GENETIC ARCHITECTURE AND MARKER-ASSISTED BREEDING FOR SALT TOLERANCE IN SOYBEAN

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri - Columbia

In Partial Fulfillment

of the Requirements for the Degree

of Doctor of Philosophy in Plant Breeding, Genetics, and Genomics

By

TUYEN DUC DO

Dr. J. Grover Shannon, Dissertation Supervisor

MAY 2018

The undersigned, appointed by the Dean of the Graduate School,

have examined the Dissertation entitled

GENETIC ARCHITECTURE AND MARKER-ASSISTED BREEDING FOR SALT TOLERANCE IN SOYBEAN

Presented by TUYEN DUC DO

A candidate for the degree of

Doctor of Philosophy in Plant Breeding, Genetics, and Genomics

And hereby certify that, in their opinion, it is worthy of acceptance.

Dr. J. Grover Shannon, Chair	Dr. Henry T. Nguyen, Co-chair
Dr. Pengyin Chen, Committee member	Dr. Dong Xu, Committee member
Dr. Tri D. Vuong, Committee member	Dr. Andrew M. Scaboo, Committee member

ACKNOWLEGEMENTS

I would sincerely like to thank everyone who has directly or indirectly shared views with or contributed in materializing my research and Ph.D. program.

Foremost, I would like to express my deep sense of gratitude and heartfelt thanks to my primary academic advisor, **Dr. J. Grover Shannon**, Professor Emeritus, Division of Plant Sciences, University of Missouri – Columbia for providing me an external grant after the first two years supported by my government to pursue the Ph.D. degree in Mizzou, for his patience and enthusiasm even difficulties in distance. His competent guidance helped me not only resolving scientific problems throughout my research program, but also developing other skills for my future career.

I would also like to thank my co-advisor, **Dr. Henry T. Nguyen**, Curators Professor, Division of Plant Sciences, University of Missouri – Columbia for giving me the opportunities to work in his lab and to learn more scientific areas beside my major, for his guidance and help in publications. My research program would have not been possible to complete without the sophisticated equipment of his lab.

Besides my advisor and co-advisor, I would like to thank the rest of my doctoral committee: **Dr. Pengyin Chen**, Endowed Professor, Division of Plant Sciences, University of Missouri – Columbia, **Dr. Dong Xu**, Professor, Department of Electrical Engineering and Computer Science, University of Missouri – Columbia, **Dr. Tri D. Vuong**, Senior Research Scientist, Division of Plant Sciences, University of Missouri – Columbia, and **Dr. Andrew M. Scaboo**, Assistant Research Professor, Division of Plant Sciences,

University of Missouri – Columbia for their valuable ideas and suggestions as members of my advisory committee.

I would like to thank all my lab mates and my friends for their help, supports and kind co-operation throughout my graduate program in Mizzou. I will be forever grateful for the friendships that I have gotten here.

I am grateful to my home institute (Cuu Long Delta Rice Research Institute), Vietnam Academy of Agricultural Sciences, Ministry of Agriculture and Rural Development (Vietnam), and the Vietnam International Education Development (VIED) fellowship program for giving me the great opportunity and financially supporting in part to study graduate program in the United States. It would have not been possible for me to complete Ph.D. program without this support.

Last but not least, I must express my very profound gratitude to my family members, my parents for giving birth to me in the first place and providing me with unfailing support, my parents-in-law, my brother and sister, my brother-in-law, and my sisters-in-law for supporting me spiritually throughout my life. Special thanks to my spouse, **Nam-Ly Thi Tran**, and my son, **Minh-Dang Duc Do**, for continuous encouragement throughout my years of study and through the process of researching for this Ph.D. program. This accomplishment would not have been possible without their love and emotional support. Thank you for your contributions.

TUYEN DUC DO

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Salinity is one of the major abiotic stresses that inhibits plant growth and causes seed yield loss in soybean. Although a major gene for salt tolerance on chromosome (Chr.) 3 was mapped, cloned and characterized, it does not fully explain genetic variability for tolerance in soybean. Two mapping approaches, quantitative trait loci (QTL) mapping and genome-wide association study (GWAS), can complement each other to identify genomic regions and molecular markers associated with traits of interest. QTL mapping is more suitable to map traits governed by rare alleles in a designed population while GWAS is better in mapping traits underlined by few genes of large effect in the natural population. This study was performed to identify additional loci and new sources for salt tolerance by using both approaches. For bi-parental QTL mapping, salt tolerance of 132 F_2 families was evaluated by accessing leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC). Their genotypes were obtained using the Illumina Infinium SoySNP6K BeadChip assay to map salt tolerant gene(s). A major locus significantly associated with LSS, CCR, LSC, and LCC was mapped to Chr. 3 with LOD scores of 19.1, 11.0, 7.7, and 25.6, respectively. In addition, a second locus associated with salt tolerance for LSC was also detected and mapped on Chr. 13 with a LOD score of 4.6 and an R^2 of 0.115. The evaluation of salt tolerance of an F_5 population derived from the same cross showed that combining salt tolerant alleles of major and minor loci significantly increased salt tolerance. On the other hand, GWAS for salt tolerance was conducted using SNPs of two datasets, SoySNP50K iSelect BeadChip and 3.7M SNP dataset (from whole-genome sequencing data), across 305 soybean accessions of a diverse panel. The known gene on Chr. 3 was confirmed by three gene-based markers (GBMs) that

integrated into both datasets. Other genomic regions significantly associated with salt tolerance were identified on Chrs. 1, 2, 5, 6, 8, 14, 18, and 19 by analyzing 3.7M SNP dataset, in which the position on Chr. 8 strongly predicted a new minor locus for salt tolerance. The genotype-phenotype correlation using three GBMs discovered six new salt tolerant sources that may carry novel gene(s) for salt tolerance. By complementation tests and segregation analysis of salt tolerance among F_2 plants developed from a cross of Fiskeby III and a salt tolerance accession, PI 468908, it was speculated that salt tolerance from PI 468908 was possibly controlled by a new gene instead of the known gene on Chr. 3. These significant loci in new salt tolerant sources coupled with significant SNP markers could be useful for marker-assisted selection in molecular breeding programs to improve salt tolerance in soybean.

Chapter 1:
INTRODUCTION

Impacts of salinity stress

The dominance of salt water across the earth has led to the widespread occurrence of salt-affected soils. Secondary salinization of soils is caused by irrigation water and de-forestation. Salt-affected soils are classified into two main categories: sodic and saline. Sodic (alkaline) soils are dominated by excess sodium on exchange sites and a high concentration of carbonate/bicarbonate anions. They have a high pH (greater than 8.5 and up to 10.8) with a high sodium absorption ratio (SAR) and poor soil structure. Saline soils are generally dominated by sodium ions; but with the dominant anions being chloride and sulphate, pH values and SARs are much lower and electrical conductivities higher than in sodic soils. Salt-affected soils contain sufficient concentrations of soluble salts to reduce the growth of most plant species (Flowers and Flowers 2005).

Statistics on the global extent of salt-affected soils vary according to data sources, but estimates in general are approximately 1 billion hectares (FAO/AGL 2000; Szabolcs 1989). Based on the Harmonized World Soil Database (HWSD), the global salt-affected land is estimated 1128 million ha, of which are 60% of saline soils, 26% of sodic soils and 14% of saline-sodic soils (Wicke et al. 2011). Globally among irrigated agricultural land, an estimated 60 million ha is affected by salinity (FAO and ITPS 2015), and 0.25–0.5 million ha salt build-up annually causes lost agricultural production (FAO 2002). By simple extrapolation, the global annual cost of salt-induced land degradation is

approximately US$ 27.3 billion because of lost crop production in salt-affected irrigated areas (Qadir et al. 2014).

The United States has large salt-affected areas of soils with 5.2 million ha (FAO and ITPS 2015) or approximately 23% of the total irrigated land. The primary salt-affected soils are located in the western areas of the country. The salt-affected agricultural areas are located in California, Arizona, North and South Dakota and the coastal regions of South Texas (Pitman and Läuchli 2002). However, the secondary salt-affected regions related to excessive dissolved-solids concentration from human activities (Fig. 1.1) are found in many states (Anning and Flynn 2014) including soybean production areas.

Effects of salinity on soybean

High salt concentration has negative impacts on plant growth, nodulation, agronomic traits, seed quality and quantity, and thus reduces the yield of soybean (Phang et al. 2008). Higher concentration of salts led to an absolute decrease in seed germination (Abel 1969; Abel and MacKenzie 1964). The seedling stage of soybean is considered to be much more sensitive to salt stress than at germination (Hosseini et al. 2002). The agronomic traits of soybean could be severely affected by high salinity, including reduction in plant height, leaf size, biomass, number of internodes, number of branches, number of pods, weight per plant, and weight of 100 seeds (Abel and MacKenzie 1964). Under greenhouse conditions, chlorophylls *a*, *b* and total chlorophyll content and relative water content were significantly reduced with increasing NaCl salinity. The highest relative electrolytic leakage and lipid peroxidation occurred at the highest salinity level (Weisany et al. 2011). In addition, salt stress affects the nodulation of soybean, reduces the efficiency

of nitrogen fixation, and decreases the number and biomass of root nodules (Delgado et al. 1994; Duzan et al. 2004; Elsheikh and Wood 1995).

Salt tolerance mechanism in soybean

Three mechanisms of salinity tolerance have been reported in crop plants (Fig. 1.2). **Tissue tolerance**, where high salt concentrations are found in leaves but are compartmentalized at the cellular and intracellular level (especially in the vacuole), a process involving ion transporters, proton pumps and synthesis of compatible solutes. **Osmotic tolerance**, which is related to minimizing the effects on the reduction of shoot growth, and may be related to as yet unknown sensing and signaling mechanisms. **Ion exclusion**, where Na^+ and Cl^- transport processes, predominantly in roots, prevent the accumulation of toxic concentrations of Na^+ and Cl^- within leaves. Mechanisms may include retrieval of Na^+ from the xylem, compartmentation of ions in vacuoles of cortical cells and/or efflux of ions back to the soil (Roy et al. 2014).

Low levels of Cl- in stems and leaves are related to salt tolerance of soybean (Abel 1969; Abel and MacKenzie 1964). Plant injury was associated more with Na^+ rather than with Cl^- concentration in leaves (Lenis et al. 2011). Moreover, the salt tolerance in soybean correlates to accumulation of Na^+ and Cl^- concentrations in root and leaves. The salt tolerant soybean maintains low concentrations of both Na^+ and Cl^- in leaves under saline conditions (Do et al. 2016; Lenis et al. 2011). The root genotype, but not the shoot genotype, determines salt tolerance in soybean in Fig. 1.3 (Do et al. 2016). This may relate to mechanism of ion exclusion, where Na^+ and Cl^- transport processes in roots reduce the accumulation of toxic concentrations of Na^+ and Cl^- (Roy et al. 2014)

Genetics and mapping salt tolerant gene in soybean

Genetic studies showed salt tolerance is controlled by a single dominant gene. Earlier analysis of eight crosses between soybean parents with contrasting differences of Cl^- accumulation in leaves and stems (Abel 1969; Abel and MacKenzie 1964) showed F_2 progenies segregated in a ratio of 3 excluders (tolerant) to 1 includer (sensitive). Similar results were confirmed by analyzing salt tolerance of the $F_{2:3}$ families (Hamwieh and Xu 2008; Lee et al. 2009). The major locus for salt tolerance was mapped at the similar position on chromosome (Chr.) 3 using bi-parental quantitative trait locus (QTL) mapping (Guan et al. 2014; Ha et al. 2013; Hamwieh et al. 2011; Hamwieh and Xu 2008; Lee et al. 2004; Qi et al. 2014; Zeng et al. 2017b). Moreover, some minor loci associated with salt tolerance or related traits were identified on Chrs. 2, 7, 9, 11, 13, 14, 15 and 18 (Chen et al. 2008; Zeng et al. 2017b). However, there are limitations of bi-parental linkage mapping, such as only parental alleles are detected and a few recombination events occur in mapping populations.

Genome-wide association study (GWAS) that exploits broader genetic diversity in a natural population is an alternative to traditional QTL mapping. Association mapping for salt tolerance in soybean was first reported based on analyzing the traits related germination of soybean under salt stress conditions and three significant genomic regions related to salt tolerance were found on Chrs. 8, 9 and 18 (Kan et al. 2015). In other studies, the major locus for salt tolerance on Chr. 3 was confirmed and additional genomic regions on Chrs. 2, 7, 8, 10, 13, 14, 16, and 20 were significantly associated with salt tolerant traits (Patil et al. 2016; Zeng et al. 2017a).

Although a major locus and some minor loci for salt tolerance were identified and the gene controlling salt tolerance was cloned and characterized (Do et al. 2016; Guan et al. 2014; Qi et al. 2014), genetic improvement of salt tolerance in soybean is limited to a few salt tolerance sources (e.g. S-100 in the U.S. and FT-Abyara in Brazil). Other previous studies showed that some new salt tolerance sources could be useful for studying and improving salt tolerance, such as Fiskeby III. It originated from Sweden, is tolerant or partially resistant to drought, iron deficiency chlorosis, toxic soil aluminum, salt, and atmospheric ozone pollution (Burton et al. 2016; Pathan et al. 2007; USDA 2011). On the other hand, a core collection of *G. max* and *G. soja* has been selected from 19,929 accessions of the USDA soybean collection based on diversity analysis using SoySNP50K. Thus, mapping salt tolerant genes from this germplasm and a genome-wide association study (GWAS) using a diverse panel of the soybean core collection was performed to seek additional loci associated with salt tolerance in soybean. The research for this dissertation entitled **"Genetic architecture and marker-assisted breeding for salt tolerance in soybean"** was conducted with the following objectives: (1) verify salt tolerance allele from cultivar Fiskeby III, (2) identify major or minor loci for salt tolerance, and (3) investigate the genetics of new salt tolerance sources.

Legends to Figures

Figure 1.1 Map of the conterminous U.S. showing share of long-term mean annual incremental-catchment yield contributed from human sources, predicted by the national SPARROW model of dissolved-solids transport

Figure 1.2 The three general salt tolerance mechanisms in crop plants

Figure 1.3 Na^+, K^+, and Cl^- contents in leaves, stems, and roots in salt tolerant near isogenic lines NILs25-T and NILs25-S and their graft hybrids NILs25-T/NILs25-S and NILs25-S/NILs25-T

References

Abel GH (1969) Inheritance of the capacity for chloride inclusion and chloride exclusion by soybeans. Crop Sci 9:697-698

Abel GH, MacKenzie AJ (1964) Salt tolerance of soybean varieties (*Glycine max* L. Merill) during germination and later growth. Crop Sci 4:157-161

Anning DW, Flynn ME (2014) Dissolved-solids sources, loads, yields, and concentrations in streams of the conterminous United States. U S Geological Survey Scientific Investigations Report 2014-5012:101 p.

Burton AL, Burkey KO, Carter TE, Jr., Orf J, Cregan PB (2016) Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III × Mandarin (Ottawa) soybean population. Theor Appl Genet 129:1113-1125

Chen HT, Cui SY, Fu SX, Gai JY, Yu DY (2008) Identification of quantitative trait loci associated with salt tolerance during seedling growth in soybean (*Glycine max* L.). Aust J Agr Res 59:1086-1091

Delgado MJ, Ligero F, Lluch C (1994) Effects of salt stress on srowth and sitrogen-sixation by pea, faba-bean, common bean and soybean plants. Soil Biol Biochem 26:371-376

Do TD, Chen H, Hien VT, Hamwieh A, Yamada T, Sato T, Yan Y, Cong H, Shono M, Suenaga K, Xu D (2016) *Ncl* synchronously regulates Na(+), K(+), and Cl(-) in soybean and greatly increases the grain yield in saline field conditions. Sci Rep 6:19147

Duzan HM, Zhou X, Souleimanov A, Smith DL (2004) Perception of *Bradyrhizobium japonicum* Nod factor by soybean [*Glycine max* (L.) Merr.] root hairs under abiotic stress conditions. Journal of Experimental Botany 55:2641-2646

Elsheikh EAE, Wood M (1995) Nodulation and N-2 fixation by soybean inoculated with salt-tolerant *Rhizobia* or salt-sensitive *Bradyrhizobia* in saline soil. Soil Biol Biochem 27:657-661

FAO (2002) Crops and Drops: Making the Best Use of Water for Agriculture. Food and Agriculture Organization of the United Nations, Rome

FAO, ITPS (2015) Status of the world's soil resources (SWSR) – main report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy:pp 124-127

FAO/AGL (2000) Extent and Causes of Salt affected Soils in Participating Countries. http://wwwfaoorg/ag/AGL/agll/prosoil/indexhtm

Flowers TJ, Flowers SA (2005) Why does salinity pose such a difficult problem for plant breeders? Agricultural Water Management 78:15-24

Guan RX, Qu Y, Guo Y, Yu LL, Liu Y, Jiang JH, Chen JG, Ren YL, Liu GY, Tian L, Jin LG, Liu ZX, Hong HL, Chang RZ, Gilliham M, Qiu LJ (2014) Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. Plant J 80:937-950

Ha BK, Vuong TD, Velusamy V, Nguyen HT, Shannon JG, Lee JD (2013) Genetic mapping of quantitative trait loci conditioning salt tolerance in wild soybean (*Glycine soja*) PI 483463. Euphytica 193:79-88

Hamwieh A, Tuyen DD, Cong H, Benitez ER, Takahashi R, Xu DH (2011) Identification and validation of a major QTL for salt tolerance in soybean. Euphytica 179:451-459

Hamwieh A, Xu DH (2008) Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. Breeding Sci 58:355-359

Hosseini MK, Powell AA, Bingham IJ (2002) Comparison of the seed germination and early seedling growth of soybean in saline conditions. Seed Science Research 12:165-172

Kan G, Zhang W, Yang W, Ma D, Zhang D, Hao D, Hu Z, Yu D (2015) Association mapping of soybean seed germination under salt stress. Mol Genet Genomics 290:2147-2162

Lee GJ, Carter TE, Jr., Villagarcia MR, Li Z, Zhou X, Gibbs MO, Boerma HR (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109:1610-1619

Lee JD, Shannon JG, Vuong TD, Nguyen HT (2009) Inheritance of salt tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.) accession PI483463. J Hered 100:798-801

Lenis JM, Ellersieck M, Blevins DG, Sleper DA, Nguyen HT, Dunn D, Lee JD, Shannon JG (2011) Differences in ion accumulation and salt tolerance among *Glycine* accessions. J Agron Crop Sci 197:302-310

Pathan MS, Lee J-D, Shannon JG, Nguyen HT (2007) Recent advances in breeding for drought and salt stress tolerance in soybean. In: Jenks MA, Hasegawa PM, Jain SM

(eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer Netherlands, Dordrecht, pp 739-773

Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, Shannon JG, Nguyen HT (2016) Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci Rep 6:19199

Phang TH, Shao G, Lam HM (2008) Salt tolerance in soybean. J Integr Plant Biol 50:1196-1212

Pitman MG, Läuchli A (2002) Global Impact of Salinity and Agricultural Ecosystems. In: Läuchli A., Lüttge U. (eds) Salinity: Environment - Plants - Molecules. Springer, Dordrecht

Qadir M, Quillerou E, Nangia V, Murtaza G, Singh M, Thomas RJ, Drechsel P, Noble AD (2014) Economics of salt-induced land degradation and restoration. Nat Resour Forum 38:282-295

Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

Roy SJ, Negrao S, Tester M (2014) Salt resistant crop plants. Curr Opin Biotechnol 26:115-124

Szabolcs I (1989) Salt-affected soils. Boca Raton, Fla. : CRC Press, ©1989.

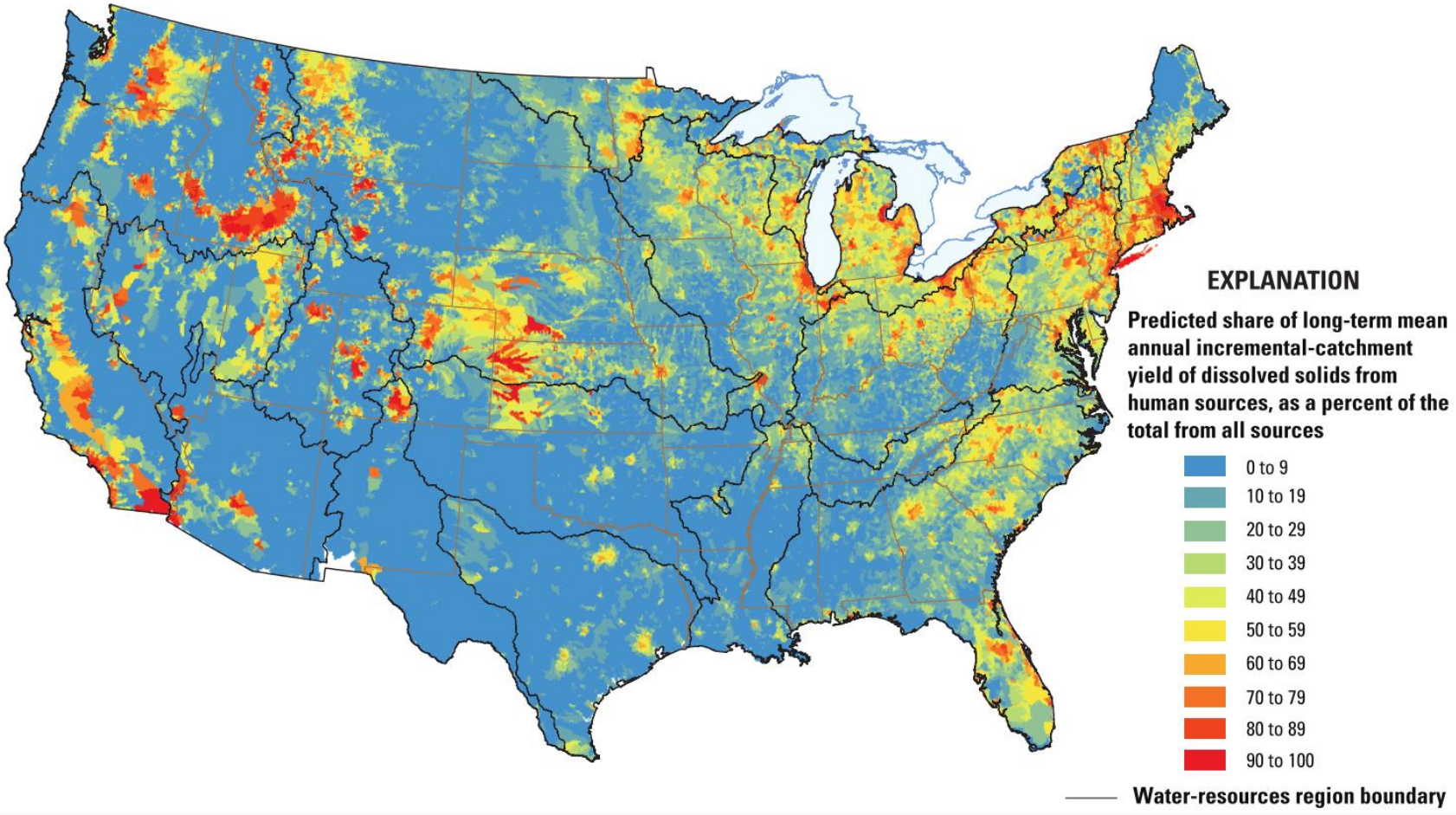
USDA (2011) Breeding plants for a high-ozone world. Agr Res 59:14-17

Weisany W, Sohrabi Y, Heidari G, Siosemardeh A, Ghassemi-Golezani K (2011) Physiological responses of soybean (*Glycine max* L.) to zinc application under salinity stress. Aust J Crop Sci 5:1441-1447

Wicke B, Smeets E, Turkenburg W, Faaij A (2011) The global technical and economic potential of bioenergy from salt-affected soils. Energy Environ Sci 4:2669-2681

Zeng A, Chen P, Korth K, Hancock F, Pereira A, Brye K, Wu C, Shi A (2017a) Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. Mol Breeding 37:30

Zeng A, Lara L, Chen P, Luan X, Hancock F, Korth K, Brye K, Pereira A, Wu C (2017b) Quantitative trait loci for chloride tolerance in 'Osage' soybean. Crop Sci 57:2345-2353



Anning and Flynn, 2014

Figure 1.1 Map of the conterminous U.S. showing share of long-term mean annual incremental-catchment yield contributed from human sources, predicted by the national SPARROW model of dissolved-solids transport

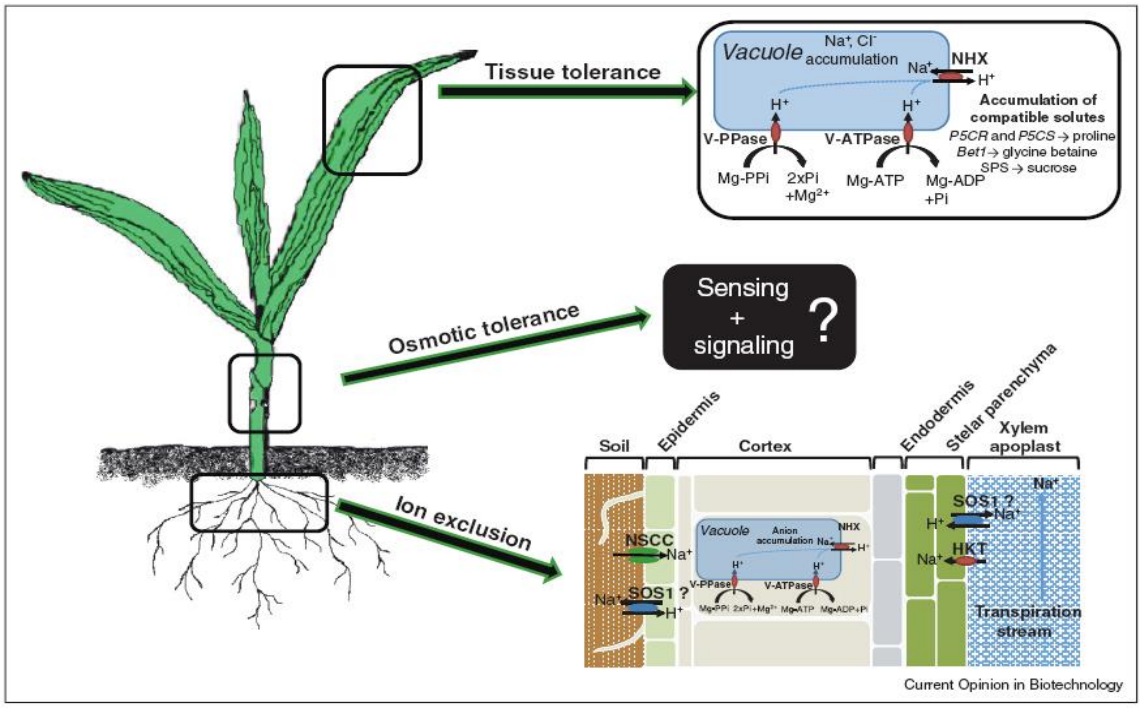
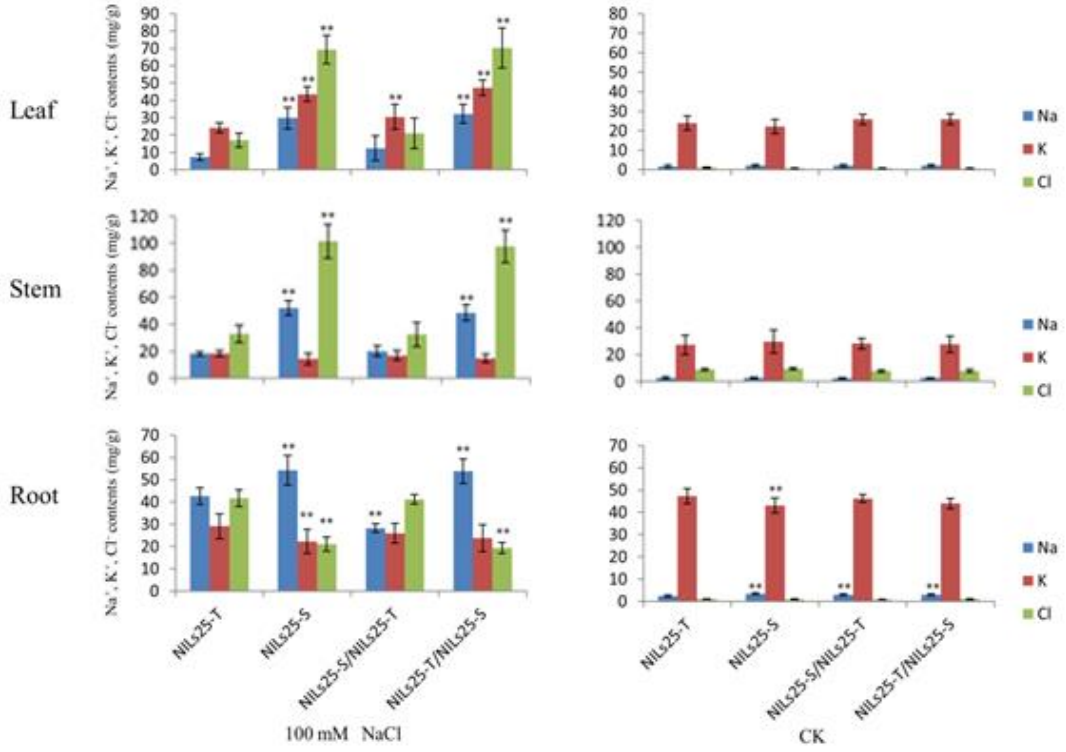


Figure 1.2 The three general salt tolerance mechanisms in crop plants (Roy et al. 2014)



Do *et al.* 2016

Figure 1.3 Na^+, K^+, and Cl^- contents in leaves, stems, and roots in salt tolerant near isogenic lines NILs25-T and NILs25-S and their graft hybrids NILs25-T/NILs25-S and NILs25-S/NILs25-T

Chapter 2:

MAPPING AND CONFIRMATION OF LOCI FOR SALT TOLERANCE IN A NOVEL SOYBEAN GERMPLASM, FISKEBY III

(This chapter was published with citation: Do TD, Vuong TD, Dunn D, Smothers S, Patil G, Yungbluth DC, Chen P, Scaboo A, Xu D, Carter TE, Nguyen HT, Grover Shannon J (2018) Mapping and confirmation of loci for salt tolerance in a novel soybean germplasm, Fiskeby III. Theor Appl Genet 131:513-524)

Abstract

Breeding soybean for tolerance to high salt conditions is important in some regions of the USA and world. Soybean cultivar Fiskeby III (PI 438471) in maturity group 000 has been reported to be highly tolerant to multiple abiotic stress conditions, including salinity. In this study, a mapping population of 132 F_2 families derived from a cross of cultivar Williams 82 (PI 518671, moderately salt sensitive) and Fiskeby III (salt tolerant) were analyzed to map salt tolerance genes. The evaluation for salt tolerance was performed by analyzing leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC) after treatment with 120 mM NaCl under greenhouse conditions. Genotypic data for the F_2 population was obtained using the Illumina Infinium SoySNP6K BeadChip assay. A major allele from Fiskeby III significantly associated with LSS, CCR, LSC, and LCC on chromosome (Chr.) 03 with LOD scores of 19.1, 11.0, 7.7 and 25.6, respectively. In addition, a second locus associated with salt tolerance for LSC was detected and mapped on Chr. 13 with a LOD score of 4.6 and an R^2 of 0.115. Three gene-based polymorphic molecular markers (Salt-20, Salt14056 and Salt11655) on Chr.03 showed a strong predictive association with phenotypic salt

tolerance in the present mapping population. These molecular markers will be useful for marker-assisted selection to improve salt tolerance in soybean.

Introduction

Salinity is an important abiotic stress factor which negatively impacts 60 million ha or approximately 20% of the total irrigated agricultural land area in the world (FAO and ITPS 2015; Squires and Glenn 2011), global crop losses are estimated at 27.3 billion US dollars, annually (Qadir et al. 2014). Salinity stress in agriculture results from application of irrigation water, which cause yield reductions of crops. The severity of the problem appears to be on the rise in agriculture, with affected farmland increased by an estimated 0.3 to 1.5 million ha annually (FAO and ITPS 2015). The trend toward increasing abiotic stress problems, including salinity stress, may be exacerbated in the future as a result of global climate change, leading to increasing environmental concern (Batlle-Sales 2011; Várallyay 2010).

In soybean, salt build-up in the soil can severely affect germination, seedling emergence, vegetative growth and development, as well as final seed yield and quality at maturity, resulting in modest to total crop loss (Blanco et al. 2007; Phang et al. 2008; Wang and Shannon 1999). In addition, salt stress reduces efficiency of nitrogen fixation and decreases the number and biomass of root nodules (Delgado et al. 1994; Elsheikh and Wood 1995; Singleton and Bohlool 1984). Salt stress may also decreases protein, free amino acids, sucrose, and starch content in mature soybean seed (El-Sabagh et al. 2015; Rabie and Kumazawa 1988). Visual symptoms of salt stress include leaf chlorosis and eventually a necrotic bleached appearance known as leaf scorching (Fig. 2.1).

Not all soybean cultivars are equally sensitive to salt stress. Although most cultivars will accumulate excess salt in a soil or nutrient solution medium, leading to toxicity, others are able to effectively exclude Cl^- accumulation from the shoot. Previous genetic studies reported a dominant gene for salt exclusion or salt tolerance in soybean. Analysis of eight soybean populations subjected to sodium chloride-induced salt stress revealed that F_2 progenies segregated for Cl^- accumulation in a ratio of 3 excluders (tolerant by excluding chloride ions from the stems and leaves) to 1 includer (sensitive in which chloride ions are transported to the stems and leaves resulting in injury) (Abel 1969; Abel and MacKenzie 1964). In addition, progenies of BC_1F_1 crosses segregated in a ratio of 1 excluder to 1 includer, consistent with a single gene hypothesis (Abel 1969). Leaf necrosis and Cl^- accumulation ratings were highly correlated in these studies and appeared pleotropic.

Using salt-induced leaf chlorosis or leaf scorching as a measure of salt tolerance in soybean, Lee et al. (2004) detected and mapped a major locus on chromosome (Chr.) 03, which explained 41% and 60% of the total phenotypic variation under greenhouse and field conditions, respectively. Subsequently, a similar major locus for salt tolerance was also identified and mapped on Chr. 03 in three additional mapping populations (Hamwieh et al. 2011; Hamwieh and Xu 2008). The locus for salt tolerance on Chr. 03 was also confirmed in a recombinant inbred line (RIL) population derived from a wild soybean [*Glycine soja* Siebold & Zucc.], accession, PI 483463, (Lee et al. 2009). A gene at this major locus, *Glyma03g32900*, was identified and cloned (Do et al. 2016; Guan et al. 2014; Patil et al. 2016; Qi et al. 2014), the functionality of this gene in controlling salt tolerance-related traits, such as leaf sodium (Na) and chloride (Cl) content still warrants further investigations.

Soybean cultivar Fiskeby III (PI 438471), originating from Sweden, has been reported to be highly or partially tolerant to drought, iron deficiency chlorosis, aluminum toxicity, salt stress, and atmospheric ozone pollution (Burton et al. 2016; Pathan et al. 2007; USDA 2011). In previous studies, Fiskeby III exhibited higher tolerance to salt stress than other salt tolerant soybean genotypes under greenhouse conditions (Lenis et al. 2011; Patil et al. 2016). The objectives of this study were to identify and map genomic location(s) for salt tolerance in cultivar Fiskeby III, and to suggest DNA markers for marker-assisted selection (MAS) for the improvement of salt tolerance in soybean. As a part of the second objective, we assessed efficacy of three developed gene-based polymorphic molecular markers (Salt-20, Salt14056 and Salt11655) in predicting phenotypic salt tolerance in the Williams 82 x Fiskeby III population.

Materials and Methods

Plant materials

Soybean cultivar Fiskeby III (PI 438471), from Sweden, is highly tolerant to salt stress (Lenis et al. 2011; Patil et al. 2016). Cultivar Williams 82 (PI 518671) is salt sensitive (Do et al. 2016; Patil et al. 2016). However, it is less sensitive to salt than cultivar Hutcheson (PI 518664, Supplementary Table S2.1) that was included as a standard sensitive check in our study.

A cross of Williams 82 (moderately sensitive) and Fiskeby III (tolerant) was accomplished at the University of Missouri, Columbia, MO, in the summer of 2012. The true F_1 hybrids were advanced to the F_2 generation in the winter nursery in Costa Rica in the winter of 2012. In summer of 2014, the F_2 seeds of the population were planted at the

Bay Farm Research Facility, Columbia, MO. Leaf tissue of F_2 seedlings was collected for DNA extraction and genotyping and individual F_2 plants were harvested separately to obtain $F_{2:3}$ families. Subsequently, $F_{2:3}$ families were utilized to screen for salt tolerance in a greenhouse. One hundred thirty-two $F_{2:3}$ families from this population were used to detect and map genomic location(s) associated with salt tolerance.

DNA extraction and marker analysis

DNA was extracted from each F_2 plant using a standard CTAB method (Doyle and Doyle 1987). Briefly, extraction buffer containing 2% CTAB, 0.1 M Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl was added to the samples. The DNA pellet was dissolved in 150 µl of 1xTE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA concentration was quantified with a spectrophotometer (NanoDrop Technologies Inc., Centreville, DE, USA) and was normalized at 50 ng/µl for marker genotyping.

Single nucleotide polymorphism (SNP) genotyping was performed at Washington University in St. Louis by using the SoySNP6K Illumina Infinium BeadChips (Illumina, Inc. San Diego, CA). The assay consisted of a series of standard protocols, such as incubation, DNA amplification, hybridization of samples to the bead assay, extension, and imaging of the bead assay. The SNP alleles were called using the GenomeStudio Genotyping Module (Akond et al. 2013; Song et al. 2013).

Genotyping of gene-based molecular markers

A previously identified gene sequence on Chr. 3 controlling salt tolerance in soybean (*Glyma03g32900)* was analyzed reference lines such as Hutcheson, Lee and Holladay including the parents in this study (Do et al. 2016; Guan et al. 2014; Qi et al.

2014; Valliyodan et al. 2016). Selected SNPs were used to develop Kompetitive Allelic Specific PCR (KASPar) assays (http://www.lgcgroup.com). Three KASPar assays were used to genotype the mapping population, including Salt-20 (M1), Salt14056 (M2), and Salt11655 (M6) at the University of Missouri, Columbia, MO that were designed on SNPs of promotor intron 3 and exon 5, respectively (Patil et al. 2016). The reaction mixture was prepared according to the protocol described by LGC Genomics, LLC (http://www.lgcgroup.com). Briefly, KASPar assays were run in a 10 µl final reaction volume containing a 5 µl KASPar master mix, 0.14 µl primer mix, 2 µl of 10–20 ng/µl genomic DNA, and 2.86 µl water. The following cycling conditions were used: 15 min at 95 °C, followed by 10 touchdown cycles of 20 s at 94 °C, 1 min at 61–55 °C (dropping 0.6 °C per cycle); and then 26 cycles of 20 s at 94 °C, 1 min at 55 °C. The fluorescent end-point genotyping method was carried out using a Roche LightCycler 480-Instrument II (Roche Applied Sciences, Indianapolis, IN, USA) as described (Patil et al. 2016).

Salt tolerance phenotyping

Leaf scorch score (LSS) and chlorophyll content ratio (CCR)

Phenotypic evaluation of the $F_{2:3}$ population for salt tolerance was conducted using the plastic cone-tainer method as previously described by Lee et al. (2008). Seven seedlings of each $F_{2:3}$ family, were grown per cone-tainer and evaluated in two replications over time in a greenhouse at the University of Missouri, Columbia, MO, using artificial lights and a 13 h photoperiod in 2015. Three soybean genotypes, Hutcheson (highly sensitive check), Fiskeby III (tolerant parent), and Williams 82 (moderately sensitive parent), were included in all phenotypic assays to ensure the consistency of the experimental conditions. At the vegetative growth stage V2 (Fehr et al. 1971), soybean seedlings in cone-tainers were

treated with salt water (120 mM NaCl) such that the salt solution filled the bottom one-third (7 cm) of each cone-tainer for 7 h/day. Electrical conductivity (EC) of salt water was monitored daily (Hamwieh et al. 2011). Individual soybean plants were visually rated for salt sensitivity or tolerance when Hutcheson showed severe leaf scorch (approximately 2 weeks after the treatment imposition). Leaf scorch was scored using a 1–5 scale, where 1 = no apparent chlorosis; 2 = slight (25% of the leaves showed chlorosis); 3 = moderate (50% of the leaves showed chlorosis and some necrosis); 4 = severe chlorosis (75% of the leaves showed chlorosis and severe necrosis); and 5 = dead (leaves showed severe necrosis and were withered) for each plant and recorded as mean for each cone-tainer as previously described (Lee et al. 2008).

Leaf chlorophyll content was quantified on individual plants of each $F_{2:3}$ family, the parents and the Hutcheson check for the topmost fully expanded leaf 1 day before and about 14 days after the initiation of the salt treatment using a chlorophyll meter (Chlorophyll meter SPAD-502, Konica Minolta). Chlorophyll content ratio (CCR) was calculated as the ratio of leaf chlorophyll content after treatment dividing leaf chlorophyll content before treatment is an indicator of as an indirect measure of the integrity and maintenance of the photosynthetic apparatus in response to stress (Ghassemi-Golezani et al. 2011; Patil et al. 2016; Weisany et al. 2011).

Leaf sodium (LSC) and chloride (LCC) content analysis

After leaf scorch and chlorophyll determinations, leaves were harvested and dried at 60^0C for 7 days. The dried leaf tissue of seven seedlings from each $F_{2:3}$ family was ground using a Thomas Model ED-5 laboratory Wiley mill (Thomas Scientific, Swedesboro, NJ, USA) and then analyzed for leaf chloride (LCC) and leaf sodium (LSC) contents at the Delta

Research Center, University of Missouri, Portageville, MO as previously described (Lenis et al. 2011). Briefly, the LCC assay was accomplished by dissolving ground leaf tissue (0.15g) in 30 ml of distilled water and agitating on an Eberbach Corporation orbital shaker (Eberbach Corporation, Ann Arbor, MI, USA) at 60 cycles per min for 1 h. Standards for calibration of 25, 50, 100 and 500 mg kg^{-1} of chloride (Cl^-) were made using Ricca Chemical Company's Primary Cl^- solution of 1000 mg kg^{-1} (Arlington, TX, USA). A standard curve was established using an ion specific electrode attached to a Fisher Scientific AR 50 dual channel pH, ion, conductivity-meter (Fischer Scientific, Pittsburg, PA, USA). After standard reference curves were established, the Cl^- in solution extracted from samples of leaves was determined for $F_{2:3}$ lines. The Cl^- quantity (LCC) in the solution was converted to Cl^- concentration by multiplying the mg kg^{-1} chloride in solution by volume of distilled water and dividing by weight of the plant sample.

Ground, dry leaf tissue (0.25 g) was used to determine leaf sodium content (LSC) by means of a modified wet acid dilution procedure with a Hach Digesdahl™ Digestion Apparatus, 115VAC, 50/60 Hz (Hach Company, Loveland, CO, USA) using H_2SO_4 and H_2O_2. Tissue concentrations of Na^+ was determined using a Perkin-Elmer™ (Wellesley, MA, USA) atomic absorption spectrophotometer as previously described (Lenis et al. 2011).

Data Analysis

The experimental design was a randomized block employing two replications. Phenotypic variation of LSS, CCR, LSC, and LCC was analyzed using ANOVA procedure and Duncan's multiple range test in IBM® SPSS® Statistics Version 20 (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM

Corp.). The frequency distribution plots for the $F_{2:3}$ families plots were developed for the four measures of salt tolerance by using Minitab 17 software (Minitab, Inc., 1829 Pine Hall Road State College, PA 16801 USA).

The midpoint of leaf scorch score (LSS) scale (2.5) was used as the cut-off to classify salt tolerance of F_2 populations into tolerant and sensitive groups. The results indicated that segregation of LSS for evaluating salt tolerance fitted a 3:1 (tolerant:sensitive) ratio (Hamwieh and Xu 2008; Lee et al. 2009). It is easier to classify salt tolerance of F_2 populations from tolerant and sensitive parents but more difficult to classify traits of F_2 populations from a cross of tolerant and moderately sensitive parents as used in the present study and unable to use for other traits. Therefore, $F_{2:3}$ families were classified into tolerant, intermediate and sensitive categories based on parental means plus/minus standard deviations as a cut-off. Phenotypic segregation of a 1:2:1 (Tolerant:Intermediate:Sensitive) ratio for all traits in $F_{2:3}$ families was tested using Chi-square analysis. Similarly, segregation of gene-based markers (GBM) in the F_2 population was also evaluated for a goodness-of-fit to a 1:2:1 ratio.

A genetic linkage map was constructed in the F_2 population using the Kosambi mapping function described in the JoinMap 4.0 software (van Ooijen 2006). A likelihood of odds (LOD) threshold score of 3.0 and a maximum genetic distance of 50 cM were utilized for an initial linkage grouping of markers. For genetic analysis, a comprehensive analysis approach, including initial interval mapping (IM), cofactor selection, genome-wide permutation test, and multi-QTL method (MQM) was performed using the MapQTL5.0 software (van Ooijen 2004). The significant threshold of the LOD score was calculated by permutation test with a large set of 1000 iterations. The epistatic interaction between genomic

regions significantly associated with salt tolerances was analyzed using a mixed-model described in the QTLnetwork 2.1 program (Yang et al. 2008).

Results

Phenotypic variance of salt tolerance

The Hutcheson check was the most salt sensitive genotype in the study (Table 2.1). In contrast, Fiskeby III exhibited salt tolerant superior to that of Williams 82 or Hutcheson, based on means of the four measures of salt tolerance employed in the study (Table 2.1; Fig. 2.1). The two replications were highly correlated for $F_{2:3}$ families and control genotypes (r >0.80 for all traits), indicating a high degree of repeatability in the phenotypic data. The heritability in four salt tolerant traits showed the highest value for CCR and the lowest values for LSC (Table 2.1). In addition, the genotypic means of the four measures of salt tolerance were highly correlated (r at least 0.40, *P<0.01*), with the highest correlation between the two traits involving color, LSS and CCR (r= 0.78, Table S2.2).

The range for LSS, CCR, LSC, and LCC among the $F_{2:3}$ families showed that some $F_{2:3}$ families had numerically higher tolerance than the salt tolerant parent, Fiskeby III (Table 2.1; Fig. 2.2). All traits showed a discontinuous distribution, with salt tolerant lines (low LSS, high CCR, low LSC and low LCC) being predominant and skewed toward salt tolerance. Phenotypes of these traits were also classified into salt tolerant, intermediate, and salt sensitive groups (Table 2.2). Based on Chi-square analyses, the phenotypic segregations well fit a F_2 ratio of 1 tolerant: 2 intermediate: 1 sensitive for LSS (χ^2=2.28 and *Pr*=0.68), CCR (χ^2=3.32 and *Pr*=0.79), and LCC (χ^2=2.69 and *Pr*=0.74), indicating that these three traits were controlled by a single gene. The LSC trait fit a F_2 ratio of 9

tolerant: 6 intermediate: 1 sensitive ratio with χ^2=3.21 and *Pr*=0.79, suggesting that this trait was controlled by two genes (Table 2.2).

Genetic linkage map

A total of 2,158 polymorphic SNP makers between the two parents, Williams 82 and Fiskeby III, were used to construct a genetic linkage map in the F_2 population. Of these 2,148 SNP loci (99.5%) were successfully mapped on the 20 linkage groups that covered 2,834 cM of the whole soybean genome and corresponded to 20 soybean chromosomes (Supplementary Fig. S2.1). The smallest number of SNP markers (71) was on Chr. 10 and the largest number (149) on Chr. 13. The average distance between two adjacent SNP markers was 1.38 cM across all chromosomes. However, largest gaps (around 40 cM) were near the center of Chrs. 10 and 11 (Table 2.3; Supplementary Fig. S2.1).

Loci associated with salt tolerance

The significant thresholds of the LOD values determined by the genome-wide permutation tests for each trait, LSS, CCR, LSC and LCC, were 2.7, 2.6, 2.4 and 2.7, respectively. The genomic region significantly associated with LSS, CCR, LSC, and LCC was identified with high LOD and high R^2 values at the same position on Chr. 03. The closest marker with the highest LOD value and R^2 for means of LSS, CCR and LSC was Gm03_40727780 (ss715585963), located at the position 38716240 (Table 2.4; Fig. 2.5). A new genomic region associated with LSC was also identified with high LOD value (4.56) and R^2 (11.5%) on Chr. 13 (Table 2.4; Fig. 2.3). The closest marker with highest LOD value, Gm13_38988256 (ss715616164) was located at the position 40167119 of the soybean physical map. No significant epistatic interactions (either additive x additive or

additive x dominant) were detected among two loci for LSC mapped on Chrs. 03 and 13 (data not shown).

Evaluation of gene-based molecular markers in F_2 population

Results of genotyping 132 F_2 plants with gene-based molecular markers (GBM) showed a clear differentiation between homozygous and heterozygous genotypes (Fig. 2.4). Because recombination between GBMs was not found in the 132 plants of the F_2 population, symbolic genotypes, AA, aa, and Aa, (A allele from Fiskeby III and an allele from Williams 82) were labeled Fiskeby III, Williams 82 and heterozygote for all GBMs. The segregation of GBMs fit a 1:2:1 ratio with χ^2=3.32 and *P=0.81* (Table 2.5).

Phenotypic data was grouped by AA, Aa, and aa categories of GBM, resulting in a significant association between GBMs markers and the four phenotypic measures of salt tolerance (Table 2.5). The AA and Aa genotypes with lower LSS, higher CCR, lower LSC and lower LCC showed higher salt tolerance comparing to aa genotype. Heterozygotes clearly showed dominance for tolerance in all four measures of salt tolerance. In addition, the results were more highly significant when GBMs (Salt-20, Salt14056 and Salt11655) that located in significant intervals containing the locus for salt tolerance with high LOD and R^2 values (Table 2.4; Fig. 2.5).

Discussion

By classifying $F_{2:3}$ families into tolerant, intermediate, and sensitive groups based on parental means, standard deviations and using the Chi-square analysis, the phenotypic segregation of LSS, CCR and LCC showed a good fit a 1:2:1 ratio while LSC best fit a 9:6:1 ratio (Table 2.2). These results suggested LSS, CCR and LCC were controlled by a

single gene while LSC was controlled by two genes. Similarly, a major locus was identified for LSS, CCR and LCC by gene mapping while two loci were mapped for LSC.

The quality and marker resolution of a genetic linkage map affects the accuracy of quantitative trait loci (QTL) mapping and the identification of candidate genes. Increasing marker density is one way to obtain higher resolution of genetic maps (Gutierrez-Gonzalez et al. 2011; Li et al. 2014). Recently, the development of high-throughput genotyping assays, such as the SoySNP6K Illumina Infinium BeadChip and SoySNP50K BeadChip provided a powerful tool for constructing high-resolution linkage maps in soybean (Akond et al. 2013; Song et al. 2013). In this study, the F_2 population was genotyped using the SoySNP6K Illumina Infinium BeadChip assays to obtain 2,158 polymorphic SNP makers. A high resolution genetic map with the average distance of 1.38 cM between two markers was constructed and used for mapping salt tolerance genes. The mapped locus for salt tolerance on Chr. 03 with high peak and a narrow marker interval reflected the efficiency of using a high-resolution map. In addition, this added precision may have allowed the detection of a new putative locus for salt tolerance on Chr.13, based on leaf analysis of shoot sodium content.

Leaf scorching, a visual trait, has been used as a main assessment for salt tolerance in many studies (Hamwieh and Xu 2008; Lee et al. 2004). Decreased chlorophyll content under salt stress has also been useful for evaluating salt tolerance (Ghassemi-Golezani et al. 2011; Hamwieh et al. 2011; Lenis et al. 2011; Patil et al. 2016). However, many non-genetic factors can generate ‘noise’ in these traits, such as variation within an experiment for light intensity, nutrient supply, air temperature, air movement, and the presence of heavy metals and alkaline salt in the medium (Hu et al. 2014; Kumar Tewari and Charan

Tripathy 1998; Resurreccion et al. 2002; Tuyen et al. 2010; Zhao et al. 2005). Thus, other traits such as leaf or shoot sodium and chloride content have been used to assess salt tolerance more directly for genetic mapping in crop plants, such as in rice (Bonilla et al. 2002; Hossain et al. 2015; Koyama et al. 2001; Lin et al. 2004; Qiu et al. 2015), wheat (Genc et al. 2010; Lindsay et al. 2004), and barley (Nguyen et al. 2013; Xue et al. 2009). Leaf sodium and chloride content are highly correlated with leaf scorching and chlorophyll content as measures of salt tolerance in soybean (Do et al. 2016; Lenis et al. 2011), and are used here in genetic mapping study for the first time. The results suggest that the major gene on Chr. 03 may control salt tolerance as measured by all four measures of tolerance.

Earlier genetic studies and QTL mapping in soybean suggested that salt tolerance is controlled by a single dominant gene (*Ncl*) by measuring chloride exclusion and leaf scorching and a major QTL is mapped to Chr. 03 (Abel 1969; Abel and MacKenzie 1964; Lee et al. 2004). The major QTL on Chr. 3 for salt tolerance was identified and confirmed in mapping populations derived from different salt tolerant sources (Ha et al. 2013; Hamwieh et al. 2011; Hamwieh and Xu 2008; Qi et al. 2014). Recently, a candidate gene, *Glyma03g32900* underlying salt tolerance was isolated and the gene function was related to a sodium transporter (Guan et al. 2014; Qi et al. 2014). Results of a separate study found that this gene controls both sodium and chloride content based on analyses of gene expression and testing near-isogenic lines (NILs) under field conditions (Do et al. 2016; Liu et al. 2016b). In this study, the gene *Glyma03g32900* controlling leaf sodium and chloride content was verified by a high correlation between gene-based markers (GBM) and the two traits, LSC and LCC. Further, the dual functions of this gene (*Glyma03g32900*) controlling two traits (LSC and LCC) should be considered for future studies. In addition,

a classical genetic analysis was also performed for LSC by using flanking markers (Gm13_37204738 and Gm13_38988256) of the putative locus on Chr. 13 and LSC was controlled by another dominant gene on Chr. 13 (data not shown).

Although several studies detected and consistently mapped the major locus for salt tolerance on Chr. 03, some evidence showed that there are additional genes underlying salt tolerance in the soybean genome. For instance, eight putative QTL significantly associated with salt tolerance under greenhouse and field conditions were reported (Chen et al. 2008). Comparison of salt tolerance performance of three sets of NILs suggested that there might be another gene that affects or interacts with the salt tolerant gene on Chr. 03 (Hamwieh et al. 2011). By association mapping of soybean seed germination under salt stress, five candidate genes located on Chrs. 08, 09, and 19 were verified in response to salt stress (Kan et al. 2015). Four soybean accessions were reported to be new sources for novel determinants of salt tolerance in genes other than the cloned gene, *Glyma03g32900* (Guan et al. 2014). Correlations between gene-based markers and phenotype showed a new soybean accession might carry novel gene(s) for salt tolerance (Patil et al. 2016). A putative locus associated with LSC mapped on Chr. 13 in the present study appears to be a novel locus that may harbor other genes underlying salt tolerance in soybean. However, a favorable allele for low LSC was from the moderately sensitive parent, Williams 82. Some $F_{2:3}$ families showed higher tolerance than salt tolerant parent (Fiskeby III) suggesting that elevated salt tolerance may be contributed by favorable alleles derived from both parents. Therefore, a greater level of salt tolerance observed in Fiskeby III might be from a different allele or was not expressed in a cross with Williams 82. Developing NILs from a cross of

Williams 82 and Fiskeby III will be useful for a further investigation of the importance of the gene(s) located on Chr. 13 in providing salt tolerance.

The interval of a genetic map significantly associated with a trait from QTL analysis and fine-mapping provided basic information to predict candidate genes (Bargsten et al. 2014). This process was more efficient with a high-density genetic map, genomic sequence data, and predicted genes in a physical map (Zhang et al. 2016). Sixty-eight candidate genes for alkaline salt tolerance were predicted in a 771.7 kb interval of a physical map on Chr. 17 (Tuyen et al. 2013). For salt tolerance, the genes located in a significant interval were predicted and cloned by map-based cloning (Do et al. 2016; Ha et al. 2013; Qi et al. 2014). Based on the genomic region significantly associated with LSC on Chr. 13, three coding sequences, including *Glyma.13g305700*, *Glyma.13g305800*, and *Glyma.13g305900*, with salt stress response/antifungal function were close to significant markers for LSC (Supplementary Fig. S2.2) (http://www.soybase.org). A further study of these new candidate genes could be useful for salt tolerance improvement in soybean.

The availability and accessibility of whole-genome sequencing data, high-resolution genetic linkage maps, candidate genes, and gene annotations are valuable genomic resources to develop functional genetic markers for genetic and breeding studies in plants (Chin et al. 2010; Galeano et al. 2012; Reinprecht and Pauls 2016). SimpleProbe and TaqMan genotyping assays developed for SNPs and mutant alleles of genes controlling oleic acid and linolenic acid content were successfully applied to genetic analysis and breeding in soybean (Shi et al. 2015). Six KASPar assays developed based on the gene *Glyma03g32900* were used to study variation of gene structural variation and salt tolerance in soybean, in which three SNP markers Salt-20 (M1), Salt14056 (M4) and Salt11655 (M6)

were shown to be highly correlated with salt tolerance phenotypes in 106 diverse soybean lines and an RIL mapping population (Patil et al. 2016). In this study, those markers were also integrated into a genetic linkage map that was employed for genetic mapping of salt tolerance. The segregations of those markers well fitted a 1:2:1 ratio and showed a high correlation with salt tolerance (Table 2.5).

In conclusion, the validation of the gene-based markers in an F_2 mapping population suggests that these markers are useful tools for tracking and selecting the salt tolerant gene on Chr. 03. By analyzing the association with salt tolerant phenotypes, gene-based markers could be used to identify promising soybean lines with novel salt tolerant gene(s). The putative locus for LSC on Chr. 13 suggests the presence of a novel gene(s) controlling salt tolerance and may be useful to stack with the known gene on Chr. 03 for improving salt tolerance in soybean.

Acknowledgements

This research was supported in part by the Missouri Soybean Merchandising Council and the Missouri Agricultural Experiment Station. Mr. Tuyen Do would like to thank Cuu Long Delta Rice Research Institute and the Vietnam Ministry of Agriculture for a graduate student scholarship.

Legends to Figures

Figure 2.1 Comparisons of salt tolerance based on leaf scorch between cultivars Fiskeby III (tolerant parent), Lee (tolerant), Williams 82 (moderately sensitive parent), and Hutcheson (sensitive check) grown under 120 mM NaCl treatment

Figure 2.2 Distribution of leaf scorch score (LSS) (panel A), chlorophyll content ratio (CCR) (panel B), leaf sodium content (LSC) (panel C), and leaf chloride content (LCC) (panel D) of 132 $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III grown under 120 mM NaCl treatment

Figure 2.3 A logarithm of the odds (LOD) plot showing the location of locus for leaf sodium content (LSC) on Chr. 03 (*panel A*) and a putative locus for leaf sodium content (LSC) on Chr. 13 (*panel B*) in $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III

Figure 2.4 Genetic segregations of three gene-based markers, Salt-20 (*panel A*), Salt14056 (*panel B*) and Salt11655 (panel C) in an F_2 population from a cross between Williams 82 and Fiskeby III

Figure 2.5 A logarithm of the odds (LOD) plot showing the location of a gene for leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC) in $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III

Supplementary Figure S2.1 A genetic linkage map was constructed in an F_2 population derived from a cross of Williams 82 and Fiskeby III

Supplementary Figure S2.2 Physical positions of the most significant markers associated with salt tolerance, *Gm13_38988256* (ss715616164), *Gm13_39054715* (ss715616173) and *Gm13_3965528* (ss715616176) and three candidate genes (*Glyma.13g305700*, *Glyma.13g305800* and *Glyma.13g305900*) (http://soybase.org) with salt stress response function in the physical map of Chr. 13

References

Abel GH (1969) Inheritance of the capacity for chloride inclusion and chloride exclusion by soybeans. Crop Sci 9:697-698

Abel GH, MacKenzie AJ (1964) Salt tolerance of soybean varieties (*Glycine max* L. Merill) during germination and later growth. Crop Sci 4:157-161

Akond M, Liu S, Schoener L, Anderson JA, Kantartzi SK, Meksem K, Song Q, Wang D, Wen Z, Lightfoot DA, Kassem MA (2013) A SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium BeadChip genotyping array. J Plant Genome Sci 1:80-89

Bargsten JW, Nap JP, Sanchez-Perez GF, van Dijk AD (2014) Prioritization of candidate genes in QTL regions based on associations between traits and biological processes. BMC Plant Biol 14:330

Batlle-Sales J (2011) Salinization: An environmental concern under climate change scenarios. In: Thomas RP (ed) Proceedings of the Global Forum on Salinization and Climate Change (GFSCC2010), Valencia, 25–29 October 2010. FAO, Rome, p 10

Blanco FF, Folegatti MV, Gheyi HR, Fernandes PD (2007) Emergence and growth of corn and soybean under saline stress. Sci Agr 64:451-459

Bonilla P, Dvorak J, Mackill D, Deal K, Gregorio G (2002) RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (*Oryza sativa* L.) using recombinant inbred lines. Philipp Agric Sci 85:68-76

Burton AL, Burkey KO, Carter TE, Jr., Orf J, Cregan PB (2016) Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III x Mandarin (Ottawa) soybean population. Theor Appl Genet 129:1113-1125

Chen HT, Cui SY, Fu SX, Gai JY, Yu DY (2008) Identification of quantitative trait loci associated with salt tolerance during seedling growth in soybean (*Glycine max* L.). Aust J Agr Res 59:1086-1091

Chin JH, Lu X, Haefele SM, Gamuyao R, Ismail A, Wissuwa M, Heuer S (2010) Development and application of gene-based markers for the major rice QTL Phosphorus uptake 1. Theor Appl Genet 120:1073-1086

Delgado MJ, Ligero F, Lluch C (1994) Effects of salt stress on srowth and sitrogen-sixation by pea, faba-bean, common bean and soybean plants. Soil Biol Biochem 26:371-376

Do TD, Chen H, Hien VT, Hamwieh A, Yamada T, Sato T, Yan Y, Cong H, Shono M, Suenaga K, Xu D (2016) *Ncl* synchronously regulates Na(+), K(+), and Cl(-) in soybean and greatly increases the grain yield in saline field conditions. Sci Rep 6:19147

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15

El-Sabagh A, Sorour S, Ueda A, Saneoka H, Barutcular C (2015) Evaluation of salinity stress effects on seed yield and quality of three soybean cultivars. Azarian J Agric 2:138-141

Elsheikh EAE, Wood M (1995) Nodulation and N-2 fixation by soybean inoculated with salt-tolerant Rhizobia or salt-sensitive *Bradyrhizobia* in saline soil. Soil Biol Biochem 27:657-661

FAO, ITPS (2015) Status of the world's soil resources (SWSR) – main report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy:pp 124-127

Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci 11:929-931

Galeano CH, Cortes AJ, Fernandez AC, Soler A, Franco-Herrera N, Makunde G, Vanderleyden J, Blair MW (2012) Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. BMC Genet 13:48

Genc Y, Oldach K, Verbyla AP, Lott G, Hassan M, Tester M, Wallwork H, McDonald GK (2010) Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress. Theor Appl Genet 121:877-894

Ghassemi-Golezani K, Taifeh-Noori M, Oustan S, Moghaddam M, Rahmani SS (2011) Physiological performance of soybean cultivars under salinity stress. J Plant Physiol Breeding 1:1-8

Guan RX, Qu Y, Guo Y, Yu LL, Liu Y, Jiang JH, Chen JG, Ren YL, Liu GY, Tian L, Jin LG, Liu ZX, Hong HL, Chang RZ, Gilliham M, Qiu LJ (2014) Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. Plant J 80:937-950

Gutierrez-Gonzalez JJ, Vuong TD, Zhong R, Yu O, Lee JD, Shannon G, Ellersieck M, Nguyen HT, Sleper DA (2011) Major locus and other novel additive and epistatic

loci involved in modulation of isoflavone concentration in soybean seeds. Theor Appl Genet 123:1375-1385

Ha BK, Vuong TD, Velusamy V, Nguyen HT, Shannon JG, Lee JD (2013) Genetic mapping of quantitative trait loci conditioning salt tolerance in wild soybean (*Glycine soja*) PI 483463. Euphytica 193:79-88

Hamwieh A, Tuyen DD, Cong H, Benitez ER, Takahashi R, Xu DH (2011) Identification and validation of a major QTL for salt tolerance in soybean. Euphytica 179:451-459

Hamwieh A, Xu DH (2008) Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. Breeding Sci 58:355-359

Hossain H, Rahman MA, Alam MS, Singh RK (2015) Mapping of quantitative trait loci associated with reproductive-stage salt tolerance in rice. J Agron Crop Sci 201:17-31

Hu H, Wang L, Wang Q, Jiao L, Hua W, Zhou Q, Huang X (2014) Photosynthesis, chlorophyll fluorescence characteristics, and chlorophyll content of soybean seedlings under combined stress of bisphenol A and cadmium. Environ Toxicol Chem 33:2455-2462

Kan G, Zhang W, Yang W, Ma D, Zhang D, Hao D, Hu Z, Yu D (2015) Association mapping of soybean seed germination under salt stress. Mol Genet Genomics 290:2147-2162

Koyama ML, Levesley A, Koebner RM, Flowers TJ, Yeo AR (2001) Quantitative trait loci for component physiological traits determining salt tolerance in rice. Plant Physiol 125:406-422

Kumar Tewari A, Charan Tripathy B (1998) Temperature-stress-induced impairment of chlorophyll biosynthetic reactions in cucumber and wheat. Plant Physiol 117:851-858

Lee GJ, Carter TE, Jr., Villagarcia MR, Li Z, Zhou X, Gibbs MO, Boerma HR (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109:1610-1619

Lee JD, Shannon JG, Vuong TD, Nguyen HT (2009) Inheritance of salt tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.) accession PI483463. J Hered 100:798-801

Lee JD, Smothers SL, Dunn D, Villagarcia M, Shumway CR, Carter TE, Shannon JG (2008) Evaluation of a simple method to screen soybean genotypes for salt tolerance. Crop Sci 48:2194-2200

Lenis JM, Ellersieck M, Blevins DG, Sleper DA, Nguyen HT, Dunn D, Lee JD, Shannon JG (2011) Differences in ion accumulation and salt tolerance among *Glycine* accessions. J Agron Crop Sci 197:302-310

Li B, Tian L, Zhang J, Huang L, Han F, Yan S, Wang L, Zheng H, Sun J (2014) Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. BMC Genomics 15:1086

Lin HX, Zhu MZ, Yano M, Gao JP, Liang ZW, Su WA, Hu XH, Ren ZH, Chao DY (2004) QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theor Appl Genet 108:253-260

Lindsay MP, Lagudah ES, Hare RA, Munns R (2004) A locus for sodium exclusion (*Nax1*), a trait for salt tolerance, mapped in durum wheat. Funct Plant Biol 31:1105-1114

Liu Y, Yu L, Qu Y, Chen J, Liu X, Hong H, Liu Z, Chang R, Gilliham M, Qiu L, Guan R (2016) *GmSALT3*, which confers improved soybean salt tolerance in the field, increases leaf Cl- exclusion prior to Na+ exclusion but does not improve early vigor under salinity. Front Plant Sci 7:1485

Nguyen VL, Ribot SA, Dolstra O, Niks RE, Visser RGF, van der Linden CG (2013) Identification of quantitative trait loci for ion homeostasis and salt tolerance in barley (*Hordeum vulgare* L.). Mol Breeding 31:137-152

Pathan MS, Lee J-D, Shannon JG, Nguyen HT (2007) Recent advances in breeding for drought and salt stress tolerance in soybean. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer Netherlands, Dordrecht, pp 739-773

Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, Shannon JG, Nguyen HT (2016) Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci Rep 6:19199

Phang TH, Shao G, Lam HM (2008) Salt tolerance in soybean. J Integr Plant Biol 50:1196-1212

Qadir M, Quillerou E, Nangia V, Murtaza G, Singh M, Thomas RJ, Drechsel P, Noble AD (2014) Economics of salt-induced land degradation and restoration. Nat Resour Forum 38:282-295

Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

Qiu XJ, Yuan ZH, Liu H, Xiang XJ, Yang LW, He WJ, Du B, Ye GY, Xu JL, Xing DY (2015) Identification of salt tolerance-improving quantitative trait loci alleles from a salt-susceptible rice breeding line by introgression breeding. Plant Breeding 134:653-660

Rabie RK, Kumazawa K (1988) Effect of salt stress on nitrogen nutrition and yield quality of nodulated soybeans. Soil Sci Plant Nutr 34:385-391

Reinprecht Y, Pauls KP (2016) Microsomal omega-3 fatty acid desaturase genes in low linolenic acid soybean line RG10 and validation of major linolenic acid QTL. Front Genet 7:38

Resurreccion AP, Makino A, Bennett J, Mae T (2002) Effect of light intensity on the growth and photosynthesis of rice under different sulfur concentrations. Soil Sci Plant Nutr 48:71-77

Shi Z, Bachleda N, Pham AT, Bilyeu K, Shannon G, Nguyen H, Li ZL (2015) High-throughput and functional SNP detection assays for oleic and linolenic acids in soybean. Mol Breeding 35:175-186

Singleton PW, Bohlool BB (1984) Effect of salinity on nodule formation by soybean. Plant Physiol 74:72-76

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8:e54985

Squires VR, Glenn EP (2011) Salination, desertification and soil erosion. In: Squires VR (ed) The role of food, agriculture, forestry and fisheries in human nutrition, vol 3. Encyclopedia of Life Support Systems (EOLSS), Australia, pp 102-123

Tuyen DD, Lal SK, Xu DH (2010) Identification of a major QTL allele from wild soybean (*Glycine soja* Sieb. & Zucc.) for increasing alkaline salt tolerance in soybean. Theor Appl Genet 121:229-236

Tuyen DD, Zhang HM, Xu DH (2013) Validation and high-resolution mapping of a major quantitative trait locus for alkaline salt tolerance in soybean using residual heterozygous line. Mol Breeding 31:79-86

USDA (2011) Breeding Plants for a High-Ozone World. Agr Res 59:14-17

Valliyodan B, Dan Q, Patil G, Zeng P, Huang J, Dai L, Chen C, Li Y, Joshi T, Song L, Vuong TD, Musket TA, Xu D, Shannon JG, Shifeng C, Liu X, Nguyen HT (2016) Landscape of genomic diversity and trait discovery in soybean. Sci Rep 6:23598

van Ooijen (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands

van Ooijen (2004) MapQTL ® 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma BV, Wageningen, Netherlands

Várallyay G (2010) The impact of climate change on soils and on their water management. Agron Res 11:385-396

Wang D, Shannon MC (1999) Emergence and seedling growth of soybean cultivars and maturity groups under salinity. Plant Soil 214:117-124

Weisany W, Sohrabi Y, Heidari G, Siosemardeh A, Ghassemi-Golezani K (2011) Physiological responses of soybean (*Glycine max* L.) to zinc application under salinity stress. Aust J Crop Sci 5:1441-1447

Xue D, Huang Y, Zhang X, Wei K, Westcott S, Li C, Chen M, Zhang G, Lance R (2009) Identification of QTLs associated with salinity tolerance at late growth stage in barley. Euphytica 169:187-196

Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu J (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24:721-723

Zhang D, Li H, Wang J, Zhang H, Hu Z, Chu S, Lv H, Yu D (2016) High-density genetic mapping identifies new major loci for tolerance to low-phosphorus stress in soybean. Front Plant Sci 7:372

Zhao DL, Reddy KR, Kakani VG, Reddy VR (2005) Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. Eur J Agron 22:391-403

Table 2.1 Variation in leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) of the parents and 132 $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III grown under 120 mM NaCl treatment in a greenhouse.

Sources	**LSS**	**CCR**	**LSC** ($g\ kg^{-1}$)	**LCC** $g\ kg^{-1}$)
Hutcheson	5.0	0.55±0.02	0.614±0.21	10.67±0.91
Fiskeby III	1.14±0.10	1.17±0.02	0.153±0.021	5.76±0.41
Williams 82	3.67±0.26	0.67±0.05	0.568±0.038	8.34±0.06
Mean of $F_{2:3}$ families	2.07	0.98	0.177	10.40
Range of $F_{2:3}$ families	1.00-4.20	0.67-1.35	0.005-0.85	3.10-13.6
H^2 ($F_{2:3}$ families)	0.38	0.49	0.12	0.47

H^2: Broad-sense heritability

Table 2.2 Chi-square tests for salt tolerant traits in an $F_{2:3}$ families from a cross between Williams 82 and Fiskeby III

Groups	LSS			CCR			LSC			LCC		
	Obs	**Cut-off**	**Mean**	**Obs**	**Cut-off**	**Mean**	**Obs**	**Cut-off**	**Mean**	**Obs**	**Cut-off**	**Mean**
Tolerant	40	<1.51	1.29	37	>1.07	1.16	69	<0.13	0.07	41	<5.3	4.35
Intermediate	64	1.51-2.7	2.09	71	0.86-1.07	0.96	58	0.13-0.62	0.25	62	5.3-8.3	6.30
Sensitive	28	>2.7	3.25	24	<0.86	0.79	5	>0.62	0.93	29	>8.3	9.97
Size		132			132			132			132	
χ^2 (1:2:1)		2.28			3.32			-			2.69	
χ^2 (9:6:1)		-			-			3.21			-	
Pr		0.68			0.81			0.79			0.74	

Obs: Observed number of $F_{2:3}$ families

Cut-off: Cut-off value based on parental means plus/minus standard deviations

-: No significant fit with expected ratio

Table 2.3 Summary of the genetic map for number of markers, chromosome length, average space between markers, and maximum space between markers in centi-morgans (cM) for each of the 20 chromosomes in the F_2 population derived from a cross between Williams 82 and Fiskeby III with SNP markers

Chromosome	No. of markers	Length of chromosome (cM)	Average space (cM)	Maximum space (cM)
1	107	127	1.20	10.4
2	127	158	1.20	14.7
3	126	130	1.00	16.3
4	122	146	1.21	14.3
5	110	145	1.33	22.8
6	134	165	1.24	8.9
7	88	146	1.69	13.9
8	118	169	1.45	12.0
9	106	150	1.43	11.3
10	71	147	2.09	40.4
11	80	141	1.78	40.0
12	75	114	1.55	13.6
13	149	180	1.22	15.5
14	99	113	1.15	12.1
15	118	142	1.22	24.7
16	96	104	1.10	11.9
17	94	146	1.57	10.5
18	142	139	1.00	10.5
19	104	143	1.39	24.9
20	82	138	1.71	15.6
Overall	2148	2843	1.38	40.4

Table 2.4 Mapping of leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) in an F_2 population from the cross of Williams 82 and Fiskeby III.

Trait	Chr.	Genetic position	LOD	R^2 (%)	Flanking markers	Interval (cM)	Additive effect	Dominance effect	h^2
LSS	Chr.03	86.9	19.08	48.2	Salt-20[a], Gm03_41135466	3.6	0.71**	-0.50**	0.38
CCR	Chr.03	86.9	11.01	31.3	Salt-20[a], Gm03_41135466	3.6	-0.10**	0.08**	0.24
LSC	Chr.03	86.9	7.69	20.6	Salt-20[a], Gm03_41135466	3.6	0.12**	0.03**	0.14
LCC	Chr.03	86.9	25.59	58.9	Salt-20[a], Gm03_41135466	3.6	2.36**	-0.67**	0.56
LSC	Chr.13	143.7	4.56	11.5	Gm13_37204738, Gm13_38988256	10.3	-0.09**	-0.07*	0.10

a: Salt-20 was mapped in the same position with Salt14056, Salt11655, Gm03_40600088, Gm03_40613405 and Gm03_40663609 on Chr.03

**: significant effect at 0.01 probability level

*: significant effect at 0.05 probability level

h^2: Narrow-sense heritability

Table 2.5 Chi-square tests for GBMs in an F_2 population and phenotype of groups for 132 $F_{2:3}$ families from a cross between Williams 82 and Fiskeby III

Phenotype of groups by GBM					Genotypes of GBM		
Groups	**LSS****	**CCR****	**LSC**** (g kg-1)	**LCC**** (g kg-1)	**GBM**	**Observed**	**Expected**
Tolerance	1.69a	1.04a	0.13a	4.69a	AA	42	33
Intermediate	1.91a	1.02a	0.14a	6.37b	Aa	61	66
Sensitive	3.05b	0.84b	0.35b	9.40c	aa	29	33
					Size	**n=132**	
					χ^2 (1:2:1)	**3.32**	
					Pr	***0.81***	

**: The different letters with mean of traits are different groups according to Duncan's multiple range test ($p < 0.01$)

Table S2.1 Variation of leaf scorch score (LSS), chlorophyll content ratio (CCR) of soybean lines grown under 120 mM NaCl treatment from initial testing salt tolerance of parental lines.

No.	Name	PI	LSS*	CRR*	Salt reaction
1	Fiskeby III	PI 438471	1.111 d	1.158 bc	Tolerant
2	Lee	PI 548656	1.190 d	1.318 a	Tolerant
3	Holladay	PI 572239	1.143 d	1.178 abc	Tolerant
4	S05-11482	PI 661090	1.238 d	1.222 abc	Tolerant
5	S10-11227		1.238 d	1.073 cd	Tolerant
6	PI 483463 (soja)	PI 483463	1.381 d	1.072 cd	Tolerant
7	William 82	PI 518671	3.667 b	0.809 efg	Moderately sensitive
8	Hutcheson	PI 518664	4.460 a	0.665 g	Sensitive
9	Magellan	PI 595362	2.921 c	0.851 ef	Moderately tolerant
10	S11-20337		1.333 d	1.085 bcd	Tolerant
11	5913	PI 088788	3.571 b	0.708 fg	Moderately sensitive
12	Peking	PI 548402	2.952 c	0.964 de	Moderately tolerant
14	Er-hei-jan	PI 437654	3.714 b	0.853 ef	Moderately sensitive
15	Pin-din-guan	PI 437690	2.778 c	0.845 ef	Moderately tolerant
16	No. 4	PI 209332	2.929 c	0.828 ef	Moderately tolerant
17	Cloud	PI 548316	1.310 d	1.237 ab	Tolerant
18	Kiio Shokuzu	PI 086006	4.381 a	0.786 fg	Sensitive
	CV(%)		14.8	8.7	-

*: The different letters with mean of traits are different groups according to Duncan's multiple range test ($p < 0.05$). CV: Coefficient of variation

Table S2.2 Pearson correlation coefficients estimated among four traits by analyzing 132 $F_{2:3}$ families from a cross between Williams 82 and Fiskeby III

	LSS	CCR	LSC
CCR	-0.784**		
LSC	0.540**	-0.448**	
LCC	0.627**	-0.396**	0.533**

**: Correlation is significant at the 0.01 level (2-tailed)

Figure 2.1 Comparisons of salt tolerance based on leaf scorch between cultivars Fiskeby III (tolerant parent), Lee (tolerant), Williams 82 (moderately sensitive parent), and Hutcheson (sensitive check) grown under 120 mM NaCl treatment

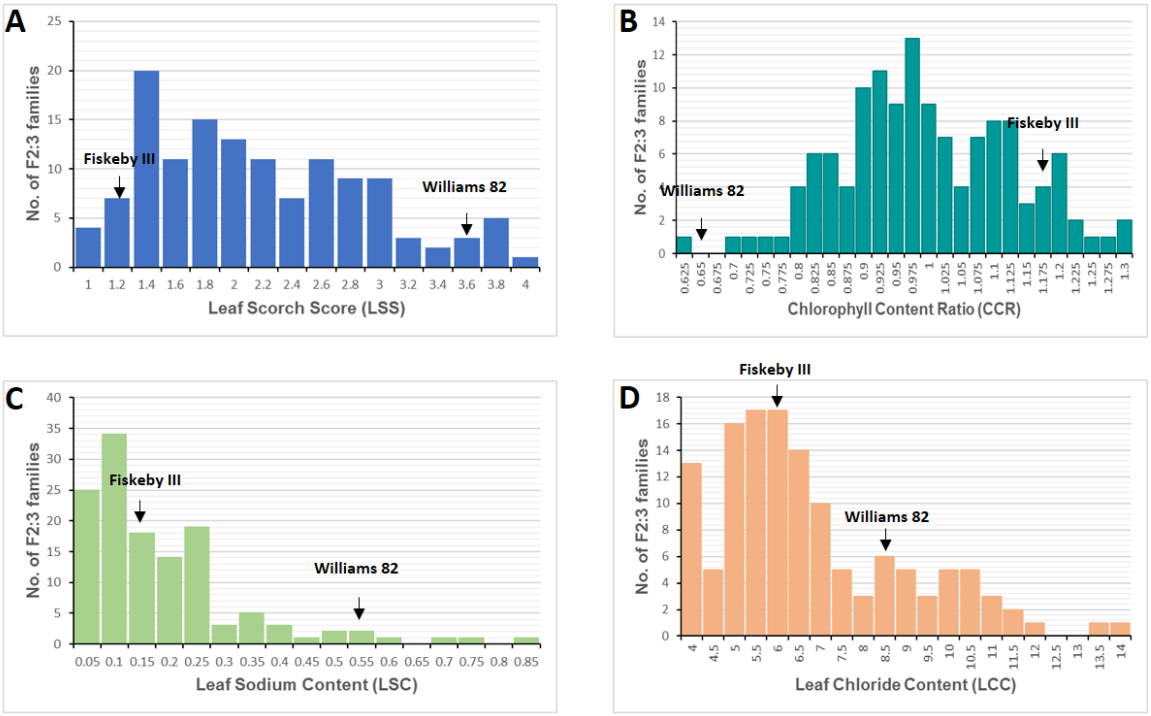


Figure 2.2 Distribution of leaf scorch score (LSS) (*panel A*), chlorophyll content ratio (CCR) (*panel B*), leaf sodium content (LSC) (*panel C*), and leaf chloride content (LCC) (*panel D*) of 132 $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III grown under 120 mM NaCl treatment

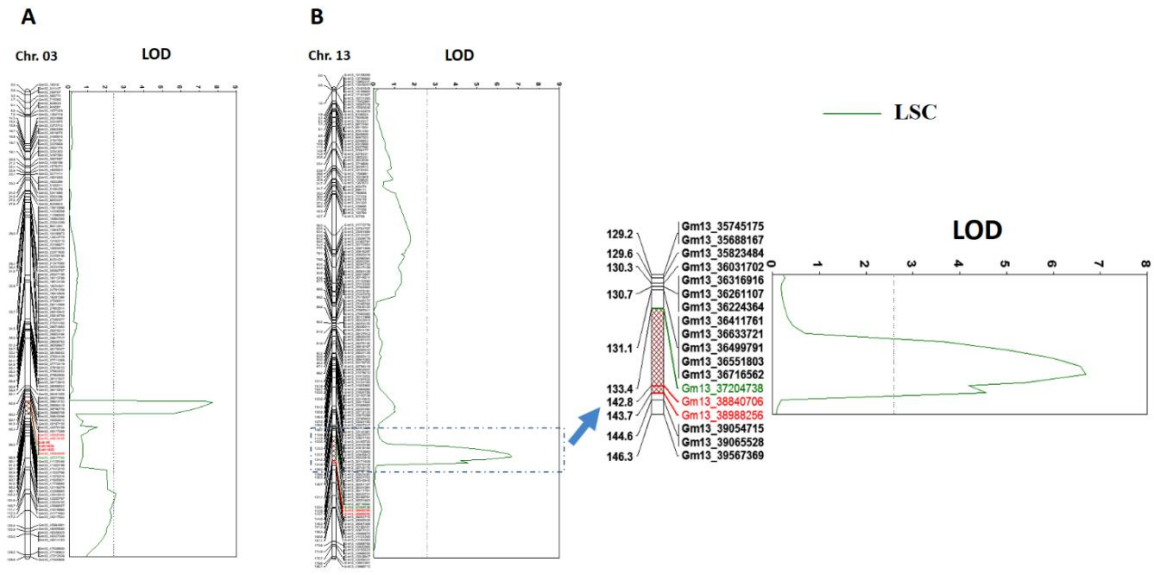


Figure 2.3 A logarithm of the odds (LOD) plot showing the location of locus for leaf sodium content (LSC) on Chr. 03 (*panel A*) and a putative locus for leaf sodium content (LSC) on Chr. 13 (*panel B*) in $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III

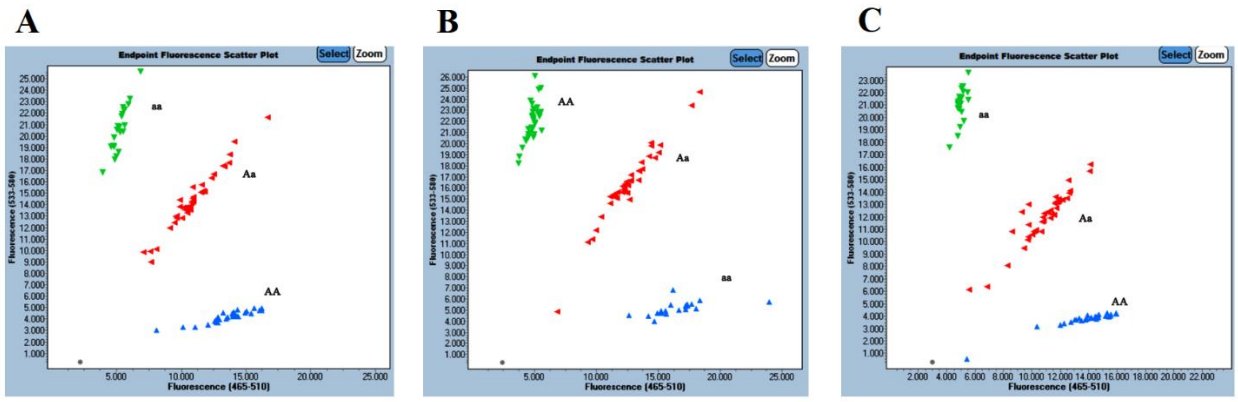


Figure 2.4 Genetic segregations of three gene-based markers, Salt-20 (*panel A*), Salt14056 (*panel B*) and Salt11655 (*panel C*) in an F_2 population from a cross between Williams 82 and Fiskeby III

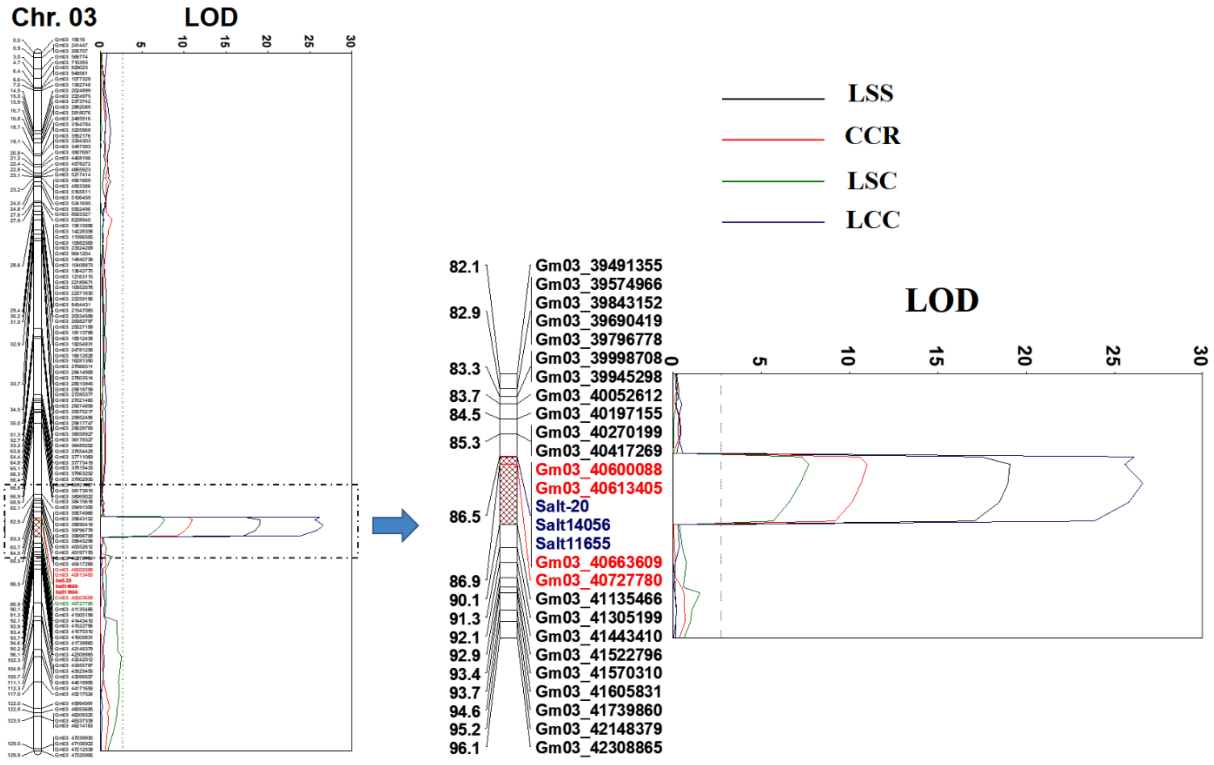


Figure 2.5 A logarithm of the odds (LOD) plot showing the location of a gene for leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC) in $F_{2:3}$ families derived from a cross between Williams 82 and Fiskeby III

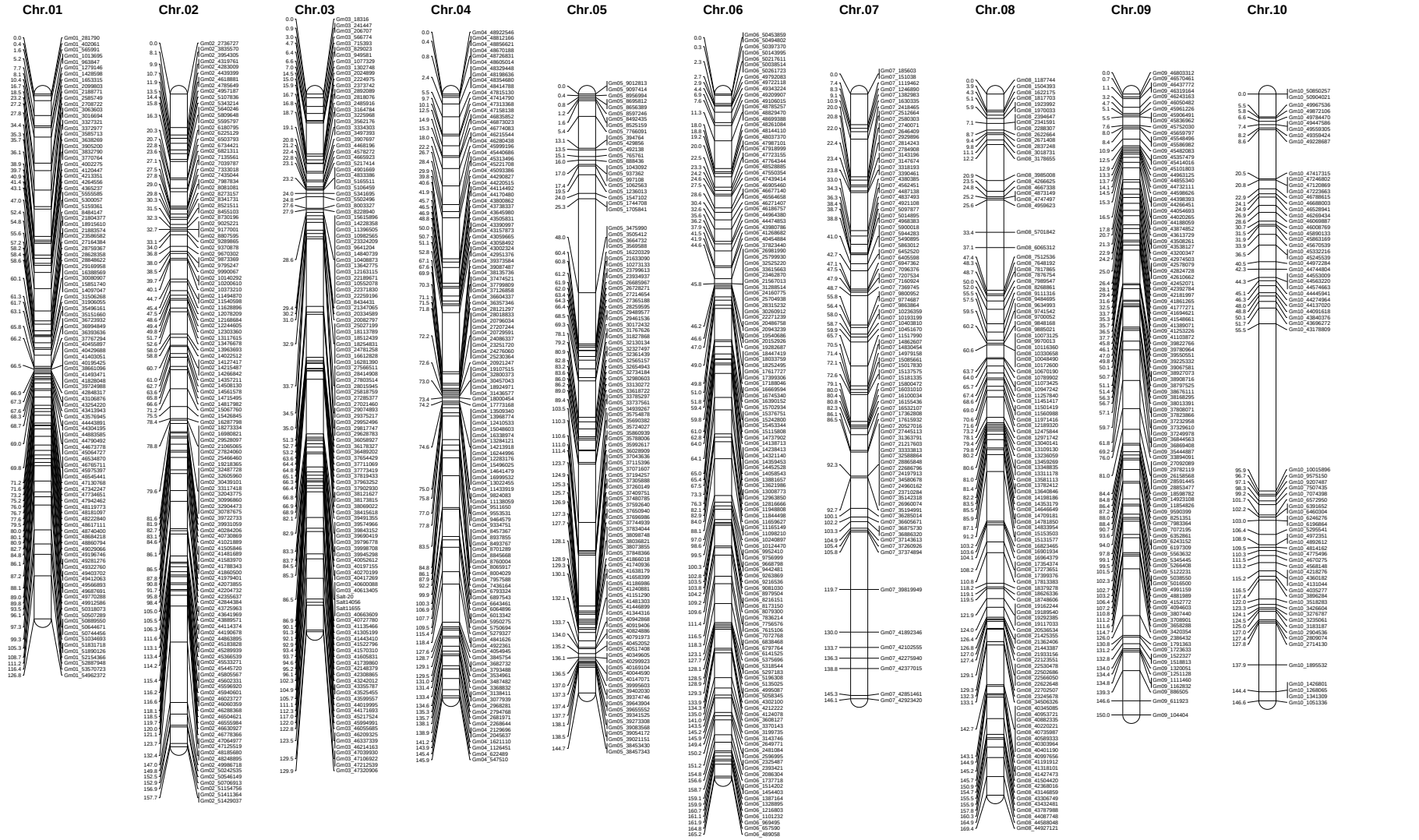


Figure S2.1 A genetic linkage map was constructed in an F_2 population derived from a cross of Williams 82 and Fiskeby III

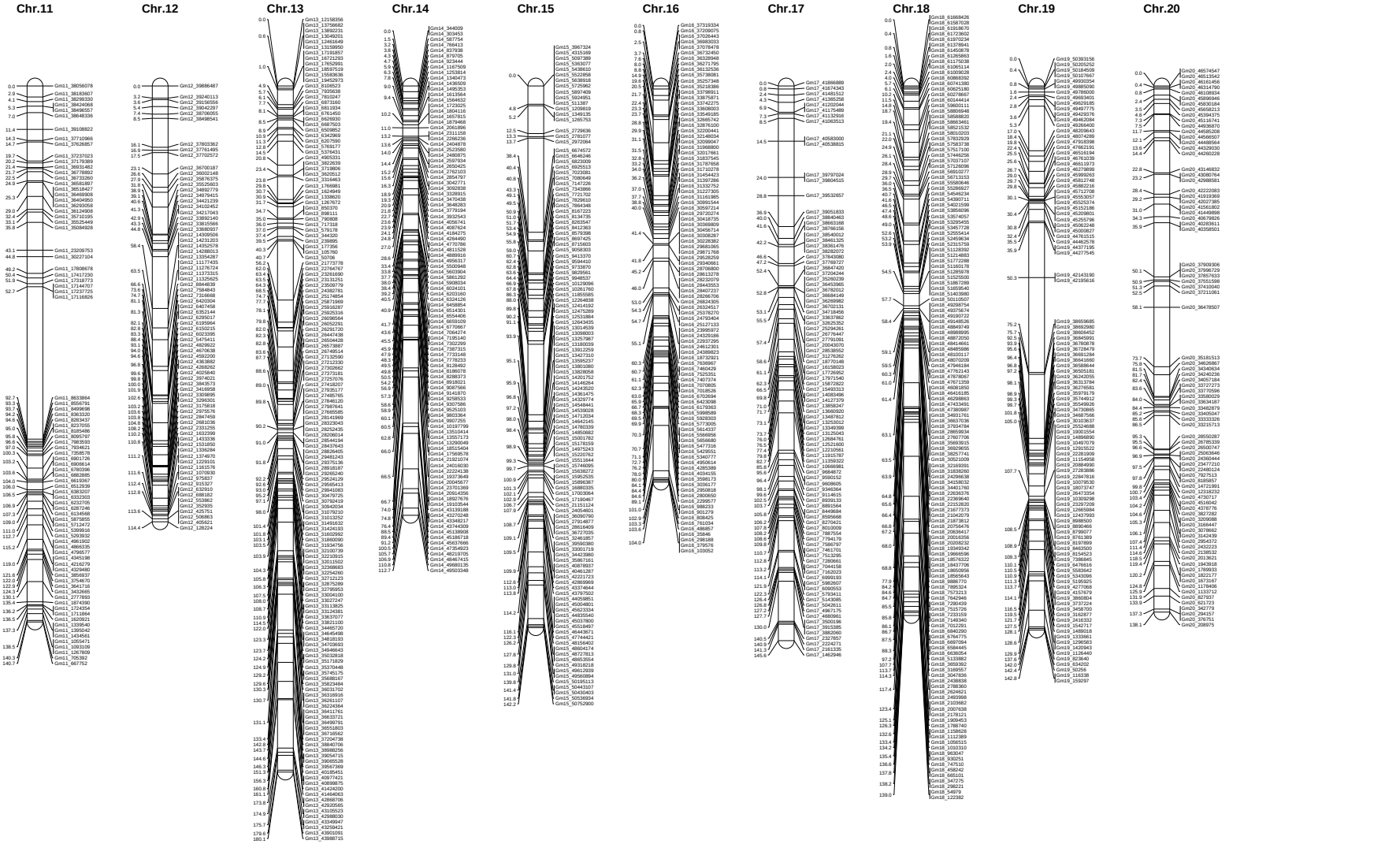


Figure S2.1 A genetic linkage map was constructed in an F_2 population derived from a cross of Williams 82 and Fiskeby III (contd…)

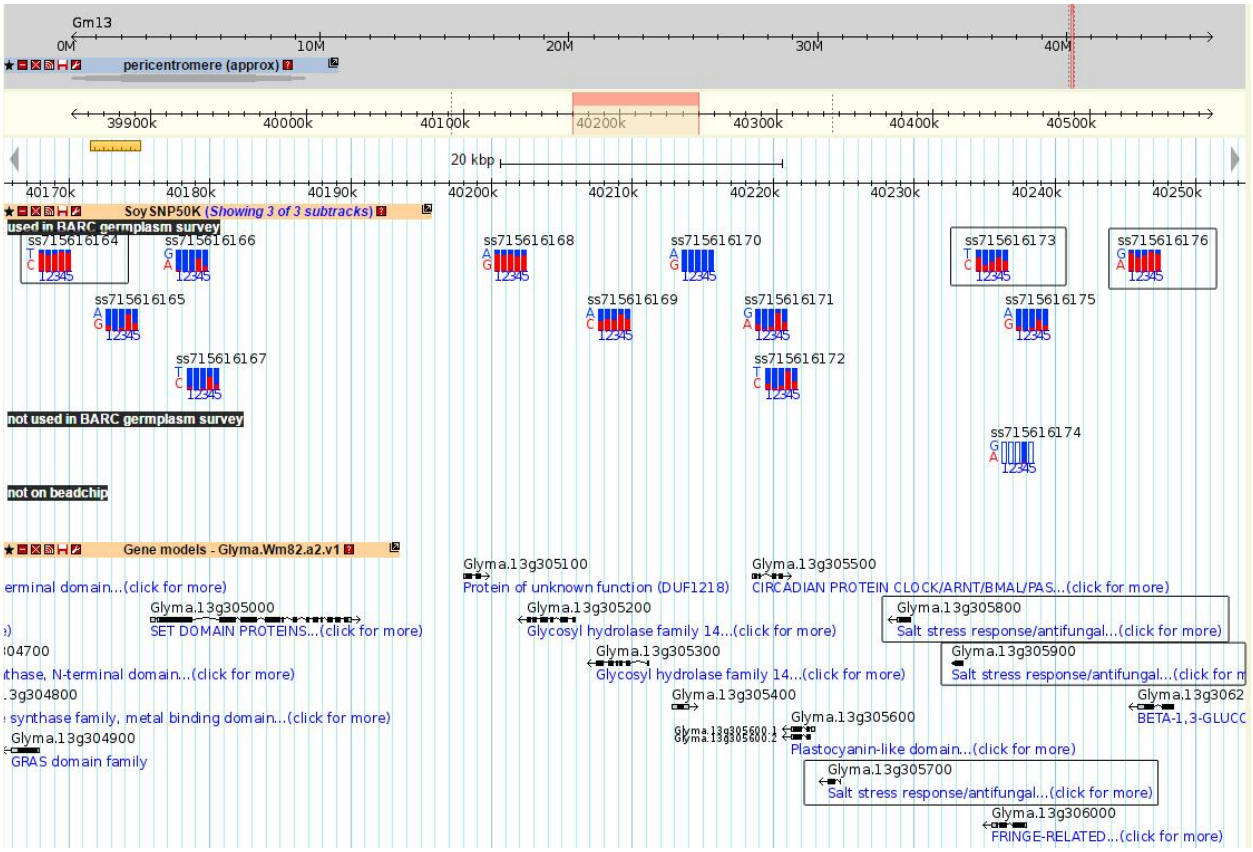


Figure S2.2 Physical positions of the most significant markers associated with salt tolerance, Gm13_38988256 (ss715616164), Gm13_39054715 (ss715616173) and Gm13_3965528 (ss715616176) and three candidate genes (*Glyma.13g305700*, *Glyma.13g305800* and *Glyma.13g305900*) (http://soybase.org) with salt stress response function in the physical map of Chr. 13

Chapter 3:

IDENTIFICATION OF NEW LOCI FOR SALT TOLERANCE IN SOYBEAN BY HIGH-RESOLUTION GENOME-WIDE ASSOCIATION MAPPING

(This chapter is a manuscript draft that is to be submitted to a scientific journal)

Abstract

Salinity is an abiotic stress that negatively affects soybean [*Glycine max* (L.) Merr.] seed yield. Although a major gene for salt tolerance was identified and consistently mapped to chromosome (Chr.) 3 by linkage mapping studies, it does not fully explain genetic variability for salt tolerance in soybean germplasm. In this study, a genome-wide association study (GWAS) was performed to map genomic regions for salt tolerance in a diverse panel of 305 soybean accessions using a single nucleotide polymorphism (SNP) dataset derived from SoySNP50K iSelect BeadChip. A second GWAS was also conducted in a subset of 234 accessions using another dataset of over 3.7M SNPs derived from whole-genome resequencing. In addition, three gene-based markers (GBM) of the known gene, *Glyma03g32900*, on Chr. 3 were also integrated into the two datasets. Salt tolerance among soybean lines was evaluated by leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC). For both association studies, a major locus for salt tolerance on Chr. 3 was confirmed by a number of significant SNPs, of which three gene-based SNP markers, Salt-20, Salt14056 and Salt11655, showed highest significant association with all four traits studied. Also, additional genomic regions were found on Chrs. 1, 8, and 18 from the second study using the 3.7M SNP dataset, in which a region identified on Chr. 8 was predicted as a new minor locus for salt tolerance

in soybean. The candidate genes harbored in this minor locus may help to reveal the molecular mechanism involved in salt tolerance and to improve tolerance in soybean cultivars. The significant SNPs will be useful for marker-assisted selection in soybean breeding programs.

Keywords: soybean, salt tolerance, genome-wide association study, marker-assisted selection.

Introduction

In crop agriculture, salinity is considered a major abiotic stress worldwide. The global extent of salt-affected soils amounts to 1,128 million ha, of which 60% are saline soils, 26% are sodic soils, and 14% are saline-sodic soils (Wicke et al. 2011). Among irrigated agricultural land, an estimated 60 million ha is affected by salinity worldwide (FAO and ITPS 2015), and 0.25–0.5 million ha salt build-up annually causes lost agricultural production (FAO 2002). The United States has large areas of soils with 5.2 million ha (FAO and ITPS 2015) or approximately 23% of the total irrigated land is salt affected. Exploiting plant salinity tolerance has been shown to be among the effective strategies to limit losses from naturally occurring salinity and from the threat of human activities coupled with global climate changes (Batlle-Sales 2011; FAO and ITPS 2015; Várallyay 2010).

Although soybean is classified as moderately salt tolerant crop plant with a threshold of 5 dS/m, salt sensitive soybean cultivars were severely affected under salt stress and did not produce seeds at a soil salinity level of 8 dS/m (Bustingorri and Lavado 2011; Papiernik et al. 2005). Soybean yield losses could result from reduced germination, low

seedling emergence, and poor plant growth and development (Blanco et al. 2007; Essa 2002; Phang et al. 2008; Wang and Shannon 1999). In addition, soybean seed protein, oil and carbohydrate content are negatively affected by salinity (El-Sabagh et al. 2015; Rabie and Kumazawa 1988).

Bi-parental quantitative trait locus (QTL) mapping has been successfully implemented to identify and confirm a major locus on chromosome (Chr.) 03 for salt tolerance in soybean (Do et al. 2018; Guan et al. 2014; Ha et al. 2013; Hamwieh et al. 2011; Hamwieh and Xu 2008; Lee et al. 2004; Qi et al. 2014). The candidate gene (*Glyma03g32900*) underlying salt tolerance was identified and related to a sodium transporter (Do et al. 2016; Guan et al. 2014; Qi et al. 2014). Near-isogenic lines carrying the salt tolerant gene were selected using molecular markers and showed high yield under the saline field conditions (Do et al. 2016; Liu et al. 2016b). Gene-based markers (GBM) were developed for marker-assisted selection (MAS) and also for identifying new tolerance genes (Do et al. 2018; Patil et al. 2016). On the other hand, minor loci for salt tolerance were identified on chromosomes (Chrs.) 2, 7, 9, 11, 13, 14, 15 and 18 (Chen et al. 2008; Do et al. 2018; Zeng et al. 2017b) and other sources may carry new gene(s) for salt tolerance (Guan et al. 2014; Hamwieh et al. 2011; Patil et al. 2016). Nevertheless, the major limitations of bi-parental linkage mapping can detect alleles from parents only and a few recombination events occur in mapping populations (Korte and Farlow 2013).

A Genome-wide association study (GWAS) presents some advantages over linkage mapping that can be applied among individuals in natural populations and exploiting broader genetic diversity (Abdurakhmonov and Abdukarimov 2008; Flint-Garcia et al. 2003; Soto-Cerda and Cloutier 2012). The concern for GWAS is spurious associations

(false positive or Type I error) that can be caused by population stratification and cryptic relatedness (Astle and Balding 2009; Balding 2006; Cappa et al. 2011; Simko and Hu 2008). However, various statistical procedures have been developed to reduce and control this issue. For instance, a Mixed Linear Model (MLM) with incorporation of population structure and a kinship matrix effectively eliminated false positives in GWAS (Chen et al. 2016; Yu et al. 2006). A Multiple-Locus Linear Mixed Model (MLMM) was then developed based on MLM by adding significant markers as covariates in a stepwise MLM to remove the confounding between testing markers and relatedness (Liu et al. 2016a; Segura et al. 2012). The MLMM, with the advantage in controlling false positives, has been successful for association mapping in *Arabidopsis*, common wheat, rice, pea, sorghum, and tomato (Angelovici et al. 2013; Desgroux et al. 2016; Dilla-Ermita et al. 2017; Jaiswal et al. 2016; Li et al. 2015; Sauvage et al. 2014).

Recently, many plant genomes have been re-sequenced using next-generation sequencing (NGS) technologies. The exploration of whole-genome re-sequencing (WGRS) data was considered as one of the requirements for GWAS (Lee et al. 2015). Sequence-based GWAS was successfully applied for mapping agronomic traits and identifying the candidate genes inside of significant genomic regions in rice, peach, and foxtail millet (Cao et al. 2016; Huang et al. 2010; Jia et al. 2013). In another GWAS study in rice, beside the confirmation of the reported genes, new genes were identified for agronomic traits using WGRS data. The results of confirming those new genes by overexpression showed high accuracy (Yano et al. 2016). Although SNP chips data in GWAS has successfully provided valuable genetic information, the higher density of SNP

data generated from WGRS could be more precise in determining the candidate genes controlling traits of interest (Cao et al. 2016; Yano et al. 2016).

An association mapping of salt tolerance was first reported in soybean when analyzing seed germination rate under salt conditions (Kan et al. 2015). The study identified three genomic regions significantly associated with the ratio of imbibition rate, the ratio of germination index, and the ratio of germination rate under salt conditions and mapped to Chrs. 8, 9, and 18 (Kan et al. 2015). The major locus for salt tolerance on Chr. 3 that has been mapped by bi-parental linkage mapping was confirmed while using SoySNP50K iSelect BeadChip and WGRS data of 106 soybean lines (Patil et al. 2016). In addition to the major locus on Chr. 3, eight additional genomic regions significantly associated with both leaf chloride concentrations and leaf chlorophyll concentrations were mapped on Chrs. 2, 7, 8, 10, 13, 14, 16, and 20 using SoySNP50K dataset and were recommended for future studies (Zeng et al. 2017a).

In this study, two SNP marker datasets, SoySNP50K iSelect BeadChip (www.soybase.org) and 3.7M SNPs developed from the soybean WGRS in the Soybean Genetics and Genomics Lab (Valliyodan and Nguyen, unpublished data) with integration of three GBMs of the salt tolerant gene (Do et al. 2018; Patil et al. 2016) were used for association mapping of salt tolerance among a diverse set of soybean plant introductions (PIs). The objectives of this study were to map additional loci for salt tolerance other than the locus on Chr. 3 and to identify new salt tolerant sources for genetic analysis and breeding to improve salt tolerance in soybean.

Materials and Methods

Plant Materials

A core set of 305 soybean plant introductions (PIs) selected from the USDA Soybean Germplasm Collection represented the most genetic diversity resulted from the SoySNP50K iSelect BeadChip analysis (Drs. Cregan and Song, personal communication). This diverse panel had 255 cultivated soybean (*Glycine max*) and 50 wild soybean (*Glycine soja*) accessions. In addition, two salt-tolerant genotypes, cultivars Lee and Fiskeby III (Abel 1969; Abel and MacKenzie 1964; Do et al. 2018; Lenis et al. 2011), two salt-sensitive genotypes, cultivars Hutcheson and Jackson (Ha et al. 2013; Hamwieh and Xu 2008), and cultivar, Williams 82, were also included as checks. According to the Germplasm Resources Information Network-National Plant Germplasm System (GRIN, https://www.ars-grin.gov/npgs/) database, this panel has a wide range of maturity groups (MG) from 000 to X and originated from 28 different countries. While conducting GWAS, a subset of 234 accessions from this panel were used for a separate association analysis based on the availability of over 3.7M SNP dataset (Valliyodan and Nguyen, University of Missouri, unpublished data).

Genotypic datasets

Over 42,000 SNP markers from the Illumina Infinium SoySNP50K iSelect BeadChip (Song et al. 2013) were accessed from the soybean database (http://www.soybase.org). Of these, a total of 37,573 SNPs was selected based on the exclusion of SNPs with greater than 5% missing data and a minor allele frequency (MAF) of less than 5%. A second SNPs dataset of over 3.7M SNPs generated from the United

Soybean Boarded-funded whole genome resequencing project in the Nguyen Lab (Valliyodan and Nguyen, University of Missouri, unpublished data). This SNP dataset was further filtered to obtain over 2,2M SNPs in the subset of 234 *G. max* soybean lines for a separate association analysis. Additionally, three GBMs, Salt-20, Salt14056 and Salt11655, that were previously reported (Do et al. 2018; Patil et al. 2016) were also incorporated into these SNP marker data sets.

Phenotyping

Soybean lines of the diverse panel were evaluated for salt tolerance under greenhouse conditions following a previously described method (Lee et al. 2008) with minor modifications. The experimental design was a randomized block with 3 replications blocking over time. Five seedlings of each line were grown per cone-tainer and evaluated in a greenhouse at the University of Missouri, Columbia, MO, using artificial lights and a 13 h photoperiod from September to December 2016. Soybean seedlings at the growth stage V2 (Fehr et al. 1971) were treated with salt water by filling 120 mM NaCl to the tank. The salt solution in the tank was kept at one-third (7 cm) from the bottom of cone-tainers for 7 h/day. When the salt-sensitive checks showed severe leaf scorch, which typically appears approx. 2 weeks after the treatment, leaf scorch score (LSS) was visually scored for each plant using a 1–5 scale, mean of LSS of each line was then calculated as previously described (Lee et al. 2008). Leaf chlorophyll content was measured for the topmost fully expanded leaf 1 day before and about 14 days after the salt treatment using a chlorophyll meter (Chlorophyll meter SPAD-502, Konica Minolta, Inc., Osaka, Japan) to calculate chlorophyll content ratio (CCR) (Do et al. 2018).

The trifoliate leaves of five seedlings for each soybean line were harvested after being read for leaf scorch and chlorophyll determinations and pooled to dry at 600C for 7 days. The dried leaf tissue was ground using a Thomas Model ED-5 laboratory Wiley mill (Thomas Scientific, Swedesboro, NJ, USA) for analyzing leaf chloride (LCC) and leaf sodium (LSC) contents at the Delta Research Center, University of Missouri, Portageville, MO (Do et al. 2018; Lenis et al. 2011). Briefly, sodium concentration was determined by processing ground leaf tissue (0.25 g) with a modified wet acid dilution and measuring in an atomic absorption spectrophotometer (Perkin-ElmerTM, Wellesley, MA, USA). On the other method, standard reference curves were established for calibration of 25, 50, 100 and 500 mg kg^{-1} of chloride. Then chloride concentration in solution of 0.15 g ground leaf tissue was measured using an ion specific electrode in a Fisher Scientific AR 50 dual channel pH, ion, conductivity-meter (Fischer Scientific, Pittsburg, PA, USA). Finally, sodium and chloride concentrations were converted to mg per kg of dry leaf tissue for leaf sodium content (LSC) and leaf chloride content (LCC).

Phenotypic data analysis

Analysis of variance (ANOVA) and the estimation of variance components of phenotypic data were performed using the PROC GLM procedure of SAS 9.4 (SAS Institute Inc. 2013). Broad-sense heritability (H^2) of four salt tolerance associated traits were calculated using the following equation for randomized block design: $H^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_e)$, where σ_g is the genotypic variance and σ_e is the error variance (Libby 1962; Sharma 2006). Soybean lines were grouped into salt tolerant and salt sensitive groups based on combining all four salt tolerant traits using Euclidean distances in NTSYSpc 2.1 (Rohlf 2000). In addition, the Pearson correlations were also calculated to measure the degree of

linear relationship between each pairs of traits and the individual hypothesis tests of the correlations were performed at $\alpha = 0.01$ using Minitab 17 (Minitab Inc. 2010).

Genome-wide association study (GWAS)

Linkage disequilibrium (LD) and correlation coefficients (r^2) were measured on all adjacent pairs within each chromosome and a fitted curve was computed based on nonlinear regression of LD (r^2)) on distance (kb) as previously defined (Remington et al. 2001). Principle components (PCs) and kinship matrix (relatedness) were applied to correct for population structure and relatedness in mixed linear models. Principal component analysis (PCA) was done for all filtered SNPs with MAF > 0.05 and the call rate > 0.95. Number of PCs were selected to add in models when an inflation factor (λ) of *p*-value close to one (Li and Zhu 2013; Lo et al. 2017) using a genotype association test with a PCs correction in SNP & Variation Suite (SVS) v8.7.0 software (Golden Helix, Bozeman, MT, USA). Identity by state (IBS) matrices indicated relatedness among soybean accessions calculated using TASSEL 5 (Bradbury et al. 2007) to construct phylogenic trees in MEGA 7 (Kumar et al. 2016) and to apply as kinship matrix in GWAS. Genome-wide associations between SNPs and salt tolerant trait were identified using the efficient mixed-model association expedited (EMMAX) and multi-locus mixed model (MLMM) with correction for population structure and relatedness in the SVS software (Golden Helix, Bozeman, MT, USA). False positives were controlled by multiple test correction with false discovery rate (FDR) ≤ 0.05 (Qu et al. 2010) and the threshold of -log10(*p*-value) for identifying significant associations was calculated at FDR = 0.05. The significant SNPs associated with salt tolerant traits were counted with larger -log10(*p*-value) than the threshold that was calculated based on P-value using the False Discovery Rate correction (Benjamini-

Hochberg) and causal SNPs surrounding the known gene for salt tolerance on Chr. 3 (Do et al. 2016; Guan et al. 2014; Qi et al. 2014).

Candidate gene of salt tolerance

The genomic data of soybean, the soybean genome assembly (Wm82.a2.v1), was integrated with the genomic scale data visualization tool by importing to the GWAS project created in the SVS software (Golden Helix, Bozeman, MT, USA). The candidate gene was searched by zooming in at position of the significant SNPs associated with salt tolerant traits. Information about the candidate gene was displayed by automatically linking with Genome Browser of phytozome website (https://phytozome.jgi.doe.gov/jbrowse/). The results were double checked by searching SoyBase Wm80 Genome Browser (https://soybase.org/gb2/gbrowse/).

Identification of new sources of salt tolerance using gene-based markers

The soybean lines from salt tolerant group based on their phenotypes were selected to evaluate matching genotypes of GBMs. The salt tolerant lines that do not match the genotypes of GBMs were identified as new sources that may carry new salt tolerant gene(s) in addition to the known gene on Chr. 3 (Do et al. 2016; Guan et al. 2014; Qi et al. 2014). Moreover, recombination between three GBMs in the salt tolerant lines was considered as new allele(s) of the locus on Chr. 3 for salt tolerance.

Results

Phenotypic variation, heritability and correlation of salt tolerant traits

The Hutcheson and Jackson checks were salt sensitive genotypes while Lee and Fiskeby III checks were salt tolerance for all traits (Table S3.1 and Fig. S3.1). This is presented in phenotypic tree (Fig. 3.2) in which the sensitive checks belonged to salt sensitive group while the salt tolerant checks were in the salt tolerant group. Phenotypic variation among 305 soybean lines was statistically significant for all traits studied, including leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC), ranging from 1-5, 0.3-1.2, 0.03-1.7 (g kg-) and 2.7-18.2 (g kg-), respectively (Table 3.1). It is interesting to note that some soybean lines showed higher salt tolerance than the salt-tolerant checks and some soybean lines were more sensitive to salt than the sensitive checks.

Correlation and heritability were estimated for the four traits in the diverse panel. The Pearson correlations indicated high linear relationship among the four traits and were significant at $\alpha = 0.01$. Among these, correlation coefficients (r^2) of CCR with the other traits were negative and ranged from -0.92 to -0.61. Conversely, the correlations between LSS, LSC and LCC were positive (Table 3.1). The broad-sense heritability was estimated based on analysis of variance for all traits (Table 3.1). The lowest value of heritability was 0.29 for LSC while higher heritability was observed in LSS, CCR and LCC as 0.82, 0.94 and 0.63, respectively.

In addition, variation among 305 soybean accessions was shown by combining the four traits to construct a phenotypic dendrogram into salt-tolerant and salt-sensitive groups.

A dissimilarity matrix that measures the degree of dissimilarity between all pairs in salt tolerance of soybean lines showed a range of Euclidean distance from 0.06 to 10.81. The cluster analysis for the Euclidean distance matrix in the dendrogram revealed two main groups of soybean lines at the lowest Euclidean value and subgroups at higher values. Thereby, 137 soybean lines, including salt-tolerant checks belonged to salt-tolerant group and the remaining lines, including salt-sensitive checks were grouped into salt-sensitive group (Fig. 3.1).

Linkage disequilibrium decay, population structure and relatedness among soybean accessions

Linkage disequilibrium (LD) decay, population structure and relatedness were analyzed for the penal of 305 soybean accessions using 37,573 SNPs of the SoySNP50K dataset and for the subset of 234 accessions using over 2.2 M SNPs of the WGRS-derived SNP dataset. An average distance between two markers were approximately 29.36 kb for the SoySNP50K dataset and 0.43 kb for the WGRS-derived SNP dataset. LD decay on all adjacent SNP pairs were presented in nonlinear curves (Fig. S3.2) with the LD blocks at an r2 of 0.2 to be 293.64 and 371.42 kb for the SoySNP50K and the WGRS-derived SNP datasets, respectively. Thus, the number of SNPs was sufficient to cover the genome-wide haplotype blocks for both datasets.

The principal component analysis (PCA) showed that variance explained by eigenvalue of each PC rapidly dropped after the first 10 PCs for both SNP datasets (Fig. 3.2A). The cumulative eigenvalues of the first three PCs were 49.7% and 28.7% of variances for the diverse panel and the subset using the SoySNP50K and the WGRS-

derived SNP datasets, respectively. For the diverse panel using 37,573 SNPs, 305 soybean accessions were separated into groups roughly corresponding to taxonomy (*G. max* and *G. soja*) and country of origin (Fig. 3.2B, 3.2C) according to the first three PCs. However, the subset of 234 *G. max* soybean accessions was less defined into groups based on the first three PCs using over 2.2 M SNPs (Fig. 3.2D).

The cryptic relatedness among soybean accessions was evaluated by kinship matrix from identity by state (IBS) from each paired soybean lines. The matrix with a range of IBS from 0.42 to 0.97 was calculated using the SoySNP50K dataset for the diverse panel to construct a phylogenetic tree showing the relationship among 305 soybean accessions (Fig. 3.3A). In this dendrogram, two main groups, *G. max* and *G. soja*, were clustered at the lowest IBS value and there were subgroups at higher IBS values. In a similar analysis, the kinship matrix with a narrower range of IBS from 0.52 to 0.89 was found by analyzing the WGRS-derived SNP dataset in the subset. The relatedness among 234 *G. max* soybean accessions was also shown in a heat map constructed from these IBS matrices (Fig. 3.3B).

The genomic inflation factors (lambda, λ) from association tests were applied to verify correction for population stratification by including PCs and kinship in the GWAS models (Table S2). The lambda values from association tests for all traits showed high values by using the general linear model (GLM) without correction for stratification ranges of 2.21 – 3.95 and 1.44 – 2.08 were determined by analyzing the SoySNP50K and the WGRS-derived SNP datasets, respectively. The lower and values closer to one with the range of 1.05 – 1.20 were calculated with population correction by PCs in P models. By adding more kinship matrices in EMMAX (Efficient mixed-model association expedited) and MLMM (Multi-locus mixed model) models, the pseudo-lambda values were from 0.93

to 1.03 for association of all traits from both genetic data sets that indicate the GWAS results were not inflated by population structure or cryptic relatedness.

GWAS for salt tolerance of the diverse panel using SoySNP50K dataset

After marker quality control and assurance, a total of 37,573 polymorphic SNPs was selected from the SoySNP50K dataset and were utilized for further analysis of LSS, CCR, LSC, and LCC in the diverse panel. Inspections of the quantile-quantile (QQ) distributions showed most SNPs matched with what was expected (Fig. S3.3). The inflation values (Table S3.2), the EMMAX and MLMM with corrections for the population stratification from the kinship matrix and the three first PCs were optimal for all four traits evaluated in this study. Numbers of SNPs significantly associated with LSS, CCR, LSC, and LCC were 44, 38, 13, and 54 based on corrections of False Discovery Rate (FDR) (Benjamini-Hochberg) with FDR ≤ 0.05. SNPs surrounding the known salt tolerance gene on Chr. 3 were the most significant SNPs associated with all four traits (Table 3.2 and Fig. 3.4B). In addition, several SNPs significantly associated with one or two traits were identified on other chromosomes using the EMMAX analysis approach, such as ss715616720 on Chr. 13 significantly associated with LSS and CCR, ss715609949 on Chr. 11 and ss715611871 on Chr. 12 with LSC, ss715592375 on Chr. 5 with LCC, and ss715592375 on Chr. 15 with LSS (Table 3.2; Fig. 3.4A). However, those significant SNPs, except the GBMs, were not detected after step 1 of MLMM.

GWAS for salt tolerance of the subset using the WGRS-derived SNP dataset

The subset of 234 *G. max* soybean accessions that had 3.7M SNPs derived from the WGRS project (Valliyodan and Nguyen, University of Missouri, unpublished data) was

chosen to further perform GWAS for salt tolerance. To ensure that population size had no effect on GWAS results, this subset was separately analyzed by reusing SoySNP50K dataset. Similar results were found compare to GWAS of salt tolerance using SoySNP50K dataset for the diverse panel of 305 soybean lines was only found for the major locus on Chr. 3 (Fig. S3.4).

The SNPs derived from the WGRS project were subjected to further quality control and assurance with MAF > 0.05 and the call rate > 0.95. Over 2.2 M polymorphic SNPs were obtained for further analysis. This dataset was employed for LD calculation, kinship construction, PCA, genotype association tests, and subsequently for GWAS of the four salt tolerant traits. Based on the QQ distributions, most SNPs were matched according to expectations (Fig. S3.5). The inflation values (Table S3.2), adding six PCs as a fixed effect and the kinship matrix as a random effect in GWAS mixed models were optimal for controlling confounding risk due to population stratification. As results of EMMAX model analysis, number of SNPs significantly associated with LSS, CCR, LSC and LCC were 217, 190, 136 and 278, respectively, based on a FDR correction (Benjamini-Hochberg) with FDR ≤ 0.05 and selecting causal SNPs surrounding the know gene on Chr. 3. The significant SNPs associated with salt tolerant traits were located on Chrs. 1, 2, 3, 5, 6, 8, 14, 15, 16, 18, 19 and 20 (Fig. 3.5A) and the most significant SNPs on each chromosome are listed in Table 3.3. SNPs inside the known gene for salt tolerance, including three GBMs on Chr. 3, were significantly associated with all of four traits with the highest -log10(*P*) values. Genomic regions on Chrs. 8 and 18 were also significantly associated with LSS, CCR, LSC and LCC, of which two adjacent peaks were shown in the significant region on Chr. 8. In addition, GWAS for salt tolerance using MLMM indicated that the significant positions in Chrs. 1, 3,

8 and 18 genomic regions were added as covariates (the most significant SNP after each step in stepwise analysis) in the stepwise analysis (Fig. 3.6). Results show, the positions on Chr. 3 and Chr. 8 appeared to be associated with all of 4 salt tolerant traits while the positions on Chr. 1 and Chr. 18 were covariates in GWAS for LSC and LCC. Thus, the genomic region on Chr. 8 (Fig. 3.5B) was suggested as a minor locus for salt tolerance in addition to the major gene on Chr. 3.

Putative candidate genes underlying salt tolerance

Besides the known salt tolerance gene on Chr. 3, additional genes were searched in the significant genomic regions associated with salt tolerant traits using Phytozone and Soybase genome browsers (https://phytozome.jgi.doe.gov/jbrowse/ and https://soybase.org/gb2/gbrowse/). The genomic intervals with the group of the significant SNPs ranged from 2,788 to 787,140 bp and were considered significant genomic regions to search for putative candidate genes. A total of 222 genes were found in the significant genomic intervals (Table S3.3) including the genes underlying the most significant SNPs on each chromosome (Table 3.3). Among these, 157 genes have predicted functions and known protein families in Phytozone and Soybase databases. Based on functional annotation, transporters (*Glyma.08g146100, Glyma.08g224400*), ion channel (*Glyma.02g204300*), membrane proton pump (*Glyma.08g225500*), universal stress protein (*Glyma.14g211300*), and Callose synthase (*Glyma.08g157400*) could be additional candidate genes related to salt tolerance in soybean. Four genes were located in two adjacent, significant genomic intervals surrounding the minor locus on Chr. 8 and others on Chrs. 2 and 14.

Identification of new sources of salt tolerance

Soybean accessions belong to salt tolerant and salt sensitive groups were classified by combining results of LSS, CCR, LSC and LCC (Fig. 3.2) and were tested with genotypes of three GBMs. To evaluate the genotype-phenotype association, genotypes of GBMs (Salt-20, Salt14056 and Salt11655) among 305 soybean accessions were classified and evaluated by compatibility with phenotypes. The GBM genotypes of salt tolerant checks (Fiskeby III and Lee) were named as mutant (mut) alleles while those of salt sensitive checks (Hutcheson and Jackson) were wide-type (wt) alleles (Table 3.4). The strong genotype-phenotype association in the diverse panel of soybean accessions showed that over 90% of accessions showed a perfect genotype-phenotype match, in which salt tolerant and salt sensitive lines carried mutant and wide-type alleles of all three GBMs, respectively. There were some salt sensitive lines with recombination of three GBMs; however, no salt sensitive lines carried mutant alleles of all three GBMs. In the salt tolerant group, 10 lines had a combination of three GBMs (Table 3.4), suggesting that these soybean accessions may be new salt tolerant sources with new alleles of the known salt tolerant gene on Chr. 3. Additionally, six other salt tolerant lines without mutant alleles of the three GBMs (Table 3.4) were predicted as salt tolerant sources carrying new gene(s) for salt tolerance in soybean.

Discussion

It has been demonstrated that association mapping is suited for the detection and characterization of quantitative traits because of broad genetic base of natural populations (Abdurakhmonov and Abdukarimov 2008; Flint-Garcia et al. 2003; Soto-Cerda and

Cloutier 2012). The development of statistical models and multiple tests to control effects of population structure and relatedness (Chen et al. 2016; Liu et al. 2016a; Segura et al. 2012; Yu et al. 2006). 2006), as well as exploration of whole-genome sequencing data made GWAS more efficient in identifying the significant genomic regions associated with the traits of interest and in predicting candidate genes (Cao et al. 2016; Yano et al. 2016). In this study, both SoySNP50K- and WGRS-derived SNP datasets were sufficient to cover the genome-wide haplotype blocks. There was no genomic inflation by population stratification found in the association analysis using EMMAX and MLMM (Figs. S3.3, S3.5; Table S3.2). Nevertheless, GWAS for detecting salt tolerance using SoySNP50K dataset only confirmed the major locus on Chr. 3. On the other hand, using the 3.7M SNP dataset discovered more significant genomic regions, including a minor locus on Chr. 8. Additionally, three GBMs of the known gene (*Glyma03g32900*) on Chr. 3 (Do et al. 2016; Guan et al. 2014; Qi et al. 2014) were significantly associated with salt tolerance by SoySNP50K-based GWASs (Fig. 3.4B) while there were many other significant SNPs in the known gene were identified by WGRS-based GWASs (Fig. S3.7). Pinpointing the known gene was only accomplished by integration of GBMs into SoySNP50K dataset. Therefore, a larger SNP dataset generated from whole-genome sequencing was better for GWAS for detecting new loci for salt tolerance.

The major locus on Chr. 3, in which the known gene for salt tolerance (*Glyma03g32900*) is located, was confirmed by previous GWAS (Patil et al. 2016; Zeng et al. 2017a). By integrating the three GBMs into the SNP marker datasets, the known gene was pinpointed with the most significant SNPs associated with all four traits by both SoySNP50K- and WGRS-based GWAS (Tables 3.2, 3.3; Figs. 3.4A, 3.5A). Also, the other

genomic regions associated with salt tolerance and related traits have been mapped by association mapping in previous studies. A GWAS analysis of 191 landraces for three germination-related traits under high salt conditions using 1142 SNPs determined 13 SNPs associated with salt tolerance on Chrs. 2, 3, 6, 8, 9, 12, 13, 14, 17 and 18 and were reported with -log10(*P*) range of 2.05-3.60 by using mixed linear model (Kan et al. 2015). In a previous association study of a panel of 283 soybean lines with 33,009 SNPs (SoySNP50K dataset), Zeng et al. (2017a) evaluated two salt tolerant traits, leaf chloride concentrations and leaf chlorophyll concentrations, confirmed the major locus on Chr. 3 and detected additional genomic regions on Chrs. 2, 7, 8, 10, 13, 14, 16 and 20 with -log10(*P*) > 4.1 and -log10(*P*) > 2.1 thresholds for a GLM and MLM, respectively. The results of those association mapping studies require confirmation because of -log10(*P*) thresholds to control spurious association did not meet the requirements for GWAS (Kan et al. 2015; Zeng et al. 2017a). In our study, the significant SNPs associated with salt tolerance using both SNP marker datasets were found on 15 chromosomes based on FDR correction (Benjamini-Hochberg) with FDR ≤ 0.05 (Tables 3.2, 3.3). Consequentially, the major locus on Chr. 3, minor locus on Chr. 8, and significant genomic regions on Chrs. 1 and 18 were confirmed and identified by a combination an association of salt tolerant traits and results of GWAS using MLMM. By comparing to previous studies, the minor locus on Chr. 8 identified in our study was 1,449,275 bp far from BARC-041663-08059 associated with germination-related traits under salt stress (Kan et al. 2015) and 17,932,879 bp far from ss715601563 associated with salt tolerance (Zeng et al. 2017a). Thus, the significant genomic region on Chr. 8 associated with salt tolerance in this study strongly predicted as new minor locus for salt tolerance in soybean.

In addition, an Cl^--tolerant QTL located between 27,665,585 and 28,206,014 bp on Chr. 13 was identified by a linkage mapping of KCl-salt tolerance (Zeng et al. 2017b). Another QTL for leaf sodium content was mapped between 38,366,685 and 40,167,119 bp on Chr. 13 (Do et al. 2018). Only one SNP marker, ss715616720, was significantly associated with leaf scorch score and chlorophyll content ratio (Table 3.2) by SoySNP50K-based GWAS using EMMAX that was undetected after one step of MLMM and WGRS-based GWAS. This could be a spurious association because of the confounding between testing markers and kinship (Liu et al. 2016a). Thus, the mapped locus for salt tolerance on Chr. 13, which was not identified in this study, may carry rare salt tolerant alleles and low frequency in the diverse panel.

Efficient tools for candidate gene prediction, Phytozone and Soybase genome browsers (https://phytozome.jgi.doe.gov/jbrowse/ and https://soybase.org/gb2/gbrowse/) enabled a search to predict putative candidate genes for iron deficiency chlorosis (Mamidi et al. 2014), nitrogen fixation traits (Dhanapal et al. 2015), soybean seed germination under salt stress (Kan et al. 2015), sudden death syndrome resistance (Zhang et al. 2015), *Phytophthora sojae* resistance (Schneider et al. 2016), and soybean cyst nematode resistance (Vuong et al. 2015; Zhang et al. 2017). Using a similar approach, *Glyma.08g146100* (EamA-like transporter family), *Glyma.08g157400* (SF9 - Callose synthase), *Glyma.08g224400* (V-type H+-transporting ATPase subunit A), and *Glyma.08g225500* (SF11 - Pyrophosphate-energized membrane proton pump) were predicted as putative candidate genes for salt tolerance that was mapped on Chr. 8 by functional nominations. Similar functional genes have been reported controlling salt tolerance or abiotic stress in plants. Callose synthase plays an important role in response to

multiple biotic and abiotic stresses, including salt stress (Chen and Kim 2009; Li et al. 2017). In addition, V-type H+-transporting ATPase subunit expressed in the roots, energizes sodium sequestration into the central vacuole and enhances salt tolerance in plants (Golldack and Dietz 2001; He et al. 2014; Zhang et al. 2012). Therefore, the putative candidate genes should be considered for post-GWAS analysis such as gene expression.

Association between salt tolerance and structure of the known gene (*Glyma03g32900*) on Chr. 3 (Guan et al. 2014; Patil et al. 2016) or gene profiling expression of this gene (Do et al. 2016) suggest other sources with novel genes for salt tolerance. Three GBMs, Salt-20, Salt14056 and Salt11655, based on *Glyma03g32900* sequence showed a high association with salt tolerance (Do et al. 2016; Patil et al. 2016) and were confirmed in this study. However, a few salt tolerant lines that were not a genotype-phenotype match carried salt sensitive alleles of these GBMs. The results suggest that those salt tolerant lines might have new salt tolerant allele(s) at known locus on Chr. 3 and in addition other lines might carry novel gene(s) for soybean salt tolerance.

In conclusion, the SNP dataset developed from the WGRS was more efficient than those of the SoySNP50K dataset for GWAS to predict minor loci and to pinpoint putative candidate genes for salt tolerance in soybean. The significant genomic regions strongly suggested a minor locus associated with salt tolerance on Chr. 8. The putative candidate genes and the significant SNPs may be helpful to study the molecular mechanism and can be useful for marker-assisted selection and molecular breeding to improve salt tolerance in soybean. Additionally, six new salt tolerant sources with predicted novel gene(s) should be further investigated to identify additional salt tolerant gene(s).

Acknowledgements

This research was partially supported by the College of Agriculture, Food and Natural Resources and the Missouri Agricultural Experiment Station. Mr. Tuyen Do would like to thank Cuu Long Delta Rice Research Institute and the Vietnam Ministry of Agriculture for a graduate student scholarship and thank Ms. Nguyen Diep Nhu Dang, an internship in Nguyen's lab, for her help in evaluation of phenotype.

Legends of Figures

Figure 3.1 Phenotypic dendrogram for salt tolerance variation among 305 soybean accessions by combining leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC)

Figure 3.2 Population structure by principal component analysis (PCA) of two SNP datasets. (A) Screen plot of the first 10 principal components (PCs) and their contribution to Eigenvalue by analyzing SoySNP50K dataset (yellow) and 3.7M SNP dataset (blue). (B) 3D scatterplot showed the first three PCs from SoySNP50K dataset corresponding to taxonomic groups for 305 soybean accessions. (C) 3D scatterplot showed the first three PCs from SoySNP50K dataset corresponding to original groups of 305 soybean accessions. (D) 3D scatterplot showed the first three PCs from 3.7M SNP dataset corresponding to 234 soybean accessions selected from the original 305 accessions

Figure 3.3 Relationships among 305 soybean accessions using SoySNP50K dataset and the subset of 234 soybean accessions using 3.7M SNP dataset. (A) Phylogenetic tree of 305 soybean accessions using SoySNP50K dataset. (B) Heatmap plot showing the relationship among 234 soybean accessions using 3.7M SNP dataset

Figure 3.4 Association mapping of salt tolerance using EMMAX with a SoySNP50K dataset for 305 diverse soybean accessions. (A) Manhattan plots showed association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf

chloride content (LCC). (B) Three GBMs associated with salt tolerance and their position inside the known gene (*Glyma03g32900*), named in SoyBase Wm82 Genome Browser version 1

Figure 3.5 Association mapping of salt tolerance using EMMAX with 3.7M SNP dataset for the subset of 234 soybean accessions. (A) Manhattan plots showed association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC). (B) The significant genomic region associated with LSC on Chr. 8. (C) One of the putative candidate gene underlying a minor locus for salt tolerance

Figure 3.6 Comparison between EMMAX and MLMM in GWAS for salt tolerance. (A) Manhattan plot showed association of SNPs with leaf sodium content (LSC) by sequence-based GWAS using EMMAX. (B) Manhattan plot showed covariates (the most significant SNP after each step in stepwise analysis) associated with LSC by sequence-based GWAS using MLMM

Supplementary Figure S3.1 Comparisons of salt tolerance based on leaf scorch between cultivars Fiskeby III and Lee (salt tolerant checks), Hutcheson and Jackson (salt sensitive checks), Williams 82 (the soybean reference cultivar), grown under a 120 mM NaCl treatment

Supplementary Figure S3.2 LD decay plot of coefficient of correlation (r2) between adjacent marker pairs and genomic distance (kb), the fitted curves are based on nonlinear regression using SoySNP50K dataset from 305 diverse genotypes (*yellow*) and using 3.7M SNP dataset in the subset from 234 genotypes selected from the original 305 accessions(*blue*)

Supplementary Figure S3.3 Quantile-quantile (QQ) plots showing the expected -log10(P) compared to the observed -log10(*P*), the results of statistical testing (EMMAX) for association across 37,573 SNPs from SoySNP50K dataset with leaf scorch score (A), chlorophyll content ratio (B), leaf sodium content (C) and leaf chloride content (D) among 305 genetically diverse soybean genotypes. Most SNPs matched with solid lines [expected -log10(*P*) = observed -log10(*P*)] were unassociated SNPs, on the other hand, sharp curves at the end presented the number of true associations

Supplementary Figure S3.4 Manhattan plots showing association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) using 37,573 SNPs from SoySNP50K dataset from the subset of 234 diverse accessions selected from the original 305 accessions

Supplementary Figure S3.5 QQ plots showing the expected -log10(P) compared to the observed -log10(*P*), the results of statistical testing (EMMAX) for association across 2,280,225 polymorphic SNPs from 3.7M SNP dataset with leaf scorch score (A), chlorophyll content ratio (B), leaf sodium content (C) and leaf chloride content (D) in the subset of 234 soybean lines selected from the original population of 305 genotypes. SNPs matched with solid lines [expected -log10(*P*) = observed -log10(*P*)] were unassociated SNPs, on the other hand, sharp curves at the end presented the number of true associations

Supplementary Figure S3.6 Significant SNPs in association with salt tolerance with the putative candidate gene (*Glyma.08G224400*) as a minor locus for salt tolerance, and LD block in this genomic region

Supplementary Figure S3.7 The significant SNPs underlying the known gene (*Glyma03g32900*) on Chr. 3 (*red color dots*) associated with salt tolerance, the known gene (*Glyma03g32900*) named in SoyBase Wm82 Genome Browser version 1, and LD block in this genomic region

References

Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int J Plant Genomics 2008:574927

Abel GH (1969) Inheritance of the capacity for chloride inclusion and chloride exclusion by soybeans. Crop Sci 9:697-698

Abel GH, MacKenzie AJ (1964) Salt tolerance of soybean varieties (*Glycine max* L. Merill) during germination and later growth. Crop Sci 4:157-161

Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, DellaPenna D (2013) Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. The Plant Cell 25:4827-4843

Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. Statist Sci 24:451-471

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781-791

Batlle-Sales J (2011) Salinization: An environmental concern under climate change scenarios. In: Thomas RP (ed) Proceedings of the Global Forum on Salinization and Climate Change (GFSCC2010), Valencia, 25–29 October 2010. FAO, Rome, p 10

Blanco FF, Folegatti MV, Gheyi HR, Fernandes PD (2007) Emergence and growth of corn and soybean under saline stress. Sci Agr 64:451-459

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Bustingorri C, Lavado RS (2011) Soybean growth under stable versus peak salinity. Sci Agric 68:102-108

Cao K, Zhou Z, Wang Q, Guo J, Zhao P, Zhu G, Fang W, Chen C, Wang X, Wang X, Tian Z, Wang L (2016) Genome-wide association study of 12 agronomic traits in peach. Nature Communications 7:13246

Cappa EP, Martínez MC, Garcia MN, Villalba PV, Marcucci Poltri SN (2011) Effect of population structure and kinship relationships on the results of association mapping tests of growth and wood quality traits in four *Eucalyptus* populations. BMC Proc 5:P23

Chen H, Wang C, Conomos Matthew P, Stilp Adrienne M, Li Z, Sofer T, Szpiro Adam A, Chen W, Brehm John M, Celedón Juan C, Redline S, Papanicolaou George J, Thornton Timothy A, Laurie Cathy C, Rice K, Lin X (2016) Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet 98:653-666

Chen HT, Cui SY, Fu SX, Gai JY, Yu DY (2008) Identification of quantitative trait loci associated with salt tolerance during seedling growth in soybean (*Glycine max* L.). Aust J Agr Res 59:1086-1091

Chen X-Y, Kim J-Y (2009) Callose synthesis in higher plants. Plant Signal Behav 4:489-492

Desgroux A, L'Anthoene V, Roux-Duparque M, Riviere JP, Aubert G, Tayeh N, Moussart A, Mangin P, Vetel P, Piriou C, McGee RJ, Coyne CJ, Burstin J, Baranger A, Manzanares-Dauleux M, Bourion V, Pilet-Nayel ML (2016) Genome-wide association mapping of partial resistance to *Aphanomyces euteiches* in pea. BMC Genomics 17:124

Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, King CA, Fritschi FB (2015) Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. Plant Genome-Us 8

Dilla-Ermita CJ, Tandayu E, Juanillas VM, Detras J, Lozada DN, Dwiyanti MS, Vera Cruz C, Mbanjo EGN, Ardales E, Diaz MG, Mendioro M, Thomson MJ, Kretzschmar T (2017) Genome-wide Association Analysis Tracks Bacterial Leaf Blight Resistance Loci In Rice Diverse Germplasm. Rice 10:8

Do TD, Chen H, Hien VT, Hamwieh A, Yamada T, Sato T, Yan Y, Cong H, Shono M, Suenaga K, Xu D (2016) *Ncl* synchronously regulates Na(+), K(+), and Cl(-) in soybean and greatly increases the grain yield in saline field conditions. Sci Rep 6:19147

Do TD, Vuong TD, Dunn D, Smothers S, Patil G, Yungbluth DC, Chen P, Scaboo A, Xu D, Carter TE, Nguyen HT, Grover Shannon J (2018) Mapping and confirmation of loci for salt tolerance in a novel soybean germplasm, Fiskeby III. Theor Appl Genet 131:513-524

El-Sabagh A, Sorour S, Ueda A, Saneoka H, Barutcular C (2015) Evaluation of salinity stress effects on seed yield and quality of three soybean cultivars. Azarian J Agric 2:138-141

Essa TA (2002) Effect of salinity stress on growth and nutrient composition of three soybean (*Glycine max* L. Merrill) cultivars. J Agron Crop Sci 188:86-93

FAO (2002) Crops and Drops: Making the Best Use of Water for Agriculture. Food and Agriculture Organization of the United Nations, Rome

FAO, ITPS (2015) Status of the world's soil resources (SWSR) – main report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy:pp 124-127

Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci 11:929-931

Flint-Garcia SA, Thornsberry JM, Buckler ESt (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Golden Helix Inc. (2017) SNP & Variation Suite v8.7.0. Bozeman, MT

Golldack D, Dietz K-J (2001) Salt-Induced Expression of the vacuolar H(+)-ATPase in the common ice plant is developmentally controlled and tissue specific. Plant Physiol 125:1643-1654

Guan RX, Qu Y, Guo Y, Yu LL, Liu Y, Jiang JH, Chen JG, Ren YL, Liu GY, Tian L, Jin LG, Liu ZX, Hong HL, Chang RZ, Gilliham M, Qiu LJ (2014) Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. Plant J 80:937-950

Ha BK, Vuong TD, Velusamy V, Nguyen HT, Shannon JG, Lee JD (2013) Genetic mapping of quantitative trait loci conditioning salt tolerance in wild soybean (*Glycine soja*) PI 483463. Euphytica 193:79-88

Hamwieh A, Tuyen DD, Cong H, Benitez ER, Takahashi R, Xu DH (2011) Identification and validation of a major QTL for salt tolerance in soybean. Euphytica 179:451-459

Hamwieh A, Xu DH (2008) Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. Breeding Sci 58:355-359

He X, Huang X, Shen Y, Huang Z (2014) Wheat V-H+-ATPase Subunit Genes Significantly Affect Salt Tolerance in *Arabidopsis* thaliana. PLoS One 9:e86982

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961

Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, Balyan HS, Gupta PK (2016) Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum* L.). PLoS One 11:e0159343

Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H, Zhu C, Lu Y, Zhou C, Fan D, Weng Q, Guo Y, Huang T, Zhang L, Lu T, Feng Q, Hao H, Liu H, Lu P, Zhang N, Li Y, Guo E, Wang S, Wang S, Liu J, Zhang W, Chen G, Zhang

B, Li W, Wang Y, Li H, Zhao B, Li J, Diao X, Han B (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat Genet 45:957

Kan G, Zhang W, Yang W, Ma D, Zhang D, Hao D, Hu Z, Yu D (2015) Association mapping of soybean seed germination under salt stress. Mol Genet Genomics 290:2147-2162

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29

Lee CY, Kim AJ, Lee S, Park YJ (2015) Concept of Genome-Wide Association Studies. In: Koh HJ, Kwon SY, Thomson M (eds) Current Technologies in Plant Molecular Breeding. Springer, Dordrecht, pp 175-204

Lee GJ, Carter TE, Jr., Villagarcia MR, Li Z, Zhou X, Gibbs MO, Boerma HR (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109:1610-1619

Lee JD, Smothers SL, Dunn D, Villagarcia M, Shumway CR, Carter TE, Shannon JG (2008) Evaluation of a simple method to screen soybean genotypes for salt tolerance. Crop Sci 48:2194-2200

Lenis JM, Ellersieck M, Blevins DG, Sleper DA, Nguyen HT, Dunn D, Lee JD, Shannon JG (2011) Differences in ion accumulation and salt tolerance among Glycine accessions. J Agron Crop Sci 197:302-310

Li G, Zhu H (2013) Genetic Studies: The Linear Mixed Models in Genome-wide Association Studies. Open Bioinforma J 7:27-33

Li S, Zhang L, Wang Y, Xu F, Liu M, Lin P, Ren S, Ma R, Guo Y-D (2017) Knockdown of a cellulose synthase gene BoiCesA affects the leaf anatomy, cellulose content and salt tolerance in broccoli. Sci Rep 7:41397

Li X, Li X, Fridman E, Tesso TT, Yu J (2015) Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. Proc Natl Acad Sci USA 112:11823-11828

Libby WJ (1962) Estimation of variance components of internode length in a cloned population of mimulus guttatus. Genetics 47:769-777

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016a) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12:e1005767

Liu Y, Yu L, Qu Y, Chen J, Liu X, Hong H, Liu Z, Chang R, Gilliham M, Qiu L, Guan R (2016b) *GmSALT3*, which confers improved soybean salt tolerance in the field, increases leaf Cl- exclusion prior to Na+ exclusion but does not improve early vigor under salinity. Front Plant Sci 7:1485

Lo M-T, Hinds DA, Tung JY, Franz C, Fan C-C, Wang Y, Smeland OB, Schork A, Holland D, Kauppi K, Sanyal N, Escott-Price V, Smith DJ, O'Donovan M, Stefansson H, Bjornsdottir G, Thorgeirsson TE, Stefansson K, McEvoy LK, Dale AM, Andreassen OA, Chen C-H (2017) Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nat Genet 49:152-156

Mamidi S, Lee RK, Goos JR, McClean PE (2014) Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). Plos One 9

Minitab Inc. (2010) Minitab 17 Statistical Software [Computer software]. State College, PA

Papiernik SK, Grieve CM, Lesch SM, Yates SR (2005) Effects of Salinity, Imazethapyr, and Chlorimuron Application on Soybean Growth and Yield. Commun Soil Sci Plan 36:951-967

Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, Shannon JG, Nguyen HT (2016) Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci Rep 6:19199

Phang TH, Shao G, Lam HM (2008) Salt tolerance in soybean. J Integr Plant Biol 50:1196-1212

Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

Qu HQ, Tien M, Polychronakos C (2010) Statistical significance in genetic association studies. Clin Invest Med 33:E266-E270

Rabie RK, Kumazawa K (1988) Effect of salt stress on nitrogen nutrition and yield quality of nodulated soybeans. Soil Sci Plant Nutr 34:385-391 v

Rohlf FJ (2000) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, Setauket, New York

Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol 165:1120-1132

Schneider R, Rolling W, Song Q, Cregan P, Dorrance AE, McHale LK (2016) Genome-wide association mapping of partial resistance to Phytophthora sojae in soybean plant introductions from the Republic of Korea. BMC Genomics 17:607

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825-830

Sharma JR (2006) Experimental field designs for plant breeding. In: Statistical and biometrical techniques in plant breeding. New Age International (P) Ltd., pp 15-29

Simko I, Hu J (2008) Population Structure in Cultivated Lettuce and Its Impact on Association Mapping. J Am Soc Hortic Sci 133:61-68

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8:e54985

Soto-Cerda B, Cloutier S (2012) Association mapping in plant genomes. In: Caliskan PM (ed) Genetic Diversity in Plants. InTech, pp 29-54

Várallyay G (2010) The impact of climate change on soils and on their water management. Agron Res 11:385-396 v

Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, Shannon JG, Nguyen HT (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. BMC Genomics 16:593

Wang D, Shannon MC (1999) Emergence and seedling growth of soybean cultivars and maturity groups under salinity. Plant Soil 214:117-124 v

Wicke B, Smeets E, Turkenburg W, Faaij A (2011) The global technical and economic potential of bioenergy from salt-affected soils. Energy Environ Sci 4:2669-2681v

Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P-c, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet 48:927-934

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208

Zeng A, Chen P, Korth K, Hancock F, Pereira A, Brye K, Wu C, Shi A (2017a) Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. Mol Breeding 37:30

Zeng A, Lara L, Chen P, Luan X, Hancock F, Korth K, Brye K, Pereira A, Wu C (2017b) Quantitative trait loci for chloride tolerance in ‘Osage’ soybean. Crop Sci 57:2345-2353

Zhang J, Wen Z, Li W, Zhang Y, Zhang L, Dai H, Wang D, Xu R (2017) Genome-wide association study for soybean cyst nematode resistance in Chinese elite soybean cultivars. Mol Breeding 37:60

Zhang JP, Singh A, Mueller DS, Singh AK (2015) Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. Plant J 84:1124-1136

Zhang M, Fang Y, Liang Z, Huang L (2012) Enhanced expression of Vacuolar H+-ATPase subunit E in the roots is associated with the adaptation of *Broussonetia papyrifera* to salt stress. PLoS One 7:e48183

Table 3.1 Statistics for leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) evaluated for salt tolerance in an association analysis of a diverse panel of 305 soybean accessions.

Trait	Min	Max	Mean	Root MSE	CV%	H^2	Pearson correlation		
							CCR	LSC	LCC
LSS	1.0	5.0	2.78	0.527	18.93	0.82	-0.922**	0.638**	0.743**
CCR	0.3	1.2	0.80	0.047	5.94	0.94		-0.606**	-0.700**
LSC (g kg^{-1})	0.03	1.70	0.45	0.306	76.98	0.29			0.692**
LCC (g kg^{-1})	2.7	18.2	7.85	1.743	22.21	0.63			

Min: minimum;

Max: maximum

Root MSE: square root of mean square error

CV(%): coefficient of variation

H^2: Broad-sense heritability

**: Correlation coefficients are significant at the 0.01 level (2-tailed)

Table 3.2 Several SNPs significantly associated with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) evaluated in a genome-wide association study using the SoySNP50K dataset for the diverse panel of 305 soybean accessions.

Marker	Chr.	Position	-log10(*P*)				Call Rate (%)	MA	MAF	Allele Refer	Cloned/Candidate genes [a]
			LSS	CCR	LSC	LCC					
Salt-20	3	38610964	23.9*	22.3*	7.2*	25.7*	99.7	C	0.40	C/G	*Glyma03g32900*[b]
Salt14056	3	38619995	28.9*	24.7*	10.3*	29.1*	100	G	0.44	C/G	*Glyma03g32900*[b]
Salt11655	3	38622492	24.4*	22.5*	7.8*	27.9*	99.7	T	0.41	G/T	*Glyma03g32900*[b]
ss715592375	5	7534622	2.5 ns	3.1 ns	2.4 ns	4.7*	98.0	G	0.22	A/G	No gene
ss715609949	11	27743052	3.8 ns	2.6 ns	5.6*	4.1ns	99.0	A	0.06	G/A	No gene
ss715611871	12	20568054	2.6 ns	2.5 ns	4.8*	3.3 ns	99.3	T	0.10	G/T	No gene
ss715616720	13	16871244	5.8*	4.9*	1.8 ns	3.2 ns	99.7	T	0.31	C/T	No gene
ss715623199	15	9138970	4.4*	2.2 ns	1.0 ns	1.3 ns	100	T	0.24	C/T	*Glyma.15g116200*
-log10(P) threshold [c]			**4.4**	**4.3**	**4.8**	**4.2**					

[a]: The candidate gene in Soybase Wm82 Genome Browser version 2 consist of significant SNP

[b]: The cloned gene for salt tolerance is named *Glyma03g32900* in SoyBase Wm82 Genome Browser version 1.

[c]: Threshold was calculated based on P-value using False Discovery Rate correction (Benjamini-Hochberg)

MA: Minor allele

MAF: Minor allele frequency

*: Significant association with salt tolerant traits

ns: None significant association with trait

Table 3.3 The most significant SNPs on each chromosome associated with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) detected in an association analysis using the WGRS-derived SNP dataset in the subset of 234 soybean accessions.

Marker	Chr.	Position	-log10(*P*)				Call Rate (%)	MA	MAF	Allele Refer	Cloned/Candidate genes [a]
			LSS	CCR	LSC	LCC					
1:4591993-SNV	1	4591993	1.1ns	2.3ns	5.5*	4.9*	100	T	0.50	T/C	No gene
2:38974662-SNV	2	38974662	4.5*	3.0ns	0.7ns	3.4ns	100	T	0.41	T/A	No gene
2:43513959-SNV	2	43513959	4.0ns	2.3ns	1.1ns	4.4*	100	C	0.49	T/C	*Glyma.02g247400*
Salt-20	3	38610964	29.0*	25.6*	7.2*	27.4*	99.6	C	0.45	C/G	*Glyma03g32900*[b]
Salt14056	3	38619995	32.4*	28.2*	8.2*	29.6*	100	G	0.46	C/G	*Glyma03g32900*[b]
Salt11655	3	38622492	27.2*	24.1*	6.2*	25.7*	99.6	T	0.45	G/T	*Glyma03g32900*[b]
5:2725777-SNV	5	2725777	1.1ns	2.3ns	5.5*	4.9*	100	T	0.50	T/C	*Glyma.05g031300*
6:3536892-SNV	6	3536892	1.1ns	2.3ns	5.5*	4.9*	100	A	0.50	A/G	No gene
6:38850839-SNV	6	38850839	4.8*	2.9ns	2.2ns	5.2ns	100	A	0.25	A/G	No gene
8:11762527-SNV	8	11762527	4.2ns	3.7ns	5.2*	6.6*	100	T	0.35	T/A	No gene
8:11859355-SNV	8	11859355	4.3*	3.4ns	3.6ns	5.3*	100	T	0.32	T/A	No gene
8:11869912-SNV	8	11869912	4.1ns	3.3ns	3.5ns	5.0*	100	T	0.3	T/C	*Glyma.08g153800*
8:11897542-SNV	8	11897542	4.1ns	3.3ns	3.3ns	4.6*	100	C	0.3	C/T	*Glyma.08g154400*
8:12168563-SNV	8	12168563	3.3ns	3.5ns	6.7*	4.4*	100	G	0.38	G/A	*Glyma.08g157400*
8:12177874-SNV	8	12177874	2.9ns	3.7ns	6.9*	3.4ns	100	A	0.48	A/T	*Glyma.08g157400*
8:12198547-SNV	8	12198547	3.8ns	4.1ns	7.6*	4.6*	100	C	0.4	C/T	No gene
8:12221598-SNV	8	12221598	3.0ns	2.8ns	6.7*	4.3ns	100	G	0.4	G/A	*Glyma.08g157700*

Marker	Chr.	Position	-log10(*P*)				Call Rate (%)	MA	MAF	Allele Refer	Cloned/Candidate genes [a]
			LSS	CCR	LSC	LCC					
8:18197526-SNV	8	18197526	3.5ns	3.2ns	3.1ns	5.0*	100	T	0.07	T/C	*Glyma.08g224400*
8:18363900-SNV	8	18363900	4.5*	3.6ns	2.2ns	4.6*	100	A	0.47	A/G	No gene
8:18378659-SNV	8	18378659	4.8*	4.5*	3.3ns	3.7ns	100	G	0.48	T/G	No gene
14:8102511-SNV	14	8102511	4.6*	3.7ns	1.2ns	2.2ns	100	T	0.49	T/C	No gene
14:47627721-SNV	14	47627721	1.2ns	0.9ns	5.6*	0.9ns	100	A	0.06	A/G	*Glyma.14g211300*
15:11833367-SNV	15	11833367	3.0ns	2.4ns	3.4ns	5.0*	100	C	0.49	C/T	*Glyma.15g143900*
16:34944055-SNV	16	34944055	4.9*	4.1ns	1.3ns	2.4ns	100	C	0.37	C/A	No gene
18:56065139-SNV	18	56065139	1.1ns	2.3ns	5.5*	4.9*	100	G	0.50	G/T	*Glyma.18g294400*
18:57153685-SNV	18	57153685	3.2ns	5.3*	0.8ns	3.2ns	100	T	0.23	T/C	No gene
18:57203235-SNV	18	57203235	4.9*	4.1ns	2.5ns	2.9ns	100	A	0.17	A/T	No gene
19:43567289-SNV	19	43567289	7.5*	6.7*	2.1ns	5.1*	100	T	0.20	T/G	*Glyma.19g175600*
20:37261341-SNV	20	37261341	3.3ns	4.6*	0.5ns	1.6ns	100	C	0.06	T/C	No gene
-log10(*P*) threshold [c]			**4.3**	**4.4**	**5.1**	**4.3**					

[a]: The candidate gene in Soybase Wm82 Genome Browser version 2 consisting of significant SNP

[b]: The cloned gene for salt tolerance is named *Glyma03g32900* in SoyBase Wm82 Genome Browser version 1.

[c]: Threshold was calculated based on P-value using False Discovery Rate correction (Benjamini-Hochberg)

MA: Minor allele

MAF: Minor allele frequency

*: Significant association with salt tolerant traits

ns: None significant association with trait

Table 3.4 New sources for salt tolerance from matching analysis of phenotype and GBM genotypes for known gene (*Glyma03g32900*) on Chr.03

PI	Name	Taxonomy	Phenotypic Group	Genotypes of GBMs			Prediction/ Suggestion
				Salt-20	Salt14056	Salt11655	
PI101404A		*G. soja*	Tolerant	Wt	Het	Wt	New allele
PI597458B		*G. soja*	Tolerant	Wt	Mut	Mut	New allele
PI342434		*G. max*	Tolerant	Wt	Mut	Wt	New allele
PI548198	T209	*G. max*	Tolerant	Wt	Mut	Wt	New allele
PI561389B	(Okura Natto)	*G. max*	Tolerant	Wt	Mut	Wt	New allele
PI407202	K15	*G. soja*	Tolerant	Wt	Mut	Wt	New allele
PI407220	K25-B	*G. soja*	Tolerant	Wt	Mut	Wt	New allele
PI424107A	74106	*G. soja*	Tolerant	Wt	Mut	Wt	New allele
PI479752	GD 50388-2	*G. soja*	Tolerant	Wt	Mut	Wt	New allele
PI407083	RB 1072	*G. soja*	Tolerant	Wt	Wt	Mut	New gene
PI468908		*G. max*	Tolerant	Wt	Wt	Wt	New gene
PI080837	Mejiro	*G. max*	Tolerant	Wt	Wt	Wt	New gene
PI417500	Escura A	*G. max*	Tolerant	Wt	Wt	Wt	New gene
PI424116	74116	*G. soja*	Tolerant	Wt	Wt	Wt	New gene
PI483460B		*G. soja*	Tolerant	Wt	Wt	Wt	New gene
PI562551	KC26	*G. soja*	Tolerant	Wt	Wt	Wt	New gene
PI438471	Fiskeby III	*G. max*	Tolerant check	Mut	Mut	Mut	
PI548656	Lee	*G. max*	Tolerant check	Mut	Mut	Mut	

PI	Name	Taxonomy	Phenotypic Group	Genotypes of GBMs			Prediction/ Suggestion
				Salt-20	Salt14056	Salt11655	
PI518664	Hutcheson	*G. max*	Sensitive check	Wt	Wt	Wt	
PI548657	Jackson	*G. max*	Sensitive check	Wt	Wt	Wt	

Wt: Wide-type allele of GBMs (Salt-20, Salt14056, Salt11655)

Mut: Mutant allele of GBMs (Salt-20, Salt14056, Salt11655)

Table S3.1 Variation of leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) of checks grown under 120 mM NaCl treatment.

Line	LSS[a]	CCR[b]	LSC ($g\ kg^{-1}$)	LCC $g\ kg^{-1}$)
Lee (tolerant check)	1.8±0.2	1.0±0.03	0.4±0.1	5.9±1.3
Fiskeby III (tolerant check)	1.0±0.1	1.1±0.01	0.2±0.1	4.9±0.1
Hutcheson (sensitive check)	4.3±0.1	0.6±0.01	0.6±0.1	9.8±1.0
Jackson (sensitive check)	4.8±0.1	0.5±0.04	0.9±0.2	12.8±2.1

a: Leaf scorch score based on a 1-5 scale
b: The ratio of leaf chlorophyll content after treatment dividing leaf chlorophyll content after treatment

Table S3.2 Genomic inflation factor (λ) of models for analyzing association with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) among 305 soybean lines using SoySNP50K dataset and the subset of 234 soybean lines selected from the original 305 genotypes using a 3.7M SNP dataset

Dataset	Population	Trait	Naïve (λ)	P model (λ)	EMMAX (λ)	The step of MLMM (λ)				
						1	2	3	4	5
SoySNP50K	Whole set	LSS	2.43	1.07	0.94	0.99	1.00	0.99	1.00	1.00
	(n=305)	CCR	3.95	1.10	0.95	0.99	1.00	1.00	0.98	0.98
		LSC	2.42	1.08	0.97	0.98	0.97	0.98	0.98	0.97
		LCC	2.21	1.16	0.95	1.01	1.00	0.99	0.98	0.98
3.7M SNPs	Subset	LSS	1.74	1.15	0.95	0.98	1.00	1.00	1.01	1.03
	(n=234)	CCR	1.98	1.20	0.97	0.99	1.02	0.94	0.95	0.96
		LSC	1.44	1.05	0.97	0.99	1.00	1.01	1.00	1.02
		LCC	2.08	1.17	0.93	0.98	0.98	0.98	0.98	0.96

Naïve: The general linear model

P model: The statistical model with correction for principal components

EMMAX: Efficient mixed-model association expedited

MLMM: Multi-locus mixed model

Table S3.3 The genes underlying the significant genomic regions associated with salt tolerance from Phytozone and Soybase databases

No	Gene name	Chr.	Start position	End position	ID	Protein family
1	*Glyma.02g143000*	2	14735028	14747294	58185	Guanylate-binding protein C N-terminal domain, Guanylate-binding protein C C-terminal domain
2	*Glyma.02g204300*	2	38958949	38970855	67290	Ion transport protein, Cyclic nucleotide-binding domain
3	*Glyma.02g204400*	2	38974176	38974403	67302	Unknown
4	*Glyma.02g247400*	2	43512032	43519329	73998	AAR2 protein
5	*Glyma.02g247500*	2	43512042	43519329	74015	Unknown
6	*Glyma.02g247000*	2	43478314	43481473	73944	Unknown
7	*Glyma.02g247100*	2	43492461	43494623	73950	Myb-like DNA-binding domain
8	*Glyma.02g247200*	2	43501098	43502907	73957	Unknown
9	*Glyma.02g247300*	2	43504539	43508138	73962	CLASP N terminal
10	*Glyma.05g031300*	5	2720956	2729817	171678	Unknown
11	*Glyma.06g046800*	6	3535731	3536501	213837	Unknown
12	*Glyma.08g145600*	8	11066152	11073328	323171	Cellulose synthase
13	*Glyma.08g145700*	8	11083230	11086950	323189	WD domain C G-beta repeat
14	*Glyma.08g145800*	8	11088887	11089249	323201	Prolamin-like (original Pfam: PF06915)
15	*Glyma.08g145900*	8	11092669	11096142	323204	Alcohol dehydrogenase GroES-like domain, Zinc-binding dehydrogenase
16	*Glyma.08g146000*	8	11103557	11106499	323213	TFIIS helical bundle-like domain
17	*Glyma.08g146100*	8	11119305	11124078	323239	Eam A-like transporter family
18	*Glyma.08g146200*	8	11125789	11129806	323260	Unknown

No	Gene name	Chr.	Start position	End position	ID	Protein family
19	*Glyma.08g146300*	8	11136129	11139057	323266	PAP_fibrillin
20	*Glyma.08g146400*	8	11139657	11142335	323301	BolA-like protein
21	*Glyma.08g146500*	8	11145877	11149248	323308	Actin
22	*Glyma.08g146600*	8	11157526	11158422	323317	Zinc finger C C3HC4 type (RING finger)
23	*Glyma.08g146700*	8	11166509	11172366	323322	Autophagy-related protein 13
24	*Glyma.08g146800*	8	11173407	11176423	323330	RNA polymerase Rpb4
25	*Glyma.08g146900*	8	11183844	11185540	323338	GRAS domain family
26	*Glyma.08g156500*	8	12100644	12103960	324766	No apical meristem (NAM) protein
27	*Glyma.08g156600*	8	12107082	12108961	324781	Unknown
28	*Glyma.08g156700*	8	12112065	12112778	324788	Unknown
29	*Glyma.08g157000*	8	12140034	12143379	324880	Ras family
30	*Glyma.08g157100*	8	12144886	12145586	324890	Unknown
31	*Glyma.08g157200*	8	12146949	12149210	324895	SF0 - RING FINGER PROTEIN 5
32	*Glyma.08g157300*	8	12156261	12158113	324901	Unknown
33	*Glyma.08g157400*	8	12162921	12193517	324907	SF9 - CALLOSE SYNTHASE 1-RELATED
34	*Glyma.08g157500*	8	12199277	12205472	324958	SF29 - P-LOOP CONTAINING NUCLEOSIDE TRIPHOSPHATE HYDROLASES SUPERFAMILY PROTEIN
35	*Glyma.08g157600*	8	12206851	12208844	324970	Unknown
36	*Glyma.08g157700*	8	12213783	12222299	324976	Citrate synthase
37	*Glyma.08g157800*	8	12226940	12231858	325023	Cyclic pyranopterin phosphate synthase / Molybdenum cofactor biosynthesis protein 1

No	Gene name	Chr.	Start position	End position	ID	Protein family
38	*Glyma.08g157900*	8	12235183	12236400	325034	VQ motif
39	*Glyma.08g158000*	8	12247781	12249854	325039	Unknown
40	*Glyma.08g158100*	8	12250248	12251336	325044	Unknown
41	*Glyma.08g158200*	8	12254113	12256668	325049	Unknown
42	*Glyma.08g158300*	8	12260220	12262810	325055	Zinc finger C C3HC4 type (RING finger)
43	*Glyma.08g158400*	8	12264351	12266109	325061	ATPase family associated with various cellular activities (AAA)
44	*Glyma.08g158500*	8	12280422	12281584	325066	Unknown
45	*Glyma.08g158600*	8	12290252	12297255	325071	SF34 - MAJOR FACILITATOR SUPERFAMILY DOMAIN-CONTAINING PROTEIN // SUBFAMILY NOT NAMED
46	*Glyma.08g158700*	8	12301021	12302256	325130	Unknown
47	*Glyma.08g222200*	8	18038277	18058503	336034	haloacid dehalogenase-like hydrolase, Cation transporter FATPase C N-terminus, Cation transporting ATPase C C-terminus, Ca2+-ATPase N terminal autoinhibitory domain, E1-E2 ATPase
48	*Glyma.08g222300*	8	18059686	18063561	336073	GDP-fucose protein O-fucosyltransferase
49	*Glyma.08g222400*	8	18068218	18073981	336083	Adaptin N terminal region, Coatomer beta C-terminal region
50	*Glyma.08g222500*	8	18075163	18077665	336098	Exo70 exocyst complex subunit
51	*Glyma.08g222600*	8	18078248	18079952	336103	Unknown
52	*Glyma.08g222700*	8	18079943	18083744	336112	Unknown
53	*Glyma.08g222800*	8	18085127	18088964	336123	Glycosyl transferase family 2
54	*Glyma.08g222900*	8	18090096	18091341	336132	Mitochondrial ATP synthase epsilon chain

No	Gene name	Chr.	Start position	End position	ID	Protein family
55	*Glyma.08g223000*	8	18093196	18102042	336138	Vps51 FVps67
56	*Glyma.08g223100*	8	18102559	18103619	336175	Unknown
57	*Glyma.08g223200*	8	18103948	18105613	336180	RNA recognition motif. (a.k.a. RRM C RBD C or RNP domain)
58	*Glyma.08g223300*	8	18107199	18109906	336189	ACT domain, D-isomer specific 2-hydroxyacid dehydrogenase C NAD binding domain, D-isomer specific 2-hydroxyacid dehydrogenase C catalytic domain
59	*Glyma.08g223400*	8	18127350	18129276	336198	Protein kinase domain
60	*Glyma.08g223500*	8	18129324	18135381	336203	Unknown
61	*Glyma.08g223600*	8	18138052	18142927	336209	Aldose 1-epimerase
62	*Glyma.08g223700*	8	18147188	18154565	336221	Protein kinase domain
63	*Glyma.08g223800*	8	18162779	18168705	336326	XPC-binding domain, UBA FTS-N domain, Ubiquitin family
64	*Glyma.08g223900*	8	18169513	18171372	336358	YGGT family
65	*Glyma.08g224000*	8	18175289	18182199	336363	Unknown
66	*Glyma.08g224100*	8	18183935	18184694	336389	Unknown
67	*Glyma.08g224200*	8	18187905	18191895	336400	Ergosterol biosynthesis ERG4 FERG24 family
68	*Glyma.08g224300*	8	18193065	18194992	336442	Eukaryotic family of unknown function (DUF1754)
69	*Glyma.08g224400*	8	18196846	18202474	336453	V-type H+-transporting ATPase subunit A (ATPeV1A, ATP6A)
70	*Glyma.08g224500*	8	18204060	18208035	336477	ATPase family associated with various cellular activities (AAA)
71	*Glyma.08g224600*	8	18208948	18211498	336490	PAP_fibrillin

No	Gene name	Chr.	Start position	End position	ID	Protein family
72	*Glyma.08g224700*	8	18211499	18213194	336500	Unknown
73	*Glyma.08g224800*	8	18220717	18224385	336507	Protein of unknown function (DUF1666)
74	*Glyma.08g224900*	8	18225217	18226261	336563	Unknown
75	*Glyma.08g225000*	8	18232440	18234931	336581	Mitochondrial carrier protein
76	*Glyma.08g225100*	8	18245614	18246213	336591	Auxin responsive protein
77	*Glyma.08g225200*	8	18246538	18247740	336596	Auxin responsive protein
78	*Glyma.08g225300*	8	18248653	18252894	336601	Unknown
79	*Glyma.08g225400*	8	18252173	18259445	336611	Unknown
80	*Glyma.08g225500*	8	18260357	18267956	336647	SF11 - PYROPHOSPHATE-ENERGIZED MEMBRANE PROTON PUMP 2-RELATED
81	*Glyma.08g225600*	8	18260573	18261444	336719	Unknown
82	*Glyma.08g225700*	8	18275851	18276717	336724	Unknown
83	*Glyma.08g225800*	8	18290918	18300538	336732	Prenyltransferase and squalene oxidase repeat
84	*Glyma.08g225900*	8	18303275	18310432	336752	Protein tyrosine kinase, Leucine Rich Repeat, Leucine rich repeat N-terminal domain
85	*Glyma.08g226000*	8	18321377	18335864	336787	START domain
86	*Glyma.08g226100*	8	18339384	18347070	336807	YT521-B-like domain
87	*Glyma.08g226200*	8	18352624	18353085	336864	Unknown
88	*Glyma.08g226300*	8	18366397	18366591	336868	Unknown
89	*Glyma.08g226400*	8	18370226	18374060	336871	Sugar (and other) transporter
90	*Glyma.08g226500*	8	18384328	18385039	336881	Pyruvate kinase C barrel domain

No	Gene name	Chr.	Start position	End position	ID	Protein family
91	*Glyma.08g226600*	8	18391180	18403568	336885	Complex I intermediate-associated protein 30 (CIA30), NmrA-like family
92	*Glyma.08g226700*	8	18405334	18420828	336936	PHD-finger, Zinc finger C ZZ type, Histone acetylation protein, TAZ zinc finger
93	*Glyma.08g226800*	8	18444007	18444965	336977	Cupin
94	*Glyma.08g226900*	8	18449457	18452338	336982	Unknown
95	*Glyma.08g227000*	8	18464236	18469579	336992	bZIP transcription factor
96	*Glyma.08g227100*	8	18470879	18474036	337000	Unknown
97	*Glyma.08g227200*	8	18494370	18499681	337007	Protein kinase domain, EF hand
98	*Glyma.08g227300*	8	18507623	18515386	337019	Domain of unknown function (DUF3546), Arsenite-resistance protein 2
99	*Glyma.08g227400*	8	18516376	18517463	337052	Core histone H2A FH2B FH3 FH4
100	*Glyma.08g227500*	8	18518682	18527092	337057	FAR1 DNA-binding domain, SWIM zinc finger, MULE transposase domain
101	*Glyma.08g227600*	8	18535293	18537847	337141	Isoprenylcysteine carboxyl methyltransferase (ICMT) family
102	*Glyma.08g227700*	8	18539712	18545318	337152	AP2 domain
103	*Glyma.08g227800*	8	18541933	18542964	337172	Unknown
104	*Glyma.08g227900*	8	18551920	18567339	337177	Unknown
105	*Glyma.08g228000*	8	18584819	18594278	337189	Vacuolar protein sorting protein 36 Vps36, EAP30 FVps36 family
106	*Glyma.08g228100*	8	18598216	18600523	337212	WD domain C G-beta repeat
107	*Glyma.08g228200*	8	18601501	18602814	337228	Cupin

No	Gene name	Chr.	Start position	End position	ID	Protein family
108	*Glyma.08g228300*	8	18607555	18613550	337233	Oxidoreductase FAD-binding domain, Oxidoreductase NAD-binding domain
109	*Glyma.08g228400*	8	18621259	18621791	337246	S locus-related glycoprotein 1 binding pollen coat protein (SLR1-BP)
110	*Glyma.08g228500*	8	18636326	18636654	337252	Unknown
111	*Glyma.08g228600*	8	18648914	18649393	337258	S locus-related glycoprotein 1 binding pollen coat protein (SLR1-BP)
112	*Glyma.08g228700*	8	18660718	18661378	337264	Serine carboxypeptidase
113	*Glyma.08g228800*	8	18684920	18691455	337270	haloacid dehalogenase-like hydrolase, E1-E2 ATPase
114	*Glyma.08g228900*	8	18695544	18698469	337302	Dof domain C zinc finger
115	*Glyma.08g229000*	8	18698657	18698812	337320	Unknown
116	*Glyma.08g229100*	8	18702845	18708468	337323	Protein kinase domain, Leucine Rich Repeat
117	*Glyma.08g229200*	8	18714460	18714990	337345	Unknown
118	*Glyma.08g229300*	8	18717424	18718265	337350	Unknown
119	*Glyma.08g229400*	8	18720040	18747811	337354	Double-stranded RNA binding motif, Helicase associated domain (HA2), Helicase conserved C-terminal domain, Oligonucleotide Foligosaccharide-binding (OB)-fold, DEAD FDEAH box helicase
120	*Glyma.08g229500*	8	18753050	18753271	337399	Unknown
121	*Glyma.08g229600*	8	18766704	18782530	337402	SART-1 family
122	*Glyma.08g229700*	8	18795227	18796825	337434	F-box domain
123	*Glyma.08g229800*	8	18800851	18805120	337444	WD domain C G-beta repeat
124	*Glyma.08g229900*	8	18806959	18808642	337475	Unknown

No	Gene name	Chr.	Start position	End position	ID	Protein family
125	*Glyma.08g230000*	8	18819571	18822355	337486	Hsp20 Falpha crystallin family
126	*Glyma.08g230100*	8	18824513	18826014	337492	Pathogenesis-related protein Bet v I family
127	*Glyma.08g230200*	8	18838443	18843225	337498	Unknown
128	*Glyma.08g230300*	8	18841911	18842087	337509	Unknown
129	*Glyma.08g230400*	8	18847360	18848525	337512	Pathogenesis-related protein Bet v I family
130	*Glyma.08g230500*	8	18862871	18864462	337518	Pathogenesis-related protein Bet v I family
131	*Glyma.08g230600*	8	18879029	18881859	337524	BURP domain
132	*Glyma.08g230700*	8	18897439	18898641	337531	Albumin I
133	*Glyma.08g230800*	8	18911649	18919032	337537	Protein of unknown function (DUF3531)
134	*Glyma.08g230900*	8	18924711	18928709	337594	WD domain C G-beta repeat
135	*Glyma.08g231000*	8	18933180	18938287	337602	HAD superfamily C subfamily IIIB (Acid phosphatase)
136	*Glyma.08g231100*	8	18949122	18952446	337609	Protein kinase domain
137	*Glyma.08g231200*	8	18962367	18968461	337615	Unknown
138	*Glyma.08g231300*	8	18969636	18977694	337625	Helicase conserved C-terminal domain, Zinc finger C-x8-C-x5-C-x3-H type (and similar), RNA recognition motif. (a.k.a. RRM C RBD C or RNP domain), DEAD FDEAH box helicase
139	*Glyma.08g231400*	8	18981579	18991368	337695	C3HC zinc finger-like
140	*Glyma.08g231500*	8	18999077	19002548	337708	DnaJ domain
141	*Glyma.08g231600*	8	19016678	19022397	337736	Metallo-beta-lactamase superfamily
142	*Glyma.08g231700*	8	19038065	19039590	337749	Unknown
143	*Glyma.08g231800*	8	19053865	19057760	337753	Remorin C C-terminal region

No	Gene name	Chr.	Start position	End position	ID	Protein family
144	*Glyma.08g231900*	8	19064682	19066584	337772	Protein of unknown function (DUF1138)
145	*Glyma.08g232000*	8	19071849	19077651	337799	SPFH domain F Band 7 family
146	*Glyma.08g232100*	8	19083679	19090848	337805	Cyclic nucleotide-binding domain, Domain of unknown function (DUF3354), Ion transport protein
147	*Glyma.14g211300*	14	47625904	47630956	618711	Universal stress protein family
148	*Glyma.14g211400*	14	47633769	47635292	618727	Unknown
149	*Glyma.15g116200*	15	9136796	9139582	641492	Unknown
150	*Glyma.15g143800*	15	11824186	11826677	645590	cAMP-regulated phosphoprotein Fendosulfine conserved region
151	*Glyma.15g143900*	15	11830914	11834773	645598	Unknown
152	*Glyma.15g144000*	15	11843111	11847542	645606	Myb-like DNA-binding domain
153	*Glyma.15g144100*	15	11848711	11854145	645612	Unknown
154	*Glyma.16g185500*	16	34731233	34738502	693125	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
155	*Glyma.16g185600*	16	34733765	34734205	693130	Unknown
156	*Glyma.16g182700*	16	34393343	34396288	692988	Leucine Rich Repeat
157	*Glyma.16g182800*	16	34408229	34410993	692993	Unknown
158	*Glyma.16g182900*	16	34436482	34437114	692999	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
159	*Glyma.16g183000*	16	34451223	34453642	693002	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
160	*Glyma.16g183100*	16	34495057	34498580	693006	Unknown
161	*Glyma.16g183200*	16	34497329	34498190	693013	Glycosyl transferase family 90

No	Gene name	Chr.	Start position	End position	ID	Protein family
162	*Glyma.16g183300*	16	34500477	34503290	693019	Leucine Rich Repeat
163	*Glyma.16g183400*	16	34513882	34516747	693025	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
164	*Glyma.16g183500*	16	34530582	34533663	693029	Leucine Rich Repeat
165	*Glyma.16g183600*	16	34564893	34567538	693035	Leucine Rich Repeat
166	*Glyma.16g183700*	16	34575120	34579101	693039	Unknown
167	*Glyma.16g183800*	16	34577861	34578791	693044	Unknown
168	*Glyma.16g183900*	16	34579110	34580142	693049	Glycosyl transferase family 90
169	*Glyma.16g184000*	16	34585888	34589687	693056	Leucine Rich Repeat
170	*Glyma.16g184100*	16	34595483	34596460	693061	Protease inhibitor Fseed storage FLTP family
171	*Glyma.16g184200*	16	34604592	34609135	693066	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
172	*Glyma.16g184300*	16	34615252	34618050	693071	Leucine rich repeat N-terminal domain, Leucine Rich Repeat
173	*Glyma.16g184400*	16	34621915	34622244	693075	Protease inhibitor Fseed storage FLTP family
174	*Glyma.16g184500*	16	34647692	34656805	693078	Leucine Rich Repeat
175	*Glyma.16g184600*	16	34658999	34659573	693083	Unknown
176	*Glyma.16g184700*	16	34665898	34668573	693088	Leucine Rich Repeat
177	*Glyma.16g184800*	16	34673796	34676594	693092	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
178	*Glyma.16g184900*	16	34680005	34680571	693096	Unknown
179	*Glyma.16g185000*	16	34681294	34682953	693102	Glycosyl transferase family 90
180	*Glyma.16g185100*	16	34686428	34688908	693108	Leucine Rich Repeat

No	Gene name	Chr.	Start position	End position	ID	Protein family
181	*Glyma.16g185200*	16	34692049	34694364	693111	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
182	*Glyma.16g185300*	16	34701489	34704287	693116	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
183	*Glyma.16g185400*	16	34712976	34715775	693119	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
184	*Glyma.16g185700*	16	34742361	34744170	693134	Protease inhibitor Fseed storage FLTP family
185	*Glyma.16g185800*	16	34747498	34755392	693138	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
186	*Glyma.16g185900*	16	34757274	34758649	693143	CAP160 repeat
187	*Glyma.16g186000*	16	34761064	34763799	693147	Glycosyl transferase family 90
188	*Glyma.16g186100*	16	34768992	34773409	693158	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
189	*Glyma.16g186200*	16	34780811	34783568	693164	Leucine Rich Repeat
190	*Glyma.16g186300*	16	34792208	34792874	693168	Unknown
191	*Glyma.16g186400*	16	34799381	34802710	693174	Leucine rich repeat N-terminal domain, Leucine Rich Repeat
192	*Glyma.16g186500*	16	34820205	34823003	693180	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
193	*Glyma.16g186600*	16	34913608	34916555	693183	Leucine Rich Repeat
194	*Glyma.16g186700*	16	34921963	34924758	693189	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
195	*Glyma.16g186800*	16	34928102	34928675	693193	Unknown
196	*Glyma.16g186900*	16	34940125	34943209	693197	Leucine Rich Repeat

No	Gene name	Chr.	Start position	End position	ID	Protein family
197	*Glyma.18g043100*	18	3666225	3670526	746042	Cytochrome c oxidase assembly protein CtaG FCox11
198	*Glyma.18g202200*	18	48297961	48298505	765444	Serine carboxypeptidase
199	*Glyma.18g202300*	18	48298637	48298855	765449	Wound-induced protein
200	*Glyma.18g279700*	18	56065123	56066082	776626	Unknown
201	*Glyma.18g279800*	18	56074203	56079359	776637	GDSL FSGNH-like Acyl-Esterase family found in Pmr5 and Cas1p (original Pfam: PF03005)
202	*Glyma.18g279900*	18	56084932	56088020	776643	GDSL FSGNH-like Acyl-Esterase family found in Pmr5 and Cas1p (original Pfam: PF03005)
203	*Glyma.18g280000*	18	56093039	56095398	776649	Unknown
204	*Glyma.18g280100*	18	56094747	56095007	776655	Unknown
205	*Glyma.18g293400*	18	57143356	57144752	778843	Unknown
206	*Glyma.18g293500*	18	57158122	57158981	778855	Unknown
207	*Glyma.18g293600*	18	57163612	57164439	778860	Unknown
208	*Glyma.18g293700*	18	57165439	57166071	778865	Unknown
209	*Glyma.18g293800*	18	57166496	57170812	778870	PAP2 superfamily
210	*Glyma.18g293900*	18	57173334	57175247	778929	Protein of unknown function (DUF674)
211	*Glyma.18g294000*	18	57176110	57177861	778935	Protein of unknown function (DUF674)
212	*Glyma.18g294100*	18	57178773	57181329	778941	Weak chloroplast movement under blue light
213	*Glyma.18g294200*	18	57197200	57198197	778952	Protease inhibitor Fseed storage FLTP family
214	*Glyma.18g294300*	18	57205035	57207996	778957	Peptidase C26
215	*Glyma.18g294400*	18	57209293	57217181	778965	Protein kinase domain

No	Gene name	Chr.	Start position	End position	ID	Protein family
216	*Glyma.19g165500*	19	42636072	42644217	803013	C5HC2 zinc finger, jmjN domain, JmjC domain C hydroxylase
217	*Glyma.19g175600*	19	43561808	43569019	805159	LNS2 (Lipin FNed1 FSmp2) lipin C N-terminal conserved region
218	*Glyma.20g132100*	20	37250648	37255612	837187	SWIM zinc finger, MULE transposase domain, FAR1 DNA-binding domain
219	*Glyma.20g132200*	20	37257212	37261238	837232	Unknown
220	*Glyma.20g132300*	20	37263193	37263944	837250	4F5 protein family
221	*Glyma.20g132400*	20	37266529	37269324	837257	Leucine Rich Repeat, Leucine rich repeat N-terminal domain
222	*Glyma.20g132500*	20	37270462	37274826	837268	Mpv17 F PMP22 family

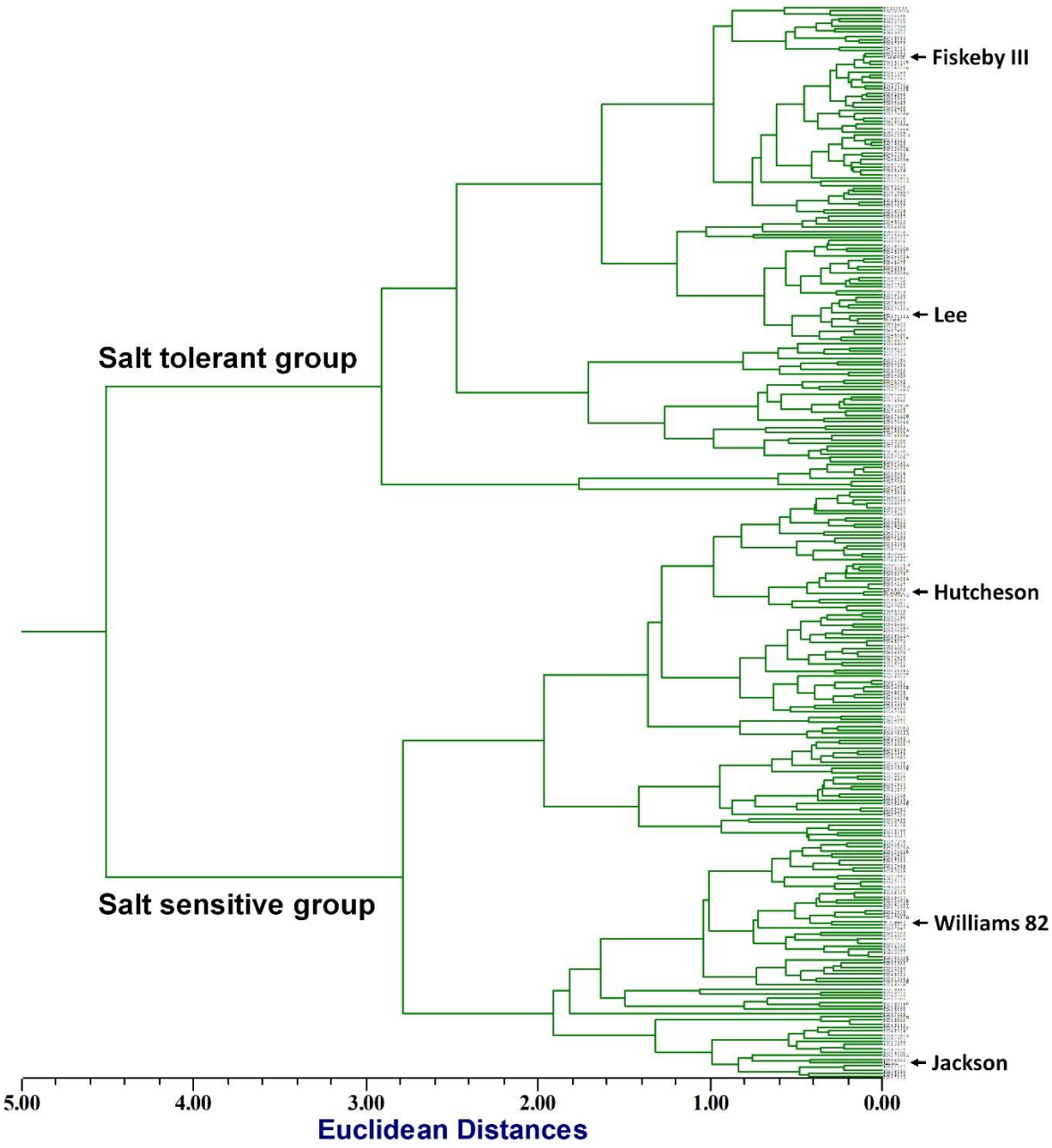


Figure 3.1 Phenotypic dendrogram for salt tolerance variation among 305 soybean accessions by combining leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC)

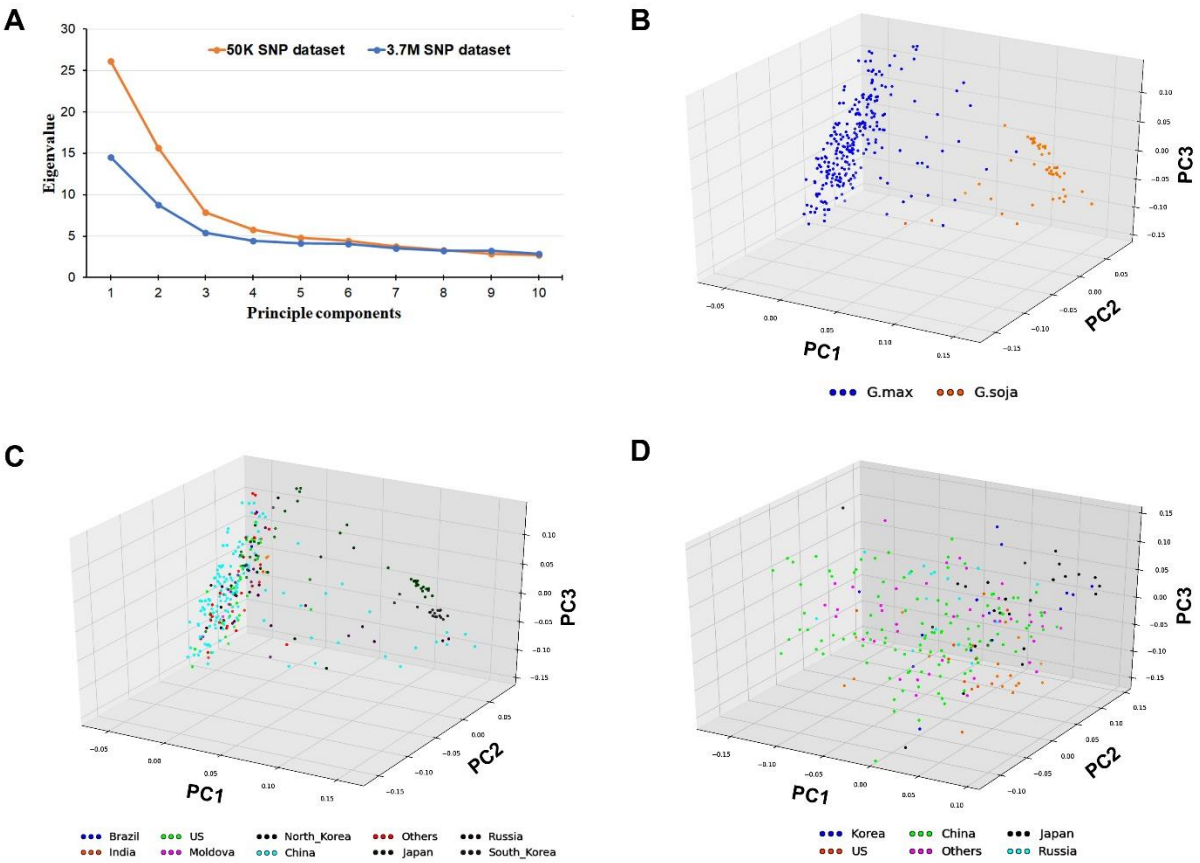


Figure 3.2 Population structure by principal component analysis (PCA) of two SNP datasets.

(A) Screen plot of the first 10 principal components (PCs) and their contribution to Eigenvalue by analyzing SoySNP50K dataset (*yellow*) and 3.7M SNP dataset (*blue*). (B) 3D scatterplot showed the first three PCs from SoySNP50K dataset corresponding to taxonomic groups for 305 soybean accessions. (C) 3D scatterplot showed the first three PCs from SoySNP50K dataset corresponding to original groups of 305 soybean accessions. (D) 3D scatterplot showed the first three PCs from 3.7M SNP dataset corresponding to 234 soybean accessions selected from the original 305 accessions

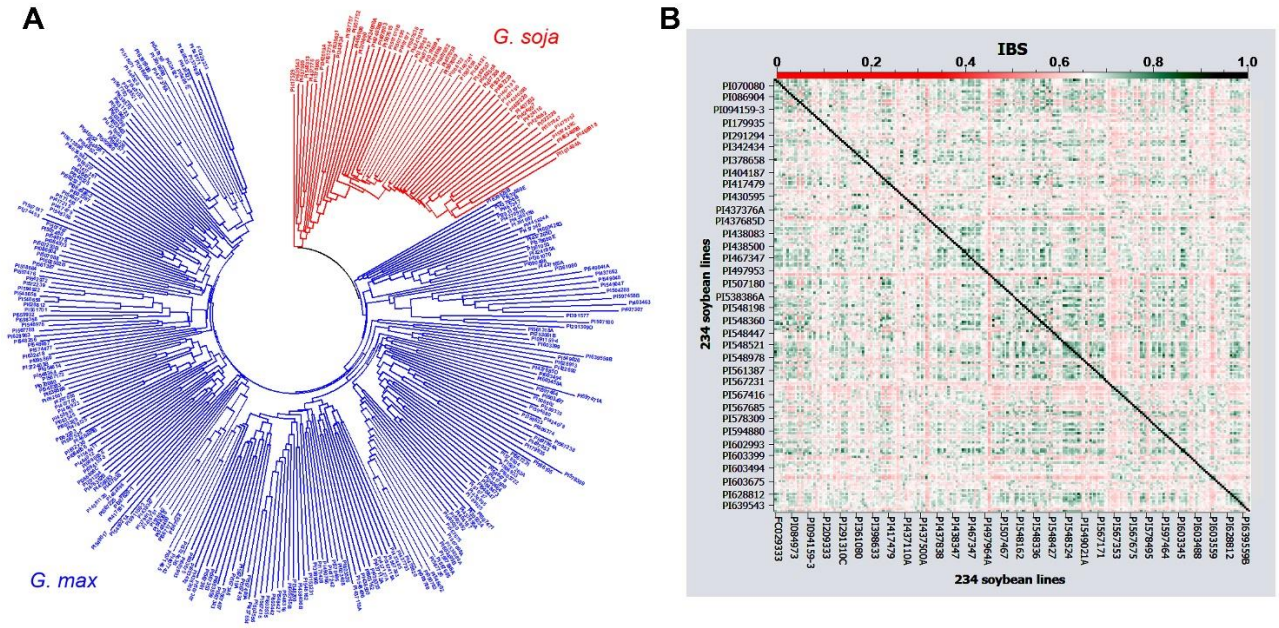


Figure 3.3 Relationships among 305 soybean accessions using SoySNP50K dataset and the subset of 234 soybean accessions using 3.7M SNP dataset.
(A) Phylogenetic tree of 305 soybean accessions using SoySNP50K dataset. (B) Heatmap plot showing the relationship among 234 soybean accessions using 3.7M SNP dataset

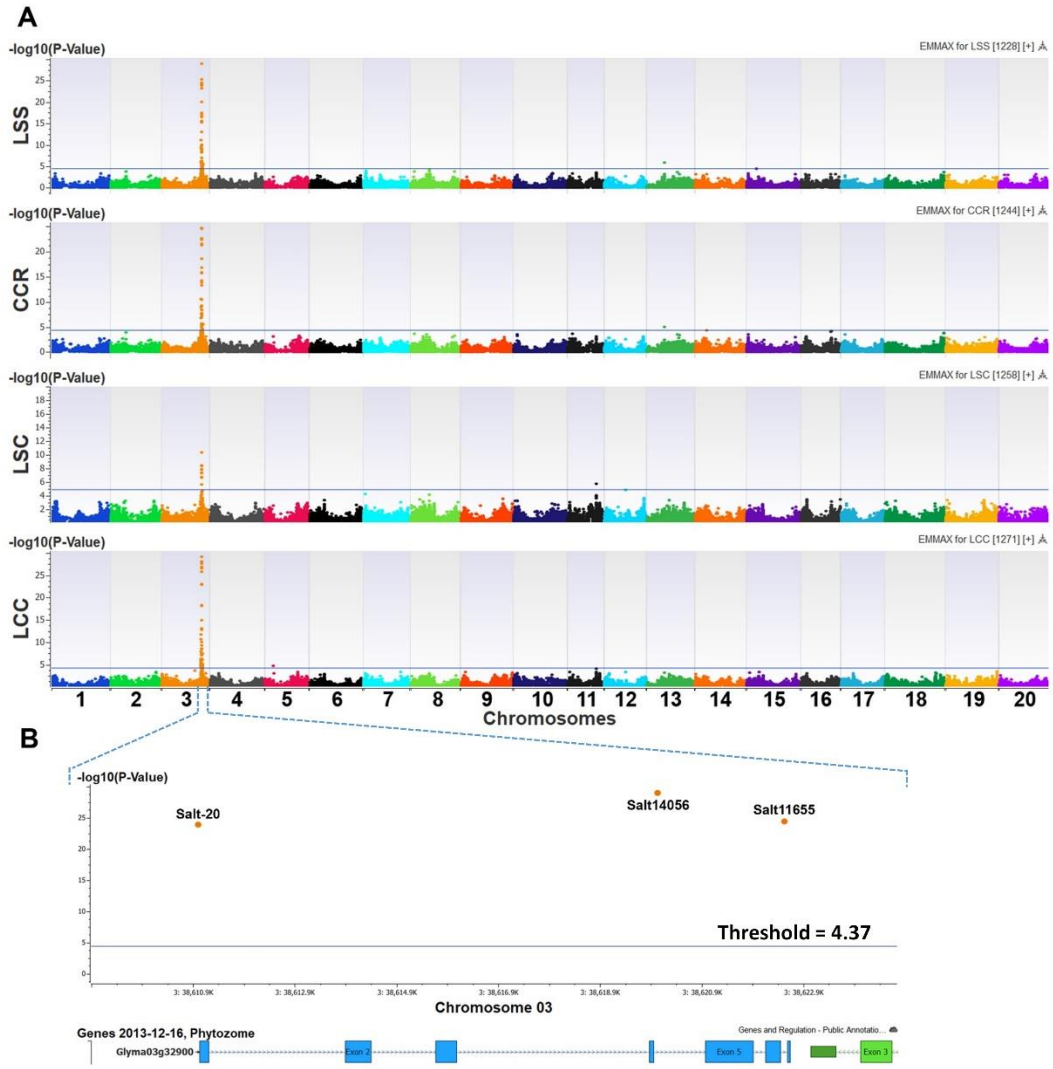


Figure 3.4 Association mapping of salt tolerance using EMMAX with a SoySNP50K dataset for 305 diverse soybean accessions.
(A) Manhattan plots showed association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC). (B) Three GBMs associated with salt tolerance and their position inside the known gene (*Glyma03g32900*), named in SoyBase Wm82 Genome Browser version 1

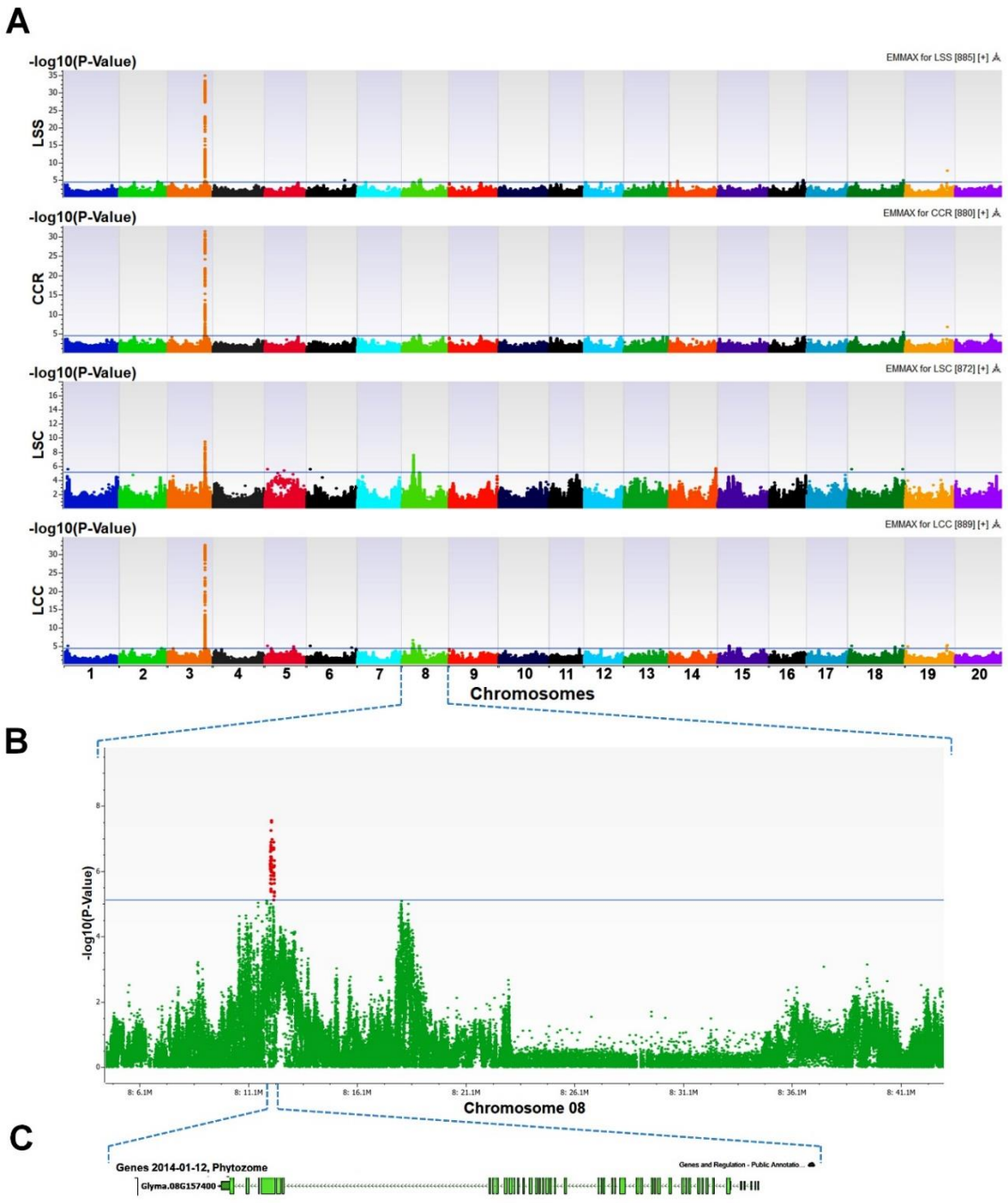


Figure 3.5 Association mapping of salt tolerance using EMMAX with 3.7M SNP dataset for the subset of 234 soybean accessions.
(A) Manhattan plots showed association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC). (B) The significant genomic region associated with LSC on Chr. 8. (C) One of the putative candidate gene underlying a minor locus for salt tolerance

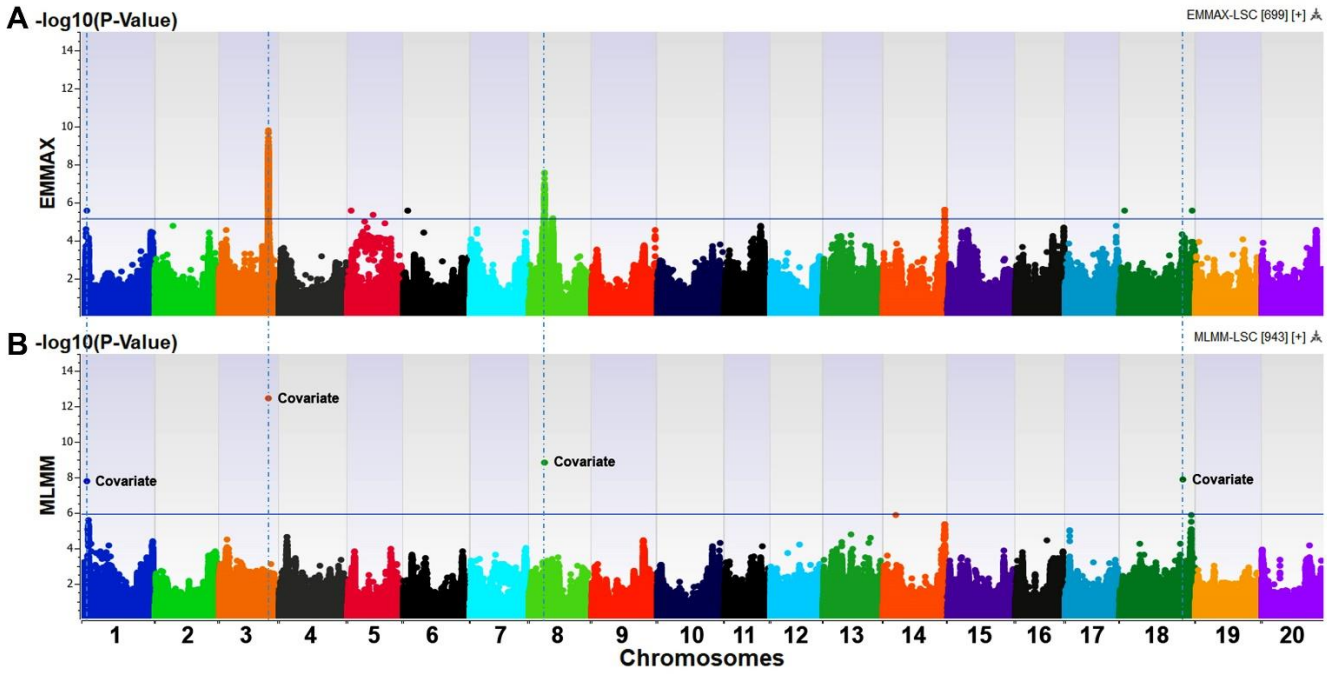


Figure 3.6 Comparison between EMMAX and MLMM in GWAS for salt tolerance. (A) Manhattan plot showed association of SNPs with leaf sodium content (LSC) by sequence-based GWAS using EMMAX. (B) Manhattan plot showed covariates (the most significant SNP after each step in stepwise analysis) associated with LSC by sequence-based GWAS using MLMM

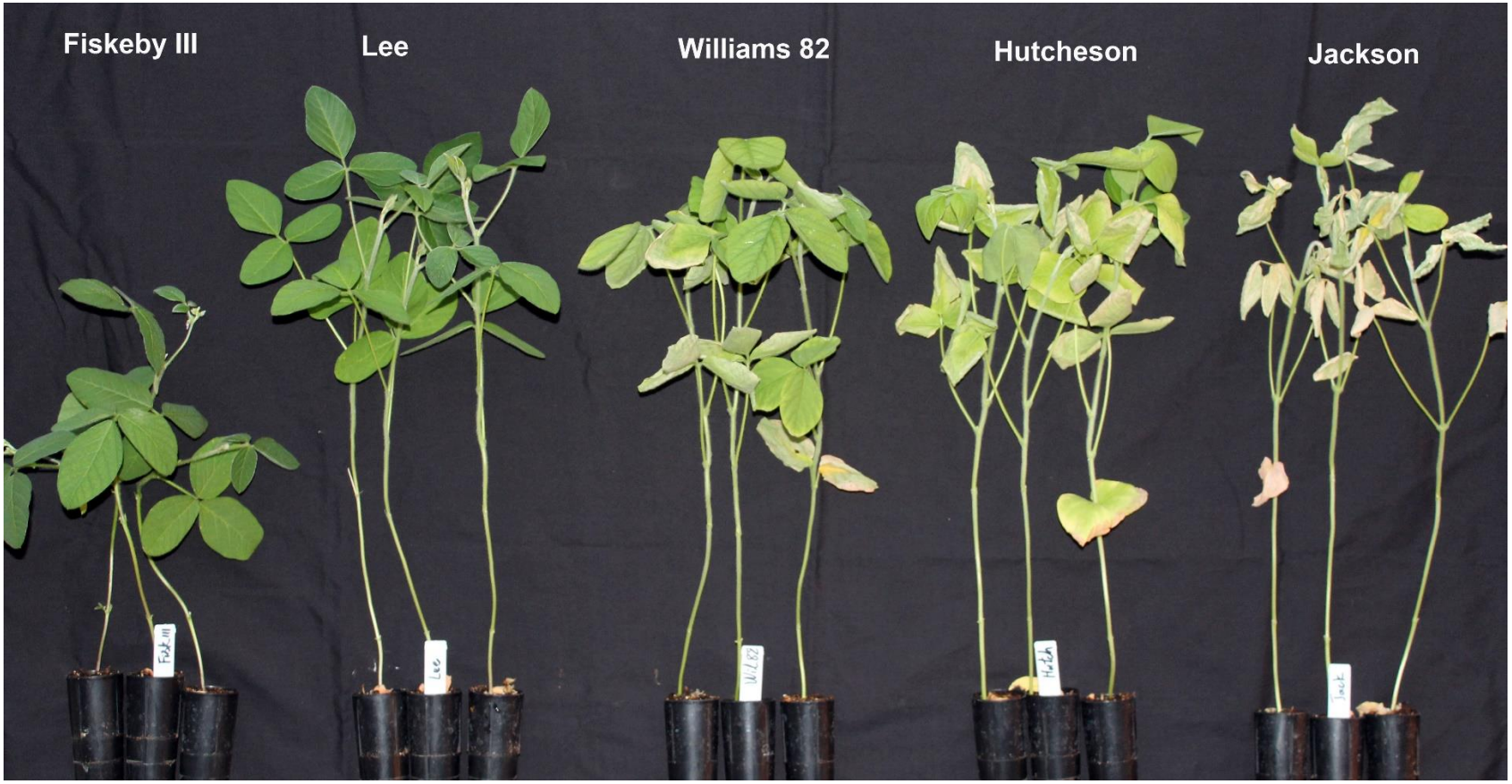


Figure S3.1 Comparisons of salt tolerance based on leaf scorch between cultivars Fiskeby III and Lee (salt tolerant checks), Hutcheson and Jackson (salt sensitive checks), Williams 82 (the soybean reference cultivar), grown under a 120 mM NaCl treatment

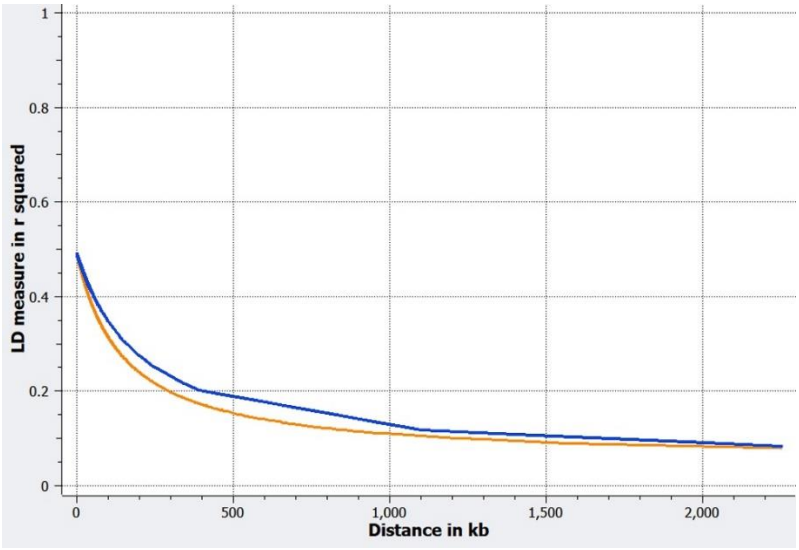


Figure S3.2 Linkage disequilibrium (LD) decay plot of coefficient of correlation (r^2) between adjacent marker pairs and genomic distance (kb)
The fitted curves are based on nonlinear regression using SoySNP50K dataset from 305 diverse genotypes (*yellow*) and using 3.7M SNP dataset in the subset from 234 genotypes selected from the original 305 accessions (*blue*)

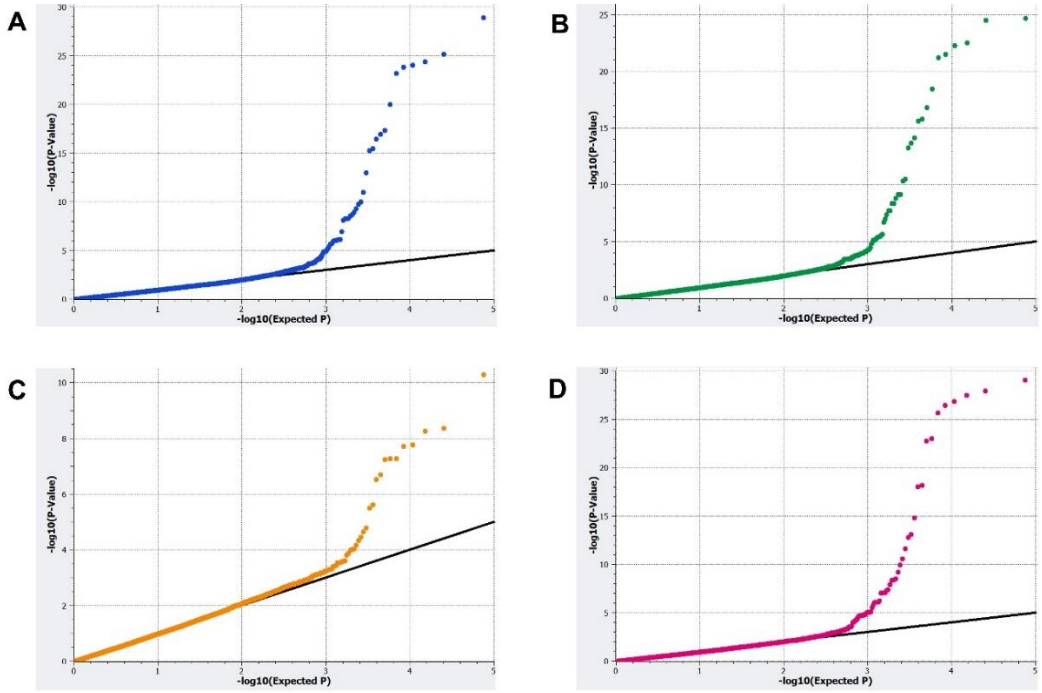


Figure S3.3 Quantile-quantile (Q-Q) plots showing the expected -log10(*P*) compared to the observed -log10(*P*), the results of statistical testing (EMMAX) for association across 37,573 SNPs from SoySNP50K dataset with leaf scorch score (A), chlorophyll content ratio (B), leaf sodium content (C) and leaf chloride content (D) among 305 genetically diverse soybean genotypes.

Most SNPs matched with solid lines [expected -log10(*P*) = observed -log10(*P*)] were unassociated SNPs, on the other hand, sharp curves at the end presented the number of true associations

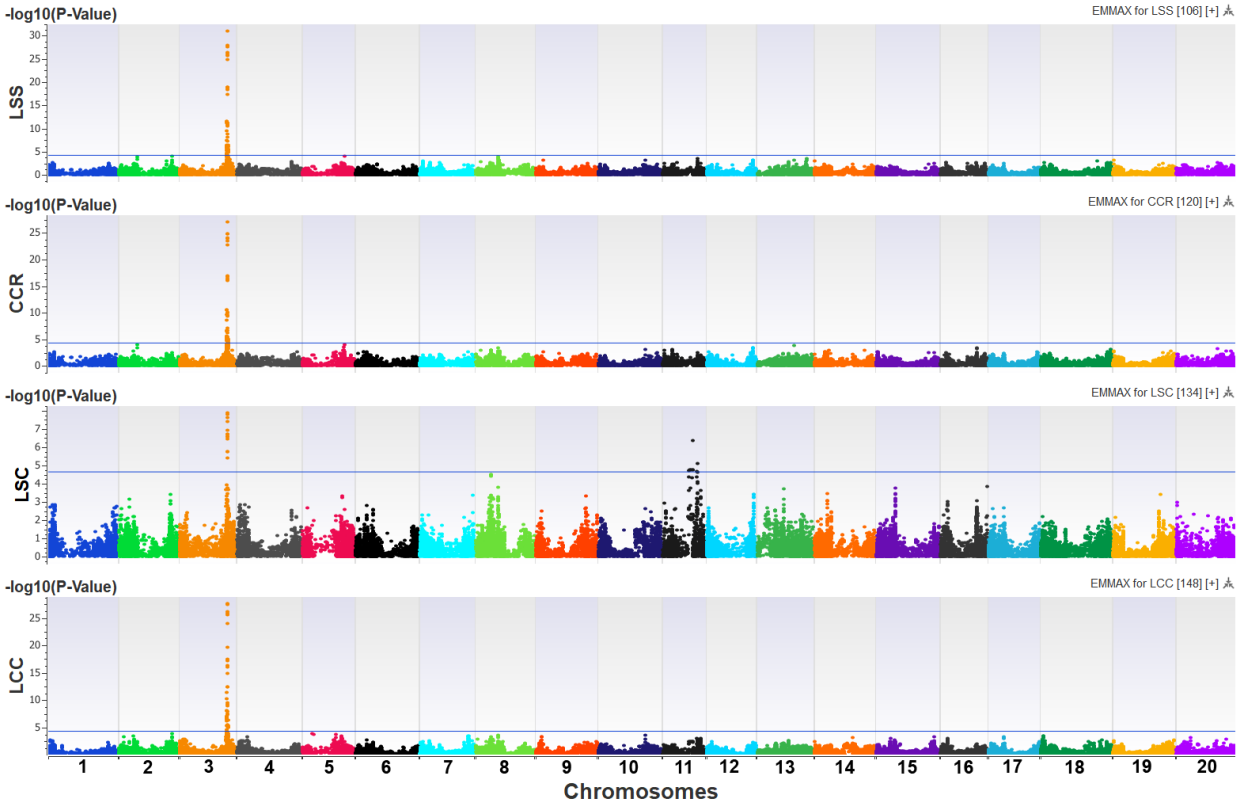


Figure S3.4 Manhattan plots showing association of SNPs distributed throughout 20 chromosomes with leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC) and leaf chloride content (LCC) using 37,573 SNPs from SoySNP50K dataset from the subset of 234 diverse accessions selected from the original 305 accessions

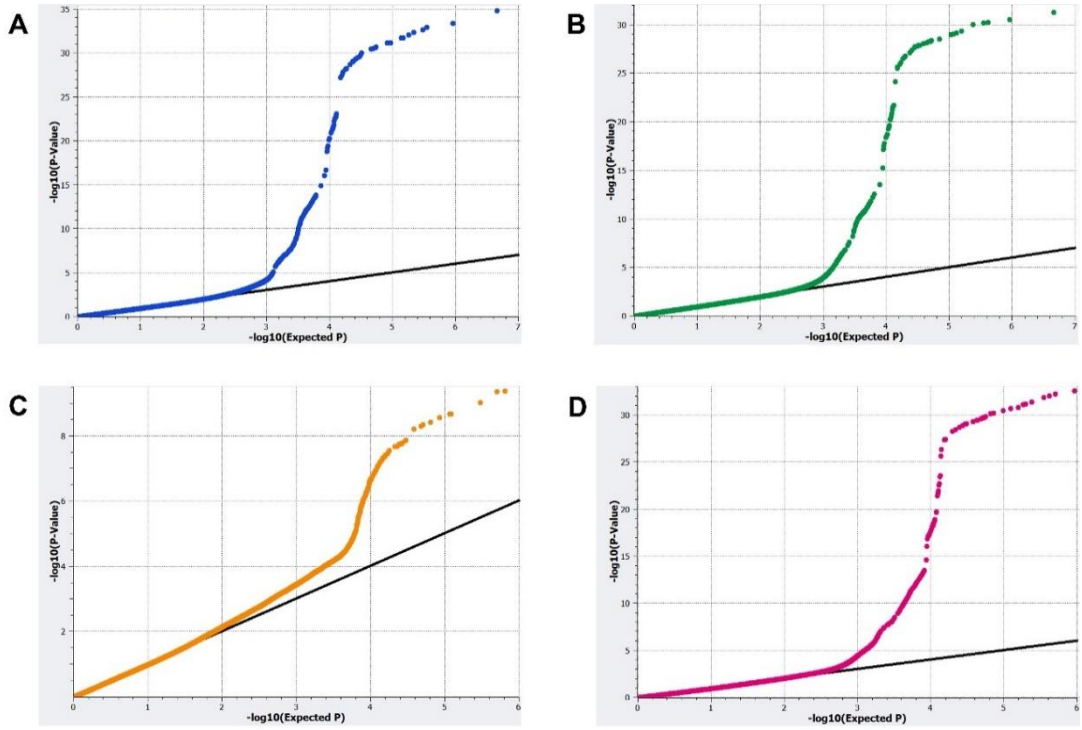


Figure S3.5 Quantile-quantile (Q-Q) plots showing the expected -log10(*P*) compared to the observed -log10(*P*), the results of statistical testing (EMMAX) for association across 2,280,225 polymorphic SNPs from 3.7M SNP dataset with leaf scorch score (A), chlorophyll content ratio (B), leaf sodium content (C) and leaf chloride content (D) in the subset of 234 soybean lines selected from the original population of 305 genotypes. SNPs matched with solid lines [expected -log10(*P*) = observed -log10(*P*)] were unassociated SNPs, on the other hand, sharp curves at the end presented the number of true associations

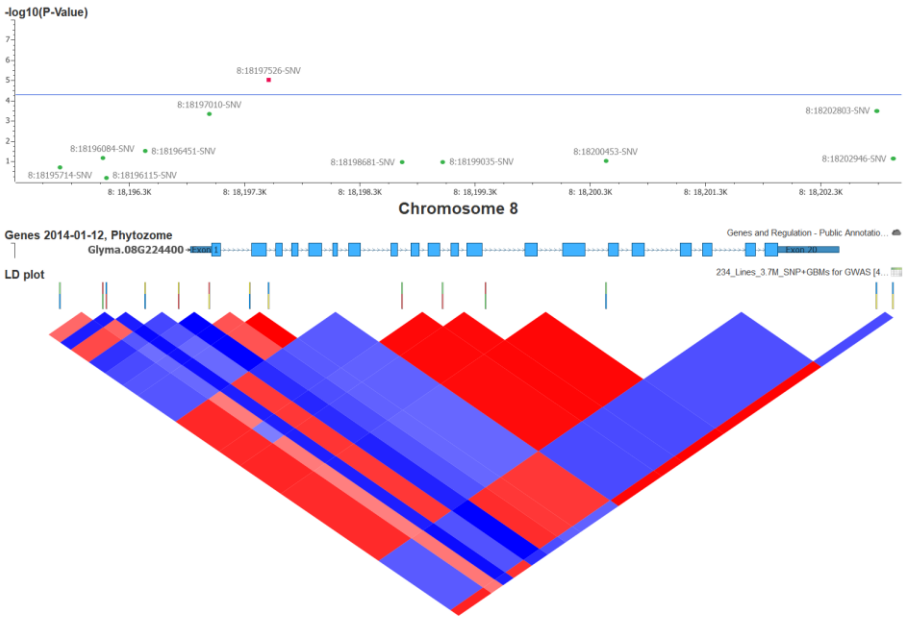


Figure S3.6 Significant SNPs in association with salt tolerance with the putative candidate gene (*Glyma.08G224400*) as a minor locus for salt tolerance, and LD block in this genomic region

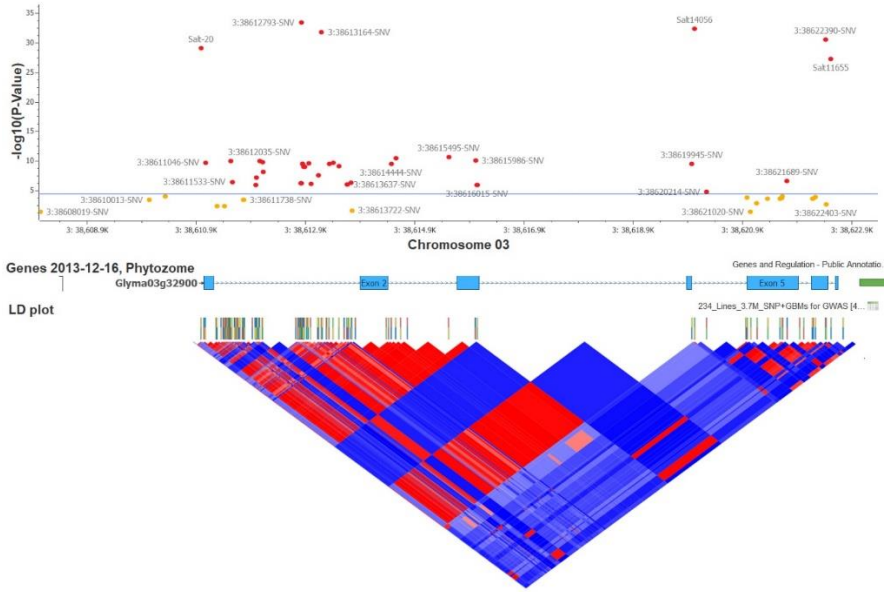


Figure S3.7 The significant SNPs underlying the known gene (*Glyma03g32900*) on Chr. 3 (*red dots*) associated with salt tolerance.
The known gene (*Glyma03g32900*) named in SoyBase Wm82 Genome Browser version 1, and LD block in this genomic region

Chapter 4:

PRELIMINARY STUDIES FOR FUTURE SALT TOLERANCE GENES

Testing the effects of mapped minor locus for salt tolerance

Rationale and goals

Other than the characterized gene (*Glyma03g32900*) on Chr. 3 (Do et al. 2016; Guan et al. 2014; Qi et al. 2014), many minor QTL and minor genomic regions associated with salt tolerance have been reported and flanking markers recommended for marker-assisted selection (MAS) in soybean (Chen et al. 2008; Do et al. 2018; Kan et al. 2015; Zeng et al. 2017a; Zeng et al. 2017b). Seven minor putative QTL significantly associated with salt tolerance loci for salt tolerance under greenhouse and field conditions were identified on Chrs. 2, 7, 9, 11, 14, 15 and 18 (Chen et al. 2008). A minor QTL for chloride tolerance was mapped on Chr. 13 in the KCl treatment of 124 $F_{4:6}$ lines from cross RA-452 × Osage (Zeng et al. 2017b). By association mapping, three genomic regions associated with germination-related traits under salt stress were identified and confirmed on Chrs. 8, 9 and 18 using GWAS across 1142 SNP markers (Kan et al. 2015). Additionally, nine genomic regions significantly associated with both leaf chloride and leaf chlorophyll concentrations and the significant SNP markers on Chrs. 2, 3, 14, 16, and 20 were recommended for MAS (Zeng et al. 2017a). Nevertheless, efforts to exploit those achievements are not reported.

Although a major gene was characterized and additional minor QTLs were mapped thus far, the development of salt tolerant soybean varieties have only focused on the major gene. Near-isogenic lines (NILs) with the major gene for salt tolerance were developed and

a trial under saline field conditions showed genotypes with tolerance had increased seed weight compared to sensitive lines (Liu et al. 2016b). In another study, salt tolerant NILs were selected by introgression of the major gene from a different donor parent, FT-Abyara. The results of a trial under saline field conditions in 2011 and 2012 showed that NILs with the major gene had increased soybean grain yield 3.6 – 5.5 times compared to a salt sensitive cultivar, Tachiyutaka (Do et al. 2016).

In our previous study, the major gene with salt tolerant allele from Fiskeby III was confirmed and a new minor locus for leaf sodium content (LSC) with a favorable allele from Williams 82 was identified (Do et al. 2018). A population of 71 F_5 lines was developed from the same cross by single-pod descent. F_5 lines carried individual genes on Chr. 3, Chr.13. Other $F_{5:6}$ lines with a combination of the two genes were classified by gene-based markers (GBM) for the gene on Chr. 3 and flanking markers for the gene on Chr. 13. The goal of this study was to evaluate salt tolerance of F_5 lines carrying individual genes on Chr. 3, Chr.13 and a combination of these two genes.

Materials and methods

Materials

71 F_5 lines from a Williams 82 × Fiskeby III cross were genotyped to select 33 homozygous lines that carried salt tolerant alleles for individual loci on Chr. 3 and Chr. 13, and a combination of salt-tolerant alleles and salt sensitive alleles for both loci (Table 4.1). The selected $F_{5:6}$ lines, Williams 82, Fiskeby III and sensitive check (Hutcheson) were used to evaluate salt tolerance under greenhouse conditions.

Three GBMs for the known gene for salt tolerance on Chr. 3 (Patil et al. 2016) and two flanking SNP markers for minor locus on Chr. 13 were used to test F_5 lines.

Methods

Leaf tissue of F_5 lines were collected for DNA extraction using CTAB method (Doyle and Doyle 1987). KASP assays of three GBMs and two flanking makers were performed according to the protocol described by LGC Genomics, LLC (http://www.lgcgroup.com).

Phenotypic evaluation of the $F_{5:6}$ lines, cultivars Williams 82, Fiskeby III, and Hutcheson for salt tolerance was conducted using the plastic cone-tainer method (Do et al. 2018; Lee et al. 2008; Patil et al. 2016). Seven seedlings of each soybean line were grown in a greenhouse at the University of Missouri, Columbia, MO, using artificial lights and a 13 h photoperiod in 2017. The experimental design was a randomized complete block design with three replications. Leaf scorch score and chlorophyll content ratio were determined as described in previous studies (Do et al. 2018; Patil et al. 2016).

Statistical analysis for the salt tolerant traits, analysis of variance (ANOVA) for unbalanced data and the least significant difference (LSD) test at $\alpha = 0.01$ were performed using the PROC GLM procedure of SAS 9.4 (SAS Institute Inc. 2013).

Results and discussion

Genotypes of three GBMs in the known gene on Chr. 3 were named as AA, homozygote of allele from salt tolerant parent (Fiskeby III); aa, homozygote of allele from moderately salt sensitive parent (Williams 82); and Aa, heterozygote. Two flanking SNP markers (*Gm13-38988256* and *Gm13-39054715*) of Chr. 13 locus were assumed to be

tightly linked with the unknown gene for salt tolerance, where BB, homozygote of allele from moderately salt sensitive parent (Williams 82); bb, homozygote of allele from Fiskeby III; and Bb, heterozygote. The genotypes of the Chr. 3 gene for the 71 soybean lines using GBMs included 48 homozygous lines of AA, 22 homozygous lines of aa, and one heterozygous line (Fig. 4.1). The genotypes of unknown gene on Chr. 13 included 28 homozygous lines of BB, 30 homozygous lines of bb, and 13 heterozygous lines (Fig. 4.1). Based on combining genotypes of both genes, 33 homozygous lines divided into four groups were used for evaluating salt tolerance. Four groups labelled as RIL_AB, RIL_Ab, RIL_aB and RIL_aa represented four genotypes, AABB, AAbb, aaBB, and aabb, respectively (Table 4.1).

The results of ANOVA showed that there were significant differences among genotypic groups and no significant difference between replications at $\alpha = 0.01$ for both salt tolerant traits (Table 4.2; Fig. 4.2A, B). Paired comparisons by post-hoc analysis (LSD) also indicated significant differences for each pair of genotypic groups for salt tolerance, except one pair of each RIL_aB and RIL_aa in chlorophyll content ratio (Table 4.1). Salt tolerance of RIL_AB (combining salt tolerant alleles of both loci) was significantly higher than that of RIL_Ab (individual salt tolerant allele of Chr. 3 locus, Fiskeby III genotype) for both LSS and CCR (Table 4.1; Figs. 4.3A, B). On the other hand, salt tolerance of RIL_ab (combining salt sensitive alleles of both loci) was significantly lower than that of RIL_aB (individual salt tolerant alleles from the Chr. 13 locus, Williams 82 genotype) for LSS only (Table 4.1). In other comparisons, salt tolerance of the RIL groups (RIL_AB and RIL_Ab) with salt tolerant allele from Fiskeby III were much higher than that of other groups for both traits (Figs. 4.3A, B). Therefore, combining two genes on Chr. 3 and Chr.

13 could likely lead to increased salt tolerance by combining the minor and unknown gene on Chr. 13 with the major gene on chromosome 3. However, the salt tolerance of RIL_AB group that combined salt tolerant alleles of both loci was not significantly higher than the tolerance level of Fiskeby III for both traits. This could be due to recombination between the flanking SNP markers and unknown gene of Chr. 13 locus occurring in some RILs of this group or noise from other genomic regions.

Conclusions and suggestions

Salt tolerant allele(s) of the minor locus on Chr. 13 might increase salt tolerance by stacking with the known gene on Chr. 3. The flanking SNP markers can be used for MAS in addition to GBMs of the known gene in breeding salt tolerant soybean.

Instead of RILs, near-isogenic lines (NILs) from residual heterozygous lines (RHLs) should be used for testing the effect of the minor locus on Chr. 13 in the future to reduce or eliminate the noise from other genomic regions. In addition, it is recommended that the genomic region where the unknown gene of Chr. 13 is confined, more flanking SNP markers should be developed and to narrow the genetic region for greater accuracy in genotyping to apply in MAS.

Preliminary testing of a line for a new salt tolerant gene(s)

Rationale and goals

Genetic analysis of F_2 segregating population is one of robust strategies for determining the inheritance of Mendelian traits. However, this method has been also applied for quantitative traits, including salt tolerance in soybean. As reported in several genetic studies for salt tolerance, the salt tolerance trait based on chloride toxicity was classified into chloride excluder (tolerant) and chloride includer (intolerant), in which the chloride excluder shows no necrosis under salt stress (Abel 1969). Eight F_2 populations from different parents were used for segregation analysis and F_2 plants segregated in ratios of 3 chloride excluder to 1 includer (Abel 1969). Classes of salt tolerance and sensitivity were identified by leaf scorch score threshold of 2.5 and the ratios of 3 tolerant to 1 sensitive found in segregation of F_2 populations using Chi-square tests (Hamwieh and Xu 2008; Lee et al. 2009). Other classifications of tolerance, intermediate and sensitivity based on means of parental traits plus/minus standard deviations were more useful in a classical genetic study for salt tolerance (Do et al. 2018).

A complementation test on segregation of F_2 population can be used to test whether two mutations occur in the same gene or not (Doebley et al. 1995; Hawley and Gilliland 2006). A salt tolerant germplasm line, Fiskeby III, with the known gene on Chr. 3 was identified by mapping using GBMs integrated with SNP data (Do et al. 2018). In our previous study, by matching of salt tolerant lines and genotypes of GBMs, six new soybean germplasm, including PI 468908, were predicted to carry novel gene(s) for salt tolerance

in soybean. Conducting the complementation tests to confirm the possibility will be meaningful for mapping new gene(s) in the future.

This study was performed with the hypothesis, novel salt tolerant gene(s) in PI 468908 were located in different chromosomal positions compared to the known gene on Chr. 3 in Fiskeby III and could complement each other for salt tolerance. To test the hypothesis, an F_2 population derived from a cross of PI 468908 and Fiskeby III (Fig. 4.4) was used for the complementation test and segregation analysis.

Materials and methods

196 F_2 plants derived from a PI 468908 × Fiskeby III cross, salt tolerant parents (PI 468908 and Fiskeby III), and sensitive check (Hutcheson) were grown in a greenhouse at the University of Missouri to evaluate salt tolerance in 2017. Similar methods and traits for evaluating salt tolerance were used; however, two evaluations of salt tolerance focusing on extreme phenotypes were performed. Briefly, the extreme salt sensitive plants were scored for leaf scorch score (LSS) and measured for chlorophyll content ratio (CCR) when the sensitive check (Hutcheson cultivar) showed leaf chlorosis (approximately two weeks after the salt treatment). In addition, the most salt tolerant plants were recorded when salt tolerant parents showed leaf chlorosis (approximately 4 weeks after the salt treatment).

The complementation of the two mutants salt tolerant phenotypes and salt sensitive phenotypes were determined by the extremes in tolerance or sensitivity. Segregating ratios of salt tolerance in F_2 population were analyzed by Chi-square tests. To focus on the extreme salt sensitive group, Hutcheson-based means of salt tolerant traits plus the standard deviation were used as a threshold to classify salt tolerant traits. The 196 F_2 plants were

classified into extremely sensitive and intermediate-tolerant categories (Do et al. 2018) and tested with a segregating ratio of 1 sensitive to 15 intermediate-tolerant using Chi-square test. For the extremely salt tolerant group, the cut-off value was based on parental means of salt tolerant traits when parents showed chlorosis symptom minus standard deviation to group 196 F_2 plants into sensitive-intermediate and extremely-tolerant categories (Do et al. 2018) and tested with a segregating ratio of 7 sensitive-intermediate to 9 tolerant.

Results and discussion

By visual evaluations of leaf scorch symptom, about 10 F_2 plants showed leaf chlorosis ten days after the salt treatment (Fig. 4.5). Number of chlorotic plants increased to 19 at the time cultivar Hutcheson showed leaf chlorosis (Table 4.3). Even though both parents were salt tolerant, salt sensitive phenotypes appeared among plants in F_2 population. Thus, the gene of PI 468908 and the known gene located on Chr. 3 from Fiskeby III complemented each other or PI 468908 carried a novel gene for salt tolerance as predicted in our previous study.

LSS and CCR determined two weeks after the salt treatment were classified into extremely sensitive and intermediate-tolerant groups. Based on Chi-square tests, CCR segregation well fitted a ratio of 1 sensitive: 15 intermediate-tolerant (χ^2 = 0.92 and *Pr* = 0.66) while a segregating ratio of LSS was close to 1:15 (χ^2 = 3.97 and *Pr* = 0.95) rather than other ratios (Table 4.3). For extremely salt-tolerant traits determined at 4 weeks after the salt treatment, salt tolerant segregations well fitted an F_2 ratio of 9 tolerant: 7 sensitive-intermediate for LSS (χ^2 = 3.18 and *Pr* = 0.86) and CCR (χ^2 = 3.11 and *Pr* = 0.81), indicating that these two traits were controlled by two genes (Table 4.4). Thus, a new gene

in PI 468908 may be located on another chromosome or distant from the known gene on Chr. 3.

Conclusions and suggestions

Salt tolerance of PI 468908 was likely controlled by a new gene rather than the known gene on Chr. 3. The position of this gene may not be on Chr. 3 or far away from location of the known gene on Chr. 3. A mapping population from PI 468908 should be considered to map the new gene in the future.

Legends to Figures

Figure 4.1 Patterns of three GBMs (Salt-20, Salt11655 and Salt14056) and two flanking markers (Gm13-38988256 and Gm13-39054715) of 71 RIL lines showing the groups of lines with homozygote of Williams 82's allele (*blue*), lines with homozygote of Fiskeby III's allele (*green*), heterozygous lines (*red*)

Figure 4.2 Plots showing variation of three experimental replications by analyzing four genotypic groups of 33 RILs for leaf scorch score (LSS) in A panel and chlorophyll content ratio (CCR) in B panel

Figure 4.3 Plots showing salt tolerant variation of four genotypic groups of 33 RILs, (A) variation of leaf scorch score (LSS) and (B) variation of chlorophyll content ratio (CCR)

Figure 4.4 Parental soybean seeds (*far left* and *far right*) and F_2 seed from cross PI 468908 × Fiskeby III (*center*)

Figure 4.5 F_2 plants ten days after salt treatment showing some plants with chlorosis symptom

References

Abel GH (1969) Inheritance of the capacity for chloride inclusion and chloride exclusion by soybeans. Crop Sci 9:697-698

Chen HT, Cui SY, Fu SX, Gai JY, Yu DY (2008) Identification of quantitative trait loci associated with salt tolerance during seedling growth in soybean (*Glycine max* L.). Aust J Agr Res 59:1086-1091

Do TD, Chen H, Hien VT, Hamwieh A, Yamada T, Sato T, Yan Y, Cong H, Shono M, Suenaga K, Xu D (2016) *Ncl* synchronously regulates Na(+), K(+), and Cl(-) in soybean and greatly increases the grain yield in saline field conditions. Sci Rep 6:19147

Do TD, Vuong TD, Dunn D, Smothers S, Patil G, Yungbluth DC, Chen P, Scaboo A, Xu D, Carter TE, Nguyen HT, Grover Shannon J (2018) Mapping and confirmation of loci for salt tolerance in a novel soybean germplasm, Fiskeby III. Theor Appl Genet 131:513-524

Doebley J, Stec A, Gustus C (1995) Teosinte Branched1 and the Origin of Maize: Evidence for Epistasis and the Evolution of Dominance. Genetics 141:333-346

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15

Guan RX, Qu Y, Guo Y, Yu LL, Liu Y, Jiang JH, Chen JG, Ren YL, Liu GY, Tian L, Jin LG, Liu ZX, Hong HL, Chang RZ, Gilliham M, Qiu LJ (2014) Salinity tolerance in soybean is modulated by natural variation in GmSALT3. Plant J 80:937-950

Hamwieh A, Xu DH (2008) Conserved salt tolerance quantitative trait locus (QTL) in wild and cultivated soybeans. Breeding Sci 58:355-359

Hawley RS, Gilliland WD (2006) Sometimes the result is not the answer: The truths and the lies that come from using the complementation test. Genetics 174:5-15

Kan G, Zhang W, Yang W, Ma D, Zhang D, Hao D, Hu Z, Yu D (2015) Association mapping of soybean seed germination under salt stress. Mol Genet Genomics 290:2147-2162

Lee JD, Shannon JG, Vuong TD, Nguyen HT (2009) Inheritance of salt tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.) accession PI483463. J Hered 100:798-801

Lee JD, Smothers SL, Dunn D, Villagarcia M, Shumway CR, Carter TE, Shannon JG (2008) Evaluation of a simple method to screen soybean genotypes for salt tolerance. Crop Sci 48:2194-2200

Liu Y, Yu L, Qu Y, Chen J, Liu X, Hong H, Liu Z, Chang R, Gilliham M, Qiu L, Guan R (2016b) *GmSALT3,* which confers improved soybean salt tolerance in the field, increases leaf Cl- exclusion prior to Na+ exclusion but does not improve early vigor under salinity. Front Plant Sci 7:1485

Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, Shannon JG, Nguyen HT (2016) Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci Rep 6:19199

Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu

X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

SAS Institute Inc. (2013) SAS® 9.4 Guide to Software Updates. 9.4 edn. Cary, NC: SAS Institute Inc

Zeng A, Chen P, Korth K, Hancock F, Pereira A, Brye K, Wu C, Shi A (2017a) Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. Mol Breeding 37:30

Zeng A, Lara L, Chen P, Luan X, Hancock F, Korth K, Brye K, Pereira A, Wu C (2017b) Quantitative trait loci for chloride tolerance in 'Osage' soybean. Crop Sci 57:2345-2353

Table 4.1 Genotype and phenotype of parents and recombinant inbred line (RIL) groups, RILs with salt tolerant allele of Chr. 3 locus (RIL_Ab), RILs with salt tolerant allele of Chr. 13 locus (RIL_aB), RILs with salt tolerant alleles of both loci (RIL_AB), and RILs with salt sensitive alleles of both loci (RIL_ab).

Line/group	Genotype	No. of lines	LSS	CCR
RIL_AB	AABB	10	1.3	1.06
RIL_Ab	AAbb	9	1.9	0.94
RIL_aB	aaBB	6	3.7	0.61
RIL_ab	aabb	8	4.5	0.51
FiskebyIII	AAbb		1.4	1.16
Williams82	aaBB		4.3	0.53
LSD (α = 0.01)			***0.555***	***0.105***

A: Salt tolerant allele of locus on Chr. 3

a: Salt sensitive allele of locus on Chr. 3

B: Assumption of salt tolerant allele of locus on Chr.13

b: Assumption of salt sensitive allele of locus on Chr.13

Table 4.2 ANOVA of leaf scorch score (LSS) and chlorophyll content ratio (CCR) among 33 RILs from a Williams 82 × Fiskeby III cross

Trait	Source	df	Sum of Squares	Mean Square	F Value	Pr > F
LSS	Genotype	3	173.91	57.97	315.87	<.0001
	Replication	2	0.46	0.23	1.25	0.292
	Error	88	16.15	0.18		
	Total	93	190.52			
CCR	Genotype	3	4.93	1.64	251.89	<.0001
	Rep	2	0.04	0.02	2.82	0.0651
	Error	88	0.57	0.01		
	Total	93	5.54			

Table 4.3 Chi-square test for leaf scorch sore (LSS) and chlorophyll content ratio (CCR) determined at 2 weeks after salt treatment in F_2 population

Group	LSS		CCR	
	Observed	**Cut-off**	**Observed**	**Cut-off**
Sensitive	19	>3.8	9	≤0.62
Intermediate-tolerant	177	≤3.8	187	>0.62
Size	**196**		**196**	
χ^2 (15:1)	**3.97**		**0.92**	
Pr	**0.95**		**0.66**	

Cut-off: Cut-off value based on mean of sensitive check plus standard deviation

Table 4.4 Chi-square test for leaf scorch sore (LSS) and chlorophyll content ratio (CCR) determined at 4 weeks after salt treatment in F_2 population

Groups	LSS		CCR	
	Observed	**Cut-off**	**Observed**	**Cut-off**
Tolerant	100	<2.9	98	>0.99
Sensitive-intermediate	96	≥2.9	98	≤0.99
Size	**196**		**196**	
χ^2 (9:7)	**3.18**		**3.11**	
Pr	**0.86**		**0.81**	

Cut-off: Cut-off value based on mean of parents minus standard deviation

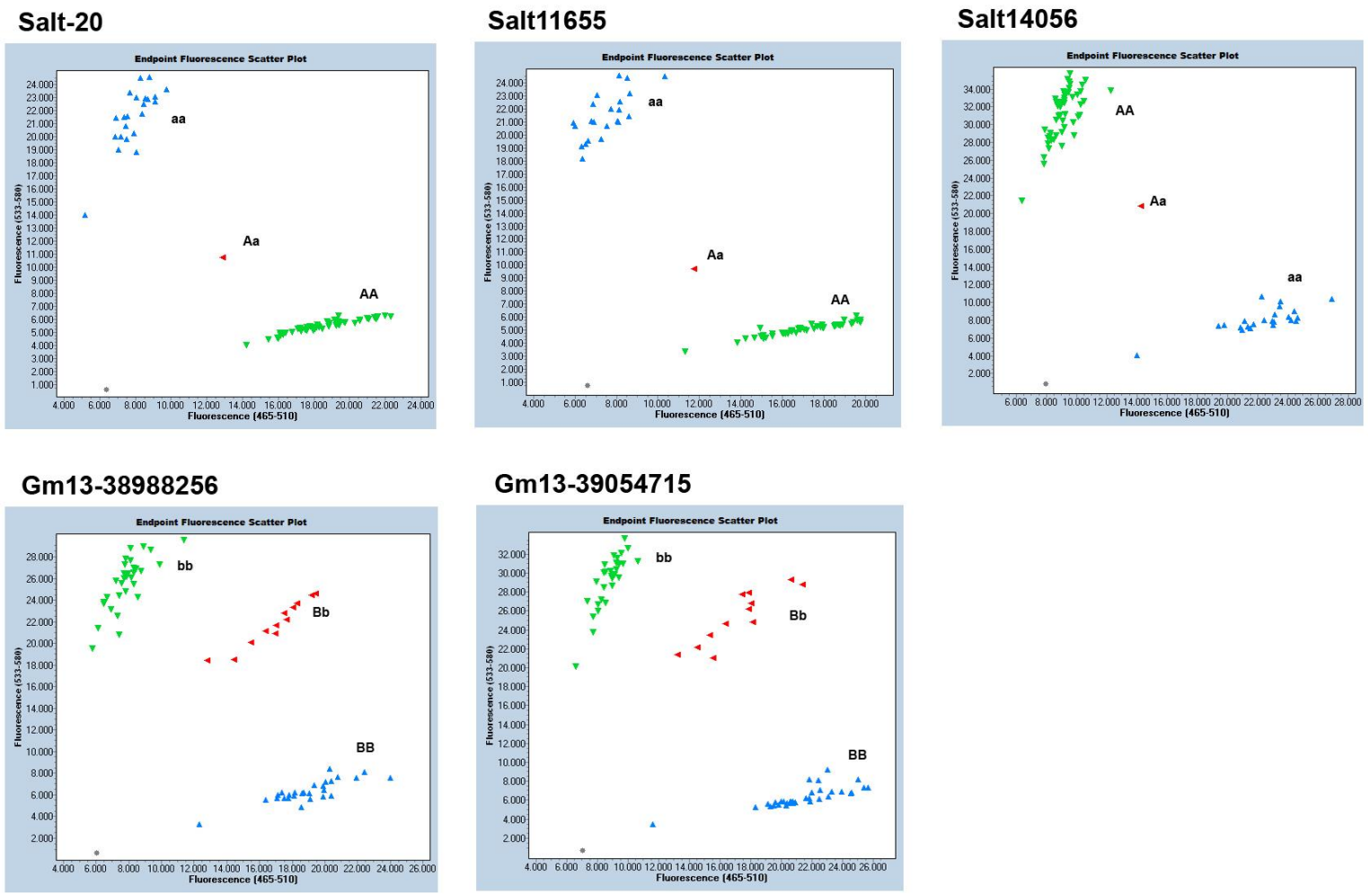


Figure 4.1 Patterns of three GBMs (Salt-20, Salt11655 and Salt14056) and two flanking markers (Gm13-38988256 and Gm13-39054715) of 71 RIL lines.
The groups of lines with homozygote of Williams 82's allele (*blue*), lines with homozygote of Fiskeby III's allele (*green*), heterozygous lines (*red*)

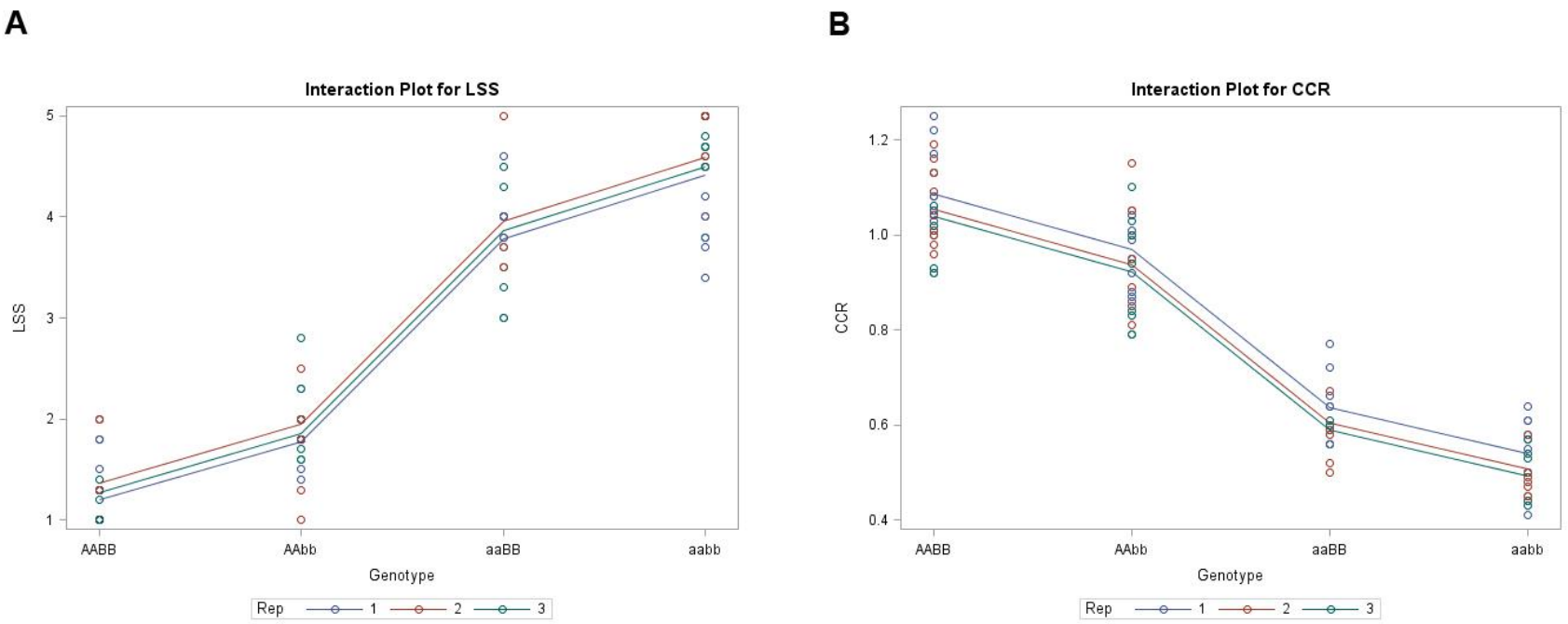


Figure 4.2 Variations of three experimental replications by analyzing four genotypic groups of 33 RILs for leaf scorch score (LSS) in A panel and chlorophyll content ratio (CCR) in B panel

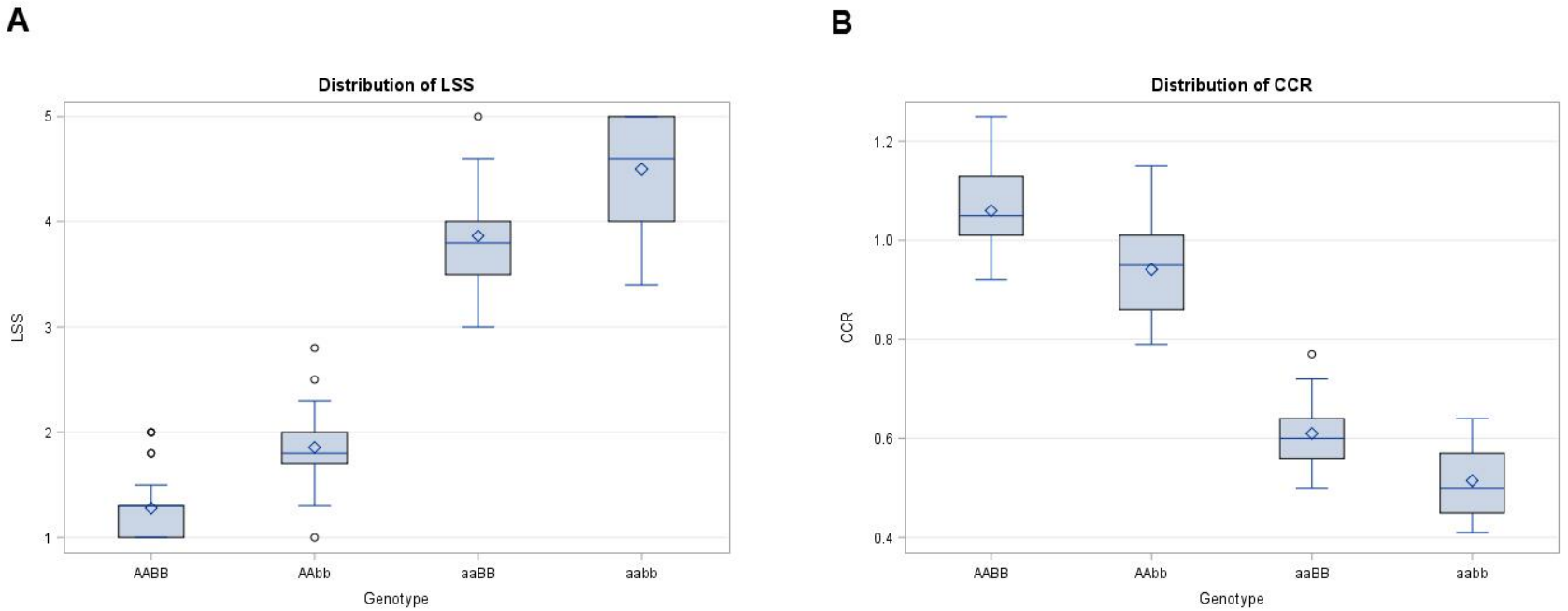


Figure 4.3 Salt tolerant variations of four genotypic groups of 33 RILs, (A) variation of leaf scorch score (LSS) and (B) variation of chlorophyll content ratio (CCR)

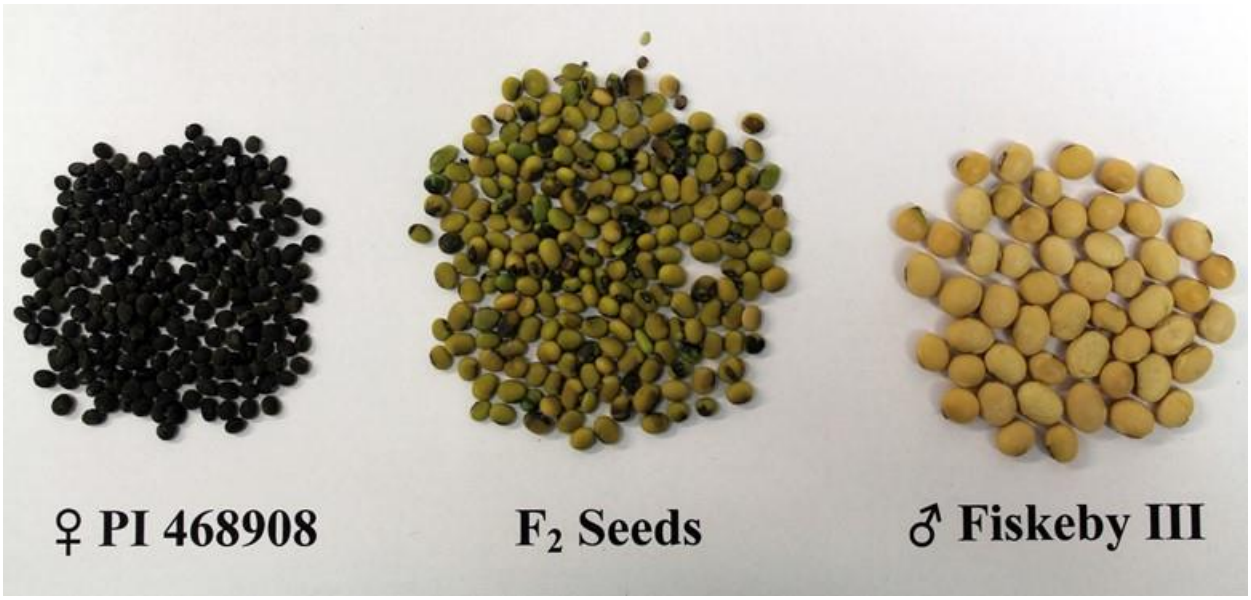


Figure 4.4 Parental soybean seeds (*far left* and *far right*) and F_2 seed from a PI 468908 × Fiskeby cross III (*center*)

Figure 4.5 F_2 plants 10 days after the salt treatment showing some plants with chlorosis symptom

Chapter 5:

OVERALL SUMMARY AND FUTURE REFLECTIONS

Overall summary

Soybean cultivar Fiskeby III (PI 438471) in maturity group 000 has been reported to be highly tolerant to multiple abiotic stress conditions including high salt tolerance in previous studies. A mapping population derived from a cross of cultivar Williams 82 (PI 518671, moderately salt sensitive) and Fiskeby III (salt tolerant) was analyzed to map salt tolerance genes. Salt tolerance of 132 $F_{2:3}$ lines was evaluated by analyzing leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC) after treatment with 120 mM NaCl under greenhouse conditions. Genotypic data of 2158 polymorphic SNP markers for the F_2 population was obtained using the SoySNP6K Illumina Infinium BeadChip assay. A major locus was significantly associated with LSS, CCR, LSC, and LCC for salt tolerance on chromosome (Chr.) 03 with LOD scores of 19.1, 11.0, 7.7 and 25.6, respectively. In this significant interval, three GBMs (Salt-20, Salt14056 and Salt11655) for the known gene showed a strong predictive association with phenotypic salt tolerance. In addition, a second locus associated with salt tolerance for LSC was detected and mapped on Chr. 13 with allele of interest from Williams 82. Based on the genomic region significantly associated with LSC on Chr. 13, three coding sequences, including *Glyma.13g305700, Glyma.13g305800,* and *Glyma.13g305900,* with salt stress response/antifungal function were close to significant markers for LSC. Thus, GBMs are useful tools for tracking and selecting the salt tolerant gene on Chr. 03. The putative locus for LSC on Chr. 13 suggests the presence of a novel gene(s) controlling salt tolerance and may be useful to stack with the known gene on Chr.

03 for improving salt tolerance in soybean. Another preliminary test of two different genes on Chrs 3 and 13 was conducted using recombinant inbred lines from the same cross. Combining of two genes on Chr. 3 and 13 significantly increased salt tolerance of RILs under greenhouse conditions. However, the unknown gene on Chr. 13 expressed a minor effect on salt tolerance in soybean. Therefore, the flanking SNP markers should be used for MAS in addition to the GBMs of the known gene on Chromosome 3 in breeding salt tolerant soybeans.

In addition to bi-parental QTL mapping, genome-wide association study (GWAS) was performed to map additional loci for salt tolerance in a diverse panel of 305 soybean accessions using 37,573 single nucleotide polymorphism (SNP) markers derived from SoySNP50K iSelect BeadChip. A second GWAS was also conducted using over 3.7M SNP dataset derived from whole-genome sequencing in a subset of 234 *G. max* accessions. Salt tolerance among soybean lines was evaluated by leaf scorch score (LSS), chlorophyll content ratio (CCR), leaf sodium content (LSC), and leaf chloride content (LCC). For both SNP datasets, the known gene for salt tolerance on Chr. 3 was also confirmed by the most significant GBMs associated with all of four traits that integrated into both datasets. In addition, genomic regions associated with salt tolerance were found on Chrs. 1, 8, and 18 by analyzing 3.7M SNP dataset, in which the position on Chr. 8 was strongly predicted as a new minor locus for salt tolerance in soybean. The candidate genes harbored in this minor locus were predicted on functional annotation in databases of Phytozone and Soybase. GWAS using dataset of 3.7M SNPs generated from whole-genome sequencing was more efficient than SoySNP50K-based GWAS to predict minor loci and pinpoint putative candidate genes for salt tolerance. Additionally, 6 salt tolerant sources with predicted novel

gene(s) were found by genotype-phenotype correlations using GBMs. By complementation tests and segregation analysis of an F_2 population, salt tolerance of PI 468908 (one of those new salt tolerant sources) was controlled by a new gene rather than the known gene on Chr. 3 and position of this gene may be on other chromosome or distant from the location of the known gene on Chr. 3. In addition to new salt tolerant sources, the putative candidate underlying minor locus on Chr. 8 and the significant SNPs may be helpful to study the molecular mechanism involved in tolerance and will be useful for marker-assisted selection to improving salt tolerance.

Future reflections

Although the effect of minor locus on Chr. 13 for salt tolerance was preliminarily confirmed using recombinant inbred lines (RILs) and candidate genes underlying this locus were predicted based on functional annotation. However, RILs were not the best option to test the effect of a minor locus and to investigate the candidate genes because of genetic noise from other genomic regions and recombining of the flanking markers with the unknown gene for salt tolerance. Near isogenic lines (NILs) from residual heterozygous line (RHL) for both genes on Chr. 3 and 13 should be considered as the best materials to investigate minor loci. Ideally, the known gene on Chr. 3 will be easily identified for selection by GBMs. Development NILs will allow the ability to focus on the minor gene on Chr. 13. A set of RHLs for the known gene and QTL region on Chr. 13 identified using GBMs and flanking markers will be genotyped by the SoySNP6K Illumina Infinium BeadChip assay to develop and select NILs. Subsequently, NILs will be used for confirming salt tolerance and investigating the candidate genes on Chr. 13.

Two candidate genes underlying the minor locus on Chr. 8 from GWAS for salt tolerance may relate to cell wall changes under salt stress and sodium sequestration into the central vacuole based on citation and functional annotation. Those two genes should be considered for analyzing correlation of gene expression and salt tolerance to identify the promising lines for functional studies.

In addition, a new gene from PI 468908 was preliminarily confirmed by complement test and segregation analysis of an F_2 population. To map and localize this gene, a mapping population from this line should be developed in next steps. Similarly, studies should be designed for other new salt tolerant sources predicted to carry new gene(s). The new gene(s) from new salt tolerant sources undetected by GWAS were probably rare alleles in nature and had low allele frequency in our population. Another idea to study all the new sources, a population by nested crossing should be considered to increase detecting rare alleles by nested-association mapping, as well as to separate this population into subpopulations for QTL mapping of salt tolerance.

VITA

Tuyen Do was born and grew up in a farming family locating in a rural area of Hung Yen province, Vietnam. He is a son of Mr. Trieu Duc Do and Mrs. Tuyen Thi Tri who do farming for living. Despite of family limited earnings and resources, his parents continued to support and encourage him to study at Khoai Chau high school, Hung Yen, Vietnam. While pursuing his high school education, he helped his parents working on their farm. Having been assisting the family with farming, Tuyen has learned and understood the impacts of agricultural production and motivated him to study in agricultural area.

In 1996, he successfully passed the entrance examination and was admitted to study in Department of Genetics, Faculty of Biology, Hanoi University of Science, Ha Noi, Vietnam. Following the completion of general courses, as a junior student he had a chance to work in Dr. Dat Dinh Trinh Laboratory, using isozyme technology to analyze genetics of some organisms, *Culex quinquefasciatus*, *Apis cerana* and *Ctenopharyngodon idella*. With his research progress and accomplishment, he was promoted to conduct research work of his thesis at Department of Molecular Biology, Agricultural Genetics Institute, Ha Noi, Vietnam, in the third year of the program. It provided him a great opportunity to continue his research work towards to his graduation. While pursuing his degree study, he learned and had good experiences in molecular biotechnology. He completed his research project focusing on the identification of genetic polymorphism of some rice varieties using RAPD marker and graduated with honor status. After his graduation in 2000, he worked in Plant Resources Center, Vietnamese Academy of Agricultural Sciences, Ha Noi, Vietnam. His responsibility was to analyze biodiversity of the Vietnamese Rice Germplasm Collection. Using isoenzyme approach, he classified a set of local rice accessions into three

subspecies based on analysis of nine enzymes at germination stage. Results of the study were presented at a symposium and published in "On-farm management of agricultural biodiversity in Vietnam", proceeding of a symposium, 6-12 December 2001, Ha Noi, Vietnam. In addition to research experiences in rice biodiversity, he also had good knowledge in plant genetic conservations, such as *in situ*, *ex situ*, and *in vitro* conservation.

In 2001, he attended a training workshop on marker-assisted selection at the Cuu Long Delta Rice Research Institute (CLRRI), Can Tho, Vietnam. He joined and worked in the CLRRI in 2002. His research interest was to improve bacterial blight resistance in rice using marker-assisted selection method. The research results were published in the Omonrice journal of the CLRRI.

In 2005, he received a DANIDA (Danish International Development Agency) scholarship and studied M.Sc. program in the University of Agricultural Sciences, Bangalore, India, under the supervision of Dr. Theertha Prasad, in conjunction with the support of Sub-Component 5, Seed component, ASPS, DANIDA, Vietnam. For his research program, he focused on analysis of yield traits in rice under low moisture using a candidate gene approach. Basically, candidate genes related to abiotic stress were used to determine correlation with rice yield components and rice yield under low moisture. The result of the study was published in Omonrice journal. In 2007, upon the completion of his M.Sc. program, he returned to his home institute to continue his work.

In 2008, he received award to become a fellow of the Japan International Research Center for Agricultural Sciences (JIRCAS) fellowship program (2008-2009), to study salt tolerance in soybean. In Japan, he has successfully identified and mapped a major quantitative trait loci (QTL) for alkaline salt tolerance in soybean. His JIRCAS fellowship

was extended one more year (2009-2010), and he was able to conduct fine-mapping of the major QTL for alkaline salt tolerant trait and isolating candidate gene for salt tolerant trait in soybean. The fascinating results led to different publications published in Theoretical and Applied Genetics, Euphytica, Breeding Science, Molecular Breeding and Scientific Reports.

In 2014, he was awarded a Vietnamese Government Scholarship to pursue his Ph.D. program with Dr. J. Grover Shannon, as advisor, and Dr. Henry T. Nguyen, as co-advisor in Division of Plant Sciences, the University of Missouri – Columbia, Missouri. His research work focused on genetics and breeding for salt tolerance in soybean. The findings of his research work led to the publications published in high impact journals, Theoretical and Applied Genetics and Scientific Reports. He completed his doctorate degree in Plant Breeding, Genetics, and Genomics in May 2018.