle version 4.1 (Browning and Browning, 2016) was used to impute missing SNPs, and ABH genotype file was generated and exported using the graphical interface of Tassel 5.0. (Bradbury *et al.*, 2007). Thereafter, a linkage map was generated using two R packages “ASMap” and “R-qtl”. Before building the linkage map, SNPs were filtered and cleaned up by removing low quality SNPs based on criteria such as “less than 10% missing data”, “duplicated SNPs”, and “SNPs with linkage distortion”.

DATA ANALYSIS

Phenotypic data were checked for normality, and data on nodules' number (not normally distributed) were transformed using Box cox transformation (SAS Institute Inc. 2011). Box cox transformation tool suggested “square root” transformation as the most adequate. No transformation was more adequate than raw data for the other phenotypic traits (NE, NFW, NDW).

R-qtl (<http://www.rqtl.org>) was used to identify significant regions that are associated to promiscuous nodulation traits. After phenotype data was added to linkage map file, the genome was scanned via the “scanone” function of Haley-Knott (hk) regression, using the argument method=“hk”. Then, multiple imputation of the genome was performed using the function “sim. geno” with method=“imp” and n. draws=64 (64 imputations). Thereafter, permutation test was performed to get a genome-wide significance threshold and p-values. Argument for that was method=“hk” with n. perm= 1000 (1000 permutations). Two significance levels (alpha=0.2 and alpha=0.05) were set as suggested in the Rqtl manuel, the location of the QTL was estimated using the Haley-Knott (hk) regression method with both “lodint” and its alternative “bayesint” functions. Later, the closest flanking markers were identified, according to both intervals calculated in the former step. Lastly, the effects of the QTLs were estimated by plotting the average phenotype for each genotype group, the argument plotPXG was used for that purpose.

RESULTS

Initially, a genotype file with over 65,000 SNPs was generated. After filtering low quality SNPs, a map with 29,620 SNPs covering all the twenty linkage groups was generated. Permutation tests generated various Lod score values for each trait. These ranged from 4.94 to 5.22 (Figure 1 and Table 1).

At least one significant QTL region was identified for all measured trait (Table 1). For the number of nodules (NN), two significant QTLs were observed: one located at the c10.loc3549 on the chromosome 10 (Lod score=5.05, and P-value=0.181). The marker S10_9400998 was the closest to that region, with S10_33852650 and S10_29297382 as flanking markers. The second QTL was located at the marker S13_11994810 on chromosome 13 (Lod score=5.05, and P-value=0.166), the flanking markers to that were S13_11994819 and S13_11994800.

For the percentage of effective nodules (NE), one QTL was found at S13_11994810 on chromosome 13 (Lod score=5.44, and P-value=0.118), with S13_11994819 and S13_15403889 as flanking markers.

The fresh weight of nodules (NFW) showed one significant QTL on chromosome 10 at the loci c10.loc3545 (Lod score=5.66, and P-value=0.103), the flanking markers were S10_33852650 and S10_37822769.

The dry weight of nodules also exhibited one significant QTL on chromosome 10 at the loci c10.loc3545 (Lod score=5.15, and P-value=0.197), with S10_33852650 and S10_9412367 as flanking markers.

TABLE 1: Qtls position, and closest flanking markers for promiscuous nodulation traits

Trait	Locus	Position	Threshold (20%)	Lod scores	P-values ($\alpha=0.2$)	Closest flanking markers
NN	c10.loc3549	3549	4.99	5.05	0.181	S10_33852650 and S10_29297382
	S13_11994810	7861	4.94	5.05	0.166	S13_11994819 and S13_11994800
NE	S13_11994810	7861	5.09	5.44	0.118	S13_11994819 and S13_15403889
NFW	c10.loc3545	3545	5.22	5.66	0.103	S10_33852650 and S10_37822769
NDW	c10.loc3545	3545	5.14	5.15	0.197	S10_33852650 and S10_9412367

NN= nodules' number, NE=percentage of effective nodules, NFW=fresh weight of nodules, and NDW=dry weight of nodules

As for the QTL effects, Figure 2a showed higher mean performance for the progenies compared to the parents' mean, suggesting over dominance QTL effect on nodules' number on chromosome 10. On the chromosome 13 (Figure 2b) the mean performance for progenies were lower compared to both parents. This suggest non-additive QTL effect on number of nodules. For nodules' effectiveness (NE), only one QTL was found, and this was on chromosome 13. Similar to the situation observed in NN on that chromosome, progenies mean performance were lower than both parents' (Figure 2c). As for the fresh weight of nodules (NFW), one QTL was found on chromosome 10, progenies had higher mean performance than both parents (Figure 2d), this situation is similar to the one observed in NN on the same chromosome 10, suggesting an over dominance QTL effect on NFW. The same apply to dry weight of nodules (NDW) (Figure 2e), with one QTL on chromosome 10, where progenies' mean performance were higher than both parents'.

a

b

FIGURE 1: Curves showing Lod scores versus chromosomes: (a) chromosomes from 1 to 10 (b) chromosomes from 11 to 20

FIGURE 2: QTLs effects for nodule traits on chromosome 10 and 13

DISCUSSION

The QTL identified on chromosome 10 for NN was the same as the one of NFW and NDW. Even though the closest flanking markers were not exactly the same, they all shared one closest flanking marker, namely S10_33852650. This is an indication that these three traits are under the control of the same genome region, suggesting a possibility of pleiotropic genes in that region. The same was observed on chromosome 13 for NN and NE, where the marker S13_11994810 was significant for both traits at position 7861. Although the flanking markers showed larger interval for NE, both traits shared the flanking marker S13_11994819. As observed previously, this suggests the probable existence of pleiotropic genes in that region, which control nodule number and effective nodules. The genotypes wondersoya and CG 2043 used in this study revealed the existence of QTLs associated with promiscuous nodulation in soybean, these findings open doors to the possibility to carry out marker assisted breeding. However, the over dominance effect of QTL consistently observed on chromosome 10 is an indication of interactions of several alleles at this region. As described by East (1908) and Shull (1908), over dominance is a result of cumulative action of divergent alleles or stimulation of divergent alleles in heterozygotes. This hypothesis was supported by Hull (1945). East (1936) emphasized further that a combination of higher number of divergent alleles exhibit higher heterosis than less divergent combinations. Hence, this study enabled understand that in promiscuous nodulation, heterozygotes can have advantage to the survival of many alleles that are recessive and harmful in homozygotes. The detection of those QTLs that enable the above mentioned hypotheses is very insightful, as it may inform on the possibility applying marker assisted selection on promiscuity character in soybean breeding, also this can serve as starting point for gene cloning.

The fact that no significant QTL was observed with $p\text{-value} = 0.05$ can be explained by the low population size (92 individuals). For instance, Semagn *et al.* (2006) reported that a total of 200 individuals is required to construct a reasonably accurate linkage map. Also some of the SNPs closer to the interesting regions could have been filtered out during data processing, leading to reduction in accuracy. It is therefore recommended that SSR markers that are polymorphic in those regions be tested to confirm these observations. Moreover, it would be useful to perform such QTL identification study on a larger F₂ population or RILs population, as this would yield more QTLs with more accuracy. It appears that a set of genes located in those genomic regions on the chromosomes 10 and 13 control the nodulation traits, this suggest the hypothesis of pleiotropic genes involved in the expression of the nodulation traits. It is recommended that those genes be cloned for proper understanding of the promiscuity character in soybean.

CONCLUSION

There are quantitative traits loci associated with promiscuous nodulation in soybean. The present study identified two loci, one on chromosome 10 (c10.loc3545) and one on chromosome 13 (S13_11994810). All nodulation traits were associated with one or both QTLs. Number of nodules (NN) was associated with both QTLs. NN shared one QTL (c10.loc3545) with NFW and NDW, and the QTL (S13_11994810) with NE. This allows inferring on the possibility of applying marker assisted selection and search for pleiotropic genes in those genomic regions. Also, the findings of this study can serve as starting point for gene cloning. Therefore, it is recommended that investigations be pursued on that matter, to identify and clone those genes to enable boosting breeding activities for the development of promiscuous soybean cultivars.

ACKNOWLEDGEMENT

Eric Agoyi is a Fellow of the Norman E. Borlaug Leadership Enhancement in Agriculture Program funded by USAID. Support for this research was provided in part by the Borlaug Leadership Enhancement in Agriculture Program (Borlaug Leap) through a grant to the university of California Davis by the United States Agency for International Development. The opinions expressed herein are those of the authors and do not necessarily reflect the views of USAID. Acknowledge goes to Prof. Diers' Lab in the National Soybean Research Laboratory (NSRL) at the University of Illinois, Urbana-Champaign, United State of America for providing lab facilities and technical support to the implementation of this research. Also we acknowledge support received from NaCCRI and CIAT-Uganda for their support during DNA extraction and processing for shipment.

References

- Abaidoo CR, Keyser HH, Singleton PW, Dulal BD. (2000). *Bradyrhizobium* spp. (TGx) isolates nodulating the new soybean cultivars in Africa are diverse and distinct from bradyrhizobia that nodulate North American soybeans. *International Journal of Systematic and Evolutionary Microbiology* 50:225-234.
- Agoyi EE, Afutu E, Tumuhairwe JB, Odong TL, Tukamuhabwa P. (2016). Screening soybean genotypes for promiscuous symbiotic association with bradyrhizobium strains. *African Crop Science Journal*. 24:49-59. DOI: <http://dx.doi.org/10.4314/acsj.v23i4.1>
- Agoyi EE, Mohammed KE, Odong TL, Tumuhairwe JB, Chigeza G, Tukamuhabwa P. (2016). Mode of inheritance of promiscuous nodulation and combining abilities in soybean genotypes. *International Journal of Agronomy and Agricultural Research (IJAAR)*, 9(1): 73-82.
- Anuar AR, Shamsuddin ZH, Ya'acob O. (1995). Contribution of legume – N by nodulated groundnuts for growth of subsequent Maize in an acid soil. *Soil Biol. Biochem.* 27(2): 595-601.
- Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, et al. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* 193, 1073–1081. doi: 10.1534/genetics.112.147710
- Browning BL and Browning SR. (2016). Genotype imputation with millions of reference samples. *Am J Hum Genet* 98:116-126. doi:10.1016/j.ajhg.2015.11.020
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, et al. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12: 499–510. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21681211>. Accessed 16 April 2019.
- Deschamps S, Liaca V and May GD. (2012). Genotyping-by-sequencing in plants. *Biology* 1, 460–483. doi: 10.3390/biology1030460.
- Doerge R. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Rev.* 3: 43-52.
- East EM. (1936). Heterosis. *Genetics* 21: 375–397.
- East EM. (1908). "Inbreeding in corn". Reports of the *Connecticut Agricultural Experiments Station* for 1907: 419–428.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, *PLoS ONE* 6 (5) (2011) e19379.
- Fu YB, Cheng B and Peterson GW. (2014). Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. *Genet. Resour. Crop Evol.* 61, 579–594. doi: 10.1007/s10722-013-0058-1.

Fu YB and Peterson GW. (2011). Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. **Plant Genome** 4, 226–237. doi: 10.3835/plantgenome2011.08.0022.

Giller KE, Cadisch G, Ehaliotis C, Adams E, Sakala W and Mafongoya P. (1997). Building soil nitrogen capital in Africa. Pp. 151–192 in 'Replenishing soil fertility in Africa', ed. by Buresh R, Sanchez P and Calhoun F. American Society for Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI.

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q and Buckler ES. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. **PLoS ONE** 9(2):E90346. doi:10.1371/journal.pone.0090346.

Gwata ET., Wofford DS, Pfahler, PL and Boote KJ. (2004). Genetics of promiscuous nodulation in soybean: Nodule dry weight and leaf color score. **Journal of Heredity** 95(2): 154-157, ISSN: 0022 - 1503.

Hadley HH and Hymowitz T. (1973). Speciation and cytogenetics. In: Soybeans: Improvement, Production, and Uses, pp. 97- 116. Edited by Caldwell BE, Madison, WI: **American Society of Agronomy**.

Hull FH. (1945). Recurrent Selection for Specific Combining Ability in Corn. **J. Am. Soc. Agron.**, 37:134-145.

IITA. (1996). Understanding nitrogen fixation in promiscuous soybean. In: **Microbiology in the service of crops and soils**. Annual Report. Pp20-23.

Jiang GL. (2013). Molecular Markers and Marker-Assisted Breeding in Plants <http://dx.doi.org/10.5772/52583> accessed 22 November 2018.

Jonah PM, Bello LL, Lucky O, Midau A, Moruppa SM. (2011). Review: The Importance of Molecular Markers in Plant Breeding Programmes. **Global Journal of Science Frontier Research** 11(5). Publisher: Global Journals Inc. (USA) ISSN: 0975-5896.

Kim K, Unfried JR, Hyten DL, Frederick RD, Hartman GL, Nelson RL, Song Q, Diers BW. (2012). Molecular mapping of soybean rust resistance in soybean accession PI 561356 and SNP haplotype analysis of the Rpp1 region in diverse germplasm. *Theor Appl Genet* 125:1339–1352.

Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715.

Liu S, Liu Y, Yang X, et al. (2014). The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes, **Nat. Commun.**, 5, 3930.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. **PLoS Genetics** 9:e1003215. doi: 10.1371/journal.pgen.1003215.

Machido DA, Olufajo OO, Yakubu SE, Yusufu SS. (2011). Enhancing the contribution of the legumes to the N-fertility of soils of the semi-arid zone of Nigeria. **African Journal of Biotechnology**, 10:1848-1853.

Malosetti M, van Eeuwijk FA, Boer MP, Casas AM, Elía M, Moralejo M, Bhat PR, Ramsay L, Molina-Cano J-L. (2011). Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. **Theor Appl Genet.** ;122(8):1605–1616. doi: 10.1007/s00122-011-1558-z. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]

Narum SR, Buerkle CA, Davey JW, Miller MR and Hohenlohe PA. (2013). Genotyping-by -sequencing in ecological and conservation genomics. **Mol. Ecol.** 22, 2841–2847. doi: 10.1111/mec.12350.

Prasanna BM. (2007). QTL mapping and its applications in crop plants. Indian Agricultural Research Institute, **PUSA**, New Delhi-110 012.

Peoples MB, Ladha JK, Herridge DR. (1995). Enhancing legume N₂- fixation through plant and Soil Management. **Plant Soil** ; 174: 85-101.

Poland JA, Rife TW. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. **Plant Genome J.** 5: 92. Available: <https://www.crops.org/publications/tpg/abstracts/5/3/92> Accessed 18 June 2019.

Pulver EL, Kueneman EA, and Ranga-Rao V. (1985). Identification of promiscuous nodulating soybean efficient in N₂ fixation. **Crop Sci.** 25:1065– 1070.

Rincker K, Hartman LG, Diers WB. (2016). Fine Mapping of Resistance Genes from Five Brown Stem Rot Resistance Sources in Soybean. **The Plant Genome** 9 doi: 10.3835/plantgenome2015.08.0063.

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES and Gardner CA. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biol.** 14(6):R55. doi:10.1186/gb-2013-14-6-r55.

Salvagiotti F, Cassman KG, Specht JE, Walters DT, Weiss A, Doberman A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybean: A review. **Field Crops Res.** 108:1-13 (<http://tiny.cc/soybeanNreview>).

Sanginga N, Dashiell K, Okogun JA, and Thottappilly G. (1997). Nitrogen fixation and N contribution by promiscuous nodulating soybeans in the southern Guinea savanna of Nigeria. **Plant and Soil** ; 195, 2, 257-266, ISSN: 0032 – 079X.

Semagn K, Bjornstad A, Ndjioudjop MN. (2006). An overview of molecular marker methods for plants. **Afr. J. Biotech.**, 5:2540–2568.

Shull GH. (1908). "The composition of a field of maize". Reports of the American Breeders Association: 296–301.

Spindel J, Wright M, Chen C, et al. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations, *Theor. Appl. Genet.*; 126, 2699–716.

Terefa H. (2011). Breeding for promiscuous soybeans at IITA, Soybean, In: Dr. Alexander S (Ed.) Molecular Aspects of Breeding, ISBN: 978-953-307-240-1, In: Tech, Available from: <http://www.intechopen.com/books/soybean-molecular-aspects-of-breeding/breeding-for-promiscuoussoybeans-at-IITA> accessed the 14th April 2019.

Uitdewilligen JG, Wolters AMA, Bjorn B, et al. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One*;8(5), e62355. doi:10.1371/journal.pone.0062355.

van Poecke RM, Maccaferri M, Tang J, Truong HT, Janssen A, van Orsouw NJ, et al. (2013). Sequence-based SNP genotyping in durum wheat. *Plant Biotechnol. J.* 11, 809–817. doi: 10.1111/pbi.12072.

Yu N, Kim M, King RZ, Harris KD, Buck WJ, Li Z, Diers WB. (2015). Fine mapping of the Asian soybean rust resistance gene **Rpp2** from soybean PI 230970. *Theor Appl Genet* 128: 387. doi:10.1007/s00122-014-2438-0.