Genetic-based dissection of arsenic accumulation in maize using a genome-wide association analysis method

Zhan Zhao1,1, Huaiheng Zhang1,1, Zhongjun Fu2, Hao Chen1, Yanan Lin1, Pengshuai Yan1, Weihua Li1, Huiling Xie1, Zhiyong Guo1, Xuehai Zhang1,• and Jihua Tang1,3,•

1Key Laboratory of Wheat and Maize Crops Science, Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou, China
2Maize Research Institute, Chongqing Academy of Agricultural Sciences, Chongqing, China
3Hubei Collaborative Innovation Center for Grain Industry, Yangtze University, Jingzhou, China

Summary
Understanding the mechanism of arsenic (As) accumulation in plants is important in reducing As's toxicity to plants and its potential risks to human health. Here, we performed a genome-wide association study to dissect the genetic basis of the As contents of different maize tissues in Xixian, which was irrigated with As-rich surface water, and Changge using an association population consisting of 230 representative maize inbred lines. Phenotypic data revealed a wide distribution and high repeatability for the As contents in maize tissues. The As concentrations in maize tissues followed the same trend in the two locations: kernels < axes < stems < bracts < leaves. In total, 15, 16 and 15 non-redundant quantitative trait loci (QTLs) associated with As concentrations were identified (P ≤ 2.04 × 10⁻⁶) in five tissues from Xixian, Changge, and the combination of the locations, respectively, explaining 9.70%–24.65% of the phenotypic variation for each QTL, on average. Additionally, four QTLs [involving 15 single nucleotide polymorphisms (SNPs)] were detected in the single and the combined locations, indicating that these loci/SNPs might be stable across different environments. The candidate genes associated with these four loci were predicted. In addition, four non-redundant QTLs (6 SNPs), including a QTL that was detected in multiple locations according to the genome-wide association study, were found to co-localize with four previously reported QTL intervals. These results are valuable to understand the genetic architecture of As mechanism in maize and facilitate the genetic improvement of varieties without As toxicity.

Keywords: arsenic accumulation, maize tissues, genetic loci, genome-wide association analysis.

Introduction
Soil heavy metal and metalloid pollution have become a worldwide environmental problem owing to human activities, such as metal processing, mining, sewage, irrigation and applications of herbicides and fertilizers (Dai et al., 2016). Arsenic (As), which is a toxic metalloid, has been classified as a Group I carcinogen (Smith et al., 2002; Wu et al., 2015). Wilson et al. (2010) reported that the background concentration of As in soil is less than 10 mg/kg, while in some mining-polluted soils, the As content is as high as 17 400 mg/kg. Because micronutrients in the soil are absorbed and accumulated by plants and enter the human body through the food chain (Dai et al., 2016), high As levels could cause serious human health risks. Sharma et al. (2016) reported that the consumption of As-contaminated wheat grains by adults and children results in increased risks of developing cancer and noncancerous health disorders. And the consumption of rice grown in soils contaminated by As and other heavy metals near mine areas can also seriously affect heavy metal levels in human blood (Cai et al., 2015). Additionally, a high soil As level has many negative impacts on plant growth, causing damage to the plant’s lateral roots and inhibiting water transport (Rahman et al., 2015; Zanella et al., 2016), thereby reducing crop quality and yield (Dai et al., 2016). It also negatively affects important plant metabolism-associated processes, such as photosynthesis, transpiration and respiration, chlorophyll biosynthesis and nucleic acid synthesis, which results in hampered plant growth (Armendariz et al., 2016; Mishra et al., 2016; Rahman et al., 2015).

Considering the seriousness of the As-associated damage to plant growth and the potential risks to human health, it is necessary to investigate the genetic mechanisms of As accumulation and tolerance in plants to reduce the toxicity. Quantitative trait locus (QTL) associated with As accumulation and tolerance in rice has been detected using different populations in previous study (Dasgupta et al., 2004; Zhang et al., 2008). Additionally, various genes involved in As accumulation in different species have been reported, such as, in tobacco, the overexpression of ‘A. thaliana’s phytochelatin synthase 1 gene increases root As accumulation (Verma et al., 2016). The overexpression of the two As-responsive rice glutaredoxin genes can reduce intracellular As accumulation and increase tolerance in A. thaliana, and significantly reduces arsenite accumulation by maintaining a glutathione pool and modulating aquaporins in yeast (Duan et al., 2016; Yang et al., 2016). In rice, the gene encoding a CRT-like transporter is important for glutathione homeostasis and As tolerance (Yang et al., 2016). Yu et al. (2016) reported A. thaliana’s inositol transporter genes could increase As accumulation in Saccharomyces cerevisiae and regulated As accumulation in A. thaliana seeds.

The genome-wide association analysis (GWAS), as a powerful tool, has been widely used to detect genetic variants related to complex quantitative traits in different plants (Atwell et al., 2010; Huang et al., 2010), such as the salt tolerance during germination...
in autotetraploid alfalfa (Norton et al., 2014), drought-related metabolic changes in maize (Zhang et al., 2016), drought tolerance in maize seedlings (Wang et al., 2016) and heavy metal tolerance in rice (Norton et al., 2014). In a recent report, six candidate genes with pleiotropic effects on stalk cell wall components in maize have been identified by main of GWAS method (Li et al., 2016a). And in A. thaliana, a gene associated with leaf arsenic accumulation was identified through GWAS (HAC1, Chao et al., 2014) and linkage mapping (ATQ1, Sanchez-Bermejo et al., 2014), respectively, the gene encoded a protein that as an arsenate reductase playing a critical role in arsenate resistance.

Maize, as an important crop over the world, is also a model plant. In the genetic dissection of As concentration in previous study, Liu et al. (2012) investigated trend in As accumulation and distribution in the different tissues of the 122 elite inbred maize lines. Ding et al. (2011) and Fu et al. (2016) used linkage mapping to unravel the genetic basis of As accumulation in two maize recombinant inbred line populations and found that As accumulation in different tissues in maize has a different molecular mechanism. However, the genetic basis of As concentration in maize is unclear up to now. Here, a GWAS with high-density single nucleotide polymorphisms (SNPs) was used to identify the natural allelic variations that contribute to As accumulation in maize tissues, such as kernels, axes, stems, bracts and leaves. The main purposes were as follows: (i) to explore the rules of As accumulation and distribution in different maize tissues of an association population; and (ii) to investigate the significant SNPs/loci and potential candidate genes associated with As variation in the tissues of maize under different soil As concentration. This research will further our understanding of the mechanisms of As accumulation and aid in the development of strategies to ensure food safety.

Results

Performance of the measured traits

In the association population, the As contents in five maize tissues were higher from the Xixian (XX) location compared with the corresponding tissues from the Change (CG) location. This might be attributed to a higher content of As in soil at XX than that at CG, and it suggested that soil As is an important determinant affecting the As content in maize tissues. The average As contents in kernels, axes, stems, bracts and leaves were 17.07, 51.22, 50.44, 56.61 and 186.65 μg/kg, at XX location, and 16.69, 43.48, 44.48, 53.08 and 157.15 μg/kg, respectively, at CG location. Combining data from the two locations, the average As contents in kernels, axes, stems, bracts and leaves were 16.88, 47.35, 47.46, 54.84 and 171.90 μg/kg, respectively (Table 1). In the association population, the As concentrations varied widely in all five tissues independent of the location and in the combined data (Figure 1, Figure S1). For example, the As contents of the combined environment in kernels, axes, stems, bracts and leaves ranged from 5.92 to 34.10 μg/kg, 27.27 to 88.86 μg/kg, 26.25 to 110.72 μg/kg, 31.06 to 91.45 μg/kg and 115.05 to 247.67 μg/kg, respectively (Table 1). The frequency of the As content’s phenotypic value for data combined from the two locations exhibited an approximately normal distribution (Figure S1). For the As concentrations in different maize tissues, the same trend, kernels < axes < stems < bracts < leaves (Figure 1), was consistent with a previous report (Fu et al., 2016). Moreover, the Pearson correlation coefficient showed that close relationships were observed among the As concentrations between same tissues across locations (Table S1). According to the variance analysis (Table 2), the As contents in the five tissues in plants of the association population showed significant variations in environment, genotype and the genotype x environment interaction. The repeatability levels for the As contents in kernels, axes, stems, bracts and leaves were 86.73%, 86.71%, 86.46%, 86.79% and 82.16%, respectively (Table 2).

GWAS

Several genotyping platforms (Illumina MaizeSNP50 BeadChip, RNA sequencing, genotyping by sequencing and Affymetrix Axiom Maize 600K array) that integrated effective imputation methods and generated 1.25 M SNPs with a minor allele frequency greater than 0.05 were used in this study (Liu et al., 2017). Three models, Q (only accounts for population structure), K (only accounts for relative kinship) and Q + K (accounts for population structure and relative kinship), were employed to perform GWAS for the As concentrations in five maize tissues at two separate locations and a combination of the two locations. To test the optimal GWAS model, quantile–quantile plots (QQ plots) for each trait under the three models were plotted (Figure S2), and the QQ plots indicated that the Q + K model was more reliable than the K model. However, the K and Q + K models were too strict in reducing the type I errors (false positives), which results in more type II errors (false negatives). The Q model’s performance significantly reduced the type I errors (false positive) compared to the other two models. Thus, in the following analysis, the GWAS results from Q model will be further elucidated.

According to the linkage disequilibrium (LD) decay of this association population (Liu et al., 2017), the 30-kb regions flanking the left and right sides of each significant SNP could

| Location | Trait | Mean ± SD* (μg/kg) | Range (μg/kg) | Kurt* | Skew* |
|----------|-------|--------------------|---------------|-------|-------|
| Xixian   | Kernel| 17.07 ± 3.66       | 7.52–31.02    | 1.04  | 0.31  |
|          | Axis  | 51.22 ± 14.13      | 24.96–92.80   | –0.09 | 0.4   |
|          | Stem  | 50.44 ± 9.53       | 33.86–88.38   | 2.63  | 1.44  |
|          | Bract | 56.61 ± 8.40       | 35.80–81.14   | –0.05 | 0.29  |
|          | Leaf  | 186.65 ± 17.98     | 139.60–234.57 | 0.09  | 0.32  |
| Changege | Kernel| 16.69 ± 7.32       | 2.59–39.46    | 0.48  | 0.39  |
|          | Axis  | 43.48 ± 7.58       | 27.78–83.30   | 3.02  | 1.02  |
|          | Stem  | 44.48 ± 19.78      | 16.97–136.36  | 3.28  | 1.57  |
|          | Bract | 53.08 ± 16.47      | 21.75–107.12  | 0.68  | 0.67  |
|          | Leaf  | 157.15 ± 42.77     | 63.52–275.18  | –0.18 | 0.47  |
| BLUP     | Kernel| 16.88 ± 4.98       | 5.92–34.10    | 0.74  | 0.38  |
|          | Axis  | 47.35 ± 10.45      | 27.27–88.86   | 0.61  | 0.57  |
|          | Stem  | 47.46 ± 13.98      | 26.25–110.72  | 3.31  | 1.59  |
|          | Bract | 54.84 ± 11.56      | 31.06–91.45   | 0.44  | 0.55  |
|          | Leaf  | 171.90 ± 27.74     | 115.05–247.67 | –0.01 | 0.54  |

*SD, Standard Deviation.
Kurt, kurtosis, which is a measure of the ‘tailedness’ of the probability distribution of a real-valued random variable.
Skew, skewness, which is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.
be defined as QTLs. Under the Q model, 79 SNP-trait associations were detected at the significance level of \( P \leq 2.04E-06 \) in five tissues (Table S2), which involving 45 SNPs. According to the definition of the QTLs, 45 QTLs were detected in total. Furthermore, these QTLs can be categorized into 28 non-redundant QTLs (all of the QTLs with overlapping flanking intervals were categorized as non-redundant QTLs). Briefly, in XX, 15 non-redundant QTLs (27 significant SNPs) were identified in five tissues, and each locus could explain phenotypic variation \( (R^2) \) ranging from 9.20% to 18.74% and a mean of 11.59%. In CG, 16 non-redundant QTLs (25 significant SNPs) were detected in five tissues, and each locus could explain phenotypic variation ranging from 9.70% to 24.65%, with a mean of 12.67%. Moreover, 15 non-redundant QTLs (27 significant SNPs) were identified in the combination of the two locations, with an \( R^2 \) ranging from 9.70 to 19.70%, and a mean of 11.87% (Figures 2 and 3, Table 3, Table S2). Furthermore, the same 27 significant SNPs from the GWAS of BLUP data could still be identified in the GWAS of LMS date, and only the \( P \) value and \( R^2 \) of each SNP has slight difference (Table S3). The detailed information for the GWAS results, including \( P \) value and \( R^2 \) of each non-redundant QTL, physical positions of peak SNPs and the most likely candidate genes and their annotations, is listed in Table S2. Manhattan plots for As concentrations in all five tissues at XX, CG and the combination are shown in Figure S3. All potential candidate genes and their functional annotations within 60 kb (30-kb upstream and downstream of the peak SNPs) of each non-redundant QTL identified from GWAS are provided in Table S2.

In Table 4, the co-localization of non-redundant QTLs identified across different locations is summarized. Overall, 38 SNP-trait associations, involving 15 significant SNPs in four non-redundant QTLs and corresponding to three tissues (axis, bract and stem), were detected simultaneously in XX, CG and the combination of the locations (Figure 2). Thus, these QTLs might be stable across different environments. Furthermore, a QTL hot spot existed at the end of chromosome 2 (230.49–230.55 Mb) (Figure 2), indicating that this hot spot could be the key factor for regulating As accumulation in maize bracts. Detailed analyses of the candidate genes underlying these four non-redundant QTLs will almost certainly provide useful information concerning As accumulation.

### Candidate genes revealed by multiple locations

Based on the GWAS, 28 non-redundant QTLs were identified in this study (Table S2). This was consistent with the quantitative nature of As content in maize being controlled by a large number of QTLs, as well as with previously studied agronomic traits (Ding et al., 2011; Fu et al., 2016; Xiao et al., 2016). Interestingly, four non-redundant QTLs that were detected simultaneously in both single-environment analysis and the joint analysis were used to identify their candidate genes based on LD decay or supporting flanking intervals. A total of nine candidate genes were found (Table S2), including five candidate genes (GRMZM2G125487, GRMZM2G125527, GRMZM2G125495, GRMZM2G426556 and GRMZM2G125507) located in the 9th non-redundant QTL (from 203.57 Mb to 203.63 Mb) on chromosome 2 that were significantly associated with the As content in the bracts. GRMZM2G125495 encodes a protein that participates in extracellular-glutamate-gated ion channel activity and is a likely candidate for the gene that transports As ions in plants. Only one gene, GRMZM2G820781, located in the 12th non-redundant QTL (from 7.65 Mb to 7.71 Mb) on chromosome 3 was significantly associated with the As content in the axes. The gene GRMZM2G820781 encodes an uncharacterized protein, which was only expressed in pericarp aleurone. Additionally, no genes in the 19th non-redundant QTL (from 84.09 Mb to 84.15 Mb) on chromosome 7 significantly associated with the As content in the stem, while GRMZM2G057317, which is nearest to the peak SNP, encodes an unknown protein that was only expressed in the embryo. Moreover, three candidate genes (GRMZM2G052457, GRMZM2G158872 and GRMZM2G452669) located in the 10th non-redundant QTL on chromosome 2 (230.49–230.55 Mb) were significantly associated with the As content in the bracts. These three genes also formed the only QTL hot spot identified in

---

**Table 2** Variance analysis of the five measured tissues across two locations in the association population

| Tissue | Variance | MS | \( F \) | \( P \) value\(^{**} \) | \( \omega^2 \) (%)\(^{**} \) |
|--------|---------|----|------|--------------------|--------------------|
| Kernel E | 51.11 | 7.86** | 5.15E-03 | 86.73 |
| G | 198.97 | 30.6*** | 2.00E-16 |
| G*E | 26.42 | 4.06*** | 2.00E-16 |
| Axis E | 20708.39 | 350.36*** | 1.72E-66 | 86.71 |
| G | 874.92 | 14.8*** | 1.65E-198 |
| G*E | 116.24 | 1.97*** | 2.37E-12 |
| Stem E | 12283.58 | 118.67*** | 4.46E-26 | 86.46 |
| G | 1575.83 | 25.22*** | 8.30E-102 |
| G*E | 213.33 | 2.06*** | 5.88E-14 |
| Bract E | 4286.86 | 83.83*** | 3.41E-19 | 86.79 |
| G | 1068.87 | 20.9*** | 1.97E-251 |
| G*E | 141.18 | 2.76*** | 8.87E-27 |
| Leaf E | 300372.56 | 516.69*** | 3.95E-91 | 82.16 |
| G | 6871.67 | 11.82*** | 1.27E-166 |
| G*E | 1225.78 | 2.11*** | 8.80E-15 |

\( ** \)significant at \( \alpha = 0.001 \).

\( * \)significant at \( \alpha = 0.01 \); E: environment, G: genotype, G × E: genotype-by-environment interaction.

\( \omega^2 \), repeatability.
this study. GRMZM2G052457 encodes an adenosylmethionine-8-amino-7-oxononanoate transaminases, GRMZM2G158872 has an unknown function, and GRMZM2G452669 encodes an alpha-galactosidase precursor. The most likely candidate genes are listed in Table 4.

**Discussion**

When using a GWAS, the power, that is the probability of detecting the causal variant, should be the first consideration. Additionally, traits are sensitive to different statistical models, which vary in their abilities to control type I or type II errors. Thus, the models can be seen as trait-dependent. For example, Yang et al. (2011) reported that the K and Q + K models performed well for flowing time, ear height and ear diameter; however, the Q + K models performed slightly better than the K model for all three traits (Yang et al., 2011). In the present study, the Q model, which was selected to illustrate the GWAS results, had a greater ability to correct false-positive associations than the K and Q + K models (Figure S2), although it could not completely control the population structure.

GWAS-based mapping has been successfully used to identify a new arsenate reductase enzyme critical for limiting As accumulation in *A. thaliana* (Chao et al., 2014). In this study, only 15 SNPs or 4 QTLs, which stably existed across different environments and significantly associated with As accumulation in different maize tissues, were identified through a GWAS.

---

**Table 3** Summary of significant loci–trait associations identified by the genome-wide association study

| CG* | XX | BLUP |
|-----|----|------|
| Number of traits† | 5/5 | 5/5 | 5/5 |
| Number of loci‡ | 16 | 15 | 15 |
| Average loci per trait§ | 3.20 ± 1.72 | 3.00 ± 2.61 | 3.00 ± 2.61 |

*CG and XX represent the two locations Changge and Xixian, respectively; BLUP, Best linear unbiased prediction.
†Number of traits having significantly associated loci/number of total detected traits.
‡Number of significant loci detected in each location on the association panel [P ≤ 2.04 × 10^{-6}, Q model].
§Average number of significant loci (or QTLs) detected per trait ± SD.
| ID | Trait | Location | Peak SNP Chr. Position (bp) | QTL interval (Mb) | P value | Candidate gene ** | Annotation |
|---|---|---|---|---|---|---|---|
| 2 | 12 | Axis | BLUP | chr3.S_7683509 3 7 683 509 | 7.65–7.71 | 2.25E-07 | 11.11 | GRMZM5G820781 | Unknown |
| 8 | 12 | Axis | CG | chr3.S_7675260 3 7 675 260 | 7.65–7.71 | 1.52E-06 | 9.81 | GRMZM5G820781 | Unknown |
| 14 | 12 | Axis | XX | chr3.S_7675260 3 7 675 260 | 7.65–7.71 | 1.77E-06 | 9.76 | GRMZM5G820781 | Unknown |
| 3 | 12 | Axis | BLUP | chr3.S_7675260 3 7 675 260 | 7.65–7.71 | 2.25E-07 | 11.11 | GRMZM5G820781 | Unknown |
| 29 | 9 | Bract | CG | chr2.S_203601399 2 203 601 399 | 203.57–203.63 | 3.56E-07 | 10.39 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 30 | 9 | Bract | BLUP | chr2.S_203601456 2 203 601 456 | 203.57–203.63 | 1.13E-07 | 11.20 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 37 | 9 | Bract | XX | chr2.S_203601399 2 203 601 399 | 203.57–203.63 | 2.64E-07 | 10.60 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 18 | 9 | Bract | BLUP | chr2.S_203601456 2 203 601 456 | 203.57–203.63 | 1.13E-07 | 11.20 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 38 | 9 | Bract | CG | chr2.S_203601399 2 203 601 399 | 203.57–203.63 | 3.56E-07 | 10.39 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 19 | 9 | Bract | BLUP | chr2.S_203602088 2 203 602 088 | 203.57–203.63 | 9.09E-07 | 9.64 | GRMZM2G125495 | Extracellular-glutamate-gated ionic channel activity |
| 40 | 10 | Bract | XX | chr2.S_230520093 2 230 520 093 | 230.49–230.55 | 1.65E-06 | 9.20 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 41 | 10 | Bract | XX | chr2.S_230520234 2 230 520 234 | 230.49–230.55 | 9.56E-07 | 9.76 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 22 | 10 | Bract | BLUP | chr2.S_230520341 2 230 520 341 | 230.49–230.55 | 3.38E-07 | 10.50 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 23 | 10 | Bract | BLUP | chr2.S_230521739 2 230 521 739 | 230.49–230.55 | 1.08E-06 | 9.68 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 24 | 10 | Bract | BLUP | chr2.S_230521763 2 230 521 763 | 230.49–230.55 | 2.48E-07 | 10.77 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 25 | 10 | Bract | BLUP | chr2.S_230522102 2 230 522 102 | 230.49–230.55 | 2.69E-07 | 10.71 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 32 | 10 | Bract | CG | chr2.S_230521763 2 230 521 763 | 230.49–230.55 | 6.78E-07 | 10.06 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 43 | 10 | Bract | XX | chr2.S_230522102 2 230 522 102 | 230.49–230.55 | 5.87E-07 | 10.14 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 26 | 10 | Bract | BLUP | chr2.S_230523364 2 230 523 364 | 230.49–230.55 | 2.46E-07 | 10.27 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 27 | 10 | Bract | BLUP | chr2.S_230523621 2 230 523 621 | 230.49–230.55 | 1.43E-06 | 9.81 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 28 | 10 | Bract | BLUP | chr2.S_230523716 2 230 523 716 | 230.49–230.55 | 2.46E-07 | 10.27 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 29 | 10 | Bract | CG | chr2.S_230523364 2 230 523 364 | 230.49–230.55 | 8.55E-07 | 9.92 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 34 | 10 | Bract | CG | chr2.S_230523364 2 230 523 364 | 230.49–230.55 | 8.58E-07 | 9.92 | GRMZM2G052457 | Adenosylmethionine-8-amino-7-oxononanoate transaminases |
| 8 | 12 | Axis | BLUP | chr3.S_7675260 3 7 675 260 | 7.65–7.71 | 7.28E-07 | 10.42 | GRMZM5G820781 | Unknown |
| 14 | 12 | Axis | XX | chr3.S_7675260 3 7 675 260 | 7.65–7.71 | 1.77E-06 | 9.76 | GRMZM5G820781 | Unknown |
| 3 | 12 | Axis | BLUP | chr3.S_7683509 3 7 683 509 | 7.65–7.71 | 2.25E-07 | 11.11 | GRMZM5G820781 | Unknown |

** Table 4: Candidate genes revealed by multiple locations **

*Published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.*

© 2017 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 16, 1085–1093
In this study, six SNPs associated with four non-redundant QTLs that significantly associated with As accumulation, as detected by GWAS, were found to co-localize (here, any two QTLs with physical positions less than 5 Mb apart were declared as co-localized) with a previously detected QTL from a linkage population (Fu et al., 2016; Table S2). The loci, located from 25.71 to 25.77 Mb on chromosome 1, co-localized with the previously reported QTL CASA1/CAS51, and only one candidate gene, GRMZM2G130987, was found. The gene encodes a protein that has P-P-bond-hydrolysis-driven protein transmembrane transporter activity, which may participate in As ion transport. The other three significant SNPs (chr2.S_203601399, chr2.S_203601456 and chr2.S_203602088) on chromosome 2 were found within the QTL BASA2/XASA2. This QTL was also detected in multiple locations in our study. Five candidate genes (GRMZM2G125487, GRMZM2G125527, GRMZM2G125495, GRMZM2G426556 and GRMZM2G125507) are located in this QTL. GRMZM2G125495 encodes a protein that has extracellular-glutamate-gated ion channel activity and is a likely candidate gene. This provides a solid reason to further study the gene's function in As accumulation. Another region, between 90.45 Mb and 90.51 Mb on chromosome 10, contained one SNP (chr10.S_90481204), which co-localized with the QTL XASA10 that includes four candidate genes (GRMZM2G166665, GRMZM2G166616, GRMZM2G166639 and GRMZM2G166694). On chromosome 10, the region from 143.13 to 143.19 Mb, which contains three genes (GRMZM2G326066, GRMZM2G021822 and GRMZM2G021885), was co-localized with CASL10. The joint use of linkage mapping and GWAS was more efficient in identifying candidate genes and detecting loci governing As accumulation in maize tissues. These co-localized QTLs from different genetic populations and the common QTLs were all detected in the same population at multiple locations, which will provide useful reference information for studies on the functional verification of As accumulation and for breeding varieties with low As concentrations.

The highest As toxicity in plants was primarily attributed to two aspects. First, a high As level can reduce phosphate uptake from the agricultural soil because of their structural similarities. Second, the presence of As (V) in plants could lead to oxidative stress by altering the synthesis of adenosine triphosphate and the phosphate group of DNA (Rosas-Castor et al., 2014). In a previous study, the induction of oxidative stress was reported as the main process underlying As toxicity in plants and the up-regulation of a set of oxidative stress-related proteins could induce the plant's responses to acute inorganic As toxicity. This regulation might be a consequence of the production of reactive oxygen species derived from As (Requejo and Tena, 2005). Furthermore, He et al. (2006) first proved that arsenic could induce NAD(P) H-quinone oxidoreductase I by activating the Nrf2/Keap1 signalling pathway and implying that As accumulation in different maize tissues may be controlled by different genetic mechanisms. However, the results will be helpful in illustrating the genetic mechanisms and cloning the genes governing As accumulation in maize. In A. thaliana, GWAS and QTL mapping are used to complement each other, overcoming their individual limitations (Tiwari et al., 2016). In maize, the combination of QTL mapping and GWAS has been used to identify candidate genes for plant ear height and ear traits (Li et al., 2016b; Xiao et al., 2016). The combination of the two methods has also been applied to detect loci governing agronomic traits in other species, like soya bean (Zhao et al., 2015) and rice (Lou et al., 2015).

| ID* | Locus† | Trait | Location | Peak SNP Chr. | Position (bp) | P-value‡ | R²(%) | Candidate gene** | Annotation†† |
|-----|--------|-------|----------|---------------|---------------|----------|-------|----------------|-------------|
| 9   | 12     | Ass   | CG       | chr3.S_7683509| 7 683 509     | 7.65-7.71 | 1.39E-07| Unknown        |             |
| 15  | 12     | Ass   | XX       | chr3.S_7683509| 7 683 509     | 7.65-7.71 | 1.38E-06| GRMZM2G057317 |             |
| 65  | 19     | Stem  | BLUP     | chr7.S_84124201| 784 124 201  | 84.09-84.15 | 4.58E-08| Unknown        |             |
| 70  | 19     | Stem  | CG       | chr7.S_84124201| 784 124 201  | 84.09-84.15 | 3.23E-07| GRMZM2G057317 |             |
| 76  | 19     | Stem  | XX       | chr7.S_84124201| 784 124 201  | 84.09-84.15 | 1.47E-07| GRMZM2G057317 |             |

*The ID of the SNP/trait associations and non-redundant QTLs.
†The physical position in base pairs for the peak SNP according to version 5b.60 of the maize reference sequence.
‡30-kb upstream and downstream of the peak SNP.
P-values of peak SNPs estimated by the Q model.
kThe phenotypic variance explained by the corresponding locus.
**Plausible biological candidate genes in the physical intervals of each QTL.
††Annotation information according to InterProScan (http://www.ebi.ac.uk/interpro/).
As the concentration (μg/kg) was determined in the different maize tissues using atomic fluorescence spectrometry (AFS-3000, Beijing Haiguang Analytical Instrument Co., Beijing, China). Each sample was measured three times, the average data of As concentration of each tissue were calculated first, and then, the average value of As concentration in each tissue for one material in the three replications was used to detect the significant loci/SNP. A two-way analysis of variance was used to analyse the data using the IBM SPSS Statistics package, and repeatability was also calculated based on the method developed by Knapp (1986).

Repeatability (w²) for each trait across the two locations was computed as follows:

\[ w^2 = \frac{\sigma_G^2}{\sigma_G^2 + (\sigma_{GE}/n) + \sigma_{r}^2/(nr)} \]

In the above formula, \(\sigma_G^2\) represents the genotypic variance, \(\sigma_{GE}\) represents the genotype × environment variance, \(\sigma_r^2\) represents the error variance, \(n\) represents the number of locations, and \(r\) represents the number of replications. The estimates of \(\sigma_G^2\), \(\sigma_{GE}\) and \(\sigma_r^2\) were analysed by an analysis of variance using the lmer function in the lme4 package in the R environment (R Core Team, 2012; version 3.1.3; http://www.r-project.org/).

The best linear unbiased prediction (BLUP) was obtained by fitting the mixed linear model in the R package lme4 (R Core Team, 2012) for the estimation of the breeding values of each line across two locations, and the formula is

\[ Y = (1|LINE) + (1|ENV) + (1|REP|sin%|LINE:ENV) + (1|LINE:ENV) \]

where \(Y\) is trait data, the parenthesis indicates the random effects, ‘|’ means groups, and ‘:’ means interactions. LINE indicates all testcrosses used; ENV indicates the environments, each of which is a combination of years and locations, and REP is the replications in one ENV. And the BLUP values were then combined to reduce the prediction bias caused by the unbalanced data. Finally, the BLUP data for the As concentrations of each tissue across two locations were also used for the GWAS. The Pearson correlation coefficients among the As contents of different tissues were implemented with SPSS software (v13.0). To test whether the major QTLs from BLUP values still be found, LSM (least square means) of the data from different locations of the same tissue were calculated in the R package lsmeans (R Core Team, 2012). And the LSM values for the As concentrations of each tissue across two locations were also used for the GWAS.

**GWAS**

According to the method described by Liu et al. (2017), 1.25 million (M) SNPs (1, 253, 814 SNPs) with a minor allele frequency ≥5% were obtained by combining genotypes from previous RNA sequencing (Fu et al., 2013) and the Illumina MaizeSNP50 BeadChip (Ganal et al., 2011) with newly identified SNPs from the Affymetrix Axiom Maize 600K array (Unterseer et al., 2014) and genotyping by sequencing technology (Elshire et al., 2011) was used for the GWAS (the SNP data are available at http://www.maizego.org/Resources.html).

A GWAS was conducted for five different tissues at XX, CG and the combination of the locations. To test the optimal GWAS model, three models, Q (only accounts for population structure), K (only accounts for relative kinship) and Q + K (accounts for population structure and relative kinship), were
implemented in the software TASSEL 3.0 (Bradbury et al., 2007) to test the statistical associations between genotype and phenotype.

Because many of the SNPs should be in linkage disequilibrium, the effective number of independent markers was also calculated using the GEC software tool in a previous study (Deng et al., 2017), and the suggested P-value of 2.04E-06 (1/dependent marker number) was used to control the genome-wide type I error rate. The P-value of each SNP was obtained from TASSEL 3.0 (Bradbury et al., 2007) and used to construct QQ and Manhattan plots for the As concentration in each tissue.

Analysis of candidate genes

Based on the reported maize B73's genome sequence (RefGen_v2), the filtered maize working gene list was downloaded from MaizeGDB (http://www.maizegdb.org) and used to identify possible candidate genes in each QTL. Candidate genes were annotated according to InterProScan (http://www.ebi.ac.uk/interpro/scan.html). In the previous study, the LD of the association population was estimated using the 1.25 M SNPs and the LD decay was 30 kb ($r^2 = 0.1$) (Liu et al., 2017). All of the potential candidate genes and their annotations within 60 kb (30-kb upstream and downstream of the peak SNP, which is the SNP with the lowest P-value) of the identified loci are listed in Table S2. For the loci without appropriate candidates, the gene nearest the peak SNP was assigned. The physical locations of the SNPs were based on B73 RefGen_v2.

Acknowledgements

We thank the research group of Jianbing Yan at Huazhong Agricultural University for materials, technology and SNP data support. This work was supported by the Major Projects in Henan Province (161100110500), Research Fund for the Doctoral Program of Higher Education of China (RFDP, 20134105110001) and the funds of the Central Plains Scholars in Henan Province. The authors declare no conflict of interest.

References

Armandez, A.L., Talano, M.A., Villasuso, A.L., Travaglia, C., Racagni, G.E., Reinoso, H. and Agostini, E. (2016) Arsenic stress induces changes in lipid signalling and evokes the stomata closure in soybean. *Plant Physiol. Biochem.* 103, 45-52.

Atwell, S., Huang, Y.S., Vilhjalmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D. et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, 465, 627-631.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633-2635.

Cai, L.M., Xu, Z.C., Qi, J.Y., Feng, Z.Z. and Xiang, T.S. (2015) Assessment of exposure to heavy metals and health risks among residents near Tonglushan mine in Hubei, China. *Chemosphere*, 127, 127-135.

Chao, D.Y., Chen, Y., Chen, J., Shi, S., Chen, Z., Wang, C., Danku, J.M. et al. (2014) Genome-wide association mapping identifies a new arsenite reductase enzyme critical for limiting arsenic accumulation in plants. *PLoS Biol.* 12, e1002009.

Dai, Y., Li, J., Liu, K., Zhao, X. and Cao, Y. (2016) Major controlling factors and prediction models for arsenic uptake from soil to wheat plants. *Ecotox Environ. Safe.*, 130, 256-262.

Dasgupta, T., Hossain, S.A., Mehang, A.A. and Price, A.H. (2004) An arsenite tolerance gene on chromosome 6 of rice. *New Phytol.* 163, 45-49.

Deng, M., Li, D., Luo, J., Xiao, Y., Liu, H., Pan, Q., Zhang, X. et al. (2017) The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol. J.* 15, 1250-1263 https://doi.org/10.1111/pbi.12712.

Ding, D., Li, W., Song, G., Qi, H., Liu, J. and Tang, J. (2011) Identification of QTLs for arsenic accumulation in maize (Zea mays L.) using a RIL population. *PLoS ONE*, 6, e25646.

Duan, G.L., Hu, Y., Schneider, S., McDermott, J., Chen, J., Sauer, N., Rosen, B.P. et al. (2016) Inositol transporters AtTNT2 and AtTNT4 regulate arsenic accumulation in Arabidopsis seeds. *Nat. Plants*, 2, 15202.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379.

Fu, J., Zheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, J., Zeng, J. et al. (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat. Commun.*, 4, 2832.

Fu, Z., Li, W., Xing, X., Xu, M., Liu, X., Li, X., H., Xue, Y. et al. (2016) Genetic analysis of arsenic accumulation in maize using QTL mapping. *Sci. Rep.*, 6, 21292.

Ganal, M.W., Dursztewitz, G., Polley, A., Berard, A., Buckler, E.S., Charcosset, A., Clarke, J.D. et al. (2011) A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, 6, e28334.

He, X., Chen, M.G., Lin, G.X. and Ma, Q. (2006) Arsenic induces NAD(P)H:quinone oxidoreductase I by disrupting the Nrf2 Keap1 Cul3 complex and recruiting Nrf2 to the antioxidant response element enhancer. *J. Biol. Chem.*, 281, 23620.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C. et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.*, 42, 961-967.

Huang, L., Wu, H. and Kuip, T.J.V.D. (2015) The health effects of exposure to arsenic-contaminated drinking water: a review by global geographical distribution. *Inter. J. Environ. Res. Pub. Health*, 25, 432.

Knapp, S.J. (1986) Confidence intervals for heritability for two-factor mating design single environment linear models. *TAG. Theor. Appl. Genet.* 72, 587-591.

Li, K., Wang, H., Hu, X., Liu, Z., Wu, Y. and Huang, C. (2016a) Genome-wide association study reveals the genetic basis of stalk cell wall components in maize. *PLoS ONE*, 11, e0158906.

Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., Ma, J. et al. (2016b) Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Front Plant Sci.*, 7, 833.

Li, Y., Wang, D., Liu, Y., Zheng, Q. and Sun, G. (2016c) A predictive risk model of groundwater arsenic contamination in China applied to the Hui River Basin, with a focus on the region’s cluster of elevated cancer mortalities. *Appl. Geochem.*, 77, 178-183.

Liu, Z., Li, W., Qi, H., Song, G., Ding, D., Fu, Z., Liu, J. et al. (2012) Arsenic accumulation and distribution in the tissues of inbred lines in maize (Zea mays L.). *Genet. Resour. Crop Ev.*, 59, 1705-1711.

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., Wang, X. et al. (2017) Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol Plant*, 10, 414-426.

Lou, Q., Chen, L., Mei, H., Wei, H., Feng, F., Wang, P., Xia, H. et al. (2015) Quantitative trait locus mapping of deep rooting by linkage and association analysis in rice. *J. Exp. Bot.*, 66, 4749-4757.

Mallick, S., Sinam, G. and Sinha, S. (2011) Study on arsenate tolerant and sensitive cultivars of Zea mays L.: differential detoxification mechanism and effect on nutrients status. *Ecotox Environ. Saf.*, 74, 1316-1324.

Mishra, S., Affeld, M., Sobolik, R., Andresen, E., Falkenberg, G. and Kupper, H. (2016) Analysis of sublethal arsenic toxicity to Ceratophyllum demersum: subcellular distribution of arsenic and inhibition of chlorophyll biosynthesis. *J. Exp. Bot.*, 67, 4639-4646.

Norton, G.J., Douglas, A., Lahner, B., Yukobava, E., Guerini-Rossi, M.L., Pinson, S.R., Tarpley, L. et al. (2014) Genome-wide association mapping of grain arsenic, copper, molybdenum and zinc in rice (Oryza sativa L.) grown at four international field sites. *PLoS ONE*, 9, e89685.

R Core Team (2012) R: A Language and Environment for Statistical Computing. Vienna, Austria: R foundation for statistical computing. URL http://www.R-project.org/.
Arsenic accumulation mechanisms in maize tissues

Wu, G.R., Hong, H.L. and Yan, C.L. (2015) Calcium mitigates Arsenic toxicity in rice seedlings by reducing Arsenic uptake and modulating the antioxidant defense and glyoxalase systems and stress markers. *Biomed. Res. Int.* 2015, 340812.

Requejo, R. and Tena, M. (2005) Proteome analysis of maize roots reveals that oxidative stress is a main contributing factor to plant arsenic toxicity. *Phytochemistry* 66, 1519–1528.

Rosas-Castor, J.M., Guzman-Mar, J.L., Hernandez-Ramirez, A., Garza-Gonzalez, M.T. and Hinojosa-Reyes, L. (2014) Arsenic accumulation in maize crop (2ea mays): a review. *Sci. Total Environ.* 488–489, 176–187.

Salt, D.E., Baxter, I. and Lahner, B. (2008) Ionomics and the study of the plant ionome. *Annu. Rev. Plant Biol.* 59, 709–733.

Sanchez-Bermejo, E., Castillio, G., del Llano, B., Navarro, C., Zardo-Fernandez, S., Martinez-Herrera, D.J., Leo-del Puerto, Y. et al. (2014) Natural variation in arsenic tolerance identifies an arsenic reductase in Arabidopsis thaliana. *Nat. Commun.* 5, 4617.

Sharma, S., Kaur, J., Nagpal, A.K. and Kaur, I. (2016) Quantitative assessment of possible human health risk associated with consumption of arsenic contaminated groundwater and wheat grains from Ropar Wetland and its environs. *Environ. Monit. Assess.* 188, 506.

Smith, A.H., Lopipero, P.A., Bates, M.N. and Steinmaus, C.M. (2002) Public health. Arsenic epidemiology and drinking water standards. *Science* 296, 2145–2146.

Tiwari, S., Sl, K., Kumar, V., Singh, B., Rao, A.R., Mithra Sv, A., Rai, V. et al. (2016) Mapping QTls for salt tolerance in rice (*Oryza sativa*) by bulked segregant analysis of recombinant inbred lines using 50K SNP chip. *PLoS ONE* 11, e0153610.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T. et al. (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 60K SNP expression array. *BMC Genom.* 15, 823.

Verma, P.K., Verma, S., Meher, A.K., Pande, V., Mallick, S., Bansiwal, A.K., Tripathi, R.D. et al. (2016) Overexpression of rice glutaredoxins (OsGrxs) significantly reduces arsenite accumulation by maintaining glutathione pool and modulating aquaporins in yeast. *Plant Physiol. Biochem.* 106, 208.

Wang, X., Wang, H., Liu, S., Xu, X., Pan, Q., Qiao, F., Raihan, M.S. et al. (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol.* 210, 1095–1106.

Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B.M., Li, L., Li, J. et al. (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526.

Yang, J., Gao, M.X., Hu, H., Ding, X.M., Lin, H.W., Wang, L., Xu, J.M. et al. (2016) OsCLT1, a CRT-like transporter 1, is required for glutathione homeostasis and arsenic tolerance in rice. *New Phytol.* 211, 658–670.

Yu, L.X., Liu, X., William, B. and Liu, X.P. (2016) Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa (*Medicago sativa L.*) using genotyping-by-sequencing. *Front. Plant Sci.* 7, 956.

Zanella, L., Fattorini, L., Brunetti, P., Roccioletti, E., Cornara, L., D’Angeli, S., Della Rovere, F. et al. (2016) Overexpression of AtPCS1 in tobacco increases arsenic and arsenic plus cadmium accumulation and detoxification. *Planta* 243, 605–622.

Zhang, J., Zhu, Y.G., Zeng, D.L., Cheng, W.D., Qian, Q. and Duan, G.L. (2008) Mapping quantitative trait loci associated with arsenic accumulation in rice (*Oryza sativa*). *New Phytol.* 177, 350–355.

Zhang, X., Warburton, M.L., Setter, T., Liu, H., Xue, Y., Yang, N., Yan, J. et al. (2016) Genome-wide association studies of drought-related metabolic changes in maize using an enlarged SNP panel. *Theor. Appl. Genet.* 129, 1449–1463.

Zhao, X., Han, Y., Li, Y., Liu, D., Sun, M., Zhao, Y., Lv, C. et al. (2015) Loci and candidate gene identification for resistance to Sclerotinia sclerotiorum in soybean (*Glycine max L.* Merr.) via association and linkage maps. *Plant J.* 82, 245–255.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1** Frequency distribution of the As contents in five maize tissues from plants at the combined locations.

**Figure S2** Quantile-quantile plots constructed using genome-wide association study results from three models (Q, K and Q + K) for the arsenic contents in five tissues across different locations.

**Figure S3** Manhattan plots for the arsenic contents in five tissues across different locations.

**Table S1** Correlation coefficients among five maize tissues in the association population.

**Table S2** List of significant loci, their detailed information and all candidate genes within the significant loci identified by the GWAS.

**Table S3** Comparison of GWAS results using the BLUP data and least square means from different locations of the same tissue.

© 2017 The Authors. *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 16, 1085–1093.