bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions

Waseem Hussain [*a], Malachy Campbell[b], Diego Jarquin[a], Harkamal Walia[a], and Gota Morota[*b]

[a]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583

[b]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Corresponding author

*Waseem Hussain

Department of Agronomy and Horticulture

University of Nebraska-Lincoln

Lincoln, Nebraska 68583 USA.

E-mail: waseem.hussain@unl.edu

Current address: International Rice Research Institute, Los Banos, Philippines

*Gota Morota

Department of Animal and Poultry Sciences

Virginia Polytechnic Institute and State University

175 West Campus Drive

Blacksburg, Virginia 24061 USA.

E-mail: morota@vt.edu

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Running title: vGWAS for wheat grain cadmium

Core Ideas:

Variance-heterogeneity mapping for grain Cadmium (Cd) concentration in bread wheat was performed.

Novel variance-heterogeneity loci were detected on chromosomes 2A and 2B.

Loci influencing both mean and variance were identified on chromosome 5A.

Identified variance-heterogeneity loci were associated with epistatic interactions.

Homoeology within the vQTL on chromosomes 2A and 2B was found.

Abbreviations

ABC transporter, ATP-binding cassette transporter; Cd, cadmium; DGLM, double generalized linear model; GLM, generalized linear model; GRM, genomic relationship matrix; HGLM, Hierarchical generalized linear model; HWW, hard-red winter wheat; mQTL, mean quantitative trait loci; mvQTL, mean-variance quantitative trait loci; QTL, quantitative trait loci; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; vQTL, variance heterogeneity quantitative trait loci; variance vGWAS heterogeneity genome-wide association studies.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Genome-wide association mapping identifies quantitative trait loci (QTL) that influence the mean differences between the marker genotypes for a given trait. While most loci influence the mean value of a trait, certain loci, known as variance heterogeneity QTL (vQTL) determine the variability of the trait instead of the mean trait value (mQTL). In the present study, we performed a variance heterogeneity genome-wide association study (vGWAS) for grain cadmium (Cd) concentration in bread wheat. We used double generalized linear model and hierarchical generalized linear model to identify vQTL associated with grain Cd. We identified novel vQTL regions on chromosomes 2A and 2B that contribute to the Cd variation and loci that affect both mean and variance heterogeneity (mvQTL) on chromosome 5A. In addition, our results demonstrated the presence of epistatic interactions between vQTL and mvQTL, which could explain variance heterogeneity. Overall, we provide novel insights into the genetic architecture of grain Cd concentration and report the first application of vGWAS in wheat. Moreover, our findings indicated that epistasis is an important mechanism underlying natural variation for grain Cd concentration.

Genome-wide association studies (GWAS) are routinely conducted to study the genetic basis of important traits in crops. GWAS link phenotypic variation with dense genetic marker data using a linear modeling framework (e.g., Nordborg and Weigel, 2008; Ingvarsson and Street, 2011; Huang and Han, 2014; Xiao et al., 2017). Standard GWAS approaches seek to identify marker-trait associations that influence the mean phenotypic values. However, differences in the variance between genotypes are also under genetic control (Shen et al., 2012). As a result, several recent studies have identified loci associated with differences in variance between genotypes (Cao et al., 2014; Corty et al., 2018). Such genetic variants that affect the variance

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

heterogeneity of traits have been referred to as variance heterogeneity quantitative trait loci (vQTL) (Rönnegård and Valdar, 2011). vQTL can be detected by searching the difference in the variability between the groups of genotypes that carry alternative alleles at a particular locus (Forsberg and Carlborg, 2017). A simple example is genotypes of wheat with difference in plant height. One genotype group is homozygous for a certain allele and manifests greater variability (including both shorter and taller plants), while the second genotype group that is homozygous for the alternative allele involves plants that are similar or uniform in height. This contrast in plant height across two allelic groups leads to genetic variance heterogeneity. Note that the mean difference between the two groups does not have to be different for variance heterogeneity to arise (Fig. 1).

Variance heterogeneity-based genome-wide association studies (vGWAS) have emerged as a new approach for identifying and mapping vQTL. vQTL contribute to variability, which is undetected through standard statistical mapping (bi-parental or association) procedures (Rönnegård and Valdar, 2011; Shen et al., 2012; Forsberg and Carlborg, 2017). It has been argued that variance heterogeneity between genotypes can be partially explained by epistasis or gene-by-environment interactions (Brown et al., 2014; Forsberg and Carlborg, 2017; Young et al., 2018). Thus, vQTL can provide insights into epistasis or phenotypic plasticity (Nelson et al., 2013; Young et al., 2018). Moreover, these vGWAS frameworks can serve as tractable approaches to reduce the search space when assessing epistasis among markers (Brown et al., 2014; Wei et al., 2016). This is because we can limit the number of interacting marker pairs $\binom{m}{2}$ to be investigated into $\binom{k}{2}$, where k is the number of markers ($k < m$) associated with vQTL or mvQTL.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Numerous studies have reported vQTL associated with diverse phenotypes, including the tendency to left-right turning and bristles (Mackay and Lyman, 2005) and locomotor handedness (Ayroles et al., 2015) in *Drosophila*; coat color (Nachman et al., 2003), circadian activity, and exploratory behavior (Corty et al., 2018) in mice; thermotolerance (Queitsch et al., 2002), flowering time (Salomé et al., 2011), and molybdenum concentration (Shen et al., 2012; Forsberg et al., 2015) in *Arabidopsis*; litter size in swine (Sell-Kubiak et al., 2015); urinary calcium excretion in rats (Perry et al., 2012); and body mass index (Yang et al., 2012; Young et al., 2018), sero-negative rheumatoid arthritis (Wei et al., 2017), and serum urate (Topless et al., 2015) in humans. In plants, vGWAS have been limited to few species, including *Arabidopsis* (Shen et al., 2012; Forsberg et al., 2015) and maize (Kusmec et al., 2017).

Methodologically, vQTL have been detected by performing statistical tests searching for unequal variance for a quantitative trait between the marker genotypes (Rönnegård and Valdar, 2012). The most common statistical tests used to identify vQTL include Levene's test (Paré et al., 2010), Brown-Forysthe test (Brown and Forsythe, 1974), squared residual value linear modeling (Struchalin et al., 2012), and correlation least squares test (Brown et al., 2014). However, these methods have certain drawbacks when applied to genetic data. For example, Levene's and Brown-Forsythe tests are sensitive to deviations from normality of residuals and have an inherent inability to model continuous covariates (Rönnegård and Valdar, 2011; Dumitrascu et al., 2019).

Double generalized linear model (DGLM) has emerged as an alternative approach to model the variance heterogeneity for genetic studies (Rönnegård and Valdar, 2011). In DGLM, sample means and residuals are modelled jointly. Here, generalized linear models (GLM) are fit by including only the fixed effects in the linear predictor(s) for the mean and then the squared residuals are used to estimate the dispersion effects. It is important to correct for population

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

structure, which can otherwise lead to spurious associations in GWAS (Patterson et al., 2006). In DGLM, population structure can be corrected by incorporating the first few principal components of a genomic relationship matrix (GRM) (Patterson et al., 2006; Price et al., 2010) as fixed covariates in the model. However, the first few principal components may not be sufficient to account for complex population structure or family relatedness (Hoffman, 2013; Sul et al., 2018). Alternatively, we can fit linear mixed models (LMM) to explicitly correct for population structure, where the whole GRM can be included to account for relationships among individuals and correct for background genotype effects. Hierarchical generalized linear model (HGLM) has been proposed as an extension of the DGLM to model random effects in the mean component (Rönnegård and Valdar, 2012; Tan et al., 2014). In HGLM, the GRM can be used to model correlated random effects and account for population structure.

We applied a vGWAS framework to examine the genetic architecture of grain cadmium (Cd) accumulation in wheat. Cd is a heavy metal that is highly toxic to human health (Menke et al., 2009). Identifying genetic variants that control low-grain Cd concentration in wheat is necessary to understand the basis for phenotypic variation in grain Cd and can help accelerate the development of low Cd wheat varieties. A recent study assessed natural variation in bread wheat grain Cd by conducting GWAS (Guttieri et al., 2015a). However, only a fraction of phenotypic variation could be explained by the top marker associations, indicating that grain Cd concentration is a complex trait that is influenced by multiple loci and/or loci with non-additive effects (Guttieri et al., 2015a). Given the genetic complexity of Cd in wheat, we hypothesized that variation in grain Cd concentration in wheat is influenced by vQTL that are likely to be involved in epistatic interactions; this would allow us to capture additional variation that is not accounted for in a standard GWAS approach.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In this study, we sought to provide additional insights into natural variation in grain Cd concentration by extending the standard GWAS to vGWAS using a hard winter wheat association mapping panel. To achieve this, we used DGLM and HGLM to perform vGWAS. Previously, Guttieri et al., (2015a) conducted the standard GWAS using this association panel and identified a single mean effect QTL (mQTL) for grain Cd concentration on chromosome 5A. In addition, we aimed to understand the basis of vQTL by searching for pairwise epistatic interactions among vQTL and mQTL. To our knowledge, the present study is the first to conduct vGWAS and identify vQTL associated with grain Cd concentration in wheat.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

MATERIALS AND METHODS

Plant Materials and Genotyping

We analyzed a publicly available dataset comprising of phenotypes for grain mineral concentration for $n = 299$ genotyped hard-red winter wheat accessions (hereafter called as HWW association panel). The details of the study are discussed in Guttieri et al., (2015a; 2015b), and access to the data is available at http://triticeaetoolbox.org/wheat/. The data are also downloadable at https://github.com/whussain2/vGWAS/tree/master/Data. Here, we focused on grain Cd concentration (mg/kg) collected across two years (2012 and 2013) in one location (Oklahoma, USA). Briefly, the experiment was laid in an augmented incomplete block design with two replications and 15 blocks within each replication. Least square means adjusted across the replications and blocks in each year were obtained for each genotype. In this study, we averaged the least square means for each genotype across two years because of non-significant genotype x year interaction (Guttieri et al., 2015a). The association panel was genotyped using a 90K iSelect Infinium array (Wang et al., 2014b). We used a filtered marker data set consisting of $m = 14,731$ single nucleotide polymorphism (SNP) markers from the 90K iSelect Infinium array as described by Guttieri et al., (2015a). All the SNP markers were physically anchored on the new reference genome of hexaploid wheat RefSeq v1.0 (International Wheat Genome Sequencing Consortium (IWGSC), 2018).

Statistical Modeling

Genome-Wide Association Mapping

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

158 Standard GWAS or mQTL analysis based on mean differences between marker genotypes for
159 grain Cd concentration was performed similar to Guttieri et al., (2015a) using the rrBLUP
160 package (Endelman, 2011) in the R environment (R Core Team 2018).

161 **Variance-Heterogeneity Genome-Wide Association Mapping**

162 We used DGLM and HGLM to perform vGWAS and detect vQTL in the current study. The
163 description of models used is given below.

164 **DGLM**

165 DGLM is a parametric approach that can be used to jointly model the mean and dispersion using
166 a GLM framework (Smyth, 1989). The DGLM works iteratively by first fitting a linear model to
167 estimate the mean effects (mQTL). The squared residuals are used to estimate the dispersion
168 effects (vQTL) using GLM with a gamma-distributed response and the log link function. This
169 process is cycled until convergence. Here, we extended the DGLM to marker-based association
170 analysis according to Rönnegård and Valdar (2011). The mean part of DGLM was as follows:

$$\mathbf{y} = \mathbf{1}\mu_m + \mathbf{X}\boldsymbol{\beta} + \mathbf{s}_j a_{mj} + \boldsymbol{\epsilon} \ \#(\mathbf{1})$$

171 where **y** is the Cd concentration (mg/kg); **1** is the column vector of 1; μ_m is the intercept; **X** is
172 $n \times 4$ covariate matrix of the top four principle components (PCs) obtained by performing
173 principal component analysis (PCA) of marker data using the SNPRelate R package (Zheng et
174 al., 2012); $\boldsymbol{\beta}$ is the regression coefficients for the covariates; $\mathbf{s}_j \in (0,2)$ is the vector containing the
175 number of reference allele at the marker *j*, a_{mj} is the effect size or allele substitution effect of the
176 *j*th marker; and $\boldsymbol{\epsilon}$ is the residual. We assumed

$$\epsilon_i \sim N(0, \mathbf{I}\sigma^2_{\epsilon_i})$$

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$log(\sigma^2_{\epsilon_i}) = \mathbf{1}\mu_v + \mathbf{s}_j a_{vj}$$

where **I** is the identity matrix; $\sigma^2_{\epsilon_i}$ is the residual variance; and $\mathbf{1}\mu_v$ and a_v are the intercept and marker regression coefficients for the variance part of the model, respectively. While we fit separate effects for the mean using a standard linear model and for the variance using the squared residuals in gamma distributed GLM with a log link function, this is equivalent to modeling $\mathbf{y} \sim N(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{s}a_{mj}, \exp(\mathbf{1}\mu_v + \mathbf{s}_j a_{vj})$ or $\boldsymbol{\epsilon} \sim N(0, \exp(\mathbf{1}\mu_v + \mathbf{s}_j a_{vj}))$ in equation (1).

The DGLM was fitted using the dglm package (https://cran.r-project.org/web/packages/dglm/index.html) in R. SNP markers were fitted one by one, and for each marker, the effect sizes, standard errors, and p-values were obtained for the mean and dispersion components. To account for multiple testing, we determined the effective number of independent tests (Meff) using the method described by Li and Ji (2005). Subsequently, a genome-wide significance threshold level ($P < 1.44 \times 10^{-5}$) was determined using the following formula:

$$\alpha_p = 1 - (1 - \alpha_e)^{\frac{1}{\text{Meff}}} \qquad (2)$$

where α_p is the genome-wide significance threshold level, α_e is the desired level of significance (0.05), and Meff = 3,495.

HGLM

To explicitly account for population structure and kinship in GWAS, LMM have been proposed as alternative methods that allow the genetic relationships between individuals to be modeled as random effects. To perform vGWAS in the LMM framework and to identify genome-wide vQTL, we used a HGLM approach. HGLM (Lee and Nelder, 1996) is a class of GLM and is a

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

196 direct extension of the DGLM that allows joint modelling of the mean and dispersion parts and
197 introduces random effects as a linear predictor for the mean (Rönnegård and Carlborg, 2007).
198 The mean part of HGLM was given as follows:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{s}_j a_{mj} + \mathbf{Zu} + \boldsymbol{\epsilon} \ \#(3)$$

199 assuming that

$$\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$$

200 where **Z** is the incident matrix of random effects of genotypes; **u** is the vector of random effects
201 with $\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$; **G** is the GRM of VanRaden (2008); and σ_u^2 is the additive genetic variance.
202 A log link function was used for the residual variance given by $\exp(\mathbf{s}_j a_{vj})$, which is equivalent
203 to modeling $\mathbf{y}|a_{mj}, \mathbf{u}, a_{vj} \sim N(\mathbf{s}_j a_{mj} + \ \mathbf{Zu}, \exp(\mathbf{s}_j a_{vj}))$.

204 We fitted HGLM using the hglm R package (Rönnegård et al., 2010b). We reformulated the term
205 **Zu** as $\mathbf{Z}^*\mathbf{u}^*$, where $\mathbf{u}^* \sim N(0, \mathbf{I}\sigma_u^2)$; $\mathbf{Z}^* = \mathbf{Z}_0\mathbf{L}$; **L** is the Cholesky factorization of the **G** matrix;
206 and $\mathbf{Z}_0$ is the identity matrix (Rönnegård et al., 2010a). Markers treated as fixed effects were fit
207 one by one, and for each marker, the effect sizes, standard errors, and p-values were obtained for
208 the mean and dispersion components. The genome-wide significance threshold level was derived
209 as described in the DGLM analysis. Circular Manhattan and quantile-quantile (QQ) plots were
210 created using the CMplot R package (https://github.com/YinLiLin/R-CMplot).

211 **Epistasis Analysis**

212 We investigated the extent of epistasis that was manifested through variance heterogeneity. All
213 the possible pairwise interaction analyses for markers that were associated with grain Cd
214 concentration were performed using the following two markers at a time epistatic model:

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{s}_j a_j + \mathbf{s}_k a_k + (\mathbf{s}_j \mathbf{s}_k) v_{jk} + \boldsymbol{\epsilon} \#(4)$$

where **y** is the vector of Cd concentration (mg/kg); **X** is the incident matrix for the first four PCs; **β** is the regression coefficients for the PCs; $\mathbf{s}_j$ and $\mathbf{s}_k$ are SNP codes for the *j*th and *k*th markers, respectively; a_j and a_k are the additive effects of the markers j and k, respectively; and v_{jk} is the additive × additive epistatic effect of the *j*th and *k*th markers. We used Bonferroni correction to account for the multiple testing. The threshold of -log10(0.05/325) = -log10 (1.54 x 10^{-04}) = 3.8 was used to declare the significance of interaction effects.

Homoeology and Candidate Gene Analysis

Homoeologous gene construction was performed as per procedure described by (Santantonio et al., 2019). Briefly, the annotated coding sequences within the 2A vQTL were aligned back onto themselves using the IWGSC RefSeq v.1.0 coupled with BLAST tool in Ensemble Plants browser (Bolser et al., 2017). For candidate gene identification for the SNP markers associated with variance heterogeneity, we used Ensembl Plants browser to retrieve the candidate genes and functional annotations (http://plants.ensembl.org/Triticum_aestivum/Info/Index) and the wheat RefSeq v1.0 annotations (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018) available at https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations. For candidate gene analysis, we first determined the positions of significant SNP markers, and the interval was defined as the distance between the lowest and highest markers based on the position of SNP. For example, if the position of the lowest SNP and highest SNP was 715,333,165 bp and 717,146,211 bp in the vQTL region on chromosome 2A, we defined 2A as the 715,333,165-717,146,211 interval for candidate gene identification. After defining the interval for the 2A (2A: 715,333,165-

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

717,146,211) and 2B (2B: 691,780,716- 701,097,263 bp) regions, we explored the intervals using Ensembl Plants browser and extracted the Gene IDs within these intervals. The Gene IDs within the defined interval on chromosomes 2A and 2B were analyzed using the IWGSC RefSeq v.1.0 (International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018) integrated genome annotations to obtain the predicted genes and functional annotations.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

RESULTS

Variance Heterogeneity GWAS Provide Additional Insights into Natural Variation in Grain Cd

Although grain Cd concentration is a highly heritable trait, recent GWAS revealed that significant loci can only explain a fraction of the variation for this trait (Guttieri et al., 2015a). We found the single genomic region on chromosome 5A affecting the grain Cd concentration (Fig. 2) from the standard GWAS analysis confirming the results of Guttieri et al., (2015). The DGLM and HGLM approaches were used to detect vQTL while controlling for population structure. The population structure based on PCA of the HWW association panel is given in Supplemental File S1: Figure S1. QQ plots (Supplemental File S1: Figure S2) show that both DGLM and HGLM had adequate control of population structure and effective control of false positives.

We classified the QTL into the following categories: mQTL, which contributes to difference in the means between marker genotypes; vQTL, which influences the variability between the genotypes; and mean-variance QTL (mvQTL), which contributes to differences in both the mean and variance between the genotypes.

Based on the DGLM, we identified two vQTL associated with the variance heterogeneity of Cd concentration. One vQTL on 2A contained four SNP markers, and one vQTL on 2B contained 17 SNP markers (Fig. 2 and Supplemental File S1: Table S1). The four SNP markers associated with the vQTL region on the chromosome 2A region spanned the physical distance of 1.81 Mb; all SNP markers were located within the 1,000 bp linkage disequilibrium (LD) block (Supplemental File S1: Figure S3). The vQTL region on 2B associated with 17 SNP markers

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

spanned the physical distance of 9.32 Mb, and the SNP markers were located within four LD blocks of sizes 0, 1, 1, and 204 kb (Supplemental File S1: Figure S4).

In addition, we identified a single mvQTL (containing four SNP markers) associated with both mean and variance heterogeneity on chromosome 5A (Fig. 2 and Table S2). The markers associated with mvQTL on chromosome 5A were identical to those obtained in the original GWAS analysis according to Guttieri et al., (2015), indicating that this region affects both the mean and the variance heterogeneity (Supplemental File S1: Figure S5). Moreover, these results showed that DGLM serves as an accurate framework to jointly detect mean and variance QTL and provides additional insights into phenotypic variation that would otherwise not be captured by standard GWAS.

The HGLM analysis revealed the same results as those obtained using DGLM and showed identical vQTL on chromosomes 2A and 2B and mvQTL on chromosome 5A associated with variance heterogeneity of Cd concentration (Fig. 2 and Supplemental File S1: Table S1). Further, we observed a potential vQTL region on 2D from the DGLM and HGLM analyses. This region was slightly below the significance threshold level but may have an implication on Cd variation given that the allopolyploid nature of wheat and the role of homoeologous gene sets on phenotypic variation (Borrill et al., 2019).

Variance Heterogeneity Loci can be Partially Explained by Epistasis

We investigated all significant markers (25 markers) associated with mvQTL on chromosome 5A and vQTL on chromosomes 2A and 2B and explored all possible pairwise additive × additive epistatic interactions. We detected significant additive × additive interactions between the markers (Fig. 3). The interaction was more evident between mvQTL on chromosome 5A and

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

vQTL on chromosomes 2A and 2B. Specifically, all the markers associated with the 5A mvQTL region revealed highly significant interactions with all the markers associated with the 2A and 2B vQTL regions. Interactions between vQTL on chromosomes 2A and 2B were also observed; however, the interactions were less evident, and only a few markers within these regions showed statistically significant interactions. Taken together, these results suggested that the vQTL and mvQTL may be manifested because of pairwise epistatic interactions.

Homoeology and Candidate Genes

Homoeology analysis between the defined regions on chromosomes 2A and 2B resulted in 22 homoeologous gene sets, consisting of 21 triplicates and only one duplicate gene set. Additional details on the homoeologous gene sets can be found in Supplemental File S2. As compared to the total number of candidate genes equal to 39 within the 1.18 Mb 2A region, 22 (58%) were homoeologous across the three genomes. Based on the annotations for the 22 homoeologous gene sets, a few of the genes encoded homeobox-leucine zipper family protein, plant peroxidase, and glycosyltransferase, which have been associated with the genetic regulation of minerals in plants (Whitt et al., 2018). For example, homeodomain-leucine zipper family protein has been functionally associated with Cd tolerance by regulating the expression of metal transporters *OsHMA2* and *OsHMA3* in rice (Ding et al., 2018; Yu et al., 2019). These genes have been found to play important roles in loading Cd onto the xylem and root-to-shoot translocation of Cd in rice. In plants, response to heavy metals involves the accumulation of reactive oxygen species (ROS) that damage DNA and cellular machinery (Kumari et al., 2008; Rascio and Navari-Izzo, 2011). In *Arabidopsis*, the peroxidase genes *At2g35380*, *PER20*, and *At2g18150* have been found to be associated with Cd responses by affecting the lignin biosynthesis in root cells under high Cd stress (Chen and Kao, 1995; van de Mortel et al., 2008). Full list of candidate genes within

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

309 the 2A and 2B region, and within the homoeologous gene sets is in Supplemental File S2. These
310 results clearly indicate that most of the genes with vQTL regions are redundant across the
311 genomes and may have significant role in the genetic regulation of grain Cd concentration in
312 wheat. However, we contend that further investigation of these regions using dense markers and
313 increased sample size is necessary to fine-map the QTL and validate potential candidate genes
314 underlying these loci and also the role of gene redundancy in generating phenotypic variation.
315

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

DISCUSSION

In the present study, we explored the genetic variants affecting variance heterogeneity of Cd. Given the complexity of genetic regulation of Cd in wheat (Guttieri et al., 2015a) and the influence of epistatic interactions, we anticipated that partial genetic regulation of Cd in wheat can be detected using methods that have been developed to identify vQTL. As reported by Rönnegård and Valdar, (2012), a potential explanation for variance-controlling QTL is epistatic interactions that are unspecified in the model. Herein, we utilized two approaches, namely, DGLM and HGLM, to detect vQTL and mvQTL associated with grain Cd concentration in wheat.

The DGLM framework is a powerful approach for vGWAS analysis. However, in DGLM, GLM is fit by including only the fixed effects in the linear predictor of mean and dispersion. Therefore, by using the DGLM approach, population structure can only be accounted for by using the first few PCs obtained from the SNP matrix; however, this may not completely account for complex population structure and family relationships (Price et al., 2010). We hypothesized that the use of random effects to model the mean component can better account for population structure and reduce spurious associations. In this approach, a random additive genetic effect is introduced to the mean component of the model that accounts for population structure and cryptic relatedness between accessions. Therefore, we performed vGWAS analysis using HGLM. Interestingly, both DGLM and HGLM approaches were effective in identifying the genetic variants controlling variability of Cd, suggesting that the loci detected with the DGLM approach are likely to be true QTL rather than artifacts from population structure. The impact of population structure on the power of DGLM and HGLM remains to be explored; further examination is warranted.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In the literature, it has been argued that variance heterogeneity can also arise by a simple mean variance relationship, which does not have biological significance (Young et al., 2018). To rule out the role of the mean-variance function in generating variance heterogeneity, we plotted the estimated effects of the top three significant associated vQTL markers at the alternate genotypes and observed that the means of all the markers were the same (Fig. 4), indicating that the effect of SNP on variance heterogeneity was not due to the consequences of mean-variance function but likely due to the genetic effects (Yang et al., 2012).

Further, variance heterogeneity can also be observed in a population when two or more alleles having different effects on the phenotype are in high LD (Cao et al., 2014; Wang et al., 2014a; Forsberg and Carlborg, 2017). To rule out the possibility of LD as a source for variance heterogeneity in grain Cd in this population, we suggest the use of high-density markers and larger sample size to identify the actual functional alleles associated with Cd, their LD patterns, and their effects on the Cd phenotype (Struchalin et al., 2010; Forsberg and Carlborg, 2017).

In QTL studies, variance heterogeneity arises because of various underlying mechanisms, such as epistatic interactions (Struchalin et al., 2010; Shen et al., 2012; Nelson et al., 2013). Epistasis gives rise to variance heterogeneity when the different allele combinations at one locus change the effect of the other loci in the genome, as shown in one pair of interacting markers (Fig. 5). Hence, identifying the loci affecting variance heterogeneity through vGWAS means that the loci are likely to be involved in epistatic interactions. To validate this assumption and investigate whether epistasis can explain the identified vQTL and mvQTL in this study, we analyzed all possible pairwise interactions between the associated markers. We detected significant epistatic interactions between the associated markers (Fig. 2), which can explain the existence of variance heterogeneity in the genotypes. Additionally, identifying vQTL through vGWAS serves as an

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

361 effective way to restrict the search space when detecting epistatic QTL. Thus, with the vGWAS
362 approach, many of the requirements necessary for conventional epistasis mapping can be
363 avoided (e.g., large sample size and extensive multiple testing corrections that reduce power).
364 However, Forsberg and Carlborg (2017) empirically showed that the presence of variance
365 heterogeneity does not always guarantee the presence of epistatic interactions that contribute to
366 the total variation of the trait; therefore, the results should be interpreted carefully when multi-
367 locus interactions are involved.

368 The genomic regions on chromosomes 2A and 2B associated with variance heterogeneity
369 revealed homoeologous gene sets with 58% genes revealing the gene redundancy mostly present
370 as three functional homoeologous copies (triplicated). This also indicates that genetic complexity
371 of Cd phenotype is not only controlled by multiple genes but may be affected by the multiple
372 homoeologs of the individual genes which warrants further investigation. Presence of multiple
373 copies of homoeologous genes may have consequence on phenotypic variation due to dosage
374 effects and or functional redundancy (Borrill et al., 2019). Dosage effect, in which the
375 phenotypic variation is amplified by the addition of each gene copies can act additively (e.g.,
376 genes controlling grain protein content (Avni et al., 2014) and grain size (Wang et al., 2018)) or
377 non-additively (e.g., genes controlling amylopectin content in wheat (Kim et al., 2003)). Non-
378 additive variation between homoeologous gene has been shown to be an important source of
379 variation in wheat. However, its relative contribution across the wheat genome as compared to
380 non-syntenic regions was proportionately less (Santantonio et al., 2019). This is in agreement
381 with our results because we observed interactions among the homoeologous genomic regions on
382 chromosomes 2A and 2B. However, this homoeologous gene interactions was less evident as
383 compared to two-way interactions found between non-syntenic vQTL regions on 2A and 2B with

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

384 the mvQTL region on 5A. The nature and functional role of homoeologous gene sets within the
385 vQTL region on 2A and 2B is not clear. However, it is increasingly feasible in wheat to examine
386 the effects of gene redundancy and explore the contribution of homoeologous genes in
387 generating phenotypic variation (Wang et al., 2018) .

388 **Conclusion**

389 We showed the potential of vGWAS for dissecting the genetic architecture of complex traits and
390 identifying novel genomic regions influencing variance heterogeneity in wheat. We provided
391 evidence that the vQTL contribute to natural variation in grain Cd concentration through non-
392 additive genetic effects. This is particularly evidenced by epistatic interactions between mvQTL
393 on chromosome 5A and vQTL on chromosomes 2A and 2B.

394 **Acknowledgements**

395 This work was supported by the National Science Foundation under Grant Number 1736192 to
396 H.W. and G.M. Data analysis was performed using the Holland Computing Center
397 computational resources at the University of Nebraska-Lincoln.

398
399

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplemental Materials

Supplemental File S1 contains Table S1 and Figures S1-S5.

Table S1: Single nucleotide polymorphism markers associated with variance heterogeneity of cadmium concentration in the hard-red winter wheat association panel.

Table S2: Single nucleotide polymorphism markers associated with the mean of cadmium concentration in the hard-red winter wheat association panel.

Figure S1: Principal component analysis of the population structure in the hard-red winter wheat association panel. The different colors represent the sub-populations of red wheat and winter wheat.

Figure S2: Quantile-quantile (QQ) plot of the outputs for the double generalized linear model and the hierarchical generalized linear model shown in the Manhattan plot.

Figure S3: Linkage disequilibrium block and annotated genes on chromosome 2A.

Figure S4: Linkage disequilibrium blocks and annotated genes on chromosome 2B.

Figure S5: Violin plot showing the differences in the mean and variance of Cadmium concentration with alternative marker allele groups.

Supplemental File S2: A list of candidate genes and homoeologous gene sets associated with the vQTL on chromosomes 2A and 2B.

Data Availability

The wheat phenotypic and genotypic data can be downloaded from (http://triticeaetoolbox.org/wheat/) and also available on the GitHub repository

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

https://github.com/whussain2/vGWAS. The R code used for the analysis is available on the GitHub repository https://github.com/whussain2/vGWAS.

Conflict of interest

The authors declare there are no competing interests.

Author's Contributions

W.H. and G.M. conceived the study. W.H. performed the data analysis and drafted the manuscript. D.J. helped the data analysis. M.C., D.J., H.W., and G.M. revised the manuscript. G.M. supervised and directed the study. All authors read and approved the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

Avni, R., R. Zhao, S. Pearce, Y. Jun, C. Uauy, F. Tabbita, T. Fahima, A. Slade, J. Dubcovsky, and A. Distelfeld. 2014. Functional characterization of GPC-1 genes in hexaploid wheat. Planta 239:313–324. doi:10.1007/s00425-013-1977-y

Ayroles, J.F., S.M. Buchanan, C. O'Leary, K. Skutt-Kakaria, J.K. Grenier, A.G. Clark, D.L. Hartl, and B.L. de Bivort. 2015. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. PNAS 112:6706–6711. doi:10.1073/pnas.1503830112

Bolser, D.M., D.M. Staines, E. Perry, and P.J. Kersey. 2017. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomic Data. A.D.J. van Dijk, ed. Springer New York, New York, NY.

Borrill, P., S.A. Harrington, and C. Uauy. 2019. Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. The Plant Journal 97:56–72. doi:10.1111/tpj.14150

Brown, A.A., A. Buil, A. Viñuela, T. Lappalainen, H.-F. Zheng, J.B. Richards, K.S. Small, T.D. Spector, E.T. Dermitzakis, and R. Durbin. 2014. Genetic interactions affecting human gene expression identified by variance association mapping. eLife 3:e01381. doi:10.7554/eLife.01381

Brown, M.B., and A.B. Forsythe. 1974. The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means. Technometrics 16:129–132. doi:10.2307/1267501

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

449 Cao, Y., P. Wei, M. Bailey, J.S.K. Kauwe, and T.J. Maxwell. 2014. A Versatile Omnibus Test
450 for Detecting Mean and Variance Heterogeneity. Genet Epidemiol 38:51–59

451 Chen, S.L., and C.H. Kao. 1995. Cd induced changes in proline level and peroxidase activity in
452 roots of rice seedlings. Plant Growth Regul 17:67–71. doi:10.1007/BF00024497

453 Corty, R.W., V. Kumar, L.M. Tarantino, J.S. Takahashi, and W. Valdar. 2018. Mean-Variance
454 QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in
455 Mice. G3: Genes, Genomes, Genetics g3.200194.2018. doi:10.1534/g3.118.200194

456 Ding, Y., S. Gong, Y. Wang, F. Wang, H. Bao, J. Sun, C. Cai, K. Yi, Z. Chen, and C. Zhu. 2018.
457 MicroRNA166 Modulates Cadmium Tolerance and Accumulation in Rice. Plant Physiol.
458 177:1691–1703. doi:10.1104/pp.18.00485

459 Dumitrascu, B., G. Darnell, J. Ayroles, and B.E. Engelhardt. 2019. Statistical tests for detecting
460 variance effects in quantitative trait studies. Bioinformatics 35:200–210.
461 doi:10.1093/bioinformatics/bty565

462 Endelman, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R
463 Package rrBLUP. The Plant Genome 4:250–255. doi:10.3835/plantgenome2011.08.0024

464 Forsberg, S.K.G., M.E. Andreatta, X.-Y. Huang, J. Danku, D.E. Salt, and Ö. Carlborg. 2015. The
465 Multi-allelic Genetic Architecture of a Variance-Heterogeneity Locus for Molybdenum
466 Concentration in Leaves Acts as a Source of Unexplained Additive Genetic Variance.
467 PLOS Genetics 11:e1005648. doi:10.1371/journal.pgen.1005648

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

468 Forsberg, S.K.G., and Ö. Carlborg. 2017. On the relationship between epistasis and genetic
469 variance heterogeneity. J Exp Bot 68:5431–5438. doi:10.1093/jxb/erx283

470 Guttieri, M.J., P.S. Baenziger, K. Frels, B. Carver, B. Arnall, S. Wang, E. Akhunov, and B.M.
471 Waters. 2015a. Prospects for Selecting Wheat with Increased Zinc and Decreased
472 Cadmium Concentration in Grain. Crop Science 55:1712–1728.
473 doi:10.2135/cropsci2014.08.0559

474 Guttieri, M.J., P.S. Baenziger, K. Frels, B. Carver, B. Arnall, and B.M. Waters. 2015b. Variation
475 for Grain Mineral Concentration in a Diversity Panel of Current and Historical Great
476 Plains Hard Winter Wheat Germplasm. Crop Science 55:1035–1052.
477 doi:10.2135/cropsci2014.07.0506

478 Hoffman, G.E. 2013. Correcting for Population Structure and Kinship Using the Linear Mixed
479 Model: Theory and Extensions. PLOS ONE 8:e75707. doi:10.1371/journal.pone.0075707

480 Huang, X., and B. Han. 2014. Natural Variations and Genome-Wide Association Studies in Crop
481 Plants. Annual Review of Plant Biology 65:531–551. doi:10.1146/annurev-arplant-
482 050213-035715

483 Hulse, A.M., and J.J. Cai. 2013. Genetic Variants Contribute to Gene Expression Variability in
484 Humans. Genetics 193:95–108. doi:10.1534/genetics.112.146779

485 Ingvarsson, P.K., and N.R. Street. 2011. Association genetics of complex traits in plants. New
486 Phytologist 189:909–922. doi:10.1111/j.1469-8137.2010.03593.x

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

International Wheat Genome Sequencing Consortium (IWGSC), IWGSC RefSeq principal investigators:, R. Appels, K. Eversole, C. Feuillet, B. Keller, J. Rogers, N. Stein, IWGSC whole-genome assembly principal investigators:, C.J. Pozniak, N. Stein, F. Choulet, A. Distelfeld, K. Eversole, J. Poland, J. Rogers, G. Ronen, A.G. Sharpe, Whole-genome sequencing and assembly:, C. Pozniak, G. Ronen, N. Stein, O. Barad, K. Baruch, F. Choulet, G. Keeble-Gagnère, M. Mascher, A.G. Sharpe, G. Ben-Zvi, A.-A. Josselin, Hi-C data-based scaffolding:, N. Stein, M. Mascher, A. Himmelbach, Whole-genome assembly quality control and analyses:, F. Choulet, G. Keeble-Gagnère, M. Mascher, J. Rogers, F. Balfourier, J. Gutierrez-Gonzalez, M. Hayden, A.-A. Josselin, C. Koh, G. Muehlbauer, R.K. Pasam, E. Paux, C.J. Pozniak, P. Rigault, A.G. Sharpe, J. Tibbits, V. Tiwari, Pseudomolecule assembly:, F. Choulet, G. Keeble-Gagnère, M. Mascher, A.-A. Josselin, J. Rogers, RefSeq genome structure and gene analyses:, M. Spannagl, F. Choulet, D. Lang, H. Gundlach, G. Haberer, G. Keeble-Gagnère, K.F.X. Mayer, D. Ormanbekova, E. Paux, V. Prade, H. Šimková, T. Wicker, Automated annotation:, F. Choulet, M. Spannagl, D. Swarbreck, H. Rimbert, M. Felder, N. Guilhot, H. Gundlach, G. Haberer, G. Kaithakottil, J. Keilwagen, D. Lang, P. Leroy, T. Lux, K.F.X. Mayer, S. Twardziok, L. Venturini, Manual gene curation:, R. Appels, H. Rimbert, F. Choulet, A. Juhász, G. Keeble-Gagnère, Subgenome comparative analyses:, F. Choulet, M. Spannagl, D. Lang, M. Abrouk, G. Haberer, G. Keeble-Gagnère, K.F.X. Mayer, T. Wicker, Transposable elements:, F. Choulet, T. Wicker, H. Gundlach, D. Lang, M. Spannagl, Phylogenomic analyses:, D. Lang, M. Spannagl, R. Appels, I. Fischer, Transcriptome analyses and RNA-seq data:, C. Uauy, P. Borrill, R.H. Ramirez-Gonzalez, R. Appels, D. Arnaud, S. Chalabi, B. Chalhoub, F. Choulet, A. Cory, R. Datla, M.W. Davey, M. Hayden, J. Jacobs,

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

510 D. Lang, S.J. Robinson, M. Spannagl, B. Steuernagel, J. Tibbits, V. Tiwari, F. van Ex,
511 B.B.H. Wulff, Whole-genome methylome:, C.J. Pozniak, S.J. Robinson, A.G. Sharpe, A.
512 Cory, Histone mark analyses:, M. Benhamed, E. Paux, A. Bendahmane, L. Concia, D.
513 Latrasse, BAC chromosome MTP IWGSC–Bayer Whole-Genome Profiling (WGP) tags:,
514 J. Rogers, J. Jacobs, M. Alaux, R. Appels, J. Bartoš, A. Bellec, H. Berges, J. Doležel, C.
515 Feuillet, Z. Frenkel, B. Gill, A. Korol, T. Letellier, O.-A. Olsen, H. Šimková, K. Singh,
516 M. Valárik, E. van der Vossen, S. Vautrin, S. Weining, Chromosome LTC mapping and
517 physical mapping quality control:, A. Korol, Z. Frenkel, T. Fahima, V. Glikson, D. Raats,
518 J. Rogers, RH mapping:, V. Tiwari, B. Gill, E. Paux, J. Poland, Optical mapping:, J.
519 Doležel, J. Číhalíková, H. Šimková, H. Toegelová, J. Vrána, Recombination analyses:, P.
520 Sourdille, B. Darrier, Gene family analyses:, R. Appels, M. Spannagl, D. Lang, I.
521 Fischer, D. Ormanbekova, V. Prade, CBF gene family:, D. Barabaschi, L. Cattivelli,
522 Dehydrin gene family:, P. Hernandez, S. Galvez, H. Budak, NLR gene family:, B.
523 Steuernagel, J.D.G. Jones, K. Witek, B.B.H. Wulff, G. Yu, PPR gene family:, I. Small, J.
524 Melonek, R. Zhou, Prolamin gene family:, A. Juhász, T. Belova, R. Appels, O.-A. Olsen,
525 WAK gene family:, K. Kanyuka, R. King, Stem solidness (SSt1) QTL team:, K. Nilsen,
526 S. Walkowiak, C.J. Pozniak, R. Cuthbert, R. Datla, R. Knox, K. Wiebe, D. Xiang,
527 Flowering locus C (FLC) gene team:, A. Rohde, T. Golds, Genome size analysis:, J.
528 Doležel, J. Čížková, J. Tibbits, MicroRNA and tRNA annotation:, H. Budak, B.A.
529 Akpinar, S. Biyiklioglu, Genetic maps and mapping:, G. Muehlbauer, J. Poland, L. Gao,
530 J. Gutierrez-Gonzalez, A. N'Daiye, BAC libraries and chromosome sorting:, J. Doležel,
531 H. Šimková, J. Číhalíková, M. Kubaláková, J. Šafář, J. Vrána, BAC pooling, BAC library
532 repository, and access:, H. Berges, A. Bellec, S. Vautrin, IWGSC sequence and data

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

533 repository and access:, M. Alaux, F. Alfama, A.-F. Adam-Blondon, R. Flores, C.
534 Guerche, T. Letellier, M. Loaec, H. Quesneville, Physical maps and BAC-based
535 sequences:, 1A BAC sequencing and assembly:, C.J. Pozniak, A.G. Sharpe, S.
536 Walkowiak, H. Budak, J. Condie, J. Ens, C. Koh, R. Maclachlan, Y. Tan, T. Wicker, 1B
537 BAC sequencing and assembly:, F. Choulet, E. Paux, A. Alberti, J.-M. Aury, F.
538 Balfourier, V. Barbe, A. Couloux, C. Cruaud, K. Labadie, S. Mangenot, P. Wincker, 1D,
539 4D, and 6D physical mapping:, B. Gill, G. Kaur, M. Luo, S. Sehgal, 2AL physical
540 mapping:, K. Singh, P. Chhuneja, O.P. Gupta, S. Jindal, P. Kaur, P. Malik, P. Sharma, B.
541 Yadav, 2AS physical mapping:, N.K. Singh, J. Khurana, C. Chaudhary, P. Khurana, V.
542 Kumar, A. Mahato, S. Mathur, A. Sevanthi, N. Sharma, R.S. Tomar, 2B, 2D, 4B, 5BL,
543 and 5DL IWGSC–Bayer Whole-Genome Profiling (WGP) physical maps:, J. Rogers, J.
544 Jacobs, M. Alaux, A. Bellec, H. Berges, J. Doležel, C. Feuillet, Z. Frenkel, B. Gill, A.
545 Korol, E. van der Vossen, S. Vautrin, 3AL physical mapping:, B. Gill, G. Kaur, M. Luo,
546 S. Sehgal, 3DS physical mapping and BAC sequencing and assembly:, J. Bartoš, K.
547 Holušová, O. Plíhal, 3DL BAC sequencing and assembly:, M.D. Clark, D. Heavens, G.
548 Kettleborough, J. Wright, 4A physical mapping, BAC sequencing, assembly, and
549 annotation:, M. Valárik, M. Abrouk, B. Balcárková, K. Holušová, Y. Hu, M. Luo, 5BS
550 BAC sequencing and assembly:, E. Salina, N. Ravin, K. Skryabin, A. Beletsky, V.
551 Kadnikov, A. Mardanov, M. Nesterov, A. Rakitin, E. Sergeeva, 6B BAC sequencing and
552 assembly:, H. Handa, H. Kanamori, S. Katagiri, F. Kobayashi, S. Nasuda, T. Tanaka, J.
553 Wu, 7A physical mapping and BAC sequencing:, R. Appels, M. Hayden, G. Keeble-
554 Gagnère, P. Rigault, J. Tibbits, 7B physical mapping, BAC sequencing, and assembly:,
555 O.-A. Olsen, T. Belova, F. Cattonaro, M. Jiumeng, K. Kugler, K.F.X. Mayer, M. Pfeifer,

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

S. Sandve, X. Xun, B. Zhan, 7DS BAC sequencing and assembly:, H. Šimková, M. Abrouk, J. Batley, P.E. Bayer, D. Edwards, S. Hayashi, H. Toegelová, Z. Tulpová, P. Visendi, 7DL physical mapping and BAC sequencing:, S. Weining, L. Cui, X. Du, K. Feng, X. Nie, W. Tong, L. Wang, Figures:, P. Borrill, H. Gundlach, S. Galvez, G. Kaithakottil, D. Lang, T. Lux, M. Mascher, D. Ormanbekova, V. Prade, R.H. Ramirez-Gonzalez, M. Spannagl, N. Stein, C. Uauy, L. Venturini, Manuscript writing team:, N. Stein, R. Appels, K. Eversole, J. Rogers, P. Borrill, L. Cattivelli, F. Choulet, P. Hernandez, K. Kanyuka, D. Lang, M. Mascher, K. Nilsen, E. Paux, C.J. Pozniak, R.H. Ramirez-Gonzalez, H. Šimková, I. Small, M. Spannagl, D. Swarbreck, and C. Uauy. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361. doi:10.1126/science.aar7191

Kim, W., J.W. Johnson, R.A. Graybosch, and C.S. Gaines. 2003. Physicochemical Properties and End-use Quality of Wheat Starch as a Function of Waxy Protein Alleles. Journal of Cereal Science 37:195–204. doi:10.1006/jcrs.2002.0494

Kumari, M., G.J. Taylor, and M.K. Deyholos. 2008. Transcriptomic responses to aluminum stress in roots of Arabidopsis thaliana. Mol Genet Genomics 279:339. doi:10.1007/s00438-007-0316-z

Kusmec, A., S. Srinivasan, D. Nettleton, and P.S. Schnable. 2017. Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature Plants 3:715–723. doi:10.1038/s41477-017-0007-7

Lee, Y., and J.A. Nelder. 1996. Hierarchical Generalized Linear Models. Journal of the Royal Statistical Society. Series B (Methodological) 58:619–678

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

578 Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of
579 a correlation matrix. Heredity (Edinb) 95:221–227. doi:10.1038/sj.hdy.6800717

580 Mackay, T.F.C., and R.F. Lyman. 2005. Drosophila bristles and the nature of quantitative genetic
581 variation. Philos Trans R Soc Lond B Biol Sci 360:1513–1527.
582 doi:10.1098/rstb.2005.1672

583 Menke, A., P. Muntner, E.K. Silbergeld, E.A. Platz, and E. Guallar. 2009. Cadmium levels in
584 urine and mortality among U.S. adults. Environ. Health Perspect. 117:190–196.
585 doi:10.1289/ehp.11236

586 van de Mortel, J.E., H. Schat, P.D. Moerland, E. Ver Loren van Themaat, S. van der Ent, H.
587 Blankestijn, A. Ghandilyan, S. Tsiatsiani, and M.G.M. Aarts. 2008. Expression
588 differences for genes involved in lignin, glutathione and sulphate metabolism in response
589 to cadmium in Arabidopsis thaliana and the related Zn/Cd-hyperaccumulator Thlaspi
590 caerulescens. Plant Cell Environ. 31:301–324. doi:10.1111/j.1365-3040.2007.01764.x

591 Nachman, M.W., H.E. Hoekstra, and S.L. D'Agostino. 2003. The genetic basis of adaptive
592 melanism in pocket mice. PNAS 100:5268–5273. doi:10.1073/pnas.0431157100

593 Nelson, R.M., M.E. Pettersson, X. Li, and Ö. Carlborg. 2013. Variance Heterogeneity in
594 Saccharomyces cerevisiae Expression Data: Trans-Regulation and Epistasis. PLOS ONE
595 8:e79507. doi:10.1371/journal.pone.0079507

596 Nordborg, M., and D. Weigel. 2008. Next-generation genetics in plants. Nature 456:720–723.
597 doi:10.1038/nature07629

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

598 Paré, G., N.R. Cook, P.M. Ridker, and D.I. Chasman. 2010. On the Use of Variance per
599 Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the
600 Women's Genome Health Study. PLoS Genetics 6:e1000981.
601 doi:10.1371/journal.pgen.1000981

602 Patterson, N., A.L. Price, and D. Reich. 2006. Population Structure and Eigenanalysis. PLoS
603 Genet 2:e190. doi:10.1371/journal.pgen.0020190

604 Perry, G.M.L., K.W. Nehrke, D.A. Bushinsky, R. Reid, K.L. Lewandowski, P. Hueber, and S.J.
605 Scheinman. 2012. Sex Modifies Genetic Effects on Residual Variance in Urinary
606 Calcium Excretion in Rat (*Rattus norvegicus*). Genetics 191:1003–1013.
607 doi:10.1534/genetics.112.138909

608 Price, A.L., N.A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population
609 stratification in genome-wide association studies. Nat. Rev. Genet. 11:459–463.
610 doi:10.1038/nrg2813

611 Queitsch, C., T.A. Sangster, and S. Lindquist. 2002. Hsp90 as a capacitor of phenotypic
612 variation. Nature 417:618–624. doi:10.1038/nature749

613 R Core Team 2018. R: A language and environment for statistical computing.e. R Foundation for
614 Statistical Computing, Vienna, Austria.

615 Rascio, N., and F. Navari-Izzo. 2011. Heavy metal hyperaccumulating plants: How and why do
616 they do it? And what makes them so interesting?. Plant Science 180:169–181.
617 doi:10.1016/j.plantsci.2010.08.016

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

618 Rönnegård, L., and Ö. Carlborg. 2007. Separation of base allele and sampling term effects gives
619 new insights in variance component QTL analysis. BMC Genetics 8:1. doi:10.1186/1471-
620 2156-8-1

621 Rönnegård, L., M. Felleki, F. Fikse, H.A. Mulder, and E. Strandberg. 2010a. Genetic
622 heterogeneity of residual variance - estimation of variance components using double
623 hierarchical generalized linear models. Genetics Selection Evolution 42:8.
624 doi:10.1186/1297-9686-42-8

625 Rönnegård, L., X. Shen, and M. Alam. 2010b. hglm: A Package for Fitting Hierarchical
626 Generalized Linear Models. The R Journal 2:20. doi:10.32614/RJ-2010-009

627 Rönnegård, L., and W. Valdar. 2011. Detecting Major Genetic Loci Controlling Phenotypic
628 Variability in Experimental Crosses. Genetics 188:435–447.
629 doi:10.1534/genetics.111.127068

630 Rönnegård, L., and W. Valdar. 2012. Recent developments in statistical methods for detecting
631 genetic loci affecting phenotypic variability. BMC Genetics 13:63. doi:10.1186/1471-
632 2156-13-63

633 Salomé, P.A., K. Bomblies, R.A.E. Laitinen, L. Yant, R. Mott, and D. Weigel. 2011. Genetic
634 Architecture of Flowering-Time Variation in Arabidopsis thaliana. Genetics 188:421–
635 433. doi:10.1534/genetics.111.126607

636 Santantonio, N., J.-L. Jannink, and M. Sorrells. 2019. Homeologous Epistasis in Wheat: The
637 Search for an Immortal Hybrid. Genetics 211:1105–1122.
638 doi:10.1534/genetics.118.301851

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

639 Sell-Kubiak, E., N. Duijvesteijn, M.S. Lopes, L.L.G. Janss, E.F. Knol, P. Bijma, and H.A.
640 Mulder. 2015. Genome-wide association study reveals novel loci for litter size and its
641 variability in a Large White pig population. BMC Genomics 16:1049.
642 doi:10.1186/s12864-015-2273-y

643 Shen, X., M. Pettersson, L. Rönnegård, and Ö. Carlborg. 2012. Inheritance Beyond Plain
644 Heritability: Variance-Controlling Genes in Arabidopsis thaliana. PLOS Genetics
645 8:e1002839. doi:10.1371/journal.pgen.1002839

646 Struchalin, M.V., N. Amin, P.H. Eilers, C.M. van Duijn, and Y.S. Aulchenko. 2012. An R
647 package "VariABEL" for genome-wide searching of potentially interacting loci by
648 testing genotypic variance heterogeneity. BMC Genetics 13:4. doi:10.1186/1471-2156-
649 13-4

650 Struchalin, M.V., A. Dehghan, J.C. Witteman, C. van Duijn, and Y.S. Aulchenko. 2010.
651 Variance heterogeneity analysis for detection of potentially interacting genetic loci:
652 method and its limitations. BMC Genetics 11:92. doi:10.1186/1471-2156-11-92

653 Sul, J.H., L.S. Martin, and E. Eskin. 2018. Population structure in genetic studies: Confounding
654 factors and mixed models. PLoS Genet 14:e1007309. doi:10.1371/journal.pgen.1007309

655 Tan, Q., J.V.B. Hjelmborg, M. Thomassen, A.K. Jensen, L. Christiansen, K. Christensen, J.H.
656 Zhao, and T.A. Kruse. 2014. Hierarchical linear modeling of longitudinal pedigree data
657 for genetic association analysis. BioMed Central.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

658 Topless, R.K., T.J. Flynn, M. Cadzow, L.K. Stamp, N. Dalbeth, M.A. Black, and T.R. Merriman.
659 2015. Association of SLC2A9 genotype with phenotypic variability of serum urate in
660 pre-menopausal women. Front. Genet. 6. doi:10.3389/fgene.2015.00313

661 VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci.
662 91:4414–4423. doi:10.3168/jds.2007-0980

663 Wang, G., E. Yang, C.L. Brinkmeyer-Langford, and J.J. Cai. 2014a. Additive, Epistatic, and
664 Environmental Effects Through the Lens of Expression Variability QTL in a Twin
665 Cohort. Genetics 196:413–425. doi:10.1534/genetics.113.157503

666 Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, M. Maccaferri, S. Salvi, S.G.
667 Milner, L. Cattivelli, A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J.
668 Plieske, International Wheat Genome Sequencing Consortium, M. Lillemo, D. Mather, R.
669 Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A.R. Akhunova, C. Feuillet, J. Salse,
670 M. Morgante, C. Pozniak, M.-C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R.
671 Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden, and E.
672 Akhunov. 2014b. Characterization of polyploid wheat genomic diversity using a high-
673 density 90,000 single nucleotide polymorphism array. Plant Biotechnol. J. 12:787–796.
674 doi:10.1111/pbi.12183

675 Wang, W., J. Simmonds, Q. Pan, D. Davidson, F. He, A. Battal, A. Akhunova, H.N. Trick, C.
676 Uauy, and E. Akhunov. 2018. Gene editing and mutagenesis reveal inter-cultivar
677 differences and additivity in the contribution of TaGW2 homoeologues to grain size and
678 weight in wheat. Theor Appl Genet 131:2463–2475. doi:10.1007/s00122-018-3166-7

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

679 Wei, W.-H., J. Bowes, D. Plant, S. Viatte, A. Yarwood, J. Massey, J. Worthington, and S. Eyre.
680 2016. Major histocompatibility complex harbors widespread genotypic variability of non-
681 additive risk of rheumatoid arthritis including epistasis. Scientific Reports 6:25014.
682 doi:10.1038/srep25014

683 Wei, W.-H., S. Viatte, T.R. Merriman, A. Barton, and J. Worthington. 2017. Genotypic
684 variability based association identifies novel non-additive loci DHCR7 and IRF4 in sero-
685 negative rheumatoid arthritis. Scientific Reports 7:5261. doi:10.1038/s41598-017-05447-
686 1

687 Whitt, L., F.K. Ricachenevsky, G. Ziegler, S. Clemens, E. Walker, F. Maathuis, P. Kear, and I.
688 Baxter. 2018. A curated list of genes that control elemental accumulation in plants.
689 bioRxiv 456384. doi:10.1101/456384

690 Xiao, Y., H. Liu, L. Wu, M. Warburton, and J. Yan. 2017. Genome-wide Association Studies in
691 Maize: Praise and Stargaze. Molecular Plant 10:359–374.
692 doi:10.1016/j.molp.2016.12.008

693 Yang, J., R.J.F. Loos, J.E. Powell, S.E. Medland, E.K. Speliotes, D.I. Chasman, L.M. Rose, G.
694 Thorleifsson, V. Steinthorsdottir, R. Mägi, L. Waite, A. Vernon Smith, L.M. Yerges-
695 Armstrong, K.L. Monda, D. Hadley, A. Mahajan, G. Li, K. Kapur, V. Vitart, J.E.
696 Huffman, S.R. Wang, C. Palmer, T. Esko, K. Fischer, J. Hua Zhao, A. Demirkan, A.
697 Isaacs, M.F. Feitosa, J. Luan, N.L. Heard-Costa, C. White, A.U. Jackson, M. Preuss, A.
698 Ziegler, J. Eriksson, Z. Kutalik, F. Frau, I.M. Nolte, J.V. Van Vliet-Ostaptchouk, J.-J.
699 Hottenga, K.B. Jacobs, N. Verweij, A. Goel, C. Medina-Gomez, K. Estrada, J. Lynn
700 Bragg-Gresham, S. Sanna, C. Sidore, J. Tyrer, A. Teumer, I. Prokopenko, M. Mangino,

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C.M. Lindgren, T.L. Assimes, A.R. Shuldiner, J. Hui, J.P. Beilby, W.L. McArdle, P. Hall, T. Haritunians, L. Zgaga, I. Kolcic, O. Polasek, T. Zemunik, B.A. Oostra, M. Juhani Junttila, H. Grönberg, S. Schreiber, A. Peters, A.A. Hicks, J. Stephens, N.S. Foad, J. Laitinen, A. Pouta, M. Kaakinen, G. Willemsen, J.M. Vink, S.H. Wild, G. Navis, F.W. Asselbergs, G. Homuth, U. John, C. Iribarren, T. Harris, L. Launer, V. Gudnason, J.R. O'Connell, E. Boerwinkle, G. Cadby, L.J. Palmer, A.L. James, A.W. Musk, E. Ingelsson, B.M. Psaty, J.S. Beckmann, G. Waeber, P. Vollenweider, C. Hayward, A.F. Wright, I. Rudan, L.C. Groop, A. Metspalu, K.-T. Khaw, C.M. van Duijn, I.B. Borecki, M.A. Province, N.J. Wareham, J.-C. Tardif, H.V. Huikuri, L. Adrienne Cupples, L.D. Atwood, C.S. Fox, M. Boehnke, F.S. Collins, K.L. Mohlke, J. Erdmann, H. Schunkert, C. Hengstenberg, K. Stark, M. Lorentzon, C. Ohlsson, D. Cusi, J.A. Staessen, M.M. Van der Klauw, P.P. Pramstaller, S. Kathiresan, J.D. Jolley, S. Ripatti, M.-R. Jarvelin, E.J.C. de Geus, D.I. Boomsma, B. Penninx, J.F. Wilson, H. Campbell, S.J. Chanock, P. van der Harst, A. Hamsten, H. Watkins, A. Hofman, J.C. Witteman, M.C. Zillikens, A.G. Uitterlinden, F. Rivadeneira, M. Carola Zillikens, L.A. Kiemeney, S.H. Vermeulen, G.R. Abecasis, D. Schlessinger, S. Schipf, M. Stumvoll, A. Tönjes, T.D. Spector, K.E. North, G. Lettre, M.I. McCarthy, S.I. Berndt, A.C. Heath, P.A.F. Madden, D.R. Nyholt, G.W. Montgomery, N.G. Martin, B. McKnight, D.P. Strachan, W.G. Hill, H. Snieder, P.M. Ridker, U. Thorsteinsdottir, K. Stefansson, T.M. Frayling, J.N. Hirschhorn, M.E. Goddard, and P.M. Visscher. 2012. *FTO* genotype is associated with phenotypic variability of body mass index. Nature 490:267–272. doi:10.1038/nature11401

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

722 Young, A.I., F.L. Wauthier, and P. Donnelly. 2018. Identifying loci affecting trait variability and
723 detecting interactions in genome-wide association studies. Nature Genetics 50:1608.
724 doi:10.1038/s41588-018-0225-6

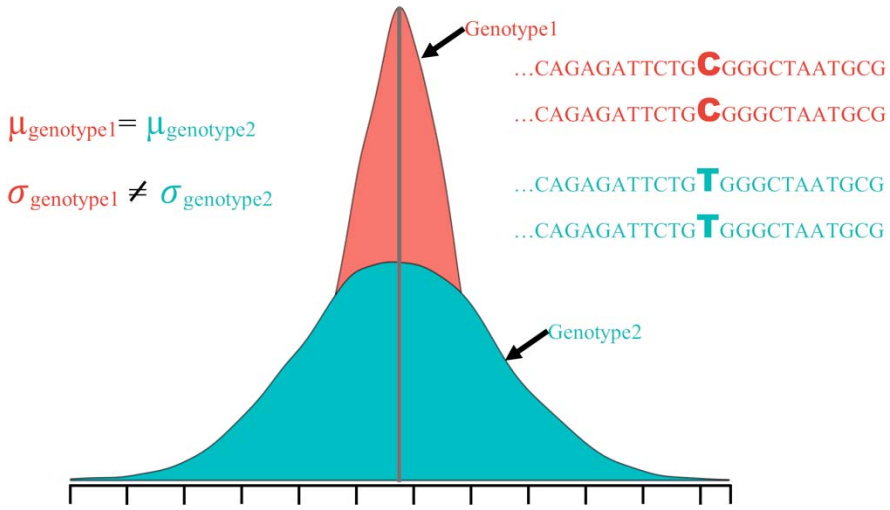
725 Yu, J., L. Wu, L. Fu, Q. Shen, L. Kuang, D. Wu, and G. Zhang. 2019. Genotypic difference of
726 cadmium tolerance and the associated microRNAs in wild and cultivated barley. Plant
727 Growth Regul 87:389–401. doi:10.1007/s10725-019-00479-1

728 Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, and B.S. Weir. 2012. A high-
729 performance computing toolset for relatedness and principal component analysis of SNP
730 data. Bioinformatics 28:3326–3328. doi:10.1093/bioinformatics/bts606

731

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

732 **Figures**

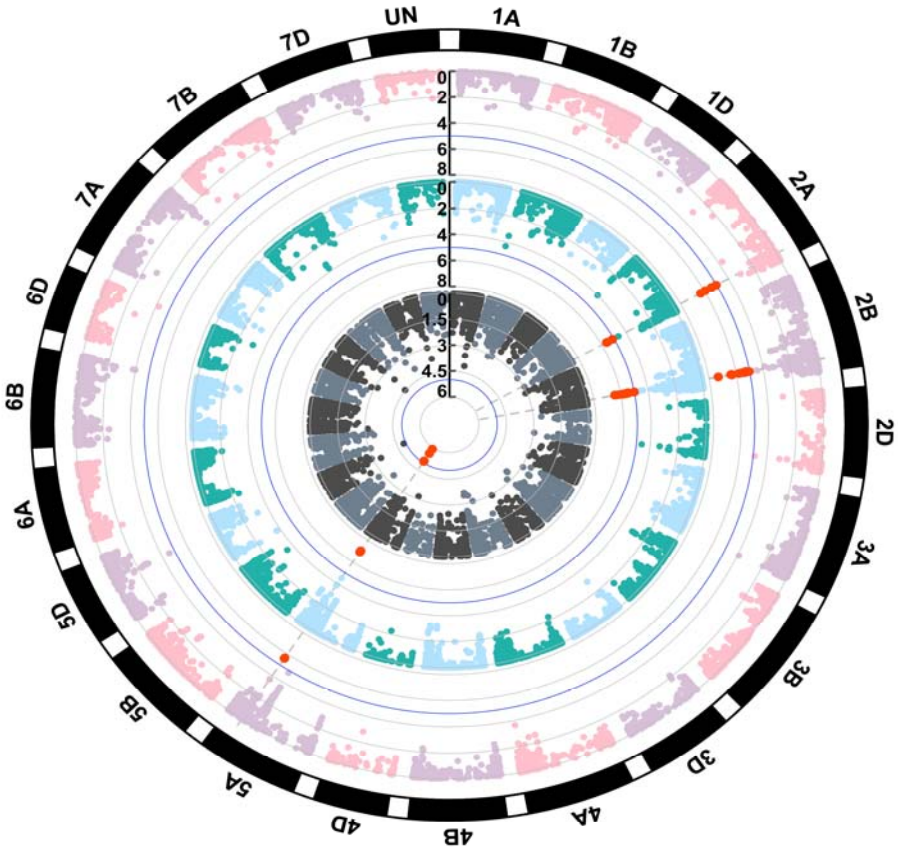


733

734 Figure 1. Illustration of variance heterogeneity of two genotype groups at a biallelic locus
735 affecting the variance not the mean. Genotypes with CC allelic combination present narrow
736 variance, whereas genotypes with TT allelic combination show greater variability. The mean
737 difference between two genotype groups is the same as shown by the solid vertical gray line.

738

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



739

740 Figure 2: Circular Manhattan plot of standard genome-wide association studies (GWAS) based
741 on mean differences (inner), and variance GWAS using double generalized linear model
742 (middle) and hierarchical generalized linear model (outer) for grain cadmium concentration in
743 the hard-red winter wheat association panel. The red dots represent the significant markers
744 associated with either mean or variance heterogeneity quantitative trait loci. The blue line in each
745 circular plot shows the cutoff for the statistical significance. The P-values in $-\log_{10}$ scale are
746 given in black vertical line.

747

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

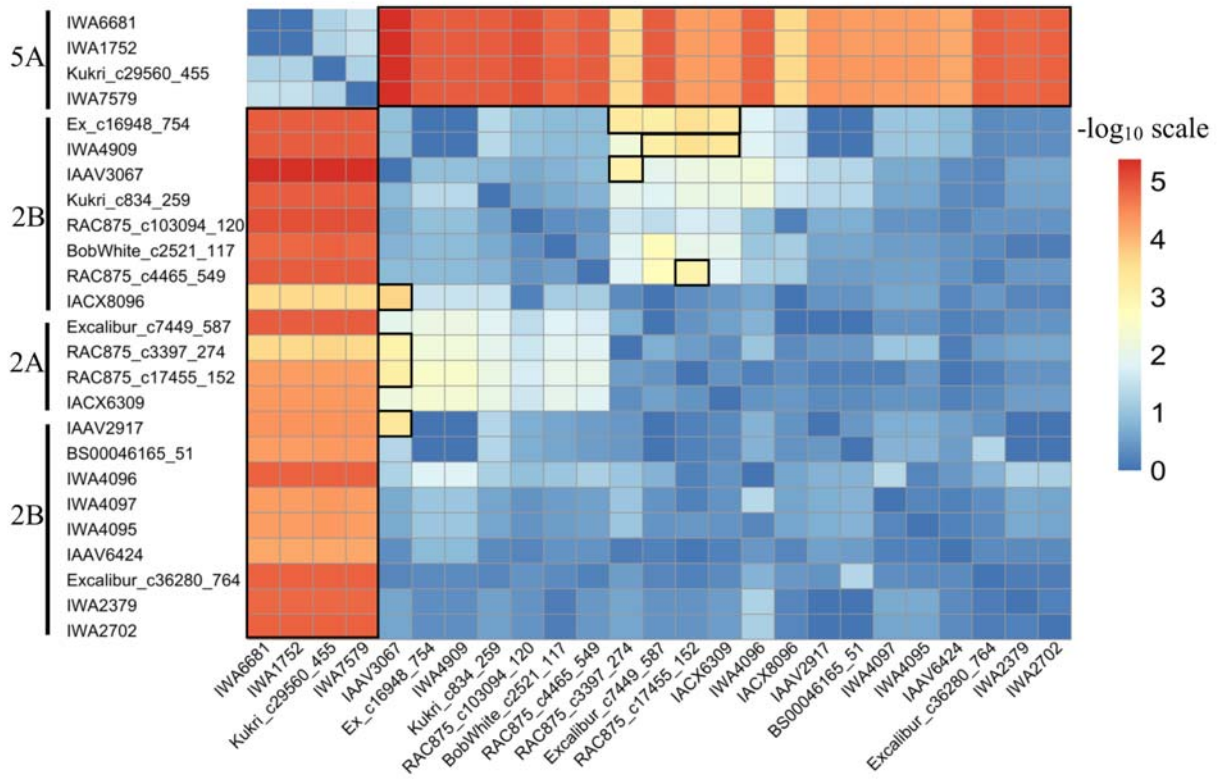
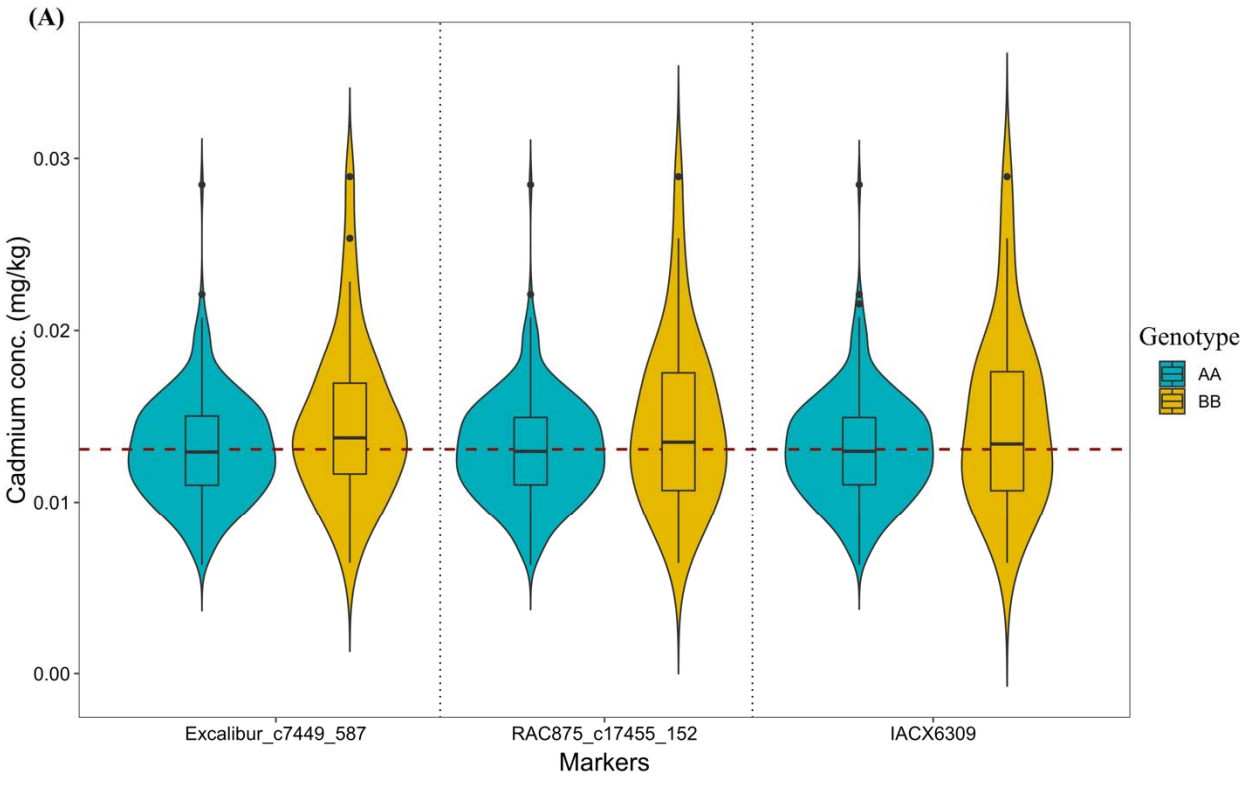
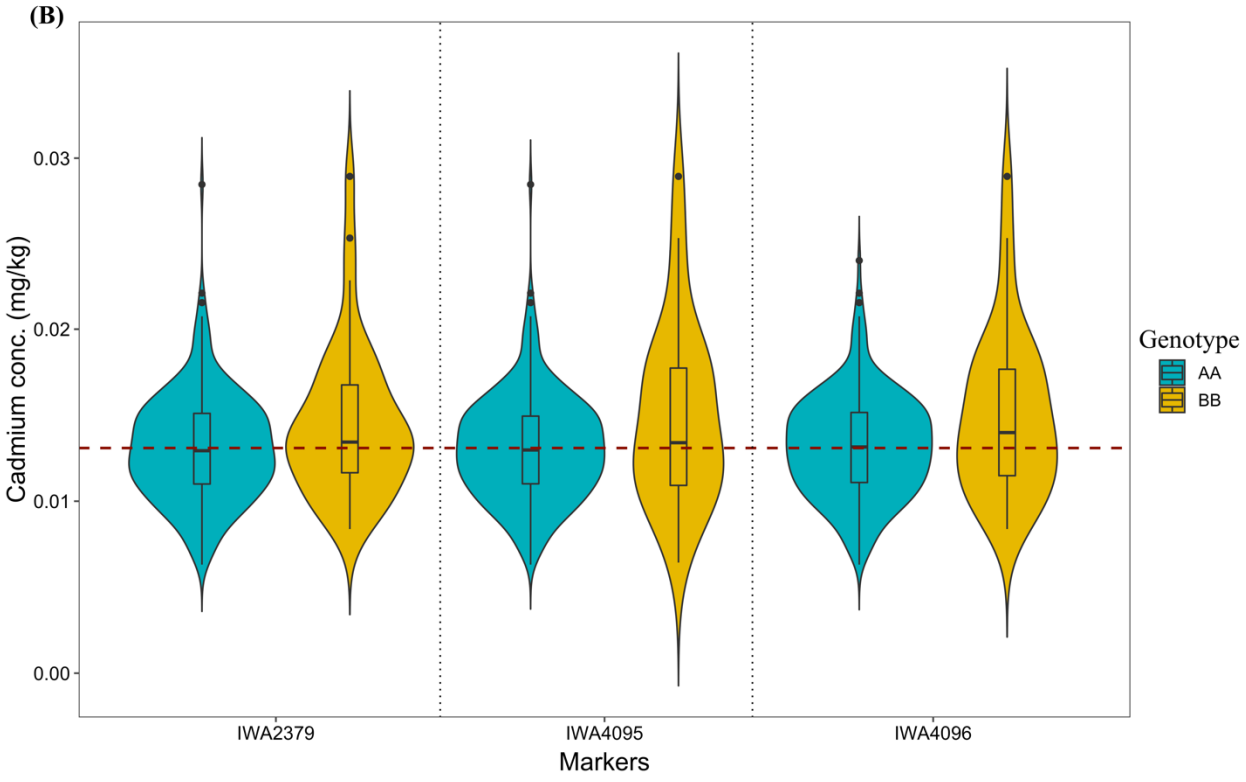


Figure 3: Heat map showing all possible pairwise epistatic interactions between the markers associated with vQTL on chromosomes 2A and 2B or mvQTL on chromosome 5A. Chromosome information of each marker is given on the left side. The heat map is sorted and color coded based on $-\log_{10}$(p-value) scale with the legend given on right side. Interactions that are significant ($-\log_{10} > 3.8$) are color coded as red or orange in color and outlined in black box.

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.





755

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4: Violin plot showing the differences in the mean and variance of grain cadmium concentration with alternative marker genotype groups coded as AA and BB for the top three significant markers associated with vQTL on (A) chromosome 2A and (B) chromosome 2B. The mean of marker genotypes AA and BB are connected by red dotted line.

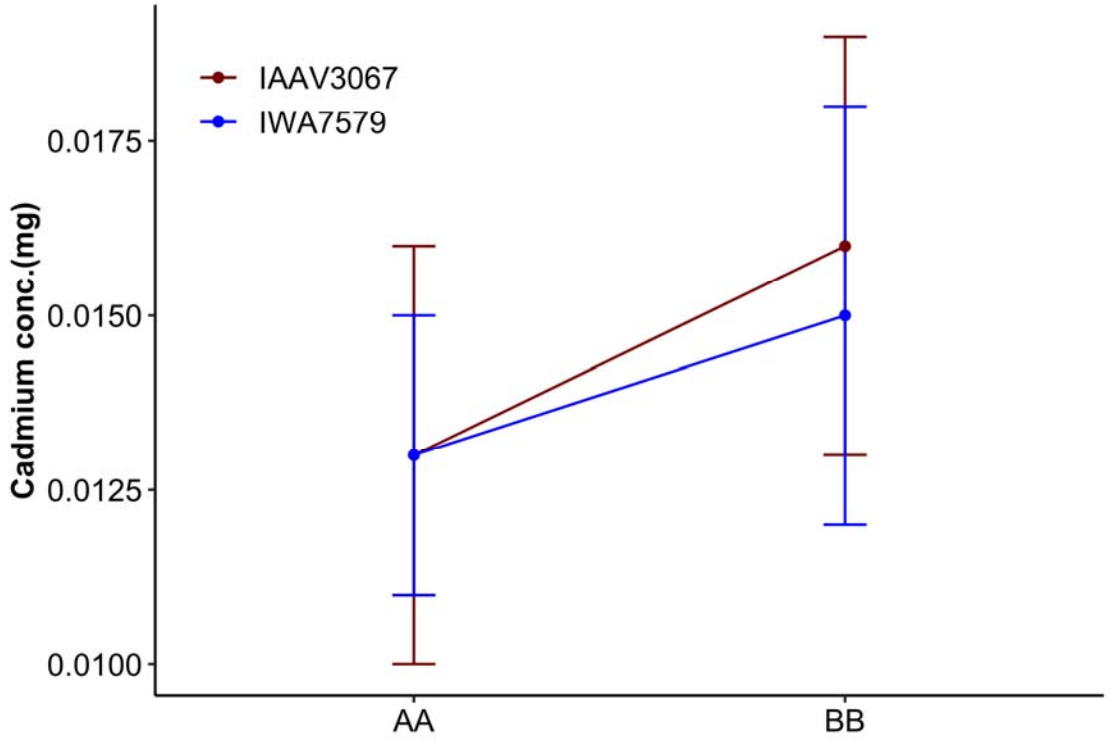


Figure 5: Epistatic interaction plot between marker pair IAAV3067 (shown in dark-red color) and IWA7579 (shown in blue color) on chromosomes 5A (mvQTL) and 2B (vQTL). The y-axis shows the phenotypic value of cadmium concentration (mg). AA and BB represent the alternate genotypes at the particular marker. Plotted points indicate two-locus genotype means ± standard deviations for the two loci represented by error bars. Large difference in the mean value of cadmium concentrations at BB genotype compared to no difference in the mean value of

bioRxiv preprint doi: https://doi.org/10.1101/668087; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

767 cadmium concentrations at AA genotype indicates the presence of interaction between the two

768 markers.