Theor Appl Genet (2013) 126:1659–1675
DOI 10.1007/s00122-013-2082-0

ORIGINAL PAPER

Association mapping combined with linkage analysis for aluminum tolerance among soybean cultivars released in Yellow and Changjiang River Valleys in China

Paul C. Korir · Jun Zhang · Kejing Wu · Tuanjie Zhao · Junyi Gai

Received: 18 March 2012 / Accepted: 1 March 2013 / Published online: 21 March 2013

Abstract Association mapping (AM) combined with linkage mapping (LM) was executed to identify molecular markers and QTL regions associated with aluminum (Al) tolerance using relative root elongation (RRE) in hydroponics as an indicator. A set of 188 soybean cultivars released in Yellow and Changjiang River Valleys and 184 recombinant inbred lines (RIL) derived from a cross KF No. 1 (tolerant) × NN1138-2 (susceptible) was used in the study. Inheritance analysis of the RIL population suggested four major genes and polygenes controlled Al-tolerance. Further, LM indicated four additive and four epistatic QTL pairs plus a collective unmapped minor QTL were responsible for Al-tolerance and explained 29.39, 18.75 and 43.07 % of the phenotypic variation (PV), respectively. In the set of released cultivars, AM identified 11 markers significant at $P < 0.03$ that explained 85.2 % of PV with six of which at $P < 0.01$ accounted for 57.9 % of PV. Ten of these eleven AM marker-QTL were mapped within range of ~2.0 cM to ~43.0 cM outside confidence interval of respective Al-tolerance QTL in previous studies. Five markers, Satt209, Sat_364, Sat_240, Sct_190 and Satt284, were located near Al-tolerance QTL regions in this and previous LM studies. Thus, the two methods confirmed these markers as being the most likely candidate regions for Al-tolerance. Allele effects relative to the population mean for the 11 QTL were estimated, and the allele A210 of Satt209 showed greatest phenotypic effect on Al-tolerance. The two most favorable alleles from each of the 11 marker loci and their carriers were identified, and accordingly the genetic constitution of Al-tolerance for the 188 cultivars was dissected as a QTL-allele matrix. Therefore, marker-assisted pairing of crosses and marker-assisted selection of progenies can be carried out to pyramid favorable alleles of all the 11 loci. This marker-assisted breeding procedure was designated as breeding by design using a QTL-allele matrix.

Communicated by H. T. Nguyen.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2082-0) contains supplementary material, which is available to authorized users.

P. C. Korir · J. Zhang · K. Wu · T. Zhao · J. Gai (✉)
Soybean Research Institute, Nanjing Agricultural University, Nanjing 210095, China
e-mail: sri@njau.edu.cn

T. Zhao · J. Gai
National Center for Soybean Improvement, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

T. Zhao · J. Gai
Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

T. Zhao · J. Gai
National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

Introduction

Aluminum (Al) toxicity is the main constraint to crop productivity in acidic soils worldwide. Acid soils (pH < 5.0), comprise approximately 40 % of the world's total land area and about 50 % of world's potential arable lands (von Uexküll and Mutert 1995). At soil pH ≤ 5.0, phytotoxic Al^{3+} solubilizes into the soil solution from non-toxic Al-oxides and Al-silicates and this ionic species damage roots, restricts plant size, and lower yields in many

crops. One proposed mechanism of Al^{3+} intoxication is the inhibition of root growth through disruption of active cell division and expansion of the root in cell wall of root apex (Delhazie and Ryan 1995). Al-induced inhibition of the root is thus the primary symptom of Al-toxicity (Kochian 1995; Delhazie and Ryan 1995), leading to poor shoot growth (Ryan et al. 1993; Ma and Furukawa 2003) and consequently low yields.

Soybean has become one of the most important legume crops due to its demand as a protein source and for biofuels. Increasing soybean production on acid soils will require agronomic and/or genetic interventions to overcome the negative effects of Al^{3+} ions. The most economical option is to develop and use Al-tolerant cultivars on these problem soils (Kochian et al. 2004). Soybean has substantial genetic variation for Al-tolerance (Foy et al. 1992) and traditional Al-resistant soybean cultivars have long been used in acid soils. In conventional breeding, improving for tolerance has relied on phenotypic selection over several years in replicated environments. However, the quantitative nature of Al-tolerance and effects from genotype-environment interactions have resulted in slow genetic progress for increased tolerance (Spehar 1995; Jena and Mackhill 2008). An understanding of genetic architecture controlling Al-tolerance can enhance efforts in development of tolerant cultivars through identification of QTL related to tolerance that can subsequently be integrated into soybean breeding programs.

Soybean plants exhibit variation in Al-toxicity tolerance in acid soils, but it is clear that molecular mechanisms remain elusive. Studies in quantitative trait loci (QTL) linkage mapping use biparental cross populations with a group of polymorphic markers to discover markers significantly associated with Al-tolerance. This method was employed (Bianchi-Hall et al. 2000; Liu and Gai 2007; Qi et al. 2008; Korir et al. 2011; Sharma et al. 2011) and results showed about two to five loci explained most of the variation in Al-tolerance levels and transgressive segregation was observed in the populations used. In these analyses, different populations using different phenotyping indicators and different statistical approaches were used to discover QTL associated with Al-tolerance. A few QTL were uncovered in single populations and were different from other populations due to the diverse parents, representing only a small fraction tolerant soybean genotypes. The observed transgressive segregation was indicative of additive and epistasis among alleles of the two parents in segregating population (Holland 2007).

In linkage analysis, only a few segregating alleles limited by the variation in the two parents in the segregating population are evaluated. Limited crossover events that have occurred in populations invariably result in poor QTL resolutions (Flint-Garcia et al. 2003). However, linkage mapping will generally provide limited insights into the analysis of complex traits because of the quantitative nature of Al-tolerance and the possible presence of additional QTL among lines in the broader natural population. Alternatively, association mapping (AM) detects gene effects and locates multiple QTL at higher resolutions in populations with multiple ancestors that have undergone several rounds of historical recombination events. AM locates QTL on the basis of linkage disequilibrium (LD) between genetic markers and the trait. Decay of LD initially present in the population is determined by the genetic distance between loci and the number of historical recombination events since it arose. Thus, AM is a powerful molecular tool for simultaneously detecting multiple alleles at each locus, even those with modest effects particularly where LD decays rapidly (Buckler and Thornsberry 2002; Hirschhorn and Daly 2005).

Choice of mapping populations and comparison of their advantages and disadvantages for association analysis is an important consideration for finding useful genetic markers for use in molecular plant breeding programs. Germplasm collections, elite breeding lines and synthetic populations are three types of populations that could be considered for AM implementation (Breseghello and Sorrells 2006a, b). Populations of cultivars are genetically stable, from which maximum relative efficiency of marker-assisted selection for quantitative traits compared with phenotypic selection is expected since markers capture a significant portion of the variation for the trait (Lande and Thompson 1990). Released cultivar populations with a large body of phenotypic data accumulated from replicated field experiments over locations and years are desirable materials for AM because a substantially higher level of polymorphism and detection of favorable alleles in the target population is expected (Breseghello and Sorrells 2006a, b). Since the cultivar population is derived from a few recent, intensely selected ancestors, the LD level is expected to be high as confirmed with inbred lines of maize (Ching et al. 2002), advanced breeding lines of soybean (Wang et al. 2008) and a collection of released cultivars and advanced lines of durum wheat (Maccaferri et al. 2010). Similarly, population structure can be prominent because it is common for closely related lines to be subjected to advance trials among locations within a production region.

The combination of AM and LM can provide both the power and resolution needed for detecting QTL of interest and might prove more successful than either strategy alone. Therefore, the two approaches should be integrated for a thorough detection of QTL potentially useful for application in MAS and introgression into elite Al-sensitive cultivars. Although some LM studies for Al-tolerance in soybean have been carried out (Bianchi-Hall et al. 2000; Liu and Gai 2007; Qi et al. 2008; Korir et al. 2011), to our knowledge, no AM strategy to identify QTL associated with

Al-tolerance has been reported in soybean, although this approach has been used in maize (Krill et al. 2010). In China, a population of soybean cultivars released in the past 40 years in the Yellow and Changjiang River Valleys (YCRV) represents a major part of current germplasm used for breeding in China except Northeast China. The objective of our study was to search for markers associated with QTL regions conferring Al-tolerance in soybean using independent AM and LM approaches and then to scrutinize the genetic architecture of a population of cultivars released in YCRV with respect to favorable tolerance alleles and their inheritance traced in pedigrees. Genome-wide AM was performed and the uncovered marker-trait associations were compared with the LM results and those from previous studies to demonstrate the novelty of AM technology.

Materials and methods

Plant material

Linkage mapping experiments were conducted with a set of 184 $F_{2:7}$-derived recombinant inbred line (RIL) NJRIKY population from a cross between Kefeng (KF) No. 1 (Al-tolerant) and Nannong (NN) 1138-2 (Al-sensitive) parents developed using single-seed descent method. NN1138-2 and KF No. 1 are released cultivars from Changjiang River Valleys and Yellow River Valleys, respectively. Previous studies showed the two parents exhibited contrasting characteristics for Al-tolerance (Liu and Gai 2007; Qi et al. 2008) and other agronomic traits (Zhang et al. 2004). The genetic linkage map for this study was constructed based on this population.

For AM analysis, a population composed of 188 cultivars released in YCRV during the past 40 years was used, which represents a major part of currently used germplasm in soybean variety development in China except Northeast China. There is overlap of genetic base of released cultivars in YCRV regions. For example, important parents, such as Youbian 30 and NN1138-2, were used widely both in Changjiang River Valleys and Yellow River Valleys. These cultivars have also been shown to exhibit a clear geographic differentiation and genetic diversity (Cui et al. 2000; Dong et al. 2004) and are used in most breeding programs.

Phenotyping

Evaluation of soybean for tolerance to Al-toxicity was done in a hydroponics solution because it is a fast and efficient procedure to determine tolerant and sensitive lines while controlling genotype-environment effects commonly encountered in the field studies. The Al-treatment concentration used in this study was determined by generating an Al-dose response of two tolerant cultivars PI 416937 (Campbell and Carter 1990; Villagarcia et al. 2001) and KF No. 1 (Liu and Gai 2007; Qi et al. 2008) and one sensitive cultivar NN1138-2 (Qi et al. 2008). This was to define the concentration that would provide the best separation between the Al-tolerant and Al-sensitive cultivars. Tap root length of each of five uniformly germinated seedlings of each cultivar was measured and then exposed to hydroponics solution containing 0, 12, 25, 40 and 60 μM $Al_2^{3+}(SO4)_3 \times 18H_2O$ for 4 days and measured again. Growth effect of Al on each cultivar was determined as root length elongation and expressed relative to root length elongation in control treatment. Concentration 25 μM provided the widest separation among the cultivars; therefore, it was chosen for screening the population.

The RIL population was phenotyped for taproot length growth along with the parents. Seeds were pre-germinated in quartz sand for 7 days at 28 °C in continuous darkness. Three uniformly germinated seedlings were rinsed and held in foam support floats suspended in 12-L plastic culture containers without Al^{3+} at pH 4.1 for acclimation to hydroponics conditions (Piñeros et al. 2002). After 2 days of acclimation, initial root length (IRL) measurement was taken using a ruler. Then the solution was replaced with nutrients consisting of 800 μM $CaNO_3 \times 4H_2O$ with modified 1/5 strength Steinberg nutrient solution (Foy et al. 1967) with 25 μM Al^{3+} at pH 4.1 (stress treatment) and the same nutrient solution without Al^{3+} (control treatment), respectively. The seedlings were grown at 25 °C under 16 h day illumination and the solutions were renewed after 2 days. After 4 days of Al-stress, final root length (FRL) measurements were taken. Net root length increase (NRL) was calculated as FRL–IRL. The RIL genotypes were evaluated in two replicated experiments using a split-plot design in two replications with cultivar as main plots and Al-level as sub-plots for each experiment. In each replication, three individual seedlings of each genotype and the parents were phenotyped in each replicated experiment. Relative root elongation (RRE), defined as NRL under Al-stress relative to its NRL under control, was used as the indicator for Al-tolerance.

The AM population of 188 cultivars was assayed in two replicated experiments similarly as for the RIL population, but with the three cultivars PI 416937, KF No. 1 and NN1138-2 as control checks.

Statistical analysis

Analysis of variance (ANOVA) was conducted on combined data for both RIL and the released cultivar population using the MIXED Model procedure of SAS (SAS Institute 2004), with RIL/cultivar as fixed effects and experiment, replication and RIL/cultivar × experiment as

random effects. The heritability value (h^2) on entry-mean basis of Al-tolerance was calculated from ANOVA using $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{g\times e}^2/y + \sigma_e^2/ry)$, where σ_g^2 is genotypic variance, σ_e^2 is error variance, $\sigma_{g\times e}^2$ is genotype × environment (or experiment) interaction variance, r is number of replications, and y is number of environments or experiments (Fehr 1987).

Genotyping

Leaf tissue was collected from 3 to 4 leaves for DNA extraction and SSR amplification for genotypic analysis. Total genomic DNA were extracted following a standard CTAB extraction protocol (Doyle and Doyle 1990) and amplification done in a 10-μL reaction mixture containing 1 μL 10 × PCR buffer, 20 ng template DNA, 0.4 μM forward and reverse primers, 1 μL containing 60 μM of each dNTP, 2 mM $MgCl_2$ and 0.5 U *Taq* DNA polymerase. Amplification conditions were initial denaturation at 95 °C for 4 min, 30 cycles of 94 °C for 40 s; 46–50 °C annealing for 60 s; 72 °C extension for 1 min; and an 8-min extension at 72 °C. The reactions were performed on a PTC-225 (or 240) DNA Thermal Cylcer (BIO-RAD, Foster city, CA, USA). PCR products were sized separated in 8 % (w/v) polyacrilymide gel electrophoresis (PAGE) at 20 W. Gels were silver stained, filmed and visualized with the visadoc 3.0 image scanning system (BIO-RAD, Foster city, CA, USA). The size of the band was recorded based on its migration distance relative to the pBR322 DNA Marker (MBI Fermentas) using Quantity One software (Version 4.4.0).

For the RIL population we used a newly improved genetic linkage map reconstructed by the National Center for Soybean Improvement and the Institute of Genetics and Developmental Biology of Chinese Academy of Sciences (Wang 2009). The map was developed using JoinMap 3.0 (Plant Research Int., Wageningen, The Netherlands) (Van-Ooijen and Voorrips 2002) and consisted of 24 linkage groups with 834 markers, including 580 SSR, 184 RFLP, 44 EST, 15 RAPD, 7 TF and four morphological and physiological markers, covering 2,307.8 cM with an average of 2.8 cM between markers. The marker order basically was consistent to the public genetic linkage map (Song et al. 2004) except LGs (Linkage Groups) D1b, F, H and I separated into two segments (Zhang et al. 2004; Wang 2009). Regarding the SSR markers, 378 were adopted from the 2003 USDA consensus map (SoyBase, http://www.soybase.org/) and the public soybean genetic linkage map (Song et al. 2004), 7 EST-SSR were from Hisano et al. (2008), while 80 BAC-SSR, 119 unigene-SSR and others were designed by Wang (2009).

The AM population was genotyped using a total of 197 SSR markers distributed over the entire 20 chromosomes. These loci were selected based on their genomic locations on the integrated genetic linkage map of soybean (Song et al. 2004) and EST-SSR soybean map (Hisano et al. 2008). Included were some SSR markers that stretch out within ~15 cM range on either side of significant QTL peak regions for Al-tolerance previously detected with linkage mapping analysis.

Mapping strategy

Three different analysis strategies were employed for the genetic analysis of Al-tolerance. Segregation analysis of NJRIKY was to detect the genetic structure, linkage mapping was to detect and locate QTL while the association mapping strategy was to scan QTL and their alleles in YCRV. The joint segregation analysis method based on major genes plus polygene mixed inheritance model (Gai 2006) was used on the phenotypic data of the RIL population to describe the genetic system of Al-tolerance in terms of major gene and polygene effects.

Quantitative trait loci mapping was performed with a mixed model-based composite interval mapping (MCIM) procedure of QTLNetwork v2.0 (Yang et al. 2007) and composite interval mapping (CIM) executed with WinQTL Cartographer 2.5 (Wang et al. 2005). QTLNetwork was used to estimate the effects of individual significant additive QTL and additive × additive epistatic QTL pair effects. Significant additive QTL were obtained from F-statistic profile obtained from 2D genome scan procedure. The critical F-value was calculated by permutation tests of 1,000 times at genome-wise significance level of 0.05. At the end of each scanning procedure, forward and backward selection steps were performed to test for the presence of hidden peaks that might result owing to the high correlation of closely linked markers and random effects. Finally, QTL peak regions that exceeded the threshold F-value of 9.4 in this study were declared significant. With CIM, a significant QTL was declared with logarithm of odds ratio (LOD) threshold of 2.5 and permutation tests of 1,000 times at a significance level of $P = 0.05$. A QTL declared significant with LOD of 2.0, but not significant with MCIM procedure, was proposed as a suggestive QTL. Narrow sense heritability ($h_{(a)}^2$), representing phenotypic variation explained by additive QTL, was obtained from the MCIM mapping procedure.

For AM, three data types are required: phenotypic trait information, genotypic data and population structure within the test population. Population substructure creates genome-wide LD between unlinked loci when the allele frequencies between the sub-populations are significantly different (Wright et al. 2005). However, it is possible to statistically control such effects in association data analysis via several methods, including structured association Q-model that takes population structure into account (Pritchard et al. 2000). Of the 197 SSR markers screened, three were monomorphic in the population and were discarded from

further analysis. The marker loci with allele frequency of <0.05 in the population were filtered and grouped together with missing alleles in population structure and LD analysis. Finally, 186 markers were highly polymorphic and their polymorphic information content (PIC) was calculated from PowerMarker v.3.0 (Liu and Muse 2005). Pairwise estimates for chromosaomal LD (r^2, correlation between alleles at two loci) were performed for the whole genome with TASSEL software (http://www.maizegenetics.net) without rapid permutation tests to control for genome-wide error rate. The presence of population structure was determined using 43 unlinked loci with the model-based clustering of STRUCTURE software (Pritchard et al. 2000) with linkage ancestry 'admixture model' with correlated allele frequencies among populations assumed. We set k (no. of subpopulations) to vary from 2 to 10 with five independent runs for each k value and performed a burn-in length of 10,000 iterations followed by another 10,000 iterations for each k. To select the optimal k, we used the posterior probabilities of the natural logarithmic likelihood of data (ln Pr (X|K) = InPD) from the five runs of each k. The percentage parentage for the run that had highest InPD was used in the Q matrix for the analysis. General linear model (GLM) with incorporation of trait data and population structure (Q) as covariate was used to discover marker-trait associations.

In the exploration of superior Al-tolerance alleles, allelic effects were estimated according to population experimental average (Agrama and Yan 2009). Although this method shows systematic difference compared with those estimated from null alleles (Breseghello and Sorrells 2006a, b), the latter method is more affected by the size of the population and, in addition, the biological interpretation of allelic effects is not straightforward. Thus, marker-trait association and allelic effects under presence of population structure (Q) were estimated with GLM model procedure of TASSEL software. The P values corresponding to marker-trait association were log transformed to $-\log_{10}$ (P value) and plotted against the respective marker loci to amplify the significant marker-trait associations in the 188 cultivars. All the alleles with significant positive and negative effects at each locus (for all significant markers $P < 0.03$ with its sum less than 1.00) were used to identify favorable Al-tolerant cultivars and to describe the genetic structure of the 188 released cultivars.

Results

Linkage mapping

Inheritance analysis

The ANOVA revealed significant genotypic difference in RRE among the RILs in each of the two experiments; however, genotypic differences between the experiments did not show significant interactions, and hence data points were pooled into single genotypic value for analysis (not shown). The genotypes showed a continuous distribution ranging from 22.2 to 98.0 %, with a mean of 57.1 ± 12.9 % and a genotypic coefficient of variability (*GCV*) of 21.7 % (Fig. 1). The data confirmed KF No. 1 (P_1), the Al-tolerant parent, was significantly more tolerant to aluminum toxicity than NN1138-2 (P_2), Al-sensitive parent, as reported previously (Qi et al. 2008; Korir et al. 2010, 2011). RRE means were 75.9 ± 3.7 % for KF No. 1 and 46.3 ± 1.2 % for NN1138-2 (Fig. 1). However, some RILs showed transgressive segregation on either side of the parents, indicating that multiple genetic factors were segregating in the population. The estimated heritability of RRE was 91.2 %, indicating there was little environmental influence and, therefore, reasonable genotypic progress in Al-tolerance can be achieved through selecting on RRE. Joint segregation analysis of RIL, P_1 and P_2 (Gai 2006), showed the inheritance model best fitting Al-tolerance was model I-10 (Table 1) best described as polygenic with four additive major genes (three equal with effects plus one with a lesser effect. The three major genes with equal effect showed a much larger additive effect than the other major gene, 0.075 vs. 0.019, in RRE, while the collective additive effect of polygenes was 0.0328 (Table 1). The major gene and polygene heritability was 81.95 and 13.91 %, respectively, indicating that the former accounted for a major part of the genotypic variation in this RIL population.

Additive QTL mapped

The mapping results coincide with those from the above-mentioned statistical analysis on inheritance of Al-tolerance. Four additive QTL were identified with both the MCIM and CIM procedures on four LGs (Chromosomes), i.e. A2 (Gm08), K (Gm09), B1 (Gm11) and L (Gm19), and contributed 5.73–8.92 % in total 29.39 % (MCIM) and 4.03–11.72 % in total 31.43 % (CIM) of phenotypic variance (PV) (Table 2). Positive alleles on A2, K and B1 were inherited from KF No. 1 and the allele on L was from NN1138-2. The QTL positions, confidence intervals and direction of additive effects of the QTL on A2, K and B1 were comparable between the two linkage mapping procedures; however, the magnitude of their additive effects was slightly different. Only the QTL on B1 was mapped differently (18.5 cM apart) by the two procedures probably due to differences in software for each of the genetic models. Although individual QTL effects were not significantly different with the two mapping algorithms, the genetic contribution R^2 values in CIM (for QTL on A2 and B1) were greater than their respective $h^2_{(a)}$ values in MCIM. This is because the genetic model of CIM determines only

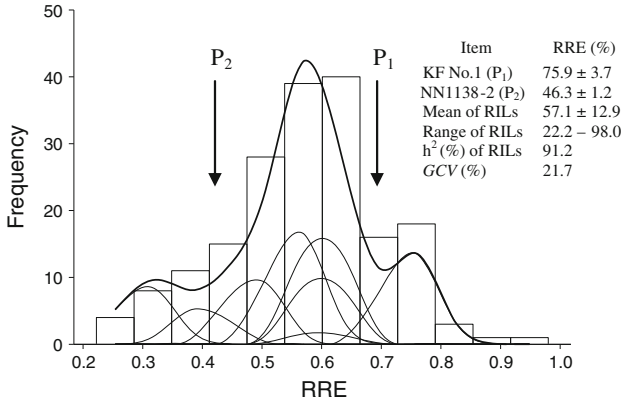

Fig. 1 Distribution of RRE among 184 RILs from the cross KF No. 1 × NN 1183-2. *RRE* relative root elongation. *Column chart* and the *solid line* represent total actual distribution and total theoretical distribution of RRE, respectively. The *fine lines* represent theoretical distribution of each genetic component under the best fitting model conditions of I-10 with four additive major genes and additive dominant polygenes (Gai 2006). Statistics of the important items are depicted in the *inset* table

additive effect (as R^2), and not epistasis nor QTL × E interactions, whereas MCIM partitions total PV into component variance due to additive ($h^2_{(a)}$), epistasis and QTL × E interactions. Hence, $h^2_{(a)}$ value is naturally expected to be lower than R^2, but instances where $h^2_{(a)}$ is greater than R^2 can be attributable to chance. These results suggest the combined use of two mapping methods improves the credibility of results and increases our confidence in the QTL positions on the linkage map.

Epistatic QTL pairs mapped

The 2D genome scan of MCIM showed that four different epistatic QTL pairs mapped on six LGs explained 3.84–5.73 % in total or 18.75 % of PV (Table 3). Most epistatic loci were not individually significantly additive and none was found to be duplicative for any epistatic QTL pair. In addition, none of any two major QTL was epistatic, indicating the major QTL were individually and significantly additive. However, the major QTL region GMpTI_D-Sat_247 on B1 interacted with a non-additive QTL locus A953_2H-K411_11 on B2, while a QTL on A2 that was mapped only 5.5 cM from the major QTL (133.8 cM) was epistatic to QTL locus on F. The presence of additive and epistatic interactions among alleles of the two parents could partly explain the tolerant and sensitive transgressive segregation observed in the RIL population.

Collective unmapped minor QTL and total genetic dissection of the RIL population

The sum of all contributions from mapped additive QTL and epistatic QTL pairs in RRE was less than the total genetic variation estimated from genotypic variances in ANOVA, similar to results found by Korir et al. (2011). Therefore, the remaining genetic variation was attributed to those QTL not detected in the mapping procedure and designated as genetic variation due to unmapped QTL or a collection of unmapped minor QTL. The total phenotypic variance was partitioned into mapped QTL, collective unmapped minor QTL, QTL × environment and environment variance components (Table 4). The heritability of RRE from the joint data analysis was 91.20 % and additive QTL effects contributed about 29.39 % while epistasis contributed about 18.75 % to PV (or 32.23 and 20.56 % contribution to genotypic variation, respectively). This indicated additive QTL effects were more important than epistatic effects; however, both were important in genetic control of Al-tolerance (Fig. 1; Table 4). However, it should be emphasized that a substantial part, 43.07 %, of the phenotypic variance or 47.23 % of the genetic variance was due to unmapped QTL. The unmapped QTL were mainly composed of small effects or minor QTL because the saturation level of the map (average distance of 2.8 cM between adjacent markers) should be sufficient to detect all the major QTL. In addition, the environment and genotype × environment (G × E or QTL × EXP here) variances contributed 8.15 and 0.65 % of the total phenotypic variance, respectively, indicating the QTL × EXP variation was not important and negligible in accurately

Table 1 Estimates of genetic parameters of optimum model I-10 for RRE among 184 RILs from a cross of KF No. 1 × NN1138-2 using segregation analysis under a major gene plus minor gene inheritance model

First-order genetic parameters estimate		Second-order genetic parameters estimate					
Mean	Genetic effect (%)	σ^2_p	σ^2_p	σ^2_{mg}	σ^2_{pg}	h^2_{mg} (%)	h^2_{pg} (%)
0.6126	$d_a = d_a = d_c = 0.075$ $d_d = 0.0190$ [d] = 0.0328	0.0338	0.0014	0.0277	0.0047	81.95	13.91

RRE relative root length elongation, the same is for later tables. d_a, d_b, d_c, d_d additive effect of the major genes, *[d]* collective dominance effect of polygenes, σ^2_p total phenotypic variance, σ^2_e environmental variance, σ^2_{mg} major gene variance, σ^2_{pg} polygene variance, h^2_{mg} (%) major gene heritability, h^2_{pg} (%) polygene heritability (Gai 2006)

estimating RRE. Thus, the QTL mapping results provided more genetic information than the segregation analysis and importance of the collective unmapped minor QTL was realized and the technology for integrating these QTL should be considered in breeding for Al-tolerance due to lack of markers.

Association analysis

Genetic variation in the released cultivar population

Tolerance to Al-toxicity is a complex trait that needs precise phenotyping to accurately determine its genetic factors (Jena and Mackhill 2008). The 188 cultivars used for the AM study are a sample of 618 cultivars released in YCRV in China (Supplemental Table 1). Tolerance levels in the AM population phenotyped for RRE was wide and the values ranged 25.11–93.73 % with an average of 59.16 % (Supplemental Table 2, frequency distribution of RRE, individual values for each accession and check not shown for saving space). The mean RRE values for the tolerant check cultivars PI 416937 and KF No. 1 were 85.81 and 74.33 %, respectively, and 44.67 % for the susceptible check cultivar NN1138-2. The significant variation among cultivars was indicative of the variation for Al-tolerance that can be expected among locations within a production region such as the YCRV. Broad sense heritability for the trait calculated from the ANOVA on an entry means basis was 91.80 % suggesting there was a reasonably low level of error in determining the RRE for the population. One important characteristic of RRE is its freedom from the confounding effects of constitutive variation inherent among cultivars; therefore the variation observed is more likely due to the actual response to Al-stress.

Population structure of the released cultivar population

For saving space, in this section data were summarized as text with the tables and figures omitted. Of the 186 polymorphic loci, the number of available alleles varied from 2 to 16 with a mean of 5.83 alleles per locus and polymorphism information content (PIC) ranged from 0.11 to 0.97 with a mean of 0.64 per locus, indicating reasonable diversity existed within the cultivar population for RRE. Genome-wide loci pairs in LD with $r^2 > 0.001$ were 14,794 and the maximum observed r^2 value was 0.395. Linkage disequilibrium decayed with distance to basal levels of approximately $r^2 < 0.1$ at distances <25 cM and extended to loci 120–150 cM apart. Population sub-structuring aims at assigning members of the population into subpopulations that maximize H–W linkage equilibrium within them. Using the model-based clustering method, maximum logarithmic posterior probability was attained at

Table 2 Quantitative trait loci (QTL) associated with RRE among 184 RILs from a cross of KF No. 1 × NN1138-2 as detected by mixed-model-based composite interval mapping (MCIM) of QTLNetworkV2.0 and composite interval mapping (CIM) of WinQTLCartV2.5

LG (Chr.)	MCIM, QTLNetwork						CIM, WinQTLCart					
	Marker interval	Site (cM)	CI (cM)	*F*	*A* (%)	$h^2_{(a)}$ (%)	Marker interval	Site (cM)	CI (cM)	LOD	*A* (%)	R^2 (%)
A2 (Gm08)	STAS8_3T-STAS8_6T	133.8	132.4–135.5	14.8	0.046	8.36	STAS8_3T-STAS_6T	133.7	133.7–134.3	5.6	0.041	11.72
B1 (Gm11)	GMpTI_D-Sat_247	30.7	29.7–32.2	23.1	0.054	8.92	Sat_128-Sat_149	49.2	47.3–51.4	3.7	0.038	10.44
K (Gm09)	GNB020-GNT010a	78.8	76.1–83.1	12.7	0.017	6.38	GNB020-GNT010a	76.8	76.0–80.6	2.2	0.026	4.03
L (Gm19)	GNB036-Satt313	27.4	26.7–27.8	14.2	−0.090	5.73	GNB036-Satt313	27.4	27.0–28.6	2.7	−0.030	5.24

LG(Chr.) linkage group (Chromosome), the same is for later tables, *CI* confidence interval for QTL peak of F/LOD at $\alpha = 0.05$, *F* *F*-value for the region of QTL peak exceeding the critical F-threshold value of 9.4 calculated by 1,000 permutations at genome-wise 0.05 significant level, *A* QTL Additive effect, $h^2_{(a)}$ the phenotypic variation explained by the additive QTL, R^2 percentage of phenotypic variance explained by QTL, *LOD* logarithm of odd value

Table 3 Epistatic QTL pairs detected with MCIM for RRE among 184 RILs from a cross of KF No. 1 × NN1138-2

Epistatic QTL pair	LG (Chr.)	Marker interval	Site (cM)	Distance (cM)	CI (cM)	*P*-value	*aa*	$h^2_{(aa)}$ (%)
1	A2 (Gm08)	STAS820T-STAS8_3T	128.3	1.04–4.05	126.1–131.3	0.001	0.030	5.73
	F (Gm13)	GNE495-GNB142	28.3	0.04–0.09	26.7–27.7			
2	A2 (Gm08)	Satt525-A117_2H	27.4	1.04–4.05	126.1–131.3	0.005	0.024	3.84
	F (Gm13)	Sat_234-A186H	18.1	1.99–0.13	37.5–38.6			
3	B1 (Gm11)	GMpTI_D-Sat_247	30.7	2.99–0.99	29.7–32.2	0.004	0.036	4.08
	B2 (Gm14)	A953_2H-K411_11	52.4	0.03–0.95	51.1–53.3			
4	C2 (Gm06)	OPW13-B131V	31.2	0.03–0.89	30.7–32.1	0.003	−0.031	5.10
	D1b (Gm02)	Satt611-GNE239	8.0	8.00–0.15	4.0–9.1			

Distance distance of the QTL to flanking markers, *CI* confidence interval of the QTL site at $\alpha = 0.05$, *aa* epistasis effect, $h^2_{(aa)}$ the phenotypic variation explained by the epistatic QTL pair

Table 4 Dissection of phenotypic variance into genetic, genotype × environment (G × E) and environment components among 184 RILs from a cross of KF No. 1 × NN1138-2 for Al-tolerance trait RRE (%)

Genetic portion				G × E	Environment	Total
Additive QTL	Epistatic QTL	Collective minor QTL	Total			
29.39 (4; 32.23)	18.75 (4; 20.56)	43.07 (47.23)	91.20	0.65	8.15	100

The numbers in parentheses in the genetic portion columns are the number of additive or epistatic QTL pairs and the percentage of genetic variance explained by the respective types of QTL obtained from the QTL Network. The collective minor QTL genetic portion is calculated from $h^2 - h^2_{(a)} - h^2_{(aa)}$, where h^2 is the heritability estimated from ANOVA and $h^2_{(a)}$ and $h^2_{(aa)}$ are additive and epistatic contribution obtained from the mapping procedure, respectively. The percentage explained by additive QTL or epistatic QTL pairs is estimated as: $(h^2_{(a)}/h^2) \times 100$ or $(h^2_{(aa)}/h^2) \times 100$

$k = 7$ where the value of alpha (degree of admixture) was 0.0385 and average F_{st} (measure of genetic differentiation among subpopulations) was 0.3615. This implies most cultivars originated from one primary ancestor, with a few admixed individuals and that total population heterozygosity was reduced ~36 % by subdividing it into the seven major clusters. A χ^2 test showed model-based clusters were significantly different ($\chi^2 = 27.67$, $\chi^2_{0.01,6} = 16.81$); therefore, the percentage parentage of individual cultivars in each cluster (the Q matrix) was used in the GLM-Q association analysis model.

Association mapping

This study tested the association of molecular markers with RRE in presence of population structure (Q) on a genome-wide level, and the estimated $-\log_{10}$ (*P* value) was plotted against the respective marker loci. At $P < 0.01$, i.e. $-\log_{10}$ (*P* value) > 2.0, six markers were associated with Al-tolerance and accounted for over half (57.9 %) of PV (Table 5). Two SSR markers, Satt209 and Satt186, on LGs A2 and D2, respectively, showed stronger association with Al-tolerance ($P < 0.0031$; $-\log_{10}$ (*P* value) > 2.5) and together explained 24.9 % of PV. The other four loci, Sat_364 (B1), Satt005 (D1b1), Sat_240 (F) and Satt284 (L), were associated with Al-tolerance at $0.0031 < P < 0.01$. A total of 11 markers explaining 85.2 % of PV were identified at $P < 0.03$, i.e. $-\log_{10}$ (*P* value) > 1.5 while at a less stringent *P* value of <0.05, a total of 20 statistically significant markers were identified (not shown). This suggests that with increases in significance level, false-positive marker-trait associations with relatively low contributions to PV could be discarded while retaining those explaining significantly high portion of PV.

Due to a predicted high rate of false positives resulting from multiple testing of sites (Storey and Tibshirani 2003), we subjected the AM markers to further testing using an independent segregating population. If AM is truly complementary to LM, then the AM markers, or any other marker(s) in LD should explain significance variance for Al-tolerance in the segregating population. Therefore, important markers relevant to MAS could be verified and novelty of AM demonstrated. Additive QTL regions in this and previous LM experiments together were detected on 10 of 11 LGs in which the AM identified at $P < 0.03$ (Table 5). Of these 10, QTL regions on LGs A2, B1, F, K and L were detected concurrently in the present and previous LM but QTL on D1b1, C2, D2, J and O corresponded only to QTL regions from previous studies. Therefore, 5 QTL regions, detected by AM at $P < 0.03$, were identified at the same time by LM in our study. QTL on LG I was unique to AM, suggesting it might be a new QTL locus

Table 5 Identified SSR markers significantly associated with Al-tolerance among 188 cultivars and comparisons with linkage mapping results of our study and previous studies

LG (Chr.)	Marker	No. of alleles	Site (cM)	R^2	QTL from linkage mapping				
					Present study			Previous studies	
					Closest marker	Site (cM)	$h^2_{(a)}$ (%)	CI (cM)	References
A2 (Gm08)	**Satt209**	**9**	**128.44**	**0.145**	STAS8_6T	133.8	8.36	121.7–123.7	Bianchii-Hall et al. (2000)
B1 (Gm11)	**Sat_364**	**8**	**84.25**	**0.082**	Sat_247 (MCIM)	30.7	8.92	39.1–41.1	Bianchii-Hall et al. (2000)
					Sat_128 (CIM)	49.2	10.44	65.4–78.0	
								44.7–46.9	Korir et al. (2011)
C2 (Gm06)	Satt286	7	101.75	0.052				126.2–145.5	Sharma et al. (2011)
D1b1 (Gm02)	**Satt005**	**7**	**75.29**	**0.086**				75.8–93.8	Qi et al. (2008)
								81.2–84.7	Korir et al. (2011)
D2 (Gm17)	**Satt186**	**10**	**105.45**	**0.104**				108.0–124.0	Qi et al. (2008)
								88.9–97.3	Korir et al. (2011)
F (Gm13)	**Sat_240**	**9**	**25.58**	**0.085**	*GNE495-GNB12*	*28.3*	*5.73*	33.2–36.1	Sharma et al. (2011)
					Sat_234-A186H	*18.1*	*3.84*	66.2–68.2	Bianchii-Hall et al. (2000)
I (Gm20)	Sat_174	6	36.59	0.067					
J (Gm16)	GMES0069	4	60.40	0.052				56.2–58.2	Bianchii-Hall et al. (2000)
K (Gm09)	Sct_190	5	77.37	0.043	GNB020	78.8	6.38	56.2–58.4	Korir et al. (2011)
L (Gm19)	**Satt284**	**6**	**38.16**	**0.077**	GNB036	27.4	5.73	49.6–57.0	Qi et al. (2008)
								27.4–31.3	Korir et al. (2011)
O (Gm10)	GMES1703	5	60.60	0.059				86.2–90.1	Korir et al. (2011)

The map positions (cM) of AM were estimated on the USDA public soybean genetic linkage map (Song et al. 2004); those of LM in "Present study" column were on the NJRIKY map; those in "Previous studies" column were on the respective author's genetic linkage maps. Markers in boldface were detected at $P < 0.01$, otherwise at $P < 0.03$. The two markers in italics form in the "Present study" column are two pairs of epistatic QTL while the other markers in the same column are additive QTL. In "Previous studies" column the markers are not presented but the same or close to the AM marker in a same row

discovered in this study. These results indicate a relative consistency in mapping QTL for Al-tolerance between AM and LM for the respective populations used (Table 5).

Allelic effects and genetic dissection of the released cultivars

Superior alleles and their carriers

Allelic effects of the 11 marker loci significantly associated ($P < 0.03$) with RRE were estimated (each locus composed of 4–10 alleles, in a total of 76 alleles) (Table 6). For each locus, the two best alleles were selected and ten typical cultivars carrying those alleles chosen (Table 7). The three best favorable alleles Satt209-A210, Satt005-A159 and Satt186-A267 showed significantly higher positive effect of 36.31, 32.84 and 30.37 %, respectively, on mean performance of RRE and their respective number of carrier cultivars of 4, 11 and 9, respectively, had mean RRE of 90.65, 84.66 and 86.46 % (Tables 6, 7). The next best marker alleles Satt209-A219 and Satt005-A186 exerted positive effects of 27.83 and 23.78 % in seven and ten carrier materials, respectively. Of the best two alleles of each of the six significant loci ($P < 0.01$), allele Satt209-A210 showed greatest phenotypic effect of 36.31 % while Sat_240-A268 had the least effect of 11.67 % and the number of carrier cultivars of each allele was 4 (mean RRE 90.65 %) and 25 (mean RRE 68.36 %), respectively, indicating the favorable alleles were concentrated in a relatively smaller number of the cultivars roughly consistent with frequency distribution in Supplemental Table 2.

Superior cultivars and their genetic dissection

Elite cultivars tolerant to Al-toxicity were identified and they were genotyped for specific alleles because they could be potential donor parents in breeding for tolerance. The 188 released cultivars were ranked in descending order of percent RRE and compared with the performance of the two Al-tolerant check cultivars PI 416937 and KF No. 1 which had mean RRE of 85.81 and 74.33 %, respectively (data not shown). Taking 74.33 % as the lowest RRE to consider cultivars as having superior Al-tolerance, 24 elite cultivars were selected and their allele constitutions determined on the basis of the two best alleles for the 11 significant ($P < 0.03$) markers (Table 8). These cultivars

Table 6 Identified markers and their allele effects significantly associated with Al-tolerance trait RRE (%) among 188 cultivars

LG (Chr.)	Marker	Allele expressed in band bp number (additive effect %)								
A2 (Gm08)	**Satt209**	210 (+36.31)	219 (+27.83)	174 (+19.64)	192 (+11.82)	165 (+9.51)	183 (+6.84)	201 (−2.26)	237 (−5.85)	156 (−19.72)
B1 (Gm11)	**Sat_364**	425 (+19.57)	437 (+12.76)	419 (+6.83)	441 (+2.88)	409 (+2.59)	401 (+0.94)	375 (−4.84)	393 (−10.85)	
C2 (Gm06)	Satt286	219 (+10.62)	228 (+4.85)	255 (−1.23)	237 (−7.84)	246 (−14.3)	264 (−20.80)	210 (−22.10)		
D1b1 (Gm02)	**Satt005**	159 (+32.84)	186 (+23.78)	177 (+14.71)	168 (+7.56)	150 (−3.77)	141 (−2.17)	195 (−6.73)		
D2 (Gm17)	**Satt186**	267 (+30.37)	258 (+20.44)	240 (+13.63)	213 (+5.93)	249 (+3.81)	222 (+2.32)	204 (−1.87)	231 (−12.96)	276 (−18.80)	195 (−19.40)
F (Gm13)	**Sat_240**	214 (+20.65)	268 (+11.67)	254 (+8.71)	200 (+5.33)	244 (+2.96)	224 (0.35)	192 (−3.13)	286 (−10.87)	166 (−16.70)
I (Gm20)	Sat_174	219 (+18.75)	237 (+5.29)	228 (−0.22)	210 (−5.24)	246 (−11.60)	255 (−19.80)			
J (Gm16)	GMES0069	162 (+13.83)	159 (+8.48)	165 (−3.45)	168 (−10.23)					
K (Gm09)	Sct_190	267 (+18.85)	258 (+6.52)	249 (−2.38)	240 (−5.94)	231 (−10.20)				
L (Gm19)	**Satt284**	285 (+23.47)	258 (+15.83)	276 (+11.34)	267 (+1.25)	249 (−5.51)	294 (−10.20)			
O (Gm10)	GMES1703a	179 (+12.73)	176 (+4.47)	167 (+2.53)	164 (−0.74)	182 (−10.30)				

Markers in boldface were detected at $P < 0.01$, otherwise at $P < 0.03$

originated from 9 provinces of China, 16 from the Yellow River eco-region and 8 from Changjiang River eco-region. About one-half of them, including the best five (with RRE > 85 %; and better performers than PI 416937) originated from Shandong province in the Yellow River eco-region. Of these five, four shared the best allele Satt209-A210, while cultivars that performed better than KF No. 1 shared most of the alleles with effects of >20 % on RRE. Generally, the elite cultivars possess multiple favorable alleles suggesting the potential for recombination in a breeding design for Al-tolerance. For example, the most tolerant cultivar N25388 might be further improved through complementary recombination with the favorable alleles of Satt186, Sat_240 and Satt284 from other tolerant cultivars in Table 8. In addition, combining favorable alleles among other cultivars would also improve Al-tolerance.

Allelic dose effect and tracing favorable alleles in cultivar pedigrees

Genetic base of the soybean cultivars is not very wide; therefore, it is important to trace desirable favorable alleles in the repertoire of ancestors to broaden the genetic base of the current high yielding cultivars. Variation in Al-tolerance ranged ~25 % to ~94 % with a mean of ~59 %, indicating most of the released cultivars have moderate tolerance (Supplemental Table 2). The genetic relationship of favorable alleles in the cultivars were traced in five major family pedigrees with 58-161, Xudou No. 1, Qihuang No.1, NN493-1 and NN1138-2 as respective ancestors (Supplemental Table 3 showing the NN1138-2 pedigree as an example for others). The two best alleles from each of the 11 major loci (Tables 5, 7) were tracked for their transmission in the five cultivar pedigrees and analyzed for the presence or absence of Al-tolerant favorable alleles of each marker. Each pedigree ancestor had its own favorable alleles that were transmitted to progenies, but only some progenies inherited these alleles suggesting some might be lost in transition. The ancestor parents showed varying levels of tolerance to aluminum toxicity, i.e. 78.5, 67.8, 81.0, 44.7 and 80.4 % for 58-161, Xudou No.1, Qihuang No.1, NN1138-2 and NN493-1, respectively, suggesting other sources of favorable alleles were contributed to progenies from other parents. The progenies tended to share the favorable alleles but likely at different frequencies due to diverse parents used in various breeding programs. For each ancestor, the pedigree families had a different number of favorable alleles, or lacked them altogether. For example, the two best alleles Satt209-A210 and Satt005-A159 were absent in NN1138-2 pedigree. The highest number of favorable alleles in an individual was 8 (cultivar code 85 in Xudou No.1 pedigree, data not shown)

Table 7 Effects of two best alleles on each locus and their carriers among 188 cultivars

Favorable allele (LG, Chr.)	Allele effect (%)	No. carriers	Mean RRE (%)	Carrier cultivar									
Satt209-A210 (A2, Gm08)	+36.31	4	90.65	85	86	87	142						
Satt209-A219 (A2, Gm08)	+27.83	7	82.06	3	15	57	58	59	88	95			
Sat_364-A425 (B1, Gm11)	+19.57	44	77.90	2	188	22	184	186	124	137	138	127	120
Sat_364-A437 (B1, Gm11)	+12.76	31	71.10	15	65	23	57	84	58	122	42	146	49
Satt286-A219 (C2, Gm06)	+10.62	26	78.46	186	124	120	97	185	4	132	125	31	27
Satt286-A228 (C2, Gm06)	+4.85	36	66.91	10	65	94	136	32	58	108	7	110	82
Satt005-A159 (D1b1, Gm02)	+32.84	11	84.66	188	144	98	125	163	39	158	85	67	117
Satt005-A186 (D1b1, Gm02)	+23.78	10	75.56	58	131	63	81	135	91	49	43	70	45
Satt186-A267 (D2, Gm17)	+30.37	9	86.46	184	137	18	30	174	175	43	70	11	
Satt186-A258 (D2, Gm17)	+20.44	30	76.57	130	71	51	88	143	106	68	72	110	73
Sat_240-A214 (F, Gm13)	+20.65	18	77.41	2	68	125	5	121	139	44	165	49	153
Sat_240-A268 (F, Gm13)	+11.67	25	68.36	71	22	51	186	162	144	143	73	42	158
Sat_174-A219 (I, Gm20)	+18.75	36	76.56	2	66	71	51	38	101	79	120	144	143
Sat_174-A237 (I, Gm20)	+5.29	32	63.05	109	184	58	123	89	108	62	112	82	31
GMES0069-A162 (J, Gm16)	+13.83	56	68.96	15	109	10	188	71	51	38	184	124	137
GMES0069-A159 (J, Gm16)	+8.48	62	63.56	2	26	66	1	14	92	101	79	3	57
Sct_190-A267 (K, Gm09)	+18.85	2	74.56	85	56								
Sct_190-A258 (K, Gm09)	+6.52	30	61.66	184	88	3	57	58	99	90	62	63	112
Satt284-A285 (L, Gm19)	+23.47	15	75.57	65	94	137	63	28	122	145	39	55	49
Satt284-A258 (L, Gm19)	+15.83	25	67.86	79	61	108	103	4	82	31	107	102	165
GMES1703a-A179 (O, Gm10)	+12.73	21	70.51	3	91	7	18	45	46	48	49	55	59
GMES1703a-A176 (O, Gm10)	+4.47	74	60.07	2	66	71	186	101	124	138	88	120	3

Mean RRE average RRE of the carrier cultivars. In carrier cultivar column are codes of cultivars whose names can be found in supplemental Table 1. Not all but only the maximum of ten carriers for each allele are listed here. Markers in boldface were detected at $P < 0.01$, otherwise at $P < 0.03$

and the average was 3.25, indicating substantial Al-tolerance potential in recombination and accumulation of favorable alleles in future breeding.

The 10 best Al-tolerant cultivars with the favorable alleles (from the ancestor parent or from other sources) averaged 87.0 % RRE and was 3.31 times the average (26.3 %) of the 10 most sensitive pedigree cultivars. The average favorable allele number of these tolerant cultivars was 2.3 times of the latter but composition of favorable alleles among the cultivars with high tolerance was generally quite different, suggesting that tolerance to Al is dose dependent on alleles. However, there were also unexpected cases, such as cultivar N23787 (Table 8, 76.4 %). N23787 had only two favorable alleles yet was almost as tolerant as cultivar N25435 (80.4 %) with five favorable alleles. In some cases, some moderately high tolerant cultivars, e.g. N07038.1 (in NN493-1 family) and N25461 (in NN1138-2 family) had only two favorable alleles, some of which were not derived from the available ancestors. On the other hand, some sensitive cultivars, e.g. N25275 (in 58-161 family) and N25427 (in Xudou No.1 family) had more favorable alleles with some derived from ancestors while the origin of others could not be traced since not all of the ancestors of this pedigree were available for analysis. However, the favorable Al-tolerant cultivars from both Yellow and Changjiang Valleys were found with favorable alleles indicating germplasm from both regions can be used as donors to broaden genetic base for higher tolerance through germplasm exchange.

Marker-assisted design of crosses based on QTL-allele matrix

Peleman and van der Voort (2003) proposed a concept of "Breeding by Design". It aims to harness all allelic variation for all genes of agronomic importance through a combination of precise genetic mapping, high-resolution chromosome haplotyping and extensive phenotyping. Accordingly, QTL mapping for favorable alleles and genetic dissection of germplasm resources for potential parental materials are two prerequisites toward "Breeding by Design". From the above association mapping results, the two kinds of genetic information can be obtained and organized in gene/QTL-allele matrix as in Table 8, which

Table 8 Genetic structure of 24 favorable Al-tolerant cultivars based on the two best alleles from each of the 11 marker loci significantly associated with Al-tolerance ($P < 0.03$) among 188 cultivars

Name	Code	Origin	RRE (%)	Marker locus-allele																					
				Satt209		Sat_364		Satt286		Satt005		Satt186		Sat_240		Sat_174		GMES0069		Sct_190		Satt284		GMES1703a	
				A210	A219	A425	A437	A219	A228	A159	A186	A267	A258	A214	A268	A219	A237	A162	A159	A267	A258	A285	A258	A179	A176
N25388	85	SD	93.73	√		√			√	√							√		√	√					√
N25019	86	SD	91.18	√													√	√			√				√
N09125.1	87	SD	89.47	√				√							√		√				√			√	
N23774	142	SD	88.19	√									√						√		√				
N02951.1	3	JX	86.46		√														√		√				√
N00609.1	15	SD	85.26		√		√											√						√	
N03286	57	SD	84.23		√		√												√		√				
N04978	58	SD	83.16		√		√		√		√						√				√				
N23688	59	SD	82.28		√														√	√					
N00666	88	SD	81.05		√								√						√		√				√
N25435	184	BJ	80.36			√						√					√		√		√				
N00617	137	JS	80.35			√						√							√			√			
N04930.1	144	BJ	79.01							√					√	√			√						
N06831.1	98	SH	78.47							√										√					
N00013.1	2	JS	78.46			√								√		√				√					√
N01197.21	127	ZJ	77.19			√					√								√						
N10497.1	188	SC	77.06			√				√									√						
N23787	22	SC	76.35			√									√										
N09117	131	SD	76.28					√			√													√	
N09580	63	SD	75.29								√										√	√			
N01599.1	130	HU	75.26										√												
N20646	65	GZ	75.22				√		√													√			
N24452	94	HE	74.48						√													√		√	
N01010	79	JS	74.32	√												√				√			√		
NN1138-2	68	JS	44.67										√	√		√				√					

BJ Beijing, *GZ* Guizhou, *HE* Henan, *HU* Hunan, *SC* Sichuan, *SD* Shandong, *SH* Shanghai, *JS* Jiangsu, *JX* Jiangxi, *ZJ* Zhejiang, √ indicates presence of the favorable allele. NN1138-2 is an intolerant check

is in fact a small sample of the matrix that breeders can design to formulate their crossing plans to combine favorable alleles into one individual. For example, it is possible to improve the elite accessions in Table 8 for their tolerance to Al-toxicity by choosing parents with their genetic constitutions complementary to each other. For example, accession "No. 85" (N25388) having 8 favorable alleles can be crossed with the accession "No. 68" (N05461) with two different favorable alleles, Satt186-A258 and Sat_240-A214. Accession "No. 137" (N00617) with another favorable allele Satt284-A285 was used as a parent to develop lines in which all the 11 loci have favorable alleles. Therefore, marker-assisted design of crosses based on a QTL-allele matrix can be practised based on the known genetic information. If the genetic information can be provided for all traits of interest in breeding programs, the marker-assisted design of crosses can be used for the improvement of multiple traits, and it can be facilitated using computer programs for analyzing large matrix data. We propose to designate this kind of marker-assisted pairing of crosses as the QTL-allele matrix method and we propose that this method has potential for improving traits in marker-assisted breeding programs.

Discussion

The relative consistency of QTL mapping results among indicators of Al-tolerance

In a recent study, Korir et al. (2011) used the same RIL population to map QTL conferring Al-tolerance under greenhouse sand culture. Three growth-related indicators for Al-tolerance, namely relative total plant dry weight (RTDW), relative shoot dry weight (RSDW) and relative root dry weight (RRDW) were analyzed for genetic architecture. RTDW showed relatively higher correlations and shared marker regions with RSDW and RRDW. Three types of QTL, i.e. four additive QTL, four epistatic QTL pairs and collective unmapped QTL, were identified for RTDW, and similar results were found for RSDW and RRDW. Additionally, one major QTL linked to marker region GMKF046-Sat_128 on LG B1 was shared by the three traits. In the present study, the popularly used RRE in hydroponics culture (Villagarcia et al. 2001) was used for a fast and feasible evaluation of Al-tolerance. As a comparison, the results obtained in our study were similar to that of the previous study by Korir et al. 2011, i.e. four additive QTL were detected, with one on LG B1 located close to that of RTDW, RSDW and RRDW and another on LG K close to that of RTDW. In addition, the proportions explaining the total phenotypic variance of the three types of QTL were also similar to the other indicators, especially RTDW. The comparisons further confirm all the four indicators can be used for evaluation of Al-tolerance of young seedlings of soybeans.

The rationality and structure of the tested population in association mapping of Al-tolerance QTL

Previous studies for Al-tolerance QTL showed the number of QTL detected was limited and depended on the genetic background of the population (Bianchi-Hall et al. 2000; Qi et al. 2008). Chinese cultivated soybean has been shown to exhibit geographic differentiation and genetic diversity based on the region in which they were developed (Cui et al. 2000; Dong et al. 2004). Our breeding materials are a collection of cultivars released over many years in YCRV, representing elite lines accumulated for genes and QTL alleles for yield and abiotic stresses. Since these breeding materials are genetically stable and well adapted to regional growing conditions, they were relevant genetic germplasm for AM of Al-tolerance.

The cultivars showed a wide variability in RRE and high heritability (90.2 %), much higher than that of RRE in a diversity panel of maize (41 %) (Vargas-Duque et al. 1994; Krill et al. 2010), indicating good control of random error during phenotyping. This suggested high potential for genetic improvement and dissection of Al-tolerance in soybean. The relatively large sample size used for AM was desirable because it would improve population differentiation (Rosenberg et al. 2001) and increase detection and allow determination of more alleles, even at low frequencies. The model-based clustering analysis revealed seven subpopulations roughly differentiated and coincided with geo-ecological adaptation and cycles of cultivar improvement. Population structure has been estimated from as low as 20 (Rosenberg et al. 2001) to about 100 (Semon et al. 2005). As low as 15 SSR markers were effective to differentiate among closely related individuals (Wilkening et al. 2006). Therefore, the population used in this study along with 43 SSR markers was thought to be applicable to detection of the population structure for Al-tolerance.

In this study, extensive LD was detected among both syntenic and nonsyntenic markers. The loci pairs with a useful level of LD (i.e., $r^2 > 0.1$, $P < 0.01$; Malysheva-Otto et al. 2006) accounted for only 5.88 % of the total loci pairs in LD (144 out of 2447 loci pairs in LD). The LD rapidly decayed to a basal value of $r^2 \geq 0.1$ occurring for distances <25 cM suggesting it should be possible to achieve resolution down to a 25-cM level. Variable LD across the genome is due to several factors such as variation in recombination rate and selection, but the most probable cause of the high-level LD in soybean is the selfing system of mating, as reported in other self-pollinated crops (Malysheva-Otto et al. 2006; Nordborg and

Tavare 2002). The extent of LD determines the number of markers to cover the genome and consequently the resolution level of mapping. For the soybean genome of ~3,000 cM, about 150–300 markers would be sufficient to conduct a genome-wide association mapping.

One constraint of AM strategy is the easy detection of spurious marker-trait association resulting from presence of population structure (Flint-Garcia et al. 2005). The GLM + Q model that utilized population structure as a covariate was used to account for effect of population structure. At $P < 0.05$, 20 markers statistically associated with RRE were reduced to 11 and 6 markers when false-positive error rate changed from 3 to 1 %, respectively. These results demonstrate that the AM was able to detect significantly more markers for Al-tolerance compared with results from the previous biparental QTL mapping studies.

Association mapping combined with linkage mapping in finding Al-tolerance loci and alleles

In this study, it was important to compare markers identified for Al-tolerance with AM against present linkage mapping results and from previous analyses to demonstrate the novelty of AM method. Four studies available in the literature by Bianchi-Hall et al. (2000), Qi et al. (2008), Korir et al. (2011) and Sharma et al. (2011) were used as examples (Table 5). At $P < 0.03$, 10 of the 11 marker loci identified with AM were also detected with either the present or previous LM while the remaining QTL on I was detected only by the AM method. For example, the present LM identified QTL on A2 at 133.3 cM while AM identified marker Satt209 at 128.44 cM position (Table 5). Previously, QTL *Altol* 1-1 (Bianchi-Hall et al. 2000) was mapped in a 121.7–123.7 cM interval on A2 and was reported in the public SoyBase website (www.soybase.org).

On B1, AM detected Sat_364 at 84.25 cM on Song et al.'s map (2004), and in our study LM detected a QTL in GmpTI_D-Sat_247 at 30.7 cM (confidence interval of 29.7–32.2 cM) (MCIM) or Sat_128-Sat_149 at 49.2 cM (47.3–51.4 cM) (CIM) on present NJRIKY map with 834 markers. Previously, Sat_364 was located ~ 43.2 cM apart from *Altol* 1-3 (47.3–51.4 cM) on B1 in Bianchi-Hall et al. (2000). Qi et al. (2008) detected a QTL in GMKF046-GMKF080 at 71.6 cM (65.4–78.0 cM) on B1 based on the old NJRIKY map with 451 markers. Recently, LM detected a QTL in GMKF046-Sat_128 at 45.9 cM (44.7–46.9 cM) (MCIM) on B1 on NJRIKY map in sand culture (Korir et al. 2011). The above results are cited in Table 5. Here, Sat_364, Sat_247, Sat_128 and Sat_149 can be found on Song et al.'s linkage map and GMKF046 linked to Sat_128 at different positions, 71.6 and 45.9 cM, on the old and present NJRIKY map, respectively. The distances between Sat_364 by AM and those by LM were about 30.24 ~ 34.52 cM on Song et al.'s genetic linkage map and 16.05 ~ 36.35 cM on present NJRIKY map. It seems that the QTL detected on B1 from LM in NJRIKY population by Qi et al. (2008), Korir et al. (2011) and in the present study are a same QTL. As for this QTL and the AM locus, they might be different due to the above mentioned distances. However, this is not necessarily true because different mapping populations, Al-tolerance evaluation methods, genetic linkage maps and genetic models of statistical analysis were used in respective studies. Therefore, further work is needed for final confirmation.

As further examples, the AM marker loci (Satt005, Satt186 and Sct_190) detected on D1b1 (75.29 cM), D2 (105.45 cM) and K (77.37 cM), respectively, corresponded to QTL detected by previous LM in sand culture for RTDW trait (Korir et al. 2011) at 82.7, 87.7 and 57.4 cM positions, respectively. With AM, Satt284 located at 38.16 cM on L, and with linkage mapping, QTL was located at 28.8 cM for RTDW (Korir et al. 2011) and 27.4 cM for RRE in present study. In Qi et al. (2008), Satt186 (D2) and Satt284 (L) were located 2.5 and 11.4 cM, respectively, outside of the confidence intervals. In Sharma et al. (2011), Satt286 (C2) and Sat_240 (F) were located 14.4 cM and 7.6 cM outside of the confidence intervals. These results further demonstrate that although there were some differences in QTL positions between AM and LM, the difference was not profoundly large for the QTL to be declared different; therefore, it can reasonably be stated that the AM study was an effective alternative to linkage analysis.

In fact, AM can detect not only loci but also the number of alleles for each locus. In the studies cited in Table 5, only two alleles for each locus in each RIL population could be detected by LM and the total number of alleles cited in the three mapping populations was limited with a maximum of 30 alleles. AM used released cultivars to determine QTL with 76 alleles detected in total with 4–10 alleles detected for each of the 11 loci associated with Al-tolerance. Therefore, the QTL mapping strategy of AM combined with LM has potential in taking advantage of AM to find more loci and alleles with the advantage of LM to locate the QTL position on respective linkage groups. Thus, there is potential using the two methods simultaneously to verify each other's results.

The QTL mapping results for Al-tolerance in this study suggested that the combination of AM and LM identified markers Satt209, Sat_364, Sat_240 and Satt284. They were the most valuable markers for MAS for Al-tolerance since they were detected with different indicators both in the present and previous studies and their contributions to phenotypic variation were among the highest. This reasoning is in tandem with Vargas-Duque et al. (1994), Storey and Tibshirani (2003) and Krill et al. (2010) that

due to a predicted high rate for false discovery in AM, it is important to test the markers identified from AM through linkage mapping to ascertain consistency. However, the one advantage of AM is that a great number of recombination events were sampled among 188 released cultivars. As a result, more and new genetic factors were discovered and it is probable that the QTL identified in our AM were better resolved than in the linkage mapping studies.

Implications to breeding by design for Al-tolerance

In plant breeding, choosing parents and designing crosses for effective recombination is the first step of a breeding plan. The second step is to isolate elite candidates in segregating populations by selection. For the first step, the QTL-allele matrix method has provided a way of marker-assisted genetic design for crossing plans which compose the major part of breeding by design. According to Gai et al. (2012), the key to a successful implementation lies on the accuracy of the obtained QTL-allele matrices. If the QTL-allele matrices are reliable, the crossing plans and progeny selections can be carried out based on marker-assisted procedures. However, if the QTL-allele matrices have only a relative precision on genetic differences among the germplasm rather than on marker alleles, it can be very successful in pairing crosses but not the marker-assisted selection among progenies. In this study, elite Al-tolerant cultivars from Yellow River Valleys and Changjiang River Valleys were found with favorable alleles indicating germplasm from both areas can be used as donors to broaden genetic base through introduction of germplasm from the other eco-region.

The application of genomic selection (GS) proposed by Meuwissen et al. (2001) applied to breeding populations using high marker densities is emerging as a solution to marker-assisted progeny selection. GS is a form of MAS that simultaneously estimates all locus or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs) for selection. The key process of GS is the calculation of GEBVs for individuals having only genotypic data using a model obtained from a "training population" with both phenotypic and genotypic data known (Habier et al. 2009; Heffner et al. 2009; Hill 2010). The predicted breeding values (GEBVs) are then used for selection of the individuals without phenotypic data in the breeding cycle. To maximize GEBV accuracy, the "training population" must be representative of candidates or lines selected in the breeding program to which GS will be applied.

In addition to marker-assisted pairing of crosses, the QTL-allele matrix procedure can be used also for progeny selection through genotyping the segregants if a precise QTL-allele matrix is available. It seems that the GS procedure and our breeding by design procedure based on a QTL-allele matrix use similar philosophies of genome-wide MAS. But they are different in that the former uses the marker-trait information from a smaller "training population" for estimating GEBVs of the selection candidates while the latter uses the marker-trait information (QTL-allele matrix) from a large germplasm population, like our study, to estimate the genetic constitutions and genotypic values of the selection candidates. The latter method is based on the information of trait-QTL allele composition and, therefore, might be more accurate and intuitionistic than the former. Moreover, whereas GS appears to perform better for traits with infinitesimal QTL than those with a few major effect QTL, it is possible that a model with fewer markers (such as the QTL-allele matrix) will be a better choice for trait improvement than GS.

Tracing of the favorable alleles by pedigree indicated each ancestor or parent had different favorable alleles that might have transmitted to progenies. However, only some progenies might have inherited the best alleles suggesting some might have been lost in process of transition. This implies that at least some of the favorable alleles were descended from ancestor sources and were transmitted down through several breeding cycles, which was up to six or seven cycles in this study. The pedigree analysis further showed that the presence of a favorable allele in a genotype did not guarantee high tolerance to aluminum; however, its presence is likely to be associated with higher tolerance than for a genotype without the allele. Further work should be done to establish that desired alleles are transmitted through a definite pedigree analysis system that would correctly identify and mark loci for use in MAS. In our study, the pedigree families had different numbers of favorable alleles, suggesting sources other than tolerant parents transmitted favorable alleles and indicates substantial Al-tolerance potential in recombination and accumulation of favorable alleles from several sources in future breeding.

In summary, AM combined with linkage mapping provides a powerful tool for uncovering potential candidate QTL-allele matrix for marker-assisted breeding.

Acknowledgments The National Key Basic Research Program (2009CB1184, 2010CB1259, 2011CB1093), the National Hightech R & D Program (2011AA10A105, 2012AA101106), the Natural Science Foundation of China (31071442) and the MOE 111 Project (B08025) supported this work.

References

Agrama HA, Yan WG (2009) Association mapping of straighthead disorder induced by arsenic in *Oryza sativa*. Plant Breed 128:551–558

Bianchi-Hall CM, Carter TE Jr, Bailey MA, Mian MAR, Rufty TW, Ashley DA, Boerma HR, Arellano C, Hussey RS, Parrott WA

(2000) Aluminum tolerance associated with quantitative trait loci derived from soybean PI 416937 in hydroponics. Crop Sci 40:538–545

Breseghello F, Sorrells ME (2006a) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Breseghello F, Sorrells ME (2006b) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Campbell KAG, Carter TE Jr (1990) Aluminum tolerance in soybean: I. Genotypic correlation and repeatability of solution culture and greenhouse screening methods. Crop Sci 30:1049–1054

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19. doi:10.1186/1471-2156-3-19

Cui Z, Carter TE Jr, Burton JW (2000) Genetic diversity patterns in Chinese soybean cultivars based on coefficient of parentage. Crop Sci 40:1780–1793

Delhazie E, Ryan PR (1995) Aluminum toxicity and tolerance in plants. Plant Physiol 107:315–321

Dong YS, Zhao L, Liu MB, Wang ZW, Jin ZQ, Sun H (2004) The genetic diversity of cultivated soybean grown in China. Theor Appl Genet 108:931–936

Doyle JJ, Doyle JI (1990) Isolation of plant DNA from fresh tissue. Focus 12:147–151

Fehr WR (1987) Principles of cultivar development, vol 1: Theory and technique. McGraw Hill Inc, New York

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

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Foy CD, Fleming AL, Burns GR, Armiger WH (1967) Characterisation of differential aluminium tolerance among varieties of wheat and barley. Soil Sci Soc Am Proc 31:513–521

Foy CD, Duke JA, Devine TE (1992) Tolerance of soybean germplasm to an acid Tatum subsoil. J Plant Nutr 15:527–547

Gai JY (2006) Segregation analysis on genetic system of quantitative traits in plants. Front Biol China 1:85–92

Gai JY, Chen L, Zhang YH, Zhao TJ, Xing GN, Xing H (2012) Genome-wide genetic dissection of germplasm resources, and implications for breeding by design in soybean. Breed Sci 61:495–510

Habier D, Fernando RL, Deckkers JCM (2009) Genomic selection using low-density marker panels. Genetics 182:343–353

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Hill WG (2010) Understanding and using quantitative genetic variation. Philos Trans R Soc B: Bio Sci 365:73–85

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

Hisano H, Shusei S, Sachiko I, Shigemi S, Tsuyuko W, Ai M, Tsunakazu F, Manabu Y, Shinobu N, Yasukazu N, Satoshi W, Kyuya H, Satoshi T (2008) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271–281

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161

Institute SAS (2004) SAS Institute. Inc. SAS user's guide. SAS Institute, Inc., Cary

Jena KK, Mackhill DJ (2008) Molecular markers and their use in marker-assisted selection in rice. Crop Sci 48:1266–1276

Kochian LV (1995) Cellular mechanisms of aluminum toxicity and resistance in plants. Annu Rev Plant Physiol Plant Mol Biol 46:237–260

Kochian LV, Hoekenga OA, Piñeros MA (2004) How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. Annu Rev Plant Biol 55(1):459–493

Korir PC, Zhao TJ, Gai JY (2010) A study on indicators and evaluation stages of aluminum tolerance in soybean. Front Agric China 4:280–286

Korir PC, Qi B, Wang Y, Zhao T, Yu D, Chen S, Gai J (2011) A study on relative importance of additive, epistasis and unmapped QTL for aluminium tolerance at seedling stage in soybean. Plant Breeding 130:551–562

Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA (2010) Association and linkage analysis of aluminum tolerance genes in maize. PLoS ONE 5(4):e9958. doi:10.1371/journal.pone.0009958

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Liu Y, Gai JY (2007) Identification of tolerance to aluminum toxin and inheritance of related root traits in soybeans (*Glycine max* (L.) Merr.). Front Agric China 1(2):119–128

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Ma JF, Furukawa J (2003) Recent progress in the research of external Al detoxification in higher plants: a minireview. J Inorg Biochem 97:46–51

Maccaferri M, Sanguineti MC, Mantovani P, Demontis A, Massi A, Ammar K, Kolmer JA, Czembor JH, Ezrati S, Tuberosa R (2010) Association mapping of leaf rust response in durum wheat. Mol Breed 26:189–228

Malysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordum vulgrae* L.). BMC Genet 7:6. doi:10.1186/1471-2156-7-6

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90

Peleman JD, van der Voort JR (2003) Breeding by design. Trend. Plant Sci 8:330–334

Piñeros MA, Magalhaes JV, Alves VM, Kochian LV (2002) The physiology and biophysics of an aluminum tolerance mechanism based on root citrate exudation in maize. Plant Physiol 129:1194–1206

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Qi B, Korir P, Zhao T, Yu D, Chen S, Gai J (2008) Mapping quantitative trait loci associated with aluminum toxin tolerance in NJRIKY recombinant inbred line population of soybean (*Glycine max*). J Integrat Plant Biol 50:1089–1095

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, MäkiTanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159:699–713

Ryan PR, Ditomaso JM, Kochian LV (1993) Aluminum toxicity in root —an investigation of spatial sensitivity and the role of the root cap. J Exp Bot 44:437–446

Semon M, Nielsen R, Jones MP, McCouch SR (2005) The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of linkage disequilibrium caused by

admixture with *O. sativa* and ecological adaptation. Genetics 169:1639–1647

Sharma AD, Sharma H, Lightfoot DA (2011) The genetic control of tolerance to aluminum toxicity in the 'Essex' by 'Forrest' recombinant inbred line population. Theor Appl Genet 122:687–694

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Spehar CR (1995) Diallel analysis for mineral element absorption in tropical adapted soybeans [*Glycine max* (L.)] Merrill. Theor Appl Genet 90:707–711

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. Proc Nat Acad Sci USA 100:9440–9445

Van-Ooijen JW, Voorrips RE (2002) JOINMAP 3.0, Software for the Calculation of Genetic Linkage Maps. Plant Research International: Wageningen

Vargas-Duque J, Pandey S, Granados G, Ceballos H, Knapp E (1994) Inheritance of tolerance to soil acidity in tropical maize. Crop Sci 34:50–54

Villagarcia MR, Carter TE, Rufty TW, Niewoehner AS, Jennette MW, Arrellano C (2001) Genotypic rankings for aluminum tolerance of soybean roots grown in hydroponics and sand culture. Crop Sci 41:1499–1507

von Uexküll HR, Mutert E (1995) Global extent, development and economic impact of acid soils. Plant Soil 171:1–15

Wang YF (2009) Genomic characterization of simple sequence repeats and establishment, integration and application of high density genetic linkage map in soybean. Ph. D. Dissertation. Nanjing Agricultural University, Nanjing, China, pp 60–75

Wang S, Basten CJ, Zeng ZB (2005) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh

Wang J, McClean PE, Lee R, Goos RJ, Helms T (2008) Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. Theor Appl Genet 116:777–787

Wilkening S, Chen B, Hemminki K, Forst A (2006) STR markers for kinship analysis. Hum Biol 78:1–8

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Yang J, Zhu J, Williams W (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23:1527–1536

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139